Doctoral Dissertation
博士論文

Measurement of time-dependent CP asymmetry in $B^0 \to K^0_S \pi^0 \gamma$ decays at the Belle II experiment
(Belle II実験における$B^0 \to K^0_S \pi^0 \gamma$崩壊過程の時間依存$CP$非対称度の測定)

A Dissertation Submitted for the Degree of Doctor of Philosophy
July 2023
令和5年7月博士（理学）申請

Department of Physics, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
物理学専攻

Yuma Uematsu
植松 祐真

Abstract

We measure the time-dependent CP asymmetries in $B^0 \to K^0_S\pi^0\gamma$ decay using $(387.5 \pm 5.8) \times 10^6$ $B\bar{B}$ pairs collected near the $\Upsilon(4S)$ resonance with SuperKEKB and Belle II from 2019 to 2022. For the measurement, we reconstruct one neutral B meson in the $B^0 \to K^0_S\pi^0\gamma$ decay channel, identify the flavor of the accompanying B meson from its decay products, and measure the proper-time difference between the two B mesons.

We obtain the following results: for the $K^0_S\pi^0$ invariant mass around $K^*(892)^0$ resonance (0.8 to 1.0 GeV/c^2),

$$S_{CP}(K^{*0}\gamma) = 0.00^{+0.27}_{-0.26}(\text{stat})^{+0.03}_{-0.04}(\text{syst}) \text{ and}$$
$$A_{CP}(K^{*0}\gamma) = -0.10 \pm 0.13(\text{stat}) \pm 0.03(\text{syst}),$$

and for the rest of the mass regions up to 1.8 GeV/c^2,

$$S_{CP}(K^0_S\pi^0\gamma) = +0.04^{+0.45}_{-0.44}(\text{stat}) \pm 0.10(\text{syst}) \text{ and}$$
$$A_{CP}(K^0_S\pi^0\gamma) = +0.06 \pm 0.25(\text{stat}) \pm 0.07(\text{syst}).$$

These results are the most precise to date as a single measurement.

Personal contribution

The measurement reported in this thesis results from the collective efforts of over 1,000 individuals from the Belle II collaboration and the SuperKEKB group. The collaborative work includes designing and constructing the detector and accelerator, managing data through distributed computing, performing calibration, and developing the analysis software framework, including tools for vertex fitting, flavor tagging, and multivariate analysis. The author's specific contributions primarily focus on the improvement of the event reconstruction algorithm and selection criteria detailed in Chapter 4, as well as the refinement of the procedure for estimating *CP* asymmetry described in Chapter 5. Details of each effort are described below.

The author has devoted effort to studying the vertex reconstruction of K_S^0. Firstly, the author has contributed to enhancing the K_S^0 reconstruction software described in Section 4.2.1. The author has contributed to implementing a new feature to eliminate background hits from $K_S^0 \to \pi^+\pi^-$ trajectories in collaborative work. The author has also fine-tuned the selection criteria inside the software to improve the reconstruction efficiency. Secondly, the author also developed an original resolution function to account for the vertex reconstruction using only one K_S^0, as described in Section 5.2.1. Thirdly, the author has contributed to calibrating the resolution function parameters. Although the development of the fast simulation software, as described in Appendix E, was a collaborative effort, the author has predominantly studied helix pulls in K_S^0 decay vertices. Additionally, the author has introduced an original model to calibrate the shape parameters of the helix pulls through cosmic-ray analysis, as outlined in Appendix F.

The author has improved the event selection in Section 4.2. The author has utilized differential evolution for the optimization of the thresholds. The author has investigated the case where multiple candidates exist in the event and has developed selection criteria that effectively keep the signal candidates, as described in Section 4.2.5. The author has optimized the selection further by tuning the background suppression thresholds depending on the quality of the flavor tagging, as described in Section 4.2.4.

The author has investigated the background originating from B decays. To accurately model the M_{bc}–ΔE distribution, the author has divided the background into two distinct components and applied separate modeling approaches for each, as described in Section 5.1.3. Additionally, the author has developed a novel method for constructing the proper-time difference model in the Monte Carlo simulation. In contrast to preceding studies where only the B-meson lifetime parameter is adjusted, this method simultaneously fits the lifetime and resolution function parameters, as described in Section 5.2.4.

Outline

Chapter 1 provides an overview of time-dependent *CP* violation in the B-meson system and emphasizes the significance of $B^0 \to K_S^0\pi^0\gamma$ decay in exploring physics beyond the Standard Model. Chapter 2 presents our experimental setup with the SuperKEKB collider and the Belle II detector. Chapter 3 outlines our data analysis approach for extracting *CP*-violation parameters of $B^0 \to K_S^0\pi^0\gamma$ decay. Chapter 4 details the reconstruction algorithm and selection criteria for $B^0 \to K_S^0\pi^0\gamma$ decay events. Chapter 5 covers the methodology and results of extracting *CP*-violating parameters and assessing systematic uncertainty. Chapter 6 explores constraints on physics beyond the Standard Model and discusses prospects. The thesis concludes in Chapter 7.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

High-energy physics experiments help us understand the fundamental building blocks of the universe, known as elementary particles. The Standard Model (SM) of elementary particles successfully explains most of the observed properties of these particles. However, there are still unanswered questions, such as the matter-antimatter asymmetry, the nature of dark matter and dark energy, quantum gravity, the hierarchy problem, and the strong *CP* problem. Various theories beyond the Standard Model (BSM) have been proposed to answer these questions, but so far, none of them have received enough experimental evidence. To gain insights into BSM physics, we studied the time-dependent *CP* asymmetry of the decay of B mesons into $K^0_S\pi^0\gamma$.

We review the mechanism of decay-time dependent *CP* violation in the neutral B-meson system within the SM, the case of $B^0 \to K^0_S\pi^0\gamma$ decays, and the possible BSM effects.

1.1 Quark-flavor mixing and *CP* violation

Quarks are the elementary particles composing hadrons, i.e., baryons (e.g., protons or neutrons in the nucleus) and mesons (e.g., pions mediating the nuclear force).

Three generations of up- and down-type quarks, $u^i_{L(R)}$ and $d^i_{L(R)}$ with $i = 1, 2, 3$ (up, charm, and top for up-type; down, strange, and bottom for down-type), are observed so far as predicted in the SM. These six types of quarks are called *flavor*.

While the quark flavor is independent of the strong interaction, the charged-current

weak interaction changes the flavor of the quark current as

$$J_W^{\mu+} = \frac{1}{\sqrt{2}} \bar{u}_L'^i \gamma^\mu d_L'^i.$$

Here $u_L'^i$ and $d_L'^i$ are the eigenstates of weak interaction, which is related to mass eigenstates u_L^i and d_L^i with unitary transformation:

$$u_L'^i = U_u^{ij} u_L^j, \qquad d_L'^i = U_d^{ij} d_L^j.$$

These two different bases do not make any difference for the neutral current, where the unitary matrices cancel each other. However, at the interaction with weak charged-current, the different bases allow transition between generations:

$$J_W^{\mu+} = \frac{1}{\sqrt{2}} \bar{u}_L^i \gamma^\mu \left(U_u^\dagger U_d \right)^{ij} d_L^j.$$

The unitary matrix $V = U_u^\dagger U_d$ is called the Cabbibo-Kobayashi-Maskawa (CKM) matrix.

Since there are three generations of quarks, the complex matrix V has three dimensions and 18 parameters. The unitary condition gives nine constraints, and the remaining nine parameters are classified into three angles of three-dimensional rotation, e.g., Euler angles θ_{ij} ($ij = 12, 13, 23$), and six complex phases. These complex phases are canceled out by the phase shift of the six quark fields except for one overall phase δ [1]. This irreducible complex phase is the only *CP* violation in the SM currently observed. We can write down the matrix as

$$\begin{aligned} V &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta} & c_{23} c_{13} \end{pmatrix}, \end{aligned}$$

where c_{ij} and s_{ij} are shorthands for $\cos\theta_{ij}$ and $\sin\theta_{ij}$. The Wolfenstein parametrization [2] clarifies the hierarchy of the magnitude and *CP* violation. By defining

$$\begin{aligned} \lambda &\coloneqq s_{12}, \\ A &\coloneqq s_{23}/\lambda^2, \text{ and} \\ \rho + i\eta &\coloneqq s_{13} e^{i\delta} / A\lambda^3, \end{aligned}$$

we can rewrite the matrix as

$$V = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right). \tag{1.1}$$

The size of the parameters are [3]

$$\lambda = 0.2250 \pm 0.0002, \quad A = 0.813^{+0.012}_{-0.006}, \quad \rho = 0.160^{+0.009}_{-0.005}, \quad \text{and} \quad \eta = 0.357^{+0.0012}_{-0.005}.$$

1.2 Decay-time dependent *CP* violation in *B*-meson system

A. I. Sanda, I. I. Bigi, and A. B. Carter pointed out that the maximum possible *CP* violation of $\mathcal{O}(\sin\delta)$ appears in decays of B^0 meson thorough the interference from B^0–$\overline{B}^0$ mixing [4–6]. In this section, we formulate the decay-time dependent *CP* violation in the neutral *B*-meson system.

1.2.1 Mixing of neutral *B*-mesons

The B^0 and $\overline{B}^0$ transit to each other via the box diagrams shown in Fig. 1.1. We formulate this time evolution in this section.

An initial state *B*-meson at a time $t = 0$ can be written as

$$|\psi(0)\rangle = a(0)\,|B^0\rangle + b(0)\,|\overline{B}^0\rangle\,.$$

Considering the decays of *B*-mesons, the state at a time t is

$$|\psi(t)\rangle = a(t)\,|B^0\rangle + b(t)\,|\overline{B}^0\rangle + \sum_i c_i(t)\,|f_i\rangle\,,$$

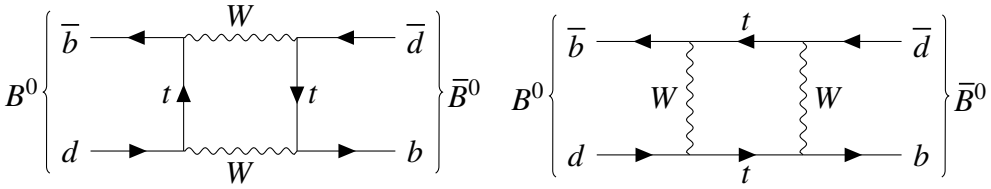


Figure 1.1: Feynman diagrams for the B^0–$\overline{B}^0$ mixing at the leading order. Only the dominant contribution from off-shell top quark is shown.

where f_i covers all the final states. The time evolution of the B-meson system, $a(t)$ and $b(t)$, satisfies the Shrödinger equation in the subspace of states:

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \boldsymbol{H}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix}. \tag{1.2}$$

The effective Hamiltonian $\boldsymbol{H}$ is not Hermitian, but it can be expressed as the sum of a Hermitian matrix and an anti-Hermitian matrix,

$$\boldsymbol{H} = \boldsymbol{M} - \frac{i}{2}\boldsymbol{\Gamma},$$

where $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ are Hermitian matrices,

$$\boldsymbol{M} = \frac{\boldsymbol{H} + \boldsymbol{H}^*}{2} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{22} \end{pmatrix},$$
$$\boldsymbol{\Gamma} = i(\boldsymbol{H} - \boldsymbol{H}^*) = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{22} \end{pmatrix}.$$

$\boldsymbol{H}$ is a transition matrix from $(\,|B^0\rangle\,,\,|\bar{B}^0\rangle)^\top$ to $(\,\langle B^0|\,,\,\langle \bar{B}^0|)$ and can be written as

$$\boldsymbol{H} = \begin{pmatrix} \langle B^0|\,\mathcal{H}_{B^0}\,|B^0\rangle & \langle B^0|\,\mathcal{H}_{B^0}\,|\bar{B}^0\rangle \\ \langle \bar{B}^0|\,\mathcal{H}_{B^0}\,|B^0\rangle & \langle \bar{B}^0|\,\mathcal{H}_{B^0}\,|\bar{B}^0\rangle \end{pmatrix},$$

where $\mathcal{H}_{B^0}$ is the effective Hamiltonian for mixing and decay. Assuming the CPT invariance of $\mathcal{H}_{B^0}$, we obtain

$$M_{11} - \frac{i}{2}\Gamma_{11} = M_{22} - \frac{i}{2}\Gamma_{22}.$$

Considering that the diagonal elements of Hermitian matrices are real, we equate real and imaginary parts separately as

$$M_{11} = M_{22} = M,$$
$$\Gamma_{11} = \Gamma_{22} = \Gamma.$$

These diagonal elements of $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ correspond to mass and total decay width. In contrast, the non-diagonal elements, M_{12} and Γ_{12}, correspond to dispersive (off-shell) and absorptive (on-shell) contributions to the mixing, as shown in Figs. 1.1 and 1.2.

The time evolution of motion is based on the mass eigenstates B_1 and B_2, which are two eigenvectors of $\boldsymbol{H}$:

$$(\,|B_1\rangle\,,\,|B_2\rangle) = \Big(\,|B^0\rangle\,,\,|\bar{B}^0\rangle\Big)\boldsymbol{P}$$

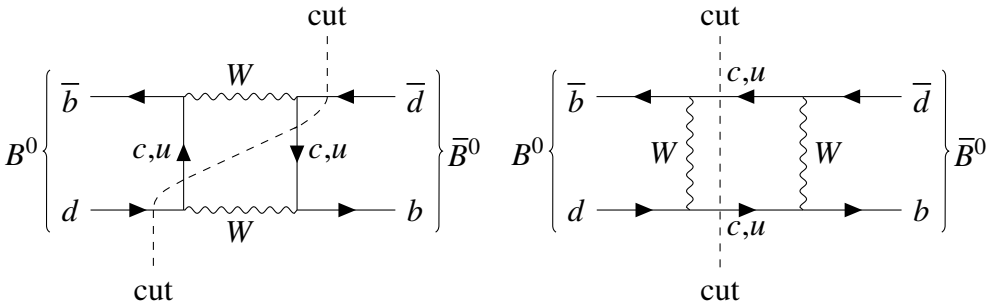


Figure 1.2: Feynman diagrams for the B^0–$\overline{B}^0$ mixing at the leading order. The dashed line shows the cut for on-shell contribution.

where

$$\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2) = \begin{pmatrix} p & p \\ q & -q \end{pmatrix},$$

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

$$|p|^2 + |q|^2 = 1.$$

We can diagonalize $\boldsymbol{H}$ with $\boldsymbol{P}$ as

$$\begin{aligned}\boldsymbol{P}^{-1}\boldsymbol{H}\boldsymbol{P} &= \begin{pmatrix} M - \frac{i}{2}\Gamma + \frac{q}{p}\left(M_{12} - \frac{i}{2}\Gamma_{12}\right) & 0 \\ 0 & M - \frac{i}{2}\Gamma - \frac{q}{p}\left(M_{12} - \frac{i}{2}\Gamma_{12}\right) \end{pmatrix} \\ &= \begin{pmatrix} m_1 - \frac{i}{2}\Gamma_1 & 0 \\ 0 & m_2 - \frac{i}{2}\Gamma_2 \end{pmatrix}.\end{aligned}$$

In the last line, we rewrite the eigenvalues with real numbers $m_1, \Gamma_1, m_2, \Gamma_2$. The solution of Eq. (1.2) is

$$\begin{aligned}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} &= \boldsymbol{P} \begin{pmatrix} e^{-im_1 t - \frac{1}{2}\Gamma_1 t} & 0 \\ 0 & e^{-im_2 t - \frac{1}{2}\Gamma_2 t} \end{pmatrix} \boldsymbol{P}^{-1} \begin{pmatrix} a(0) \\ b(0) \end{pmatrix} \\ &= \begin{pmatrix} g_+(t) & -\frac{p}{q} g_-(t) \\ -\frac{q}{p} g_-(t) & g_+(t) \end{pmatrix} \begin{pmatrix} a(0) \\ b(0) \end{pmatrix},\end{aligned}$$

where

$$g_\pm(t) = \frac{1}{2}\left(e^{-im_2 t - \frac{1}{2}\Gamma_2 t} \pm e^{-im_1 t - \frac{1}{2}\Gamma_1 t}\right).$$

Now we can write the time evolution of initially pure B^0 or $\overline{B}^0$ state as

$$\left(|B^0(t)\rangle\,,\,|\overline{B}^0(t)\rangle\right) = \left(|B^0\rangle\,,\,|\overline{B}^0\rangle\right)\begin{pmatrix} g_+(t) & -\frac{p}{q}g_-(t) \\ -\frac{q}{p}g_-(t) & g_+(t)\end{pmatrix},$$

which satisfy $|B^0(0)\rangle = |B^0\rangle$ and $|\overline{B}^0(0)\rangle = |\overline{B}^0\rangle$.

1.2.2 Decay-time dependent *CP* violation

Time-dependent decay rate to a final state f and its *CP*-conjugate decay can be calculated from transition amplitude:

$$\begin{aligned}
\Gamma\left(B^0 \to f; t\right) &\propto \left|\langle f|\,\mathcal{H}\,|B^0(t)\rangle\right|^2 \\
&= \left|g_+(t)\,\langle f|\,\mathcal{H}\,|B^0\rangle - \frac{q}{p}g_-(t)\,\langle f|\,\mathcal{H}\,|\overline{B}^0\rangle\right|^2, \\
\Gamma\left(\overline{B}^0 \to \overline{f}; t\right) &\propto \left|\langle \overline{f}|\,\mathcal{H}\,|\overline{B}^0(t)\rangle\right|^2 \\
&= \left|-\frac{p}{q}g_-(t)\,\langle \overline{f}|\,\mathcal{H}\,|B^0\rangle + g_+(t)\,\langle \overline{f}|\,\mathcal{H}\,|\overline{B}^0\rangle\right|^2,
\end{aligned}$$

where $\overline{f}$ stands for the *CP*-conjugate of state f. Considering the *CP*-eigen final state f_{CP} with eigenvalue $\xi_{CP} = \pm 1$, we can replace the transition amplitude as

$$\begin{aligned}
\langle f_{CP}|\,\mathcal{H}\,|B^0\rangle &= \mathcal{M}, \\
\langle \overline{f_{CP}}|\,\mathcal{H}\,|B^0\rangle &= \xi_{CP}\mathcal{M}, \\
\langle \overline{f_{CP}}|\,\mathcal{H}\,|\overline{B}^0\rangle &= \overline{\mathcal{M}}, \\
\langle f_{CP}|\,\mathcal{H}\,|\overline{B}^0\rangle &= \xi_{CP}\overline{\mathcal{M}}.
\end{aligned}$$

The time-dependent decay rates are

$$\begin{aligned}
\Gamma\left(B^0 \to f_{CP}; t\right) &\propto \left|\langle f_{CP}|\,\mathcal{H}\,|B^0(t)\rangle\right|^2 \\
&= \frac{1}{2}|\mathcal{M}|^2 e^{-\Gamma t}\left\{\left(1+|\lambda|^2\right)\cosh\frac{1}{2}\Delta\Gamma t + 2\xi_{CP}\,\mathrm{Re}\,[\lambda]\sinh\frac{1}{2}\Delta\Gamma t\right. \\
&\quad \left.+\left(1-|\lambda|^2\right)\cos\Delta m_d t - 2\xi_{CP}\,\mathrm{Im}\,[\lambda]\sin\Delta m_d t\right\}
\end{aligned}$$

and

$$\begin{aligned}\Gamma\left(\overline{B}^0 \to \overline{f_{CP}}; t\right) &\propto \left|\langle \overline{f_{CP}}|\,\mathcal{H}\,|\overline{B}^0(t)\rangle\right|^2 \\ &= \frac{1}{2}\left|\frac{p}{q}\right|^2 |\mathcal{M}|^2 e^{-\Gamma t}\left\{\left(1+|\lambda|^2\right)\cosh\frac{1}{2}\Delta\Gamma t + 2\xi_{CP}\,\mathrm{Re}\,[\lambda]\sinh\frac{1}{2}\Delta\Gamma t\right. \\ &\quad \left.-\left(1-|\lambda|^2\right)\cos\Delta m_d t + 2\xi_{CP}\,\mathrm{Im}\,[\lambda]\sin\Delta m_d t\right\},\end{aligned}$$

where we define

$$\begin{aligned}\Delta m_d &\coloneqq m_2 - m_1, \\ \Delta\Gamma &\coloneqq \Gamma_2 - \Gamma_1, \\ \lambda &\coloneqq \frac{q}{p}\frac{\overline{\mathcal{M}}}{\mathcal{M}}.\end{aligned}$$

Δm_d and $\Delta\Gamma$ can be written in terms of M_{12} and Γ_{12}, as q/p:

$$\begin{aligned}\Delta m_d &= 2\,\mathrm{Re}\left[\frac{q}{p}\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\right] \\ &= \sqrt{2|M_{12}|^2 - \frac{1}{2}|\Gamma_{12}|^2 + \sqrt{\left(2|M_{12}|^2 - \frac{1}{2}|\Gamma_{12}|^2\right)^2 + 4\left(\mathrm{Re}\left[\Gamma_{12}^* M_{12}\right]\right)^2}}, \\ \Delta\Gamma &= -4\,\mathrm{Im}\left[\frac{q}{p}\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\right] \\ &= -4\frac{\mathrm{Re}\left[\Gamma_{12}^* M_{12}\right]}{\Delta m_d}.\end{aligned}$$

Here, we take $\Delta m_d > 0$ and determine the sign of q/p accordingly. The off-shell contribution M_{12} shown in Fig. 1.1 is dominated by the top-quark contribution and proportional to m_t^2 [7], while the on-shell contribution Γ_{12} is expected to scale with m_b:

$$\left|\frac{\Gamma_{12}}{M_{12}}\right| \simeq O\left(\frac{m_b^2}{m_t^2}\right) \sim 10^{-3}.$$

We can write down Δm_d, $\Delta\Gamma$, and $|q/p|$ up to first order of $|\Gamma_{12}/M_{12}|$ as

$$\begin{aligned}\Delta m_d &\simeq 2|M_{12}|, \\ \Delta\Gamma &\simeq -2|M_{12}|\,\mathrm{Re}\left[\frac{\Gamma_{12}}{M_{12}}\right], \\ \left|\frac{q}{p}\right| &\simeq 1 - \frac{1}{2}\,\mathrm{Im}\left[\frac{\Gamma_{12}}{M_{12}}\right].\end{aligned}$$

We drop the first-order term as well and use $\Delta\Gamma \simeq 0$ and $|q/p| \simeq 1$ in the following discussion.

The decay-time dependent *CP* asymmetry can be written as

$$\begin{aligned}\mathcal{A}_{CP}(t) &\coloneqq \frac{\Gamma\big(\overline{B}^0 \to \overline{f_{CP}}; t\big) - \Gamma\big(B^0 \to f_{CP}; t\big)}{\Gamma\big(\overline{B}^0 \to \overline{f_{CP}}; t\big) + \Gamma\big(B^0 \to f_{CP}; t\big)} \\ &= \frac{\big(|\lambda|^2 - 1\big)\cos\Delta m_d t + 2\xi_{CP}\,\mathrm{Im}\,[\lambda]\sin\Delta m_d t}{|\lambda|^2 + 1} \\ &= A_{CP}\cos\Delta m_d t + S_{CP}\sin\Delta m_d t,\end{aligned}$$

where

$$A_{CP} \coloneqq \frac{|\lambda|^2 - 1}{|\lambda|^2 + 1} \quad \text{and} \quad S_{CP} \coloneqq \xi_{CP}\frac{2\,\mathrm{Im}\,[\lambda]}{|\lambda|^2 + 1}. \tag{1.3}$$

Here, $A_{CP} \neq 0$ if and only if $|\mathcal{M}| \neq |\overline{\mathcal{M}}|$. Therefore, we call A_{CP} direct *CP* asymmetry. Let us consider this condition more specifically. Since the *CP*-transformation only changes the complex phase of a single diagram contribution in $\mathcal{M}$, we need at least two diagram contributions to satisfy the condition. Let us consider the case where $\mathcal{M}$ is a sum of two amplitudes,

$$\mathcal{M}_1 = |\mathcal{M}_1|e^{i\theta_1}e^{i\delta_1} \quad \text{and} \quad \mathcal{M}_2 = |\mathcal{M}_2|e^{i\theta_2}e^{i\delta_2},$$

where $\theta_{1,2}$ is *CP*-odd (weak) phase and $\delta_{1,2}$ is *CP*-even (strong) phase. We can write the magnitude of amplitude as

$$\begin{aligned}|\mathcal{M}|^2 &= |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + 2|\mathcal{M}_1||\mathcal{M}_2|(\sin\theta_{12}\sin\delta_{12} - \cos\theta_{12}\cos\delta_{12}), \\ |\overline{\mathcal{M}}|^2 &= |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 - 2|\mathcal{M}_1||\mathcal{M}_2|(\sin\theta_{12}\sin\delta_{12} + \cos\theta_{12}\cos\delta_{12}),\end{aligned}$$

where $\theta_{12} = \theta_1 - \theta_2$ and $\delta_{12} = \delta_1 - \delta_2$. Therefore, we need at least two amplitudes with different weak and strong phases for non-zero A_{CP}. On the contrary, non-zero S_{CP} could arise even if $|\mathcal{M}| = |\overline{\mathcal{M}}|$. Therefore, we call S_{CP} indirect *CP* asymmetry. We also refer to S_{CP} as mixing-induced *CP* asymmetry since it arises from the phase in B^0–$\overline{B}^0$ mixing, q/p, even if the phase difference in decay, $\overline{\mathcal{M}}/\mathcal{M}$, is small. Note that the sign of S_{CP} depends on the *CP* eigenvalue of the final state. Moreover, we can confirm that A_{CP} and S_{CP} are bounded in a unit circle by definition,

$$A_{CP}^2 + S_{CP}^2 \leq 1.$$

The equality holds if λ is purely imaginary.

In the SM, considering that $M_{12}^* = \langle \overline{B}^0 | \mathcal{H}_{B^0} | B^0 \rangle$ is dominated by the diagrams in Fig. 1.1,

$$\frac{q}{p} \simeq \sqrt{\frac{M_{12}^*}{M_{12}}} \simeq \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}.$$

At the leading order, the *CP*-violating phase only appears in V_{td} and V_{ub} as shown in Eq. (1.1). Therefore, we can approximate q/p using

$$\phi_1 \coloneqq \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right] \simeq \pi - \arg\left[V_{td}\right]$$

as

$$\frac{q}{p} \simeq \frac{V_{td}}{V_{td}^*} \simeq e^{-2i\phi_1}.$$

In case $|\mathcal{M}| = |\overline{\mathcal{M}}|$ and the weak phase of $\mathcal{M}$ is negligible, we can write *CP*-violation parameters as

$$A_{CP} = 0 \quad \text{and} \quad S_{CP} = -\xi_{CP} \sin 2\phi_1.$$

The size of the *CP* violation $\sin 2\phi_1$ is measured in $B^0 \to c\overline{c}K^0$ decays, domninatly $B^0 \to J/\psi\, K_S^0$ ($\xi_{CP} = -1$) and $B^0 \to J/\psi\, K_L^0$ ($\xi_{CP} = +1$). The world average is [8]

$$\sin 2\phi_1 = 0.699 \pm 0.017.$$

1.3 Decay-time dependent *CP* violation of $B^0 \to K_S^0 \pi^0 \gamma$ decays

The final state with γ is not the *CP* eigenstate, so the formulation in Section 1.2 does not hold as it is. We derive the *CP*-violation parameters first and then discuss the interesting suppression in the SM.

1.3.1 *CP* violation in radiative decays

For the final state with photon, we have two different final states for each polarization: $f_{CP}\gamma_R$ and $f_{CP}\gamma_L$. These are not *CP* eigenstates; rather, they transform with each other. Within the arbitrariness of the *CP* phase, we can relate the two states as

$$\langle \overline{f_{CP}\gamma_R} | = \xi_{CP} \langle f_{CP}\gamma_L | \text{ and}$$
$$\langle \overline{f_{CP}\gamma_L} | = \xi_{CP} \langle f_{CP}\gamma_R | .$$

We can define the *CP* eigenstates as

$$(\langle f_{CP}\gamma_+|\,,\,\langle f_{CP}\gamma_-|) = (\langle f_{CP}\gamma_R|\,,\,\langle f_{CP}\gamma_L|)\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},$$

with *CP* eigenvalues $\pm\xi_{CP}$. Transition amplitudes to these *CP* eigenstates can be written as

$$\begin{aligned}
\langle f_{CP}\gamma_\pm|\,\mathcal{H}\,|B^0\rangle &= \mathcal{M}_\pm,\\
\langle\overline{f_{CP}\gamma_\pm}|\,\mathcal{H}\,|B^0\rangle &= \pm\xi_{CP}\mathcal{M}_\pm,\\
\langle\overline{f_{CP}\gamma_\pm}|\,\mathcal{H}\,|\bar{B}^0\rangle &= \overline{\mathcal{M}_\pm},\\
\langle f_{CP}\gamma_\pm|\,\mathcal{H}\,|\bar{B}^0\rangle &= \pm\xi_{CP}\overline{\mathcal{M}_\pm}.
\end{aligned}$$

We can relate these *CP*-eigen amplitudes to helicity amplitudes:

$$\begin{aligned}
(\mathcal{M}_+,\mathcal{M}_-) &= (\mathcal{M}_R,\mathcal{M}_L)\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},\\
\left(\overline{\mathcal{M}_+},\overline{\mathcal{M}_-}\right) &= \left(\overline{\mathcal{M}_R},\overline{\mathcal{M}_L}\right)\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{M}_{R(L)} &= \langle f_{CP}\gamma_{R(L)}|\,\mathcal{H}\,|B^0\rangle\,,\\
\overline{\mathcal{M}_{R(L)}} &= \langle\overline{f_{CP}\gamma_{R(L)}}|\,\mathcal{H}\,|\bar{B}^0\rangle\,.
\end{aligned}$$

Note that $\overline{\mathcal{M}_{R(L)}}$ corresponds to the amplitude emitting the photon with left (right) helicity. Using $\Delta\Gamma \simeq 0$, one can write the time-dependent transition amplitudes as

$$\begin{aligned}
\langle f_{CP}\gamma_\pm|\,\mathcal{H}\,|B^0(t)\rangle &= g_+(t)\mathcal{M}_\pm \mp \frac{q}{p}g_-(t)\xi_{CP}\overline{\mathcal{M}_\pm}\\
&\simeq e^{-iMt-\frac{1}{2}\Gamma t}\left\{\mathcal{M}_\pm\cos\frac{1}{2}\Delta m_d t \pm i\frac{q}{p}\xi_{CP}\overline{\mathcal{M}_\pm}\sin\frac{1}{2}\Delta m_d t\right\}, \qquad (1.4)\\
\langle\overline{f_{CP}\gamma_\pm}|\,\mathcal{H}\,|\bar{B}^0(t)\rangle &= g_+(t)\overline{\mathcal{M}_\pm} \mp \frac{p}{q}g_-(t)\xi_{CP}\mathcal{M}_\pm\\
&\simeq e^{-iMt-\frac{1}{2}\Gamma t}\left\{\overline{\mathcal{M}_\pm}\cos\frac{1}{2}\Delta m_d t \pm i\frac{p}{q}\xi_{CP}\mathcal{M}_\pm\sin\frac{1}{2}\Delta m_d t\right\}. \qquad (1.5)
\end{aligned}$$

The time-dependent decay rates of B^0 without separating the photon polarization is,

using $|q/p| \simeq 1$,

$$\begin{aligned}
\Gamma\Big(B^0 \to f_{CP}\gamma; t\Big) &= \Gamma\Big(B^0 \to f_{CP}\gamma_+; t\Big) + \Gamma\Big(B^0 \to f_{CP}\gamma_-; t\Big) \\
&\propto \left|\langle f_{CP}\gamma_+|\,\mathcal{H}\,|B^0(t)\rangle\right|^2 + \left|\langle f_{CP}\gamma_-|\,\mathcal{H}\,|B^0(t)\rangle\right|^2 \\
&\simeq \frac{1}{2}e^{-\Gamma t}\left\{\left(|\mathcal{M}_+|^2 + \left|\overline{\mathcal{M}_+}\right|^2 + |\mathcal{M}_-|^2 + \left|\overline{\mathcal{M}_-}\right|^2\right)\right. \\
&\quad + \left(|\mathcal{M}_+|^2 - \left|\overline{\mathcal{M}_+}\right|^2 + |\mathcal{M}_-|^2 - \left|\overline{\mathcal{M}_-}\right|^2\right)\cos\Delta m_d t \\
&\quad \left. - 2\xi_{CP}\,\mathrm{Im}\left[\frac{q}{p}\left(\mathcal{M}_+^*\overline{\mathcal{M}_+} - \mathcal{M}_-^*\overline{\mathcal{M}_-}\right)\right]\sin\Delta m_d t\right\} \\
&= \frac{1}{2}e^{-\Gamma t}\left\{\left(|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 + \left|\overline{\mathcal{M}_R}\right|^2 + \left|\overline{\mathcal{M}_L}\right|^2\right)\right. \\
&\quad + \left(|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 - \left|\overline{\mathcal{M}_R}\right|^2 - \left|\overline{\mathcal{M}_L}\right|^2\right)\cos\Delta m_d t \\
&\quad \left. - 2\xi_{CP}\,\mathrm{Im}\left[\frac{q}{p}\left(\mathcal{M}_R^*\overline{\mathcal{M}_L} + \mathcal{M}_L^*\overline{\mathcal{M}_R}\right)\right]\sin\Delta m_d t\right\}.
\end{aligned}$$

We can write similarly for $\overline{B}^0$,

$$\begin{aligned}
\Gamma\Big(\overline{B}^0 \to \overline{f_{CP}\gamma}; t\Big) &\propto \left|\langle\overline{f_{CP}\gamma_+}|\,\mathcal{H}\,|\overline{B}^0(t)\rangle\right|^2 + \left|\langle\overline{f_{CP}\gamma_-}|\,\mathcal{H}\,|\overline{B}^0(t)\rangle\right|^2 \\
&\simeq \frac{1}{2}e^{-\Gamma t}\left\{\left(|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 + \left|\overline{\mathcal{M}_R}\right|^2 + \left|\overline{\mathcal{M}_L}\right|^2\right)\right. \\
&\quad - \left(|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 - \left|\overline{\mathcal{M}_R}\right|^2 - \left|\overline{\mathcal{M}_L}\right|^2\right)\cos\Delta m_d t \\
&\quad \left. + 2\xi_{CP}\,\mathrm{Im}\left[\frac{q}{p}\left(\mathcal{M}_R^*\overline{\mathcal{M}_L} + \mathcal{M}_L^*\overline{\mathcal{M}_R}\right)\right]\sin\Delta m_d t\right\}.
\end{aligned}$$

Therefore, *CP* asymmetry parameters can be written as

$$A_{CP} = -\frac{|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 - \left|\overline{\mathcal{M}_R}\right|^2 - \left|\overline{\mathcal{M}_L}\right|^2}{|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 + \left|\overline{\mathcal{M}_R}\right|^2 + \left|\overline{\mathcal{M}_L}\right|^2}, \tag{1.6}$$

$$S_{CP} = \frac{2\xi_{CP}\,\mathrm{Im}\left[\frac{q}{p}\left(\mathcal{M}_R^*\overline{\mathcal{M}_L} + \mathcal{M}_L^*\overline{\mathcal{M}_R}\right)\right]}{|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 + \left|\overline{\mathcal{M}_R}\right|^2 + \left|\overline{\mathcal{M}_L}\right|^2}. \tag{1.7}$$

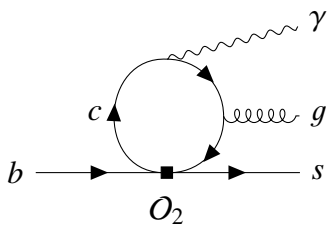


Figure 1.3: $b \to s\gamma g$ loop diagram with effective operator O_2^c.

In case of no direct *CP* violation, $|\mathcal{M}_{R(L)}| = |\overline{\mathcal{M}_{R(L)}}|$ and $A_{CP} = 0$, we can write S_{CP} as

$$S_{CP} \simeq \frac{\xi_{CP}\,\mathrm{Im}\left[e^{-2i\phi_1}\left(\mathcal{M}_R^*\overline{\mathcal{M}_L} + \mathcal{M}_L^*\overline{\mathcal{M}_R}\right)\right]}{|\mathcal{M}_R|^2 + |\mathcal{M}_L|^2}. \tag{1.8}$$

The form of the two terms in the numerator indicates that interference can occur only among final states with the same photon helicity.

1.3.2 *CP* violation in $b \to s\gamma$ transitions

Using the operator product expansion, we can write the leading contributions of the effective Hamiltonian for $b \to s\gamma$ transitions as

$$\mathcal{H}_{\text{eff}} \simeq -\frac{4G_F}{\sqrt{2}} V_{ts}^* V_{tb}\left[C_7 O_7 + C_7' O_7' + C_2 O_2^c\right],$$

where

$$O_7 = \frac{e}{4\pi^2}\bar{s}\sigma_{\mu\nu}\left(\frac{1+\gamma^5}{2}\right)bF^{\mu\nu} \text{ and} \tag{1.9}$$

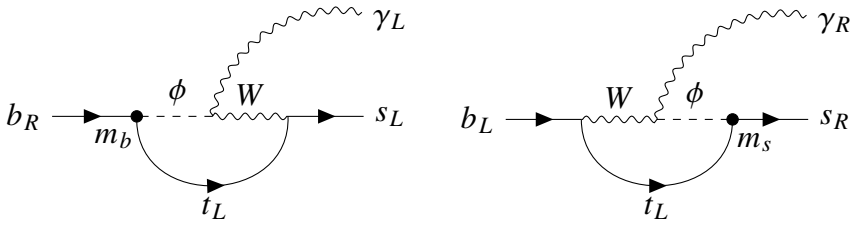
$$O_7' = \frac{e}{4\pi^2}\bar{s}\sigma_{\mu\nu}\left(\frac{1-\gamma^5}{2}\right)bF^{\mu\nu}, \tag{1.10}$$

representing the short-distance (SD) photon emission, and

$$O_2^c = \frac{V_{cs}^* V_{cb}}{V_{ts}^* V_{tb}}\left[\bar{s}\gamma_\mu(1-\gamma_5)c\right]\left[\bar{c}\gamma^\mu(1-\gamma_5)b\right], \tag{1.11}$$

contributing to the long-distance (LD) photon emission through the $b \to s\gamma g$ loop diagram shown in Fig. 1.3.

The chirality flip of $b \to s$ in O_7 and O_7' reflects the chirality structure of the leading contribution in the SM, as shown in Fig. 1.4 [7]. The angular momentum conservation



(a) $b_R \to s_L\gamma_L$ (b) $b_L \to s_R\gamma_R$

Figure 1.4: Feynman diagrams for $b \to s\gamma$ transition at the leading order in the SM (R_ξ gauge).

in the two-body decay requires the same helicity for s and γ in the b rest frame. Thus, only the helicity structures of $b_R \to s_L\gamma_L$ and $b_L \to s_R\gamma_R$ are allowed. Moreover, since the W boson only interacts with the left-handed current, the mass term, or equivalently, the interaction with the Goldstone boson (ϕ in Fig. 1.4), corresponds one-to-one to the helicity. Therefore, in the SM,

$$\frac{C_7'}{C_7} \simeq \frac{m_s}{m_b} \simeq 0.02. \tag{1.12}$$

Considering only the SD contribution, $\overline{\mathcal{M}_{R(L)}}$ and $\mathcal{M}_{R(L)}$ are proportional to $C_7^{(\prime)}$ and $C_7^{(\prime)*}$. We can further simplify Eq. (1.8) as

$$S_{CP} \simeq \frac{2\xi_{CP}\,\mathrm{Im}\left[e^{-2i\phi_1}C_7C_7'\right]}{|C_7|^2 + \left|C_7'\right|^2}. \tag{1.13}$$

We can evaluate the suppression of S_{CP} from Eq. (1.12) as

$$S_{CP} \simeq -2\xi_{CP}\frac{m_s}{m_b}\sin 2\phi_1 \sim -\xi_{CP}0.03,$$

known as AGS oscillation [9, 10].

On the other hand, the LD contribution is free from the helicity suppression we observe in SD. The evaluation of this contribution suffers from considerable uncertainty in the hadronic matrix elements. From naive dimensional analysis using soft-collinear effective theory, this contribution is suppressed by only $\Lambda_{\mathrm{QCD}}/m_b$ and could be as large as $O(0.1)$ [11, 12]. In the case of $B^0 \to K^{*0}\gamma$, the O_2^c contribution is explicitly calculated to be $O(0.01)$ using QCD sum rules on the light cone, revealing its suppression factor as $(\Lambda_{\mathrm{QCD}}/m_b)^2$ [13–15]. The total S_{CP} for $K^{*0}\gamma$ is evaluated in Ref. [15] as

$$S_{CP}(K^{*0}\gamma) = (-2.9_{\mathrm{SD}} + 0.6_{\mathrm{LD}} \pm 1.6) \times 10^{-2}.$$

The same thing is also calculated using perturbative QCD [16],

$$S_{CP}(K^{*0}\gamma) = (-3.5 \pm 1.7) \times 10^{-2},$$

and both agree with each other within uncertainty. On the other hand, the estimation of the total S_{CP} for inclusive $B^0 \to K_S^0\pi^0\gamma$ decays is not precise, suffering from potentially large LD contribution of $O(g_s/4\pi)$ as calculated in the inclusive $B \to X_s\gamma$ decay [11], sizeable uncertainty in the hadronic matrix elements, and the interference between the resonances, including resonances with different spin and parity. The $M_{K\pi}$ dependent S_{CP} for inclusive $K_S^0\pi^0\gamma$ is estimated in Ref. [12] as

$$\left|S_{CP}(K_S^0\pi^0\gamma)\right| \leq 0.08.$$

Contrary to S_{CP}, the precise estimate on A_{CP} is not provided even for exclusive $K^{*0}\gamma$, either in Refs. [15] or [16]. One difficulty is evaluating the uncertainty of the strong phases for $\mathcal{M}_L$, which affects A_{CP} as we see in Section 1.2.2. Another reason is that the calculation is not motivated because S_{CP} is more sensitive to BSM physics. In the rough estimation, we approximate $A_{CP} \simeq 0$ by ignoring the LD contribution, where $V_{ts}^*V_{tb}$ is the only phase contributing to the amplitude.

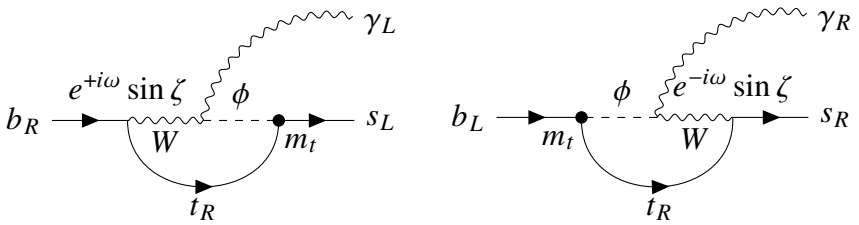
1.4 Expected effects of physics beyond the Standard Model

As we discuss in Section 1.3, S_{CP} is suppressed by m_s/m_b in the SM. The important thing is that this suppression arises from the helicity suppression in SD, which can easily be alleviated in many BSM scenarios. In particular, if the newly introduced particle with mass m_{new} flips its spin on an internal line, the suppression factor becomes m_{new}/m_b. For this reason, the measurement of S_{CP} in $b \to s\gamma$ decay is referred to as the "null test" of the SM. In the following discussion, we ignore the LD contributions for simplicity [17]. Note that this brings 2–10% uncertainty on the estimate of S_{CP}.

1.4.1 Left-right symmetric model

The left-right symmetric model (LRSM) [18] is one of the possible BSM models where S_{CP} can be enhanced [9]. In this model, the SM W, noted as W_1, is a mixture of W_L and W_R as

$$\begin{pmatrix} W_1^+ \\ W_2^+ \end{pmatrix} = \begin{pmatrix} \cos\zeta & e^{-i\omega}\sin\zeta \\ -\sin\zeta & e^{-i\omega}\cos\zeta \end{pmatrix} \begin{pmatrix} W_L^+ \\ W_R^+ \end{pmatrix},$$



(a) $b_R \to s_L\gamma_L$ (b) $b_L \to s_R\gamma_R$

Figure 1.5: Feynman diagrams contributing to $b \to s\gamma$ in the LRSM.

where W_L (W_R) couples to the left-handed (right-handed) fermion current. The additional contributions to C_7 and C_7' are shown in Fig. 1.5. The contribution from W_R enhances the C_7' through the diagram shown in Fig. 1.5(b), but it does not enhance C_7 (Fig. 1.5(a)) due to the small value of $\sin\zeta$ [19]. The effect is calculated in Ref. [19] as

$$\frac{C_7'}{C_7} \simeq 1017\frac{g_R^2}{g_L^2}\left(\frac{80.38\,\mathrm{GeV}/c^2}{M_{W_2}}\right)^2 V_{ts}^{R*}e^{-i\omega}, \tag{1.14}$$

where M_{W_2} is the mass of W_2, g_R and g_L are the coupling constants for W_R and W_L, respectively, and V_{ts}^R is the CKM matrix element for W_R. Even within the limitation of $M_{W_2} \gtrsim 5\,\mathrm{TeV}$ from the direct search result at the ATLAS experiment [20], S_{CP} can be ± 0.16 as depicted in Fig. 1(a) of Ref. [19].

1.4.2 Minimal supersymmetric standard model

One of the most attractive expansions of the SM is the model respecting supersymmetry, which transforms fermions to bosons and vice versa. In particular, the Minimal Supersymmetric SM (MSSM) [21] is the smallest expansion of the SM, introducing another Higgs doublet and the supersymmetric partners of all the SM particles. In the MSSM, the naturalness and the hierarchy problem are solved by protecting the divergence of the Higgs mass. It also provides the dark matter candidate and helps the grand unification of the interactions.

In the MSSM, the squark propagator can change its flavor and chirality and significantly contribute to C_7' [17]. With the mass insertion approximation, the chirality-flipping $b \to s$ transition term, $\left(\delta_{RL}^d\right)_{23}$, represents the dominant contribution to C_7' through the diagram shown in Fig. 1.6. Considering only the contribution from $\left(\delta_{RL}^d\right)_{23}$

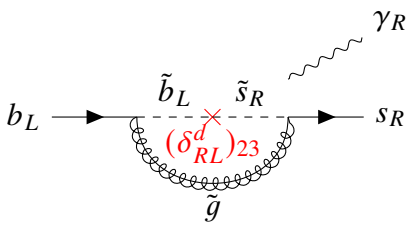


Figure 1.6: Feynman diagrams contributing to $b_L \to s_R\gamma_R$ in the MSSM.

and assuming $C_7 \simeq C_7^{\rm SM}$, $C_7' \simeq C_7'^{\rm MSSM}$, and $m_{\tilde{g}} \simeq m_{\tilde{q}}$, the effect can be calculated as

$$\frac{C_7'}{C_7} \simeq \frac{-66.39\,\text{TeV}/c^2}{m_{\tilde{q}}}\left(\delta^d_{RL}\right)_{23}. \tag{1.15}$$

The largest possible MSSM contribution within the current theoretical and experimental constraints is evaluated in the parameter scan [22] as

$$C_7'^{\rm MSSM} \sim -0.066 - 0.0042i.$$

Using Eq. (1.13) with $C_7 \simeq C_7^{\rm SM} \simeq -0.32$ (e.g., in Ref. [19]) and $C_7' \simeq C_7'^{\rm MSSM}$, the largest possible S_{CP} is obtained to be

$$S_{CP} \sim -0.27\xi_{CP}.$$

This is 10 times larger than the SM expectation.

1.5 Measurement of S_{CP} and A_{CP} at B-factory

In the B-factory, a pair of B mesons is produced from

$$e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0, \tag{1.16}$$

which is suitable for studying the time-dependent CP violation [4–6]. The $B^0\overline{B}^0$ pair generated in the above process is in S-wave and quantum-entangled so that one B meson is B^0 ($\overline{B}^0$) and the other should be $\overline{B}^0$ (B^0) at the same time. Let us assume one neutral B meson, $B_{\rm sig}$, is decaying into our signal, $K^0_S\pi^0\gamma$. By specifying the other neutral B meson, $B_{\rm tag}$, as either B^0 or $\overline{B}^0$ from its decay products, we can determine the time-evolution of $B_{\rm sig}$ from the time when $B_{\rm tag}$ decays. The decay-time dependence can be measured by boosting $\Upsilon(4S)$ in the asymmetric energy collision. Experiments to study $B\overline{B}$ property by generating $B\overline{B}$ pairs through the above process are referred to as B-factory experiments.

1.5.1 Decay-time dependent *CP* violation measurement

We briefly discuss the time evolution of $B_{\rm sig}$ and $B_{\rm tag}$ and the extraction of time-dependent *CP* asymmetry.

The $B_{\rm sig}$ and $B_{\rm tag}$ are generated in the decay of $\Upsilon(4S)$ and entangled at a time $t = 0$ as

$$\left|B(0)\right\rangle_{\rm sig} \otimes \left|B(0)\right\rangle_{\rm tag} = \frac{1}{\sqrt{2}}\Big(\left|B^0\right\rangle_{\rm sig} \otimes \left|\bar{B}^0\right\rangle_{\rm tag} - \left|\bar{B}^0\right\rangle_{\rm sig} \otimes \left|B^0\right\rangle_{\rm tag}\Big).$$

The time evolution can be written as

$$\begin{aligned}\left|B(\tau_{\rm sig})\right\rangle_{\rm sig} \otimes \left|B(\tau_{\rm tag})\right\rangle_{\rm tag} &= \frac{1}{\sqrt{2}}\Big(\left|B^0(\tau_{\rm sig})\right\rangle_{\rm sig} \otimes \left|\bar{B}^0(\tau_{\rm tag})\right\rangle_{\rm tag} - \left|\bar{B}^0(\tau_{\rm sig})\right\rangle_{\rm sig} \otimes \left|B^0(\tau_{\rm tag})\right\rangle_{\rm tag}\Big) \\ &\simeq \frac{1}{\sqrt{2}} e^{-iMT-\frac{1}{2}\Gamma T}\Big(- i \sin\frac{1}{2}\Delta m_d \Delta\tau \Big[\frac{p}{q}\left|B^0\right\rangle_{\rm sig} \otimes \left|B^0\right\rangle_{\rm tag} - \frac{q}{p}\left|\bar{B}^0\right\rangle_{\rm sig} \otimes \left|\bar{B}^0\right\rangle_{\rm tag}\Big] \\ &\qquad + \cos\frac{1}{2}\Delta m_d \Delta\tau \Big[\left|B^0\right\rangle_{\rm sig} \otimes \left|\bar{B}^0\right\rangle_{\rm tag} - \left|\bar{B}^0\right\rangle_{\rm sig} \otimes \left|B^0\right\rangle_{\rm tag}\Big]\Big),\end{aligned}$$

where $\tau_{\rm sig}$ and $\tau_{\rm tag}$ are proper times of $B_{\rm sig}$ and $B_{\rm tag}$, we define $T = \tau_{\rm sig} + \tau_{\rm tag}$ and $\Delta\tau = \tau_{\rm sig} - \tau_{\rm tag}$, and we use $\Delta\Gamma \simeq 0$. One can confirm that only the pair of B^0 and $\bar{B}^0$ are allowed at $\tau_{\rm sig} = \tau_{\rm tag}$.

Using this expression, we can consider the decays of $B\bar{B}$ pair. Let us assume that $B_{\rm sig}$ decays to a *CP* eigenstate, f_{CP}, at a time $\tau_{\rm sig}$, and $B_{\rm tag}$ decays to the flavor-specific final state, $f_\pm$, at a time $\tau_{\rm sig}$, which satisfies

$$\begin{pmatrix}\left\langle f_+\right| \\ \left\langle f_-\right|\end{pmatrix} \mathcal{H} \begin{pmatrix}\left|B^0\right\rangle & \left|\bar{B}^0\right\rangle\end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a^*\end{pmatrix}. \tag{1.17}$$

The time-dependent decay rate can be written as

$$\begin{aligned}&\Big|\Big(\left\langle f_{CP}\right|_{\rm sig} \otimes \left\langle f_\pm\right|_{\rm tag}\Big)\mathcal{H}_{\rm sig} \otimes \mathcal{H}_{\rm tag}\Big(\left|B(\tau_{\rm sig})\right\rangle_{\rm sig} \otimes \left|B(\tau_{\rm tag})\right\rangle_{\rm tag}\Big)\Big|^2 \\ &\qquad \simeq \frac{|a|^2}{4} e^{-\Gamma T}\Big(\left|\overline{\mathcal{M}}\right|^2 + \left|\mathcal{M}\right|^2\Big)(1 \pm A_{CP}\cos\Delta m_d\Delta\tau \pm S_{CP}\sin\Delta m_d\Delta\tau).\end{aligned} \tag{1.18}$$

In the case of our signal final state, $f_{CP}\gamma$, we add contributions from two *CP*-eigen final states, $f_{CP}\gamma_\pm$. We obtain a formula similar to Eq. (1.18):

$$\begin{aligned}&\sum_{h\in\pm}\Big|\Big(\left\langle f_{CP}\gamma_h\right|_{\rm sig} \otimes \left\langle f_\pm\right|_{\rm tag}\Big)\mathcal{H}_{\rm sig} \otimes \mathcal{H}_{\rm tag}\Big(\left|B(\tau_{\rm sig})\right\rangle_{\rm sig} \otimes \left|B(\tau_{\rm tag})\right\rangle_{\rm tag}\Big)\Big|^2 \\ &\simeq \frac{|a|^2}{4} e^{-\Gamma T}\Big(\left|\overline{\mathcal{M}_R}\right|^2 + \left|\overline{\mathcal{M}_L}\right|^2 + \left|\mathcal{M}_R\right|^2 + \left|\mathcal{M}_L\right|^2\Big)(1 \pm A_{CP}\cos\Delta m_d\Delta\tau \pm S_{CP}\sin\Delta m_d\Delta\tau).\end{aligned} \tag{1.19}$$

Table 1.1: Previous measurements of S_{CP} and A_{CP} in $B^0 \to K^0_S\pi^0\gamma$ decay. The uncertainties are statistical and systematic.

collaboration	$M_{K\pi}$ [GeV/c^2]	S_{CP}	A_{CP}
Belle (2006) [23] $535 \times 10^6 B\overline{B}$	(0.8, 1.0)	$-0.32^{+0.36}_{-0.33} \pm 0.05$	$-0.20 \pm 0.24 \pm 0.05$
	[0, 0.8], [1.0, 1.8)	$+0.50 \pm 0.61 \pm 0.29$	$-0.20 \pm 0.37 \pm 0.13$
	[0, 1.8)	$-0.10 \pm 0.31 \pm 0.07$	$-0.20 \pm 0.20 \pm 0.06$
BaBar (2008) [24] $467 \times 10^6 B\overline{B}$	(0.8, 1.0)	$-0.03 \pm 0.29 \pm 0.03$	$+0.14 \pm 0.16 \pm 0.03$
	(1.1, 1.8)	$-0.78 \pm 0.59 \pm 0.09$	$+0.36 \pm 0.33 \pm 0.04$

In the experiment, we can measure $\Delta\tau$ from the decay vertex distance of $B_{\rm sig}$ and $B_{\rm tag}$. However, we cannot measure T with enough precision, which requires the decay vertex reconstruction of $\Upsilon(4S)$. We can obtain the PDF for $\Delta\tau$ and $q_{\rm true} = \pm 1$, corresponding to $\left\langle f_\pm \right|_{\rm tag}$, by integrating T over the range $[|\Delta\tau|, \infty)$, corresponding to $\tau_{\rm sig} \in [0, \infty) \wedge \tau_{\rm tag} \in [0, \infty)$:

$$P(\Delta\tau, q_{\rm true}) = \frac{1}{4\tau_{B^0}} e^{-\frac{|\Delta\tau|}{\tau_{B^0}}} [1 + q_{\rm true}\{A_{CP}\cos(\Delta m_d \Delta\tau) + S_{CP}\sin(\Delta m_d \Delta\tau)\}], \quad (1.20)$$

including the correct normalization. Here we replace the total decay width Γ with the lifetime, $\tau_{B^0} = 1/\Gamma$.

1.5.2 Previous measurements of S_{CP} and A_{CP} using $B^0 \to K^0_S\pi^0\gamma$

The time-dependent *CP* violation in $B^0 \to K^0_S\pi^0\gamma$ transition was measured in the preceding *B*-factory experiments, Belle [23] and BaBar [24]. The reported results are summarized in Tab. 1.1. Their uncertainties are typically around 0.3, which is still large. The precise measurement for these parameters is crucial for both testing the SM and probing BSM physics.

Chapter 2

Setup of the Belle II experiment

The Belle II experiment [25] and the SuperKEKB accelerator [26], located in KEK, Tsukuba, Japan, is the successor of the Belle experiment and the KEKB accelerator. By colliding 7 GeV e^- and 4 GeV e^+, we selectively generate $\Upsilon(4S)$ at the invariant mass of 10.58 GeV with the Lorentz boost factor of $\beta\gamma = 0.287$.

2.1 SuperKEKB

SuperKEKB is the accelerator with the world's highest instantaneous luminosity. It consists of a linear accelerator, a positron damping ring, and main storage rings as shown in Fig. 2.1. As of Jul. 2022, $L = 4.7 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ has been recorded, and we have accumulated 428 $\mathrm{fb^{-1}}$. Among them, 423 $\mathrm{fb^{-1}}$ data are taken with sufficient detector performance, containing 362 $\mathrm{fb^{-1}}$ at the $\Upsilon(4S)$ resonance, 42 $\mathrm{fb^{-1}}$ below $\Upsilon(4S)$ resonance, and 19 $\mathrm{fb^{-1}}$ above $\Upsilon(4S)$ resonance.

Under strong limits from the beam–beam interaction, the luminosity of a ring collider can be written as

$$L \propto \gamma_{\pm} \left(\frac{I_{\pm} \xi_{y\pm}}{\beta^*_{y\pm}} \right), \tag{2.1}$$

where β^*_y stands for the vertical beta function at the beam collision point (called interaction point, IP), I for beam current, ξ_y for vertical beam–beam parameter, and γ for Lorentz factor of beam particles. Indices + and − stand for positron and electron beams, respectively.

Our target is to accumulate an integrated luminosity of 50 $\mathrm{ab^{-1}}$ with peak luminosity of $6 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$. We plan to achieve this 30 times higher luminosity than the KEKB record by increasing beam currents and squeezing β^*_y to 0.3 mm. To alleviate the lumi-

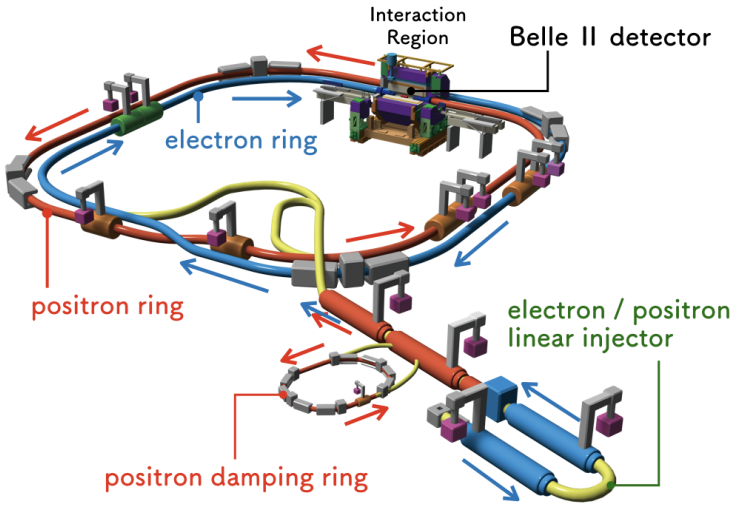


Figure 2.1: A composition of SuperKEKB [27].

nosity decrease by the hourglass effect, we adopt a large crossing angle of 83 mrad. We name this nano-beam scheme after the expected vertical beam size of $\sigma_y \simeq 50\,\text{nm}$. As of Jul. 2022, we achieved $\beta_y^* = 0.8\,\text{mm}$.

Off-orbit beam particles and radiated photons can scatter in the materials of the detector or accelerator. Particles from those secondary scatterings can intrude into the detection volume to induce the background signals in the detector, other than the signals of particles from the e^+e^- collisions we are interested in measuring. Such beam-related background hits are referred to as beam background [28]. The high luminosity of SuperKEKB leads to many beam background hits, contaminating the physics events such as photon detection or trajectory finding.

2.2 The Belle II detector and data acquisition system

The Belle II detector is a multi-purpose detector consisting of sub-detectors such as trackers, particle identifiers, an electromagnetic calorimeter, and a 1.5 T superconducting solenoid coil with a magnetic return yoke. The detector is hermetic, covering a solid angle of 3.6π. A drawing for the detector is shown in Fig. 2.2.

The coordinate system is defined as follows: z-axis is along the Belle II solenoid axis, written in the black-dashed line, directed towards the forward side (right-hand side of Fig. 2.2); x-axis is horizontal and perpendicular to the z-axis and directed towards the

right in the figure; y-axis is vertical and directed out of the paper to the top of the Belle II detector. The cylindrical coordinate (d, φ, z)[1] and the spherical coordinate (r, θ, φ)[2] are also used.

2.2.1 Detectors to measure decay vertex position

We locate the vertex detector (VXD) at the radial position of 1.4 cm, which enables us to measure the production point of a charged particle with a resolution of 14 μm [29]. The requirements for VXD are fine position resolution, a small material budget, and radiation hardness. The beam pipe is also made of beryllium to suppress the material budget. We place the silicon pixel detectors (PXD) [30] for the inner two layers (L1 and L2) to achieve a fine position resolution, and the silicon strip detectors (SVD) [31] for the outer four layers (L3 to L6) covering a total area of 1.2 m^2 (drawn in light-blue and pink in Fig. 2.2). The 3D models of VXD and the layout of the sensors are shown in Figs. 2.3 and 2.4.

(1) Pixel detector

PXD is decisive for the decay vertex position measurement. The pixel size is $50 \times (50\text{–}85)$ μm^2 depending on the pixel position. Only two L2 sensor ladders are installed out of the designed 12 until Jul. 2022. The sensors are fabricated based on the depleted field-effective transistor (DEPFET) technology [32], which allows us to make them as thin as 50 μm thick. The hits of charged particles are reconstructed by combining the neighboring pixels with S/N > 3. The data size of PXD is reduced by sending hits only in the region of interest (ROI) decided by the track position reconstructed with SVD and other detectors.

[1] $x = d\cos\varphi,\ y = d\sin\varphi$

[2] $x = r\sin\theta\cos\varphi,\ y = r\sin\theta\sin\varphi,\ z = r\cos\theta$

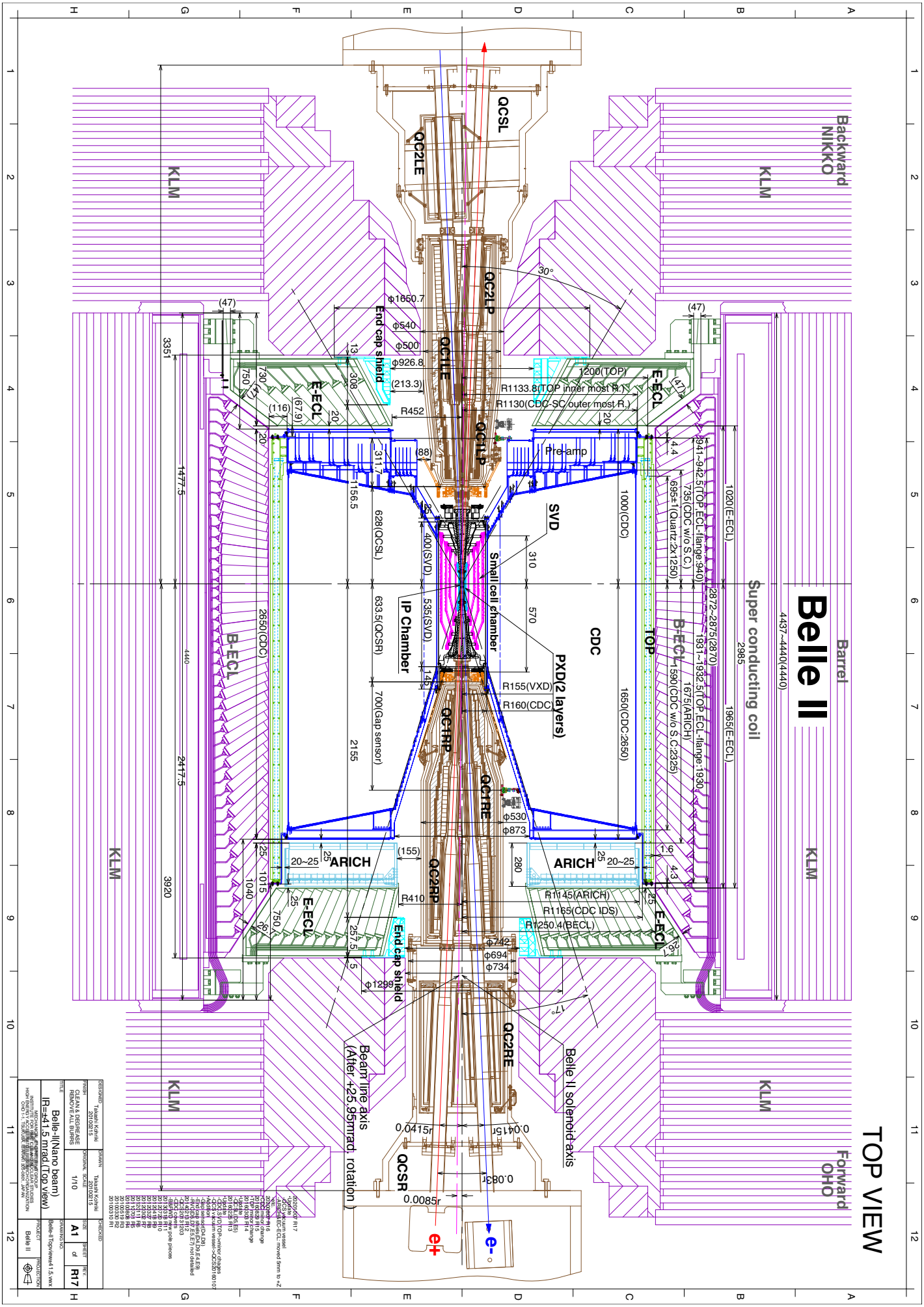


Figure 2.2: A drawing of the Belle II detector in the longitudinal section (top view) [33].

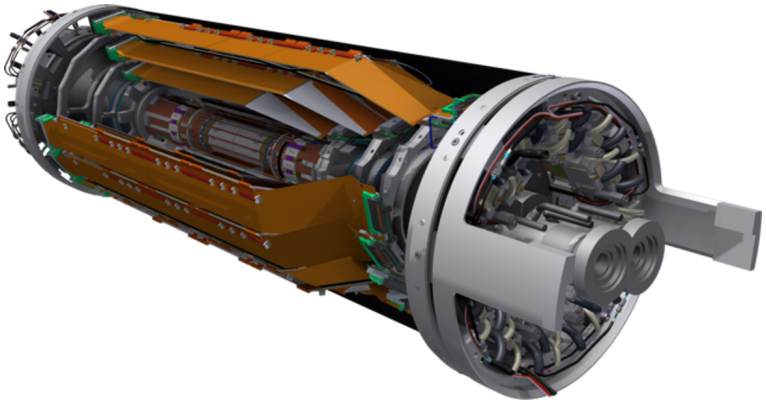


Figure 2.3: 3D model of VXD [31].

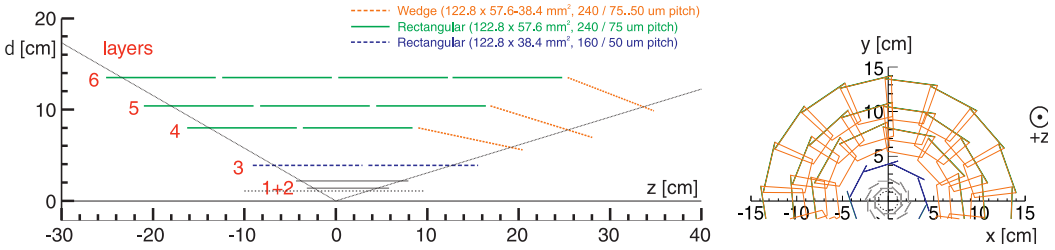


Figure 2.4: Schematic views of VXD in d–z cross-section (left) and x–y (or d–φ) cross-section (right) [31]. Inner two layers (grey) for PXD and outer four layers for SVD: blue-dashed for L3 small rectangular sensors, green for L4–6 large rectangular sensors, orange-dashed for L4–6 trapezoidal sensors.

(2) Silicon vertex detector

SVD covers cylindrical space of $4\,\text{cm} < d < 15\,\text{cm}$ from the IP. The essential roles of SVD are the ROI decision for PXD by reconstructing trajectories of charged particles with high position resolution, the reconstruction of particles with low momenta, and the decay vertex measurement of long-lived particles like K^0_S. The geometry is shown in Fig. 2.4.

SVD comprises four layers of double-sided silicon strip detectors (DSSDs). The layers are named L3 to L6, with L3 being the innermost layer and L6 being the outermost layer. The material budget is about 0.7% of the radiation length per layer. We define a local Cartesian coordinate on each DSSD plane with u- and v-axis corresponding to d–φ and z in the cylindrical coordinate. There are three types of sensors used in SVD: small rectangular sensors in L3, large rectangular sensors in the barrel region of L4, L5, and L6, and trapezoidal sensors installed slantwise in the forward region of L4, L5, and L6.

	Small	Large	Trapezoidal
No. of u-strips	768	768	768
u-strip pitch	50 μm	75 μm	50–75 μm
No. of v-strips	768	512	512
v-strip pitch	160 μm	240 μm	240 μm
Thickness	320 μm	320 μm	300 μm

Table 2.1: The main characteristics of the three types of DSSDs. Only readout strips are considered for the number of strips and strip pitch. All sensors have one intermediate floating strip between two readout strips.

The main characteristics of these sensors are summarized in Tab. 2.1.

On each side of DSSDs, we combine the neighboring strips with S/N $>$ 3 to form the hits of the charged particles with the position, charge deposit, and hit time. Then we combine the u-side and v-side hits to reconstruct the 3D position.

2.2.2 Detectors to measure energy and momentum

The energies and momenta of decay products are essential information to reconstruct particle decay, e.g., the kinematic fit of the decay vertex. The momenta of charged stable particles such as $\pi^{\pm}$ are measurable from the curvatures of trajectories in the magnetic field. Such trajectories are called tracks, parametrized by five helix parameters at the point of the closest approach (POCA) to the origin of the coordinate:

- d_0: d of POCA in the cylindrical coordinate with the sign assigned from the direction of the momentum,
- ϕ_0: the angle between the momentum projected in x–y (or d–φ) plane and x-axis.
- ω: curvature in x–y plane with the sign assigned by the charge of the particle,
- z_0: z of POCA, and
- $\tan\lambda$: the slope of the track on the $d\varphi$–z cylindrical plane.

The energies of γ and e are measurable from the intensity of photons emitted in the electromagnetic shower. We locate the central drift chamber (CDC, drawn in blue in Fig. 2.2) just outside VXD to detect and measure the tracks, and the electromagnetic calorimeter (ECL, drawn in purple (dark green) for the barrel (endcap) region noted as B-ECL (E-ECL) in Fig. 2.2) outside CDC to detect and measure the electromagnetic shower.

(1) Central drift chamber

CDC is a cylindrical drift chamber covering a radius of 16.80 cm $< d <$ 111.14 cm. The primary purposes of CDC are track finding, momentum measurement, and decision of event triggering for all charged particles. CDC contains 56 layers of sense wires in nine superlayers: the innermost superlayer contains eight layers of sense wires with smaller azimuthal gaps to reduce the hit rate and occupancy; the other superlayers have six layers. The wires are strained either along the z-axis (axial wires) or with a twisted angle (stereo wires), alternatingly for superlayers. We combine the information from axial and stereo wires in the offline analysis to reconstruct 3D trajectories. After the track finding in CDC, the combinatorial Kalman filter adds the 3D hit position in VXD to the tracks [34]. A found track is then fitted with the CDC and VXD hits considering the detector material effects starting from the IP.

(2) Electromagnetic calorimeter

ECL is a total absorption electromagnetic calorimeter covering $12.4° < \theta < 155.1°$ except for two 1°-wide gaps between the barrel and endcaps. ECL measures the energy and solid angle of γ and e with 8,736 CsI(Tl) crystals (6,624 in the barrel region and 2,112 in endcap regions). The crystal length is 30 cm, corresponding to 16.1 radiation length, and allows some shower particles to penetrate the crystal (shower leakage). The statistical fluctuation of the shower leakage is the dominant uncertainty in the energy measurement higher than 1 GeV. The width of the crystal is 5.5×5.5 cm^2 and 1.5 times larger than the Moliere radius. The signals from neighboring crystals are merged to form an ECL cluster. A cluster of 5×5 crystals with the particle-injected crystal at the center is large enough to reconstruct the shower without leakage in the transverse direction.

2.2.3 Detectors for particle identification

Identifying particle types is also essential for correctly reconstructing the event, especially for specifying the accompanying B meson as B^0 or $\bar{B}^0$. The e and γ can be identified from the characteristic electromagnetic shower in ECL. Other charged-stable particles need dedicated detectors. The separation of K and π mass is based on the different Cherenkov radiation angles. We place the time-of-propagation (TOP) counter in the barrel region between CDC and ECL (drawn in pear-green in Fig. 2.2) and the aerogel ring-imaging Cherenkov (ARICH) detector in the forward endcap region (drawn in pale-blue in Fig. 2.2). The μ can be characterized by the penetration through iron plates in

the outermost K_L^0 and μ (KLM) detector (drawn in purple in Fig. 2.2). We also utilize $\mathrm{d}E/\mathrm{d}x$ measurements in SVD and CDC to identify low-momentum charged particles.

(1) Time-of-propagation counter

The TOP counter is a Cherenkov detector which consists of quartz as a radiator and the micro-channel plate PMTs as a sensor. By using the two-dimensional position information of $(5.3\,\mathrm{mm})^2$ anode pixels and the precise timing information with the transit time spread as small as 30 ps, we can reconstruct the folded Cherenkov ring image and hence the radiation angles to separate particles such as K and π.

(2) Aerogel ring-imaging Cherenkov detector

ARICH is also a Cherenkov detector which consists of an aerogel as a radiator and the hybrid avalanche photo-detector (HAPD) as a sensor. HAPD has a readout pixel size of $(4.9\,\mathrm{mm})^2$ and is placed 20 cm away from the aerogel, enabling us to detect the Cherenkov ring image and separate K and π.

(3) K_L^0 and μ detector

KLM consists of an alternating sandwich of 4.7 cm-think iron plates and 4.4 cm-thick active detector elements, i.e., resistive plate chambers or plastic scintillators. KLM can detect showers from hadrons such as K_L^0, which can interact with nuclei in the iron plate via the strong interaction. We can separate the signals of μ from those of hadrons with their more substantial penetration without the strong interaction. The iron plate also plays a role as a support structure of the detector and a return yoke of the magnetic field.

2.2.4 Event triggering and data acquisition systems

Some detectors, such as VXD, rely on an external trigger signal for data readout. To efficiently capture the events of interest from e^+e^- collisions within the limited data bandwidth, we distribute the global trigger signal (online or level-1 trigger) to all sub-detectors. Figure 2.5 depicts the overall architecture of the trigger distribution and data acquisition systems. The trigger decision is made on the FPGA logic, utilizing the online data from CDC, ECL, TOP, and KLM. For trigger distribution, we employ front-end timing switches (FTSWs) that not only distribute triggers to the front-end readout electronics but also to the common pipeline platform for electronics readout (COPPER)

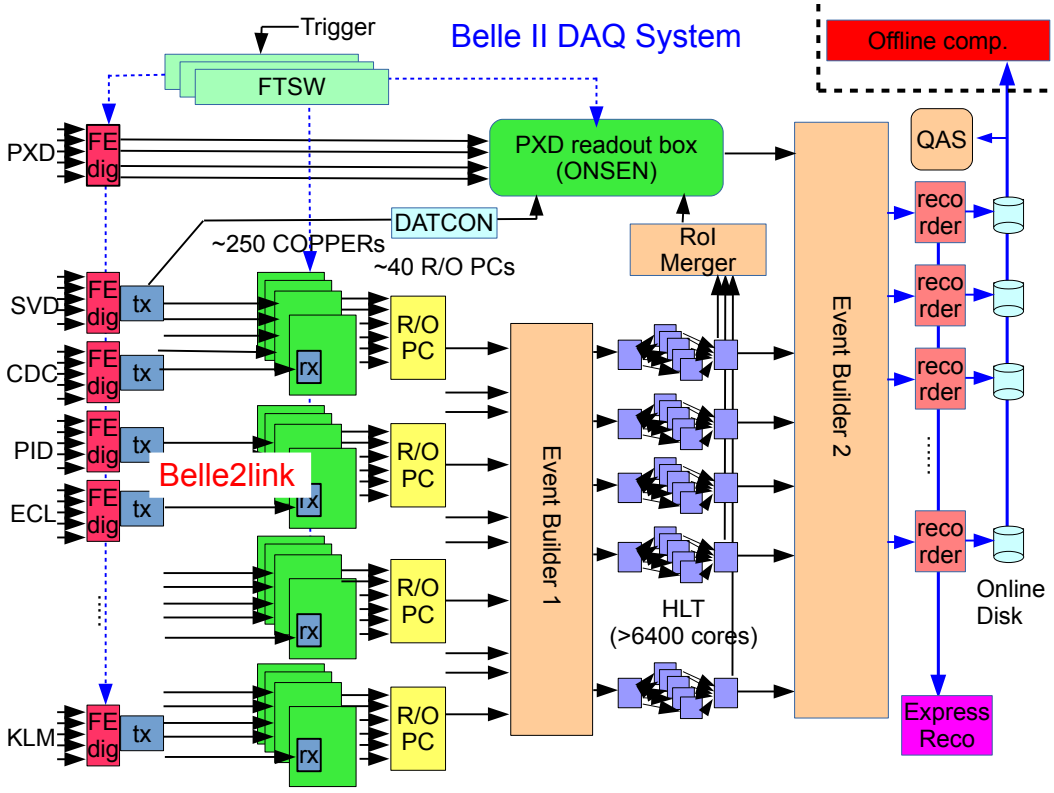


Figure 2.5: A schematic view of the Belle II DAQ system [35]. Front-end readout electronics (FE dig) for each sub-detectors are first connected to COPPER via Belle2link. The data from COPPER are first merged in readout PCs (R/O PC) and then merged in the event builders.

boards in each sub-detector. This enables FTSWs to handle trigger distribution while preventing buffer overflow in the front-end electronics and COPPER. At Event Builder 1, we merge the data from each sub-detector, excluding PXD, on an event-by-event basis. Subsequently, we apply physics-level event selections (high-level trigger, HLT) based on the entire software reconstruction of the event. The PXD data is reduced in the online selector node (ONSEN), where the created ROIs are applied. These ROIs are generated by two sources: the data concentrator (DATCON) with only SVD data and HLT using all detectors except PXD. In Event Builder 2, the reduced PXD data is added to the event, which is then stored in the online disks.

Chapter 3

Analysis strategy

As discussed in Section 1.5, to measure S_{CP} and A_{CP} in $B^0 \to K^0_S\pi^0\gamma$ using $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ events, we need to:

1. reconstruct one neutral B from the targeting final state $K^0_S\pi^0\gamma$ ($B_{\rm sig}$),
2. identify the charge sign of b quark in the other B ($B_{\rm tag}$) from its decay products referred to as flavor q, where $q = +1(-1)$ means that $B_{\rm sig}$ is $\overline{B}^0(B^0)$ at the decay time of $B_{\rm tag}$, and
3. measure the difference between the proper decay times of $B_{\rm sig}$ and $B_{\rm tag}$, referred to as proper-time difference Δt.

Figure 3.1 illustrates the above procedure. The ideal PDF of $\Delta\tau$ and $q_{\rm true}$ is given in Eq. (1.20).

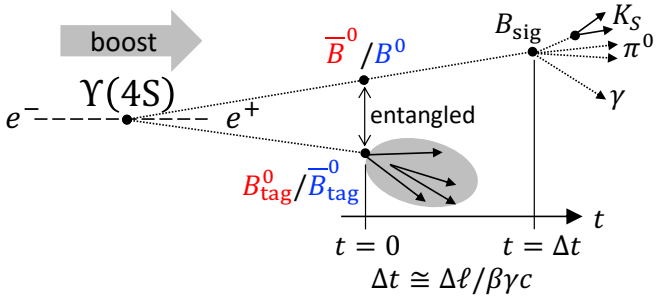


Figure 3.1: A schematic overview of time-dependent *CP* asymmetry measurement.

3.1 Reconstruction

We reconstruct $B^0 \to K^0_S \pi^0 \gamma$ decays where the K^0_S decays into $\pi^+\pi^-$ and the π^0 decays into $\gamma\gamma$.

$K^0_S \to \pi^+\pi^-$ candidates are formed by combining two oppositely-charged tracks. K^0_S decays with a significant displacement from the IP due to the long lifetime of $c\tau = 2.7\,\text{cm}$. This signature makes it easy to discriminate K^0_S from fake candidates formed by a random combination of tracks originating from the IP. We select K^0_S candidates using a binary boosted-decision-tree (BDT) classifier [36], which combines information from multiple variables nonlinearly, and the $\pi^+\pi^-$ invariant mass.

$\pi^0 \to \gamma\gamma$ candidates are formed by combining two ECL clusters with no associated tracks. We select π^0 candidates using $\gamma\gamma$ invariant mass. Since the beam backgrounds dominate low-energy ECL clusters, we also apply momentum thresholds and use a BDT classifier.

For prompt energetic γ, we select the most energetic ECL cluster in the center-of-mass (CM) frame with no associated tracks. To reject the dominant background from the asymmetric energy decays of energetic π^0 or η mesons, we evaluate π^0 or η probability using a BDT classifier.

We reconstruct B_{sig} candidates by combining K^0_S, π^0, and γ candidates. We select B_{sig} candidates using beam-energy-constrained mass M_{bc} and energy residual from the beam measurement ΔE, defined as

$$M_{\text{bc}} \coloneqq \sqrt{\left(E^*_{\text{beam}}/2\right)^2 - \mathbf{p}^{*2}_B} \quad \text{and} \quad \Delta E \coloneqq E^*_B - \left(E^*_{\text{beam}}/2\right), \tag{3.1}$$

where E^*_{beam} is the energy of the beam collision at the CM frame and E^*_B and $\mathbf{p}^*_B$ are the energy and momentum of the B_{sig} candidate in the CM frame. We classify the B_{sig} candidates with the $K^0_S\pi^0$ invariant mass: $0.8\,\text{GeV}/c^2 < M_{K\pi} < 1.0\,\text{GeV}/c^2$ for $K^*(892)^0$ resonance, referred to as mass region 1 (MR1) or $K^*(892)^0\gamma$ channel; and the rest for the inclusive decays, referred to as non-MR1 or $K^0_S\pi^0\gamma$ channel.

The dominant background comes from continuum $e^+e^- \to q\overline{q}$ events, where $q\overline{q}$ is $u\overline{u}$, $c\overline{c}$, $d\overline{d}$, $s\overline{s}$, and additionally $\tau^+\tau^-$ is considered. Since the generated particles are lighter than B-meson, they have larger back-to-back momenta in the CM frame and leave jet-like signatures in the detector. We suppress these continuum backgrounds with a BDT classifier using the observables related to the event topology.

3.2 Flavor tag

The determination of the flavor q relies on the charge of B decay products. Considering the dominant decay chain of b quarks such as $b \to c\ell^-\overline{\nu}$ ($\ell = e, \mu$) or $b \to c \to s$, negatively charged leptons and kaons indicate $\overline{B}^0$ and vice versa. We use a combination of category-based BDT classifiers provided by the software called flavor tagger [37]. The combiner BDT returns a combined single value $q \cdot r$ as a classifier between B^0 and $\overline{B}^0$, where $q = \pm 1$ is a sign to indicate the flavor and r is the reliability of the flavor assignment ranging from 0 (no flavor information) to 1 (unambiguous flavor assignment). The $q = +1(-1)$ is defined to be $B_{\text{tag}} = B^0(\overline{B}^0)$, in other words $B_{\text{sig}} = \overline{B}^0(B^0)$ at the proper decay time of B_{tag}. The tagging performance depends on r, so we split B^0 candidates into seven bins according to r, called r-bin. The bin edges are

$$(0, 0.1, 0.25, 0.45, 0.6, 0.725, 0.875, 1). \tag{3.2}$$

The numbers one to seven are assigned from the lowest to the highest r-bin.

3.3 Δt measurement and event categories

Thanks to the asymmetric energy collision, $\Upsilon(4S)$, or equivalently, the CM frame, is boosted. Since the momenta of $B^0\overline{B}^0$ are small, only 0.3 GeV/c in the CM frame, we can approximate the boost factors of B mesons with that of $\Upsilon(4S)$ and measure the proper-time difference from the distance of the two decay vertices along the boost direction:

$$\Delta t = \frac{\ell^{CP} - \ell^{\text{tag}}}{\beta\gamma c}, \tag{3.3}$$

where ℓ^{CP} and ℓ^{tag} are the decay vertex positions of B_{sig} and B_{tag} projected to the boost direction, and Δt is referred to as the measured proper-time difference.[1] We correct the bias from this approximation of the boost factor by a response function described in Section 5.2.1(1).

We measure ℓ^{CP} and ℓ^{tag} by kinematic fits (see Section 4.5). Since the only tracks used for B_{sig} vertex fit are $\pi^+\pi^-$ from K^0_S and displaced from the IP, the vertex position resolution varies from around 10 µm to 100 µm, mainly depending on the innermost VXD hit layer. This variation is considerable compared to the average flight length of

[1]The measured proper-time difference is not the difference of proper decay times but the difference of decay times in the $\Upsilon(4S)$ rest frame, as we cannot measure $\Upsilon(4S)$ decay vertices.

B^0, 130 μm. Therefore, it is crucial to accurately comprehend the vertex resolution to correctly measure the time-dependent *CP* asymmetry.

Our understanding and estimation of the vertex positions and resolutions are limited. Therefore, we introduce two event categories using the quality of the kinematic fits and adopt the following strategy to estimate S_{CP} and A_{CP}:

- We classify the reconstructed events into two categories: one is time-differential (TD) events where the good vertex measurement quality is assured; the other is the remaining events, called time-integrated (TI) events. The classification criteria are described in Section 4.5.
- We analyze both TD and TI events in the simultaneous *CP*-violation parameters fit. The TD events are modeled using Eq. (1.20). The TI events are modeled using the same PDF but integrated over $\Delta\tau$ to remove Δt dependence:

$$P(q_{\text{true}}) = \frac{1}{2}\left(1 + q_{\text{true}} A_{CP} \frac{1}{1 + x_d^2}\right), \tag{3.4}$$

  where $x_d = \Delta m_d \cdot \tau_{B^0}$.

By including the TI events in the analysis, we can improve the precision of A_{CP} measurement.

3.4 Control sample

B^+ decay provides a null test of time-dependent *CP* asymmetry measurement and is suitable for the control sample[2]. We analyze $B^+ \to K_S^0 \pi^+ \gamma$ as the control sample (channel), where we can imitate the decay vertex resolution of the signals by ignoring the π^+ track (see Appendix B.2). The branching fraction of this control sample is twice higher than the signals considering the isospin. The analysis of the control sample is summarized in Appendix B.

[2]Another merit of using B^+ decay is that one can specify B^+ or B^- from its decay products, $\pi^\pm$ in this case. However, we randomly select a single candidate in this analysis, so the charge of $\pi^\pm$ is unreliable.

Chapter 4

Event reconstruction and selection

In this chapter, we describe the reconstruction and selection criteria summarized in Section 3.1.

4.1 Experimental data and simulated samples

We analyze the e^+e^- collision data taken near the $\Upsilon(4S)$ resonance at Belle II from 2019 to summer 2022, which amounts to a total integrated luminosity of 362 fb^{-1} and $(387.5 \pm 5.8) \times 10^6$ $B\overline{B}$ pairs. We apply a loose event selection in HLT, where the events are tagged as *hadron*, namely $B\overline{B}$- or $q\overline{q}$-like. We further apply loose event selections in the offline software:

- More than one ECL cluster in the CDC acceptance with $E > 0.2\,\text{GeV}$.
- Total deposited energy in the ECL is smaller than 9 GeV.
  The energy is calculated for charged particles with $p_\text{T} > 0.1\,\text{GeV}$, $|\Delta d| < 0.5\,\text{cm}$, $|\Delta z| < 2\,\text{cm}$, and more than 20 CDC hits, and for photons in the CDC acceptance with $E > 0.1\,\text{GeV}$. Here, Δd and Δz are track's xy- and z-distance from IP.

We use Monte Carlo simulated samples to develop the analysis method. The primary simulation sample corresponds to $1\,\text{ab}^{-1}$ integrated luminosity combining the generic events of $e^+e^- \to B^0\overline{B}^0$, B^+B^-, $u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$ and $\tau^+\tau^-$. The $B^0\overline{B}^0$, without our signal, and B^+B^- combined sample is referred to as $B\overline{B}$ background sample. The $u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$, and $\tau^+\tau^-$ combined sample is referred to as the $q\overline{q}$ or continuum sample. The additional $2\,\text{ab}^{-1}$ of $e^+e^- \to B^0\overline{B}^0$ and B^+B^- is also combined with the $1\,\text{ab}^{-1}$ generic MC for modeling M_bc–ΔE and Δt in $B\overline{B}$ background. The following samples for specific decays are also generated:

- 2M events of $B^0 \to K^*(892)^0(\to K^0_S(\to \pi^+\pi^-)\pi^0)\gamma$ decays referred to as signal MC sample, used for selection optimization, $M_{\rm bc}$–ΔE modeling, and Δt resolution modeling.
- 10M events of $B^+ \to K^*(892)^+\gamma$ decays referred to as control MC sample, used for selection optimization, $M_{\rm bc}$–ΔE modeling, and Δt resolution modeling.
- 20M events of $B^0 \to (\bar{s}d)\gamma$ and $B^+ \to (\bar{s}u)\gamma$ decays referred as mixed and charged $b \to s\gamma$ MC samples, used to study non-MR1 signals and to model $B\bar{B}$ background distribution for Punzi effect correction.

For the event generation, we use EvtGen [38] for hadronic decays, KKMC [39] followed by the fragmentation by Pythia [40] for $q\overline{q}$, and Tauola [41] for τ decays. In the EvtGen, the X_s mass spectrum in $B \to X_s\gamma$ decay is generated based on the Kagan–Neubert model [42]. The detector response is simulated by Geant4 [43].

We analyze data and MC with the Belle II analysis software framework [44].

4.2 Event reconstruction and selection

We use differential evolution [45] to determine eight cut thresholds (seven daughter selections in Tab. 4.1 and one $B_{\rm sig}$ selection) at the maximum Figure of Merit (FoM). The differential evolution is a method where cut thresholds are determined by changing the cut value randomly, and a set of cut thresholds is updated when the FoM gets higher than before the change. We use significance $N_{\rm sig}/\sqrt{N_{\rm sig}+N_{\rm bkg}}$ as FoM, where $N_{\rm sig}$ and $N_{\rm bkg}$, the number of signal and background for 1 ab^{-1}-equivalent data size, are counted with the following requirements:

- in the signal-enhanced region, i.e., $5.27 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2$ and $-0.2 < \Delta E < 0.1\,{\rm GeV}$,
- in MR1, and
- with at least 1 VXD hit for each K^0_S daughter.

After the first loop of optimization, we optimize the B^0 selection in each r-bin, as summarized in Tab. 4.2

For the evaluation of FoM, we use signal MC for counting signal candidates and 1 ab^{-1} generic MC for counting background candidates. We downsize the signal MC samples to 1/10 to speed up the evaluation of FoM. This adjustment does not deteriorate the precision of the FoM, as it is primarily constrained by the number of background candidates in the generic MC.

Table 4.1: Daughter selections and optimized thresholds.

K_S^0	invariant mass	$\lvert M_{\pi^+\pi^-} - m_{K_S^0}\rvert < 34\,\mathrm{MeV}/c^2$
	V^0 probability	$P_{V^0} > 0.75$
	Λ veto	$P_\Lambda < 0.99$
π^0	invariant mass	$\lvert M_{\gamma\gamma} - 134\rvert < 30\,\mathrm{MeV}/c^2$
	momentum	$p_{\gamma\gamma} > 430\,\mathrm{MeV}/c$
γ	π^0 probability	$P_{\pi^0} < 0.83$
	η probability	$P_\eta < 0.79$

The signal candidates are regarded as truth-matched if the K_S^0 and prompt γ candidates are matched to those from the signal decay in the generator truth information. We do not require the truth-matching of π^0 since the K_S^0 provides the vertex information sufficient for the analysis.

4.2.1 Reconstruction of K_S^0

We reconstruct $K_S^0 \to \pi^+\pi^-$ decays by combining two oppositely-charged tracks. Since the $\pi^\pm$ tracks from K_S^0 originate away from the IP, we use a dedicated reconstruction tool that removes the material effect between the IP and the K_S^0 decay vertex and refits these tracks to measure their momenta correctly. We have improved this tool by expanding its functionality to remove the hits between the IP and the K_S^0 decay vertex. Since these hits are more likely to be background hits, the accuracy of the vertex reconstruction has improved with this modification. We have also tuned the selection used in the tool since we observed 20% signal inefficiency. By loosening the vertex fit χ^2 cut used in the tool and instead introducing the selection on the invariant mass, we have salvaged 10% of the signal with only a few % increase in the data size and the calculation time.

We require the $\pi^+\pi^-$ invariant mass $M_{\pi^+\pi^-}$ to be within $\pm 34\,\mathrm{MeV}/c^2$ from the PDG value, $m_{K_S^0} = 497.611\,\mathrm{MeV}/c^2$ [46]. K_S^0 decays with a significant displacement from the IP due to the long lifetime of $c\tau = 2.7\,\mathrm{cm}$. Such long-lived neutral particles decaying into two oppositely-charged particles are referred to as V^0. We utilize this feature in the BDT classifier P_{V^0} to separate K_S^0 candidates against fake candidates formed by a random combination of tracks originating from the IP, requiring $P_{V^0} > 0.75$. We veto V^0-like Λ baryon decay, $\Lambda \to p^+\pi^-$, which survives after V^0 selection, using another BDT classifier. We require loosely $P_\Lambda < 0.99$. The distributions of variables used for K_S^0 selection are shown in Fig. 4.1. We confirm that the FoM is at a plateau around the thresholds.

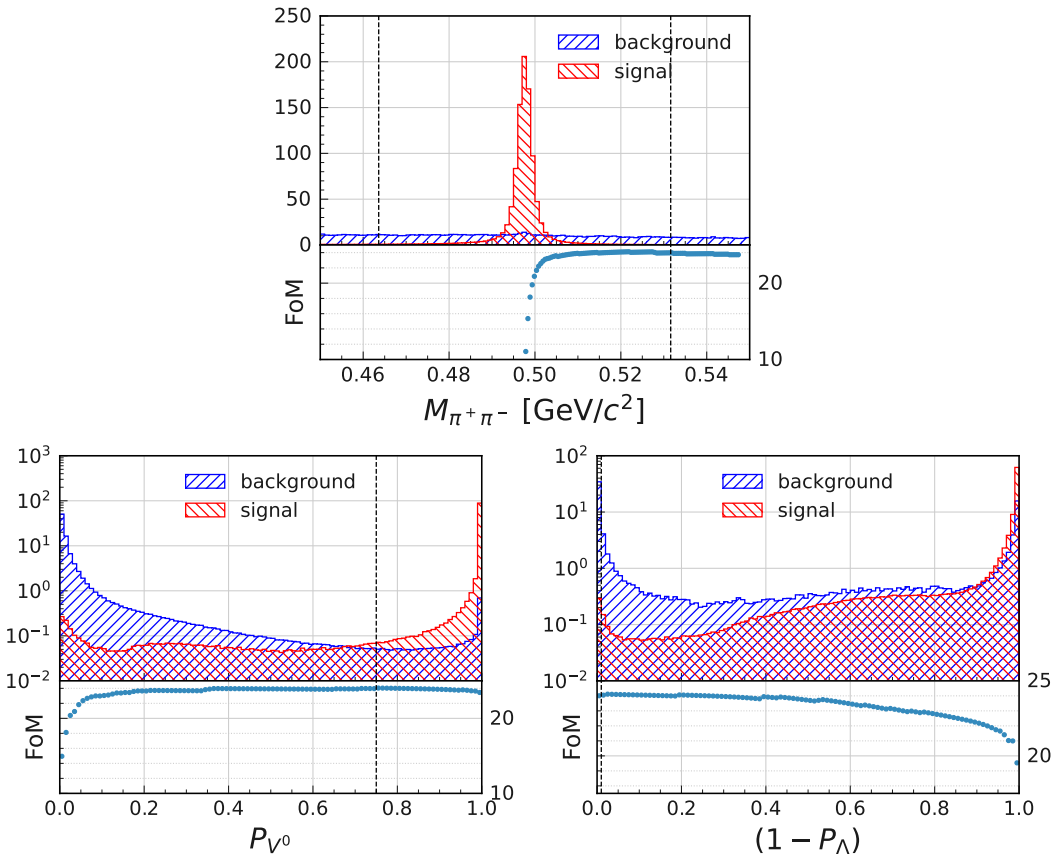


Figure 4.1: The comparison of the invariant mass, V^0 probability, and Λ veto between K_S^0 from signal B^0 decay in 2M signal MC (signal, red) and fake K_S^0 in 100 fb^{-1} generic MC (background, blue). The entries for each component are normalized to 1,000 for the invariant mass and 100 for BDT classifiers. The FoM of B^0 candidates in 1 ab^{-1} is also shown in the lower subplot. Since the selections of the invariant mass are applied on the absolute residual, FoM is shown only on the right side ($M_{\pi^+\pi^-} > 0.497611$ GeV/c^2) in the top plot. The vertical black dashed line represents the threshold we determine.

The hit position uncertainties in PXD are underestimated, and the pull widths of helix parameters are larger than 1 in MC. We correct uncertainties on track helix parameters so that the pull distributions are normalized in MC [47]. The correction factors depend on the layer of the innermost VXD hit, the track momentum, and the track angle. We also calibrate the data–MC difference of this correction using the cosmic ray analysis.

We observe slightly higher track momenta in data due to the imperfect magnetic field map. Therefore, we apply a global scaling factor of $0.99987^{+0.00038}_{-0.00057}$ to the momenta of $\pi^{\pm}$ in the experimental data, which is obtained by comparing the peak positions of the invariant mass of $D^0 \to K^-\pi^+$ decay in MC and data. The measurement uncertainty is accounted as a source of systematic uncertainty in Section 5.5.1.

4.2.2 Reconstruction of π^0

We reconstruct $\pi^0 \to \gamma\gamma$ decay by combining two ECL clusters with no associated tracks. We require the ECL clusters to have more than 1.5 ECL crystal hits[1] and energy higher than 20 MeV in the barrel, 22.5 MeV in the forward endcap, and 20 MeV in the backward endcap.

The energy distribution of the ECL clusters has a downward tail due to the shower leakage in the crystal. Even though we correct the peak position of the cluster energy, the combined invariant mass is shifted downward, as shown in Fig. 4.2. We fit around the peak with a single Gaussian and determined the peak position as 134 MeV/c^2. We select π^0 candidates with $\gamma\gamma$ invariant mass $M_{\gamma\gamma}$ in the range of 134 ± 30 MeV/c^2. We also reject dominant fake π^0 from beam background γ by requiring the momentum in the laboratory frame to be larger than 430 MeV/c. The distributions of variables used for π^0 selection are shown in Fig. 4.3. We confirm that the FoM is at a plateau around the thresholds.

For the experimental data, we apply MC/data scaling factors to the cluster energy in bins of cluster energies up to 2 GeV. The scaling factor is determined by comparing the invariant mass of $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ between data and MC. The obtained scale factors for π^0 daughter γ's are typically 1.01 with an uncertainty of 0.0005, which is accounted as a source of systematic uncertainty in Section 5.5.1.

[1] If the ECL clusters are overlapped at the crystal, the number of ECL crystal hits is weighted according to the energy splitting and returns a non-integer number.

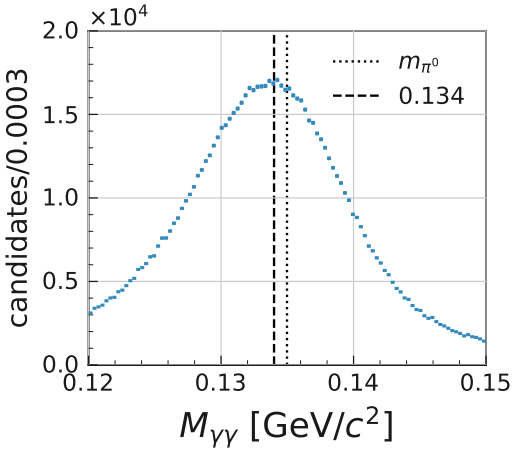


Figure 4.2: The $M_{\gamma\gamma}$ distribution of true π^0 in 2M signal MC.

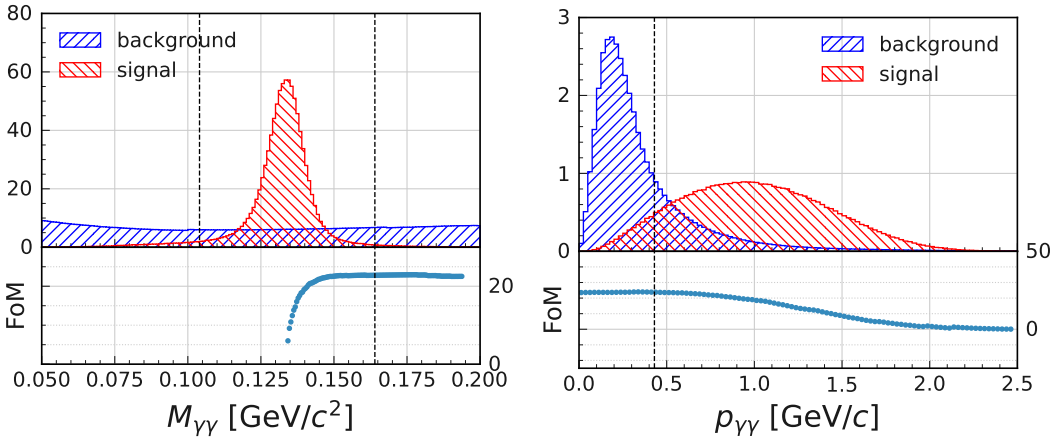


Figure 4.3: The comparison of the invariant mass and momentum between π^0 from signal B^0 decay in 2M signal MC (signal, red) and fake π^0 in 100 fb^{-1} generic MC (background, blue). The entries for each component are normalized to 1,000 for the invariant mass and 40 for momentum. The FoM of B^0 candidates in 1 ab^{-1} at each threshold is also shown in the lower subplot. Since the selections of the invariant mass are applied on the absolute residual, FoM is shown only on the right side ($M_{\gamma\gamma} > 0.134\,\mathrm{GeV}/c^2$) in the left plot. The vertical black dashed line represents the threshold we determine.

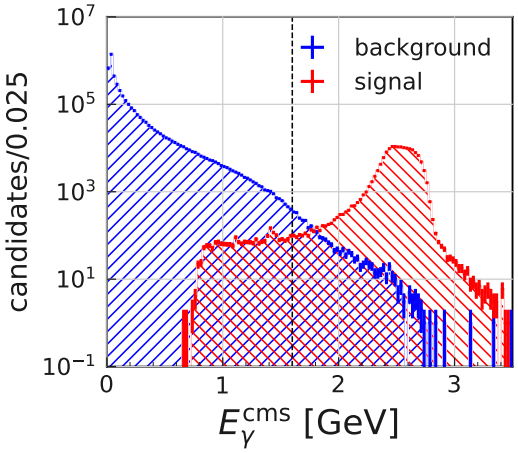


Figure 4.4: $E_\gamma^{\rm cms}$ distributions for signal and the other γ candidates from 1/11 of signal MC samples. The dashed line shows the threshold, 1.6 GeV.

4.2.3 Reconstruction of γ

We expect prompt energetic γ mainly comes from two-body decay, $B^0 \to K^*(892)^0\gamma$ or through other resonances. In such a case, the energy in the CM frame $E_\gamma^{\rm cms}$ is about $m_{B^0}c^2/2$ and significantly higher than other γ's in B decays, as shown in Fig. 4.4. Therefore, we select the most energetic ECL cluster in the CM frame with no associated tracks and require $E_\gamma^{\rm cms} > 1.6\,\text{GeV}$. Energetic γ candidates in background processes dominantly originate from the asymmetric energy decay of energetic π^0 or η mesons. To veto such γ, we combine prompt γ candidates with all the other γ candidates in the rest of the event one by one and form possible π^0 and η candidates. We use BDT classifiers as the π^0 and η probability for these candidates, and we use the highest probability for vetoing the γ candidates from π^0 or η. We require π^0 probability to be smaller than 0.83 and η probability to be smaller than 0.79. The distribution of π^0 and η probability used for the veto is shown in Fig. 4.5. We confirm that the FoM is at a plateau around the thresholds.

For the experimental data, we need to scale the cluster energy as we did for π^0. We use γ in $B^0 \to K^+\pi^-\gamma$ decays as the calibration mode. We determine the global scale factor without energy dependence by comparing ΔE between data and MC. The analysis is summarized in Appendix A. The obtained MC/data scale factor is 1.0022 ± 0.0005. The measurement uncertainty is accounted as a source of systematic uncertainty in Section 5.5.1.

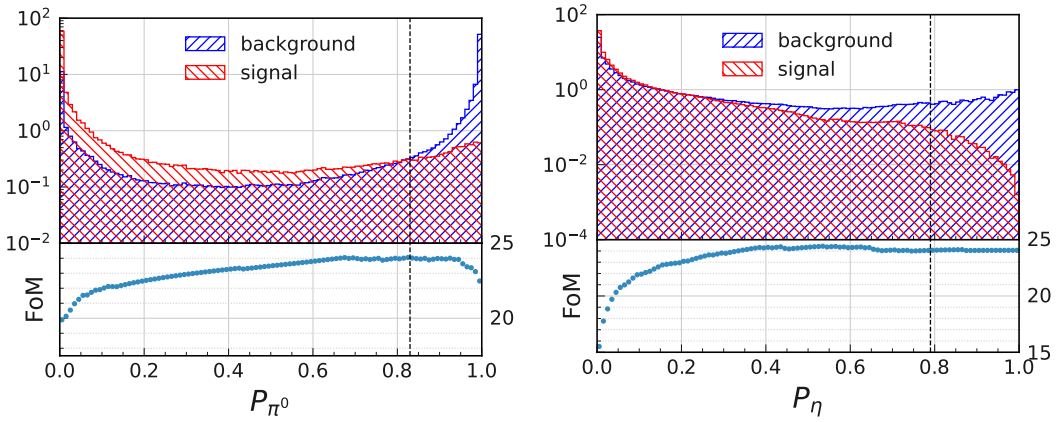


Figure 4.5: The comparison of π^0 and η probability between γ from signal B^0 decay in 2M signal MC (signal, red) and γ not from B^0 or B^+ in 100 fb^{-1} generic MC (background, blue). The entries for each component are normalized to 100. The FoM of B^0 candidates at each threshold in 1 ab^{-1} is also shown in the lower subplot. The vertical black dashed line represents the threshold we determine.

4.2.4 Reconstruction of B^0

We reconstruct $B^0 \to K^0_S\pi^0\gamma$ decays by combining K^0_S, π^0, and γ candidates. We select the B^0 candidates satisfying $5.20 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2$ and $-0.5 < \Delta E < 0.5\,{\rm GeV}$.

We also require the $K^0_S\pi^0$ invariant mass to be $0.6 < M_{K\pi} < 1.8\,{\rm GeV}/c^2$. The lower edge corresponds to the kinematical boundary, and the upper edge is required to separate D or B resonances. The $M_{K\pi}$ distribution is shown in Fig. 4.6. Restricting the B^0 candidates in MR1 also helps to mitigate the observed correlation between $M_{K\pi}$ and ΔE or r-bin.

4.2.5 Single candidate selections

We have multiple B^0 candidates in 76.7% of the events before $M_{K\pi}$ selection, with an average multiplicity of 4.07. The multiplicity from π^0 is dominant, 4.03, and mainly from fake π^0. For efficient single candidate selection, we use a BDT classifier C_{π^0} to remove fake π^0 mainly from beam background γ. We prepare the same number of true and fake π^0's from 1/11 of signal MC sample with selections[2] described in Section 4.2.2 (220,608 candidates for each). We use 80% of this sample for training and leave 20%

[2]We use the thresholds determined in the second-to-last iteration of optimization, but the difference is slight: $|M_{\gamma\gamma} - 134\,{\rm MeV}/c^2| < 27\,{\rm MeV}/c^2$ and $p_{\gamma\gamma} > 420\,{\rm MeV}/c$.

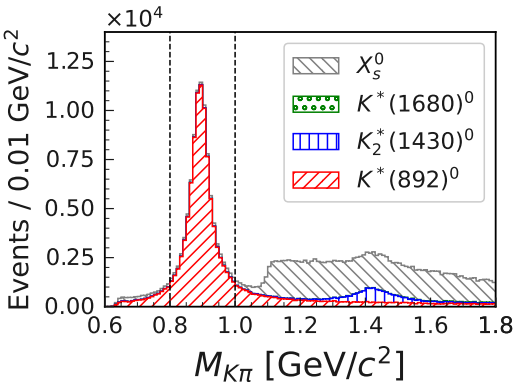


Figure 4.6: The signal $M_{K\pi}$ distribution in $b \to s\gamma$ MC sample and the visualization of mass regions. The distribution is not smooth, reflecting the X_s mass spectrum based on the Kagan–Neubert model [42].

for testing the performance. We use the following nine variables related to π^0 and its daughters for training:

- energy of both γ's,
- five properties of ECL cluster of the less energetic γ, namely, azimuthal angle, number of hits, E_1/E_9 and E_9/E_{21} (the ratios of energy deposit in 1×1, 3×3, and $5\times5-4$ (at the corners) square bundle of crystals), and MVA output for γ–K_L^0 separation using Zernike moments [48],
- helicity angle and invariant mass of π^0.

Figure 4.7 shows C_{π^0} distributions for signal and background in the BDT test sample. We select single π^0 with the highest C_{π^0} value.

If multiple K_S^0 candidates exist in the event, we choose the one with the highest V^0 probability.

We retain 87.8% of the fully-matched (π^0 truth matching required) signal candidates and 99.7% of the signal candidates defined in Section 4.2.

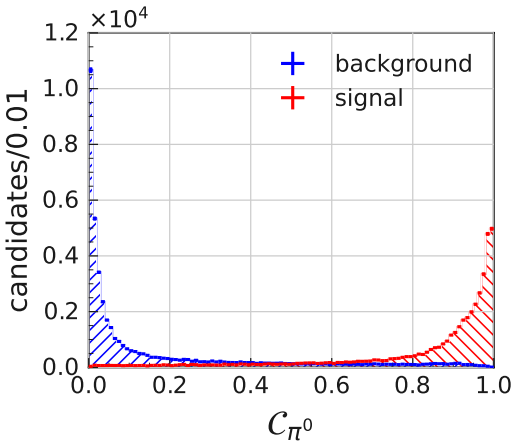


Figure 4.7: C_{π^0} distributions for signal and background in the BDT test sample.

4.3 Background suppression

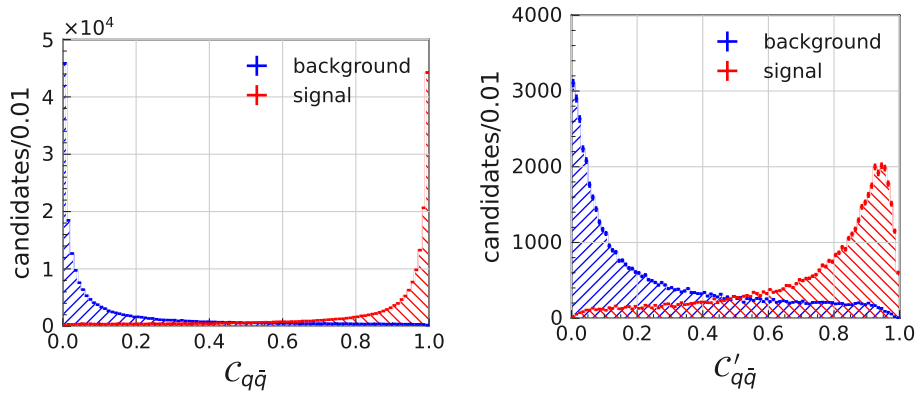
The dominant background comes from $e^+e^- \to q\overline{q}$ ($u\overline{u}$, $c\overline{c}$, $d\overline{d}$, $s\overline{s}$, and $\tau^+\tau^-$). We use a BDT classifier $C_{q\overline{q}}$ to separate the signal from $q\overline{q}$ backgrounds, which is called continuum suppression later on. We prepare the same number of signal B^0 candidates in signal MC sample and B^0 candidates from $q\overline{q}$ samples, up to single candidate selections applied[3] (962,269 candidates for each). We use 80% of this sample for training and leave 20% for testing the performance.

We define the Rest-Of-Events (ROE) using tracks with at least 1 CDC hit and $p^{\mathrm{cms}} < 3.2\,\mathrm{GeV}/c$ and ECL clusters with $p > 50\,\mathrm{MeV}/c$ and $p^{\mathrm{cms}} < 3.2\,\mathrm{GeV}/c$ for evaluation of event-shape variables. The following 29 event-shape variables are used as input:

- R_2: the ratio between the second and zeroth Fox–Wolfram moments [49],
- $\cos\theta_{\mathrm{T}}$: the cosine of the angle between the thrust axis of B_{sig} and the thrust axis of ROE [50],
- $T_O\,(T_B)$: the thrust scalar of ROE (B_{sig}) [51, 52],
- E_{T}: sum of the magnitude of transverse momenta of all the particles in the event,
- M^2_{miss}: squared missing mass of the event defined as

$$M^2_{\mathrm{miss}} = \left(E_{\Upsilon(4S)} - \sum_i E_i\right)^2 - \sum_i |\mathbf{p}_i|^2, \tag{4.1}$$

[3]We use the thresholds determined in the second-to-last iteration of optimization, but the difference is slight: $|M_{\pi^+\pi^-} - m_{K^0_S}| < 36\,\mathrm{MeV}/c^2$, $P_{V^0} > 0.76$, $P_\Lambda < 0.99$, $|M_{\gamma\gamma} - 134\,\mathrm{MeV}/c^2| < 27\,\mathrm{MeV}/c^2$, $p_{\gamma\gamma} > 420\,\mathrm{MeV}/c$, $P_{\pi^0} < 0.93$, and $P_\eta < 0.81$.



(a) $C_{q\bar{q}}$ for MR1 (b) $C'_{q\bar{q}}$ for non-MR1

Figure 4.8: $C_{q\bar{q}}$ and $C'_{q\bar{q}}$ distributions for signal and background in the BDT test sample.

Table 4.2: The lower threshold of $C_{q\bar{q}}$ ($C'_{q\bar{q}}$) in each r-bin for MR1 (non-MR1) $B_{\rm sig}$ candidates.

r-bin	1	2	3	4	5	6	7
MR1	0.93	0.92	0.86	0.86	0.91	0.87	0.65
non-MR1	0.74	0.70	0.72	0.63	0.75	0.60	0.50

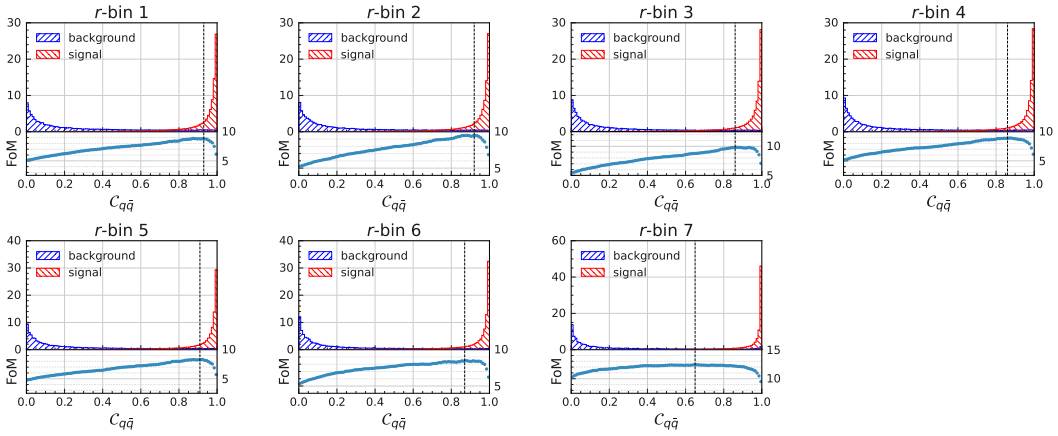
where $E_{\Upsilon(4S)}$ is the energy of $\Upsilon(4S)$ determined from the beam energies, E_i and $\mathbf{p}_i$ are the energy and momentum of a decay product, and the summation of i runs over all the decay products,

- 14 Kakuno Super Fox–Wolfram moments [53],
- 9 CLEO cones [54].

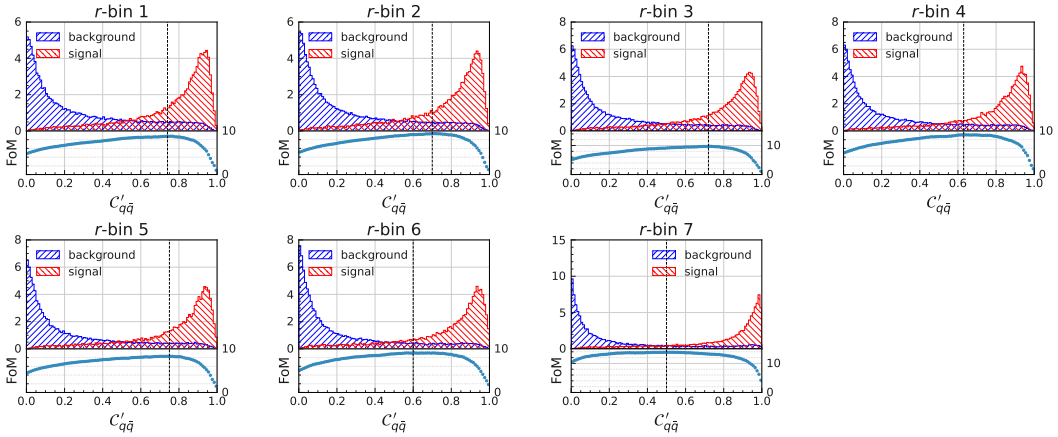
Figure 4.8(a) shows $C_{q\bar{q}}$ distributions for signal and background in the BDT test sample. One can confirm the clear separation of signal and background.

As the flavor tagger is designed for $B\overline{B}$ events, a high value of r indicates lower contamination from the $q\overline{q}$ background. The $C_{q\bar{q}}$ distribution of signal and background in each r-bin is shown in Fig. 4.9(a), and it's evident that the highest r-bin provides the best separation. Moreover, the signal-to-background ratio varies across r-bins. To address these two r-bin dependencies, we have set different thresholds for each r-bin by maximizing FoM. The results are summarized in Tab. 4.2.

We utilize signal kinematics in continuum suppression for better signal–background separation. This brings an undesirable correlation between $M_{K\pi}$ and $C_{q\bar{q}}$. After the r-bin-dependent $C_{q\bar{q}}$ selection, $M_{K\pi}$ is correlated to r-bin as well, which makes the frac-



(a) $C_{q\overline{q}}$ for MR1



(b) $C'_{q\overline{q}}$ for non-MR1

Figure 4.9: $C_{q\bar{q}}$ and $C'_{q\overline{q}}$ distributions for signal and background in each r-bin. The entries for each component are normalized to 100. The FoM in 1 ab^{-1} is also shown in the lower subplot. The vertical black dashed line represents the threshold we determine.

tion estimation in Section 5.1 complicated and possibly brings bias to the analysis. This correlation is not visible in the MR1 signal due to the limited $M_{K\pi}$ range but significant in the non-MR1 signal. To suppress this correlation, we prepare another BDT classifier $C'_{q\overline{q}}$ for non-MR1 $B_{\rm sig}$ candidates, removing R_2, T_B, and $E_{\rm T}$ from the input and using only particles in the ROE to calculate CLEO cones. Figure 4.8(b) shows $C'_{q\overline{q}}$ distributions for signal and background in the BDT test sample. The comparison of signal and background in each r-bin is shown in Fig. 4.9(b). We optimize the thresholds for the non-MR1 signal candidates in the 20M $b \to s\gamma$ MC sample. The optimized thresholds are summarized in Tab. 4.2 together with MR1. The loosened selections reflect the worse separation power and the lower signal yield.

The continuum suppression retains 77% (74%) of the signal while rejecting 95% (89%) of the background for MR1 (non-MR1) candidates.

4.4 Flavor tagging

We use a BDT classifier $q \cdot r \in [-1, 1]$ for flavor tagging. The reliability of the flavor assignment r is binned with the edges summarized in Eq. (3.2). The performance of the flavor tagger is defined as the following. We first define the wrong-tagging fraction w and the efficiency ε in a flavor-specific way:

$$w^i_\pm = P(q = \mp 1 | r \in R_i \wedge q_{\rm true} = \pm 1), \tag{4.2}$$

$$\varepsilon^i_\pm = P(r \in R_i | q_{\rm true} = \pm 1), \tag{4.3}$$

where R_i is the range of i-th r-bin. Then we define the flavor average and flavor asymmetry of the wrong-tagging fraction and the efficiency:

$$w^i = \frac{w^i_+ + w^i_-}{2}, \tag{4.4}$$

$$\Delta w^i = w^i_+ - w^i_-, \tag{4.5}$$

$$\mu^i = \frac{\varepsilon^i_+ - \varepsilon^i_-}{\varepsilon^i_+ + \varepsilon^i_-}, \tag{4.6}$$

$$\varepsilon^i = \frac{\varepsilon^i_+ + \varepsilon^i_-}{2}. \tag{4.7}$$

Using these values, we can include the effect of incorrect flavor tagging into the PDF of Eq. (1.20) as

$$\begin{aligned} P(\Delta\tau, q|r \in R_i) &= \frac{\varepsilon^i_q(1-w^i_q)P(\Delta\tau, q_{\rm true}=q) + \varepsilon^i_{-q}w^i_{-q}P(\Delta\tau, q_{\rm true}=-q)}{\sum_\pm \varepsilon^i_\pm P(q_{\rm true}=\pm 1)} \\ &= \left(1+\mu^i\frac{A_{CP}}{1+x_d^2}\right)^{-1}\frac{1}{4\tau_{B^0}}e^{-\frac{|\Delta t|}{\tau_{B^0}}}\Big[1-q\Delta w^i + q\mu^i(1-2w^i) \\ &+ \{q(1-2w^i)+\mu^i(1-q\Delta w^i)\}(A_{CP}\cos\Delta m_d\Delta t + S_{CP}\sin\Delta m_d\Delta t)\Big], \end{aligned} \tag{4.8}$$

where $P(q_{\rm true})$ and x_d is the same as Eq. (3.4). Since μ^i is close to 0, we ignore $O(\mu^i)$ terms in the PDF and obtain

$$P(\Delta\tau, q) = \frac{1}{4\tau_{B^0}}e^{-\frac{|\Delta\tau|}{\tau_{B^0}}}[1-q\Delta w+q(1-2w)(A_{CP}\cos\Delta m_d\Delta\tau+S_{CP}\sin\Delta m_d\Delta\tau)]. \tag{4.9}$$

The possible effect of the non-zero μ^i is evaluated in Section 5.5.2 as a source of systematic uncertainty.

The tagging performance in data is evaluated using the flavor-specific and large-statistics calibration channel, $B^0/\overline{B}^0 \to D^{(*)\mp}h^\pm$ decays, where $h^\pm$ represents $\pi^\pm$ or $K^\pm$ [55]. Figure 4.10 compares the performance in the calibration channel in data and MC, signal MC samples, and $B\overline{B}$ background MC samples. We confirm the agreement among these MC samples, either in MR1 or non-MR1 candidates. We can also confirm the agreement between the calibration channel data and MC. We use the performance obtained from the signal MC for MC analysis and the performance from the calibration channel data for data analysis.

We do not expect any CP asymmetry in the $q\overline{q}$ background. However, the flavor tagger may bring asymmetry. For this check, we use an $M_{\rm bc}$–ΔE sideband defined as an orange trapezoidal area shown in Fig. 4.11. This sideband is selected to be enriched in $q\overline{q}$ events while avoiding contributions from the signal and $B\overline{B}$ background. We do not apply the continuum suppression for this $q\overline{q}$-enhanced sideband. We check the CP asymmetry in sideband MC and data, as shown in Fig. 4.12. We observe no asymmetry and confirm that we can safely use the same Δt PDF for $q = \pm 1$ in $q\overline{q}$ background.

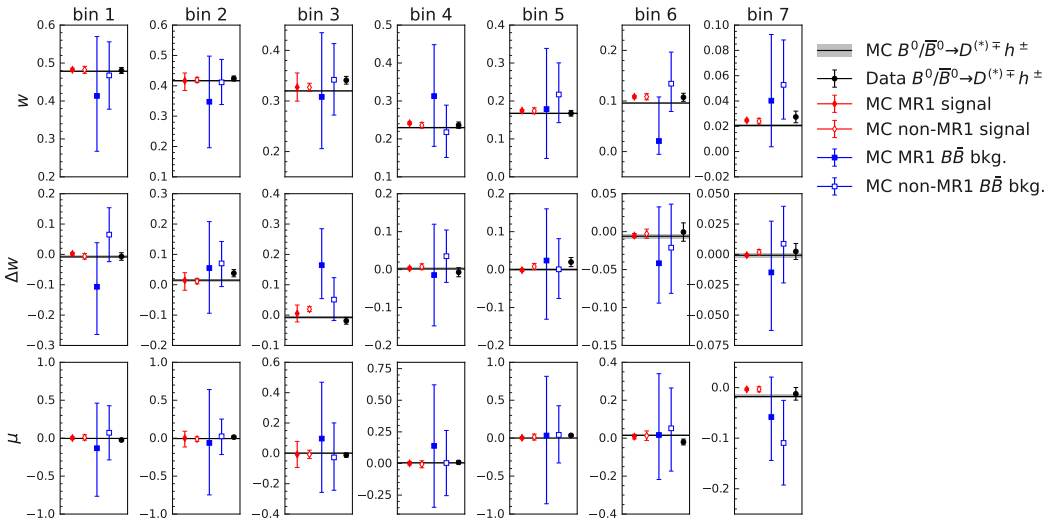


Figure 4.10: The comparison of w, Δw, and μ between $B^0 \to K^0_S \pi^0 \gamma$ in MC and $B^0/\bar{B}^0 \to D^{(*)\mp} h^\pm$ in MC and data in each r-bin.

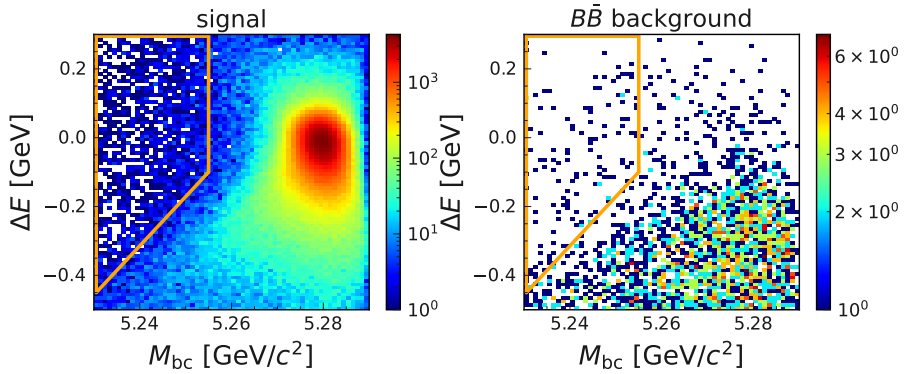
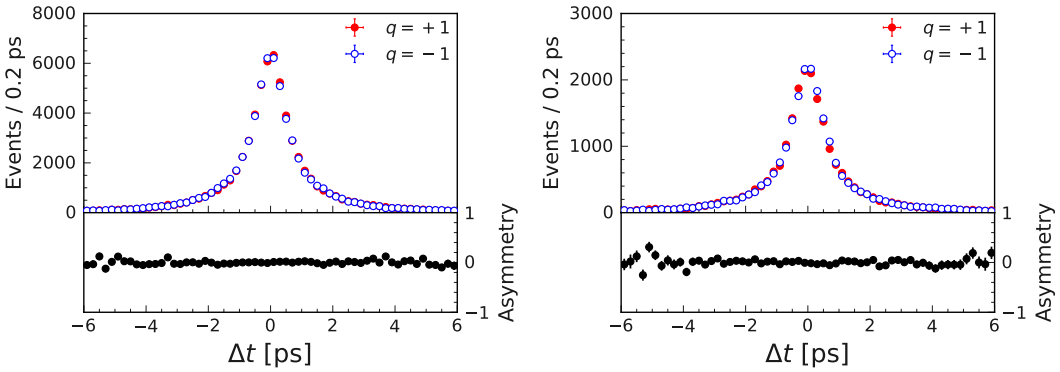


Figure 4.11: $M_{\rm bc}$–ΔE distributions of B^0 candidates in MR1. Signal candidates from signal MC (left) and background candidates from $B\bar{B}$ 3 ab^{-1} MC. The sideband is edged with orange lines.



(a) sideband MC

(b) sideband data

Figure 4.12: The flavor asymmetry of Δt distribution in sideband MC and data.

4.5 Vertex reconstruction

We measure the decay vertex positions of reconstructed B ($B_{\rm sig}$) and the other B ($B_{\rm tag}$) by kinematic fits, as illustrated in Fig. 4.13. We consider the geometry and kinematics of the entire decay chain, such as the flight of B^0 and K^0_S mesons, for the $B_{\rm sig}$ decay vertex fit [56]. We constrain $\Upsilon(4S)$ decay vertex to IP. We reconstruct $B_{\rm tag}$ decay vertex fit using the rest of the tracks in the event satisfying the following conditions:

- $\Delta d < 0.5\,$cm and $|\Delta z| < 2\,$cm,
- at least 1 PXD hit, 1 SVD hit, and 1 CDC hit, and
- $p > 50\,{\rm MeV}/c$.

We also apply the helix uncertainty correction described in Section 4.2.1 to these tracks. We constrain $B_{\rm tag}$ decay vertex to tube region extended from $\Upsilon(4S)$ decay region with $B_{\rm tag}$ momentum, which can be estimated from $E_{\Upsilon(4S)}$ and $B_{\rm sig}$ momentum. We require the fit of the decay vertices to be successful as one of the criteria.

We model the vertex resolution with the response function called resolution functions. We utilize the χ^2, the number of degrees of freedom $\nu_{\rm d.f.}$, and the uncertainty σ_ℓ of the vertex fit to construct the resolution functions. For the $B_{\rm sig}$ vertex fit, we use the fit χ^2 for the entire decay chain referred to as $\chi^2_{B^0}$ and the χ^2 calculated for the $K^0_S \to \pi^+\pi^-$ tracks referred to as $\chi^2_{K^0_S}$. For the $B_{\rm tag}$ vertex fit, we use the reduced χ^2 referred to as $(\chi^2/\nu_{\rm d.f.})^{\rm tag}$. The uncertainty is referred to as σ^{CP}_ℓ for $B_{\rm sig}$, $\sigma^{\rm tag}_\ell$ for $B_{\rm tag}$.

Based on the resolution function study with the signal MC (see Section 5.2.1), we require the following $B_{\rm sig}$ vertex quality selection, which categorizes events into TD or TI as explained in Section 3.3:

- Both π^+ and π^- from K^0_S have at least 1 VXD hit.

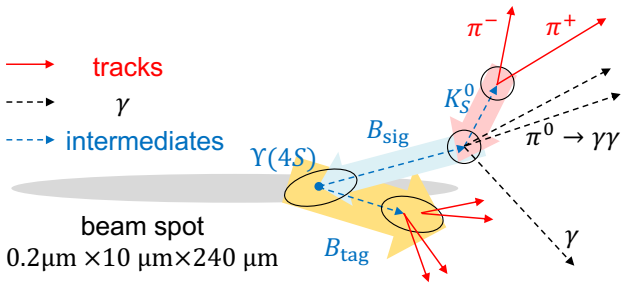
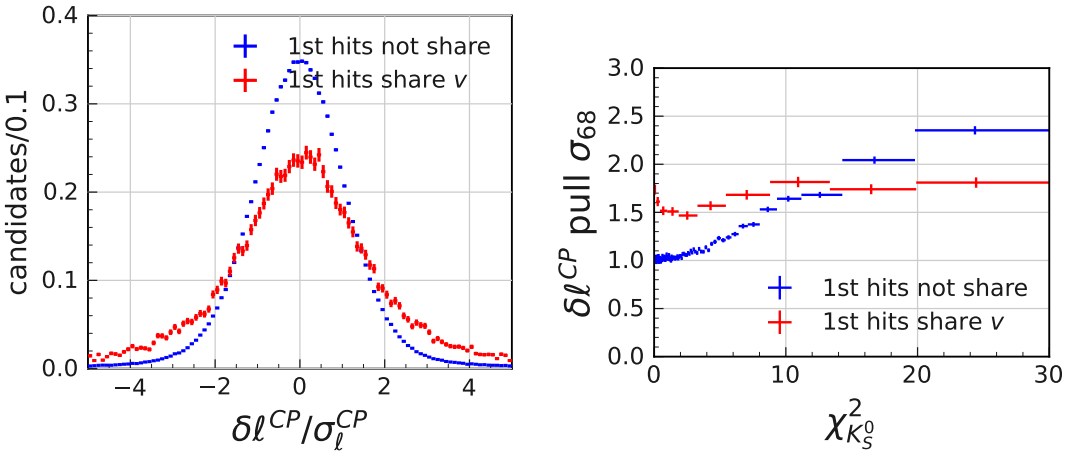


Figure 4.13: The schematic explanation for $B\bar{B}$ decay vertices reconstruction.



(a) comparison of ℓ^{CP} pulls

(b) comparison of $\chi^2_{K^0_S}$ dependence

Figure 4.14: The pulls of ℓ^{CP} and its σ_{68} (half-width between 16% and 84% quantiles) in the slice of $\chi^2_{K^0_S}$ for the cases of K^0_S daughters share their innermost hits or not.

- The innermost hits of π^+ and π^- from K^0_S do not share ν-side SVD hit.
  If K^0_S decays near the sensor, the innermost hits of daughter $\pi^\pm$ may be clusterized as a single hit and shared by the π^+ and π^- tracks. Due to the small pixel size of PXD and the narrow strip pitch on the SVD u-side, these hits are unlikely to be merged and do not bring a significant bias even if they are merged. However, due to the relatively wider strip pitch in ν-side, we observe a deterioration of the B_{sig} vertex resolution using the K^0_S with innermost ν-side hits shared, as shown in Fig. 4.14(a). The $\chi^2_{K^0_S}$ dependent resolution model cannot describe this difference as we observe different dependencies in Fig. 4.14(b). Therefore, we reject these candidates for consistent resolution modeling.
- $\chi^2_{K^0_S} < 30$ and $\chi^2_{B^0} < 100$.
  This requirement also suppresses the candidates with different behavior of resolution. The cut on the $\chi^2_{K^0_S}$ is dominant, which is also used in the resolution model, and the cut on the $\chi^2_{B^0}$ is to reject the extreme cases considering the total fit quality. We optimize the thresholds by checking the agreement of the resolution model in both χ^2 slices.
- $\sigma^{CP}_\ell < 500\,\mu\text{m}$.
  This rejects candidates with the poorest resolution and is loose enough to achieve 99% efficiency.

We require the same $B_{\rm tag}$ vertex quality as $B^0 \to K^0_S K^0_S K^0_S$ analysis [47],

- $(\chi^2/\nu_{\rm d.f.})^{\rm tag} < 100$ and
- $\sigma_\ell^{\rm tag} < 500\,\mu$m.

4.6 Selection summary

The selection criteria and efficiencies for MR1 are summarized in Tab. 4.3. The total signal efficiency in MR1 is (14.88±0.04)% for TD, (7.96±0.04)% for TI, and (22.84±0.04)% for TD and TI combined, including detector acceptance and pre-selections.

In the data, we have found 369 (449) TD events and 179 (220) TI events in MR1 (non-MR1) channel.

Table 4.3: The statistics of selections in 2M signal MC. Before single candidate selections, we count the events that include the surviving candidates. After single candidate selections, we count the truth-matched candidates. Percentages in parentheses show the efficiency of the corresponding cut.

selection	no. of signal	efficiency	
no. of events generated	2,000,000	100.0%	
after pre-selections	1,273,552	63.7%	(63.7%)
$P_{\pi^0} < 0.83$	1,177,980	58.9%	(92.5%)
$P_{\eta} < 0.79$	1,168,042	58.4%	(99.2%)
$\lvert M_{\pi^+\pi^-} - m_{K^0_S}\rvert < 34\,\mathrm{MeV}/c^2$	1,149,357	57.5%	(98.4%)
$P_{V^0} > 0.75$	1,115,458	55.8%	(97.1%)
$(1 - P_{\Lambda}) > 0.01$	1,112,348	55.6%	(99.7%)
$\lvert M_{\gamma\gamma} - 134\rvert < 30\,\mathrm{MeV}/c^2$	1,059,288	53.0%	(95.2%)
$p_{\gamma\gamma} > 430\,\mathrm{MeV}/c$	1,003,920	50.2%	(94.8%)
Single candidate selections	1,001,319	50.1%	(99.7%)
Successful B_{sig} decay vertex fit	995,486	49.8%	(99.4%)
MR1	708,478	35.4%	(71.2%)
continuum suppression	544,778	27.2%	(76.9%)
$K^*(892)^+\gamma$ veto	540,215	27.0%	(99.2%)
$5.27 < M_{\mathrm{bc}} < 5.29\,\mathrm{GeV}/c^2$	505,862	25.3%	(93.6%)
$-0.2 < \Delta E < 0.1\,\mathrm{GeV}$	456,717	22.8%	(90.3%)
at least one VXD hit	366,552	18.3%	(80.3%)
innermost v-hit not shared	348,563	17.4%	(95.1%)
$\chi^2_{K^0_S} < 30$	335,270	16.8%	(96.2%)
$\chi^2_{B^0} < 100$	335,028	16.8%	(99.9%)
$\sigma_{\ell}^{CP} < 500\,\mu\mathrm{m}$	332,114	16.6%	(99.1%)
$(\chi^2/\nu_{\mathrm{d.f.}})^{\mathrm{tag}} < 100$	298,273	14.9%	(89.8%)
$\sigma_{\ell}^{\mathrm{tag}} < 500\,\mu\mathrm{m}$	297,578	14.9%	(99.8%)
TD	297,578	14.9%	(65.2%)
TI	159,139	8.0%	(34.8%)

Chapter 5

Measurement of CP asymmetries

We model the PDF for the measured Δt and q in the TD events (see Section 4.5) by the weighted mean of the component-wise PDFs. The signal and $B\overline{B}$ background PDFs are modeled with $\Delta\tau$ and q PDF convoluted with the resolution functions, and the $q\overline{q}$ background PDF is modeled in the sideband shown in Fig. 4.11:

$$\begin{aligned} P_{\mathrm{TD}}(\Delta t, q; A_{CP}, S_{CP}) = f_{\mathrm{sig}} \int \mathrm{d}\Delta\tau P_{\mathrm{sig}}(\Delta\tau, q; A_{CP}, S_{CP}) R_{\mathrm{sig}}(\Delta t - \Delta\tau) \\ + f_{B\bar{B}} \int \mathrm{d}\Delta\tau P_{B\bar{B}}(\Delta\tau, q) R_{B\bar{B}}(\Delta t - \Delta\tau) \\ + f_{q\bar{q}} \frac{1}{2} P_{q\bar{q}}(\Delta t), \end{aligned} \tag{5.1}$$

where $P_{\mathrm{sig}}(\Delta\tau, q; A_{CP}, S_{CP})$ is $P(\Delta\tau, q)$ in Eq. (4.9). $P_{B\bar{B}}$ is the same function shape but with $S_{CP} = A_{CP} = 0$ and different lifetime parameter $\tau_{B\bar{B}}$:

$$P_{B\bar{B}}(\Delta t, q) = \frac{1}{4\tau_{B\bar{B}}} e^{-\frac{|\Delta\tau|}{\tau_{B\bar{B}}}} [1 - q\Delta w]. \tag{5.2}$$

We evaluate the possible impact of the non-zero CP violation parameters in Section 5.5.6 as a source of systematic uncertainty.

We fit only A_{CP} for the TI events using the time-integrated PDF,

$$P_{\mathrm{TI}}(q; A_{CP}) = f_{\mathrm{sig}} \frac{1}{2} \left[1 - q\Delta w + q(1 - 2w) A_{CP} \frac{1}{1 + x_d^2} \right] + f_{B\bar{B}} \frac{1 - q\Delta w}{2} + f_{q\bar{q}} \frac{1}{2}. \tag{5.3}$$

By performing the unbinned maximum likelihood fit to the measured proper-time difference Δt and the measured flavor q with

$$L(A_{CP}, S_{CP}) = \prod_i P_{\mathrm{TD}}(\Delta t_i, q_i; A_{CP}, S_{CP}) \prod_j P_{\mathrm{TI}}(q_j; A_{CP}), \tag{5.4}$$

we can extract the *CP*-violation parameters S_{CP} and A_{CP}.

5.1 Event-by-event fractions

We obtain the event-by-event fractions $f_{\rm sig}$, $f_{B\bar{B}}$, and $f_{q\bar{q}}$ in Eqs. (5.1) and (5.3) from the total fractions $F_{\rm sig}$ and $F_{B\bar{B}}$ determined in the unbinned maximum likelihood fit to the $M_{\rm bc}$–ΔE 2D distribution:

$$\begin{aligned} P(M_{\rm bc}, \Delta E; F_{\rm sig}, F_{B\bar{B}}, \chi_{q\bar{q}}, m_{q\bar{q}}) = {} & F_{\rm sig} P_{\rm sig}(M_{\rm bc}, \Delta E) \\ & + (1 - F_{\rm sig}) F_{B\bar{B}} P_{B\bar{B}}(M_{\rm bc}, \Delta E) \\ & + (1 - F_{\rm sig})(1 - F_{B\bar{B}}) P_{q\bar{q}}(M_{\rm bc}, \Delta E; \chi_{q\bar{q}}, m_{q\bar{q}}). \end{aligned} \tag{5.5}$$

For the fit, we use the MR1, non-MR1, and control $B_{\rm sig}$ candidates separately, with the selection up to $K^*(892)^+\gamma$ veto applied. The fit range is broader than the signal-enhanced region, $5.23 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2$ and $-0.4 < \Delta E < 0.3\,{\rm GeV}$. The TD and TI categories are combined since we do not observe differences in $M_{\rm bc}$ and ΔE distributions.

5.1.1 Distribution shape modeling

Both $M_{\rm bc}$ and ΔE are the kinematical observables of B mesons. Therefore, we observe peaks or bumps if the reconstructed B mesons are related to the true B mesons. Since γ energy measured in ECL has a long tail, the 2D distribution also has a tail and causes $M_{\rm bc}$–ΔE correlation. On the other hand, if the candidates are reconstructed by combining daughters from both sides of B mesons or $q\bar{q}$ background, there is no peaking structure and hence the correlation.

We model the $M_{\rm bc}$–ΔE distribution using the following two methods.

1. *Gaussian kernel density estimation*

   We use Gaussian kernel density estimation (KDE) to construct correlated two-dimensional distribution from the MC samples. The details are described in Appendix C.

2. *Factroized modeling with empirical functions*

   We can model the uncorrelated two-dimensional distribution with a product of one-dimensional distributions. We use the ARGUS function for $M_{\rm bc}$ and the 1st-

or the 2nd-order polynomial for ΔE distributions. We refer to this method as ARGUS×pol1 or ARGUS×pol2.

The ARGUS function for M_{bc} is parametrized as

$$f_{\mathrm{ARGUS}}(x = M_{\mathrm{bc}}; \chi, c = E^*_{\mathrm{beam}}/2) = \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} \cdot \frac{x}{c^2}\sqrt{1 - \frac{x^2}{c^2}} \exp\left\{-\frac{1}{2}\chi^2\left(1 - \frac{x^2}{c^2}\right)\right\}, \tag{5.6}$$

where the normalization factor $\Psi(\chi)$ can be written with the cumulative distribution and probability density function of the standard normal distribution, $\Phi(x)$ and $\phi(x)$, as

$$\Psi(\chi) = \Phi(\chi) - \chi\phi(\chi) - \frac{1}{2}, \tag{5.7}$$

the endpoint c is taken from the event-by-event beam energy measurement, and χ is the only free parameter in the fit.

The 2nd-order polynomial for ΔE is parametrized as

$$f_{\mathrm{pol2}}(x = \Delta E; m, r) = c_2(r)x^2 + c_1(m, r)x + c_0(m, r), \tag{5.8}$$

where

$$c_2(r) = -\frac{6r}{(1 + r)(H - L)^3}, \tag{5.9}$$

$$c_1(m, r) = m - (H + L)c_2(r), \text{ and} \tag{5.10}$$

$$c_0(m, r) = \frac{1}{H - L} - \frac{H^2 + HL + L^2}{3}c_2(r) - \frac{H + L}{2}c_1(m, r), \tag{5.11}$$

including the normalization in the range $[L, H]$. The free parameters m and r satisfy

$$m = \frac{f_{\mathrm{pol2}}(H) - f_{\mathrm{pol2}}(L)}{H - L}, \tag{5.12}$$

$$r = \frac{\int_L^H f_{\mathrm{pol2}}(x)\mathrm{d}x}{\int_L^H m(x - L) + f_{\mathrm{pol2}}(L)\mathrm{d}x} - 1. \tag{5.13}$$

In other words, m represents the average slope in the range, and r represents the deviation from the linear function. We can smoothly change the function to the 1st-order polynomial by setting $r = 0$, and m represents the slope.

Table 5.1: The non-MR1 signal shape parameters determined in the fit to $b \to s\gamma$ mixed MC sample. The fit uncertainties are estimated using the Hesse matrix.

shape parameters	fit results
F_{peak}	0.8704 ± 0.0019
χ_{sig}	12.96 ± 0.18
m_{sig}	-0.86 ± 0.07
r_{sig}	0.205 ± 0.034

5.1.2 Distribution of signal candidates

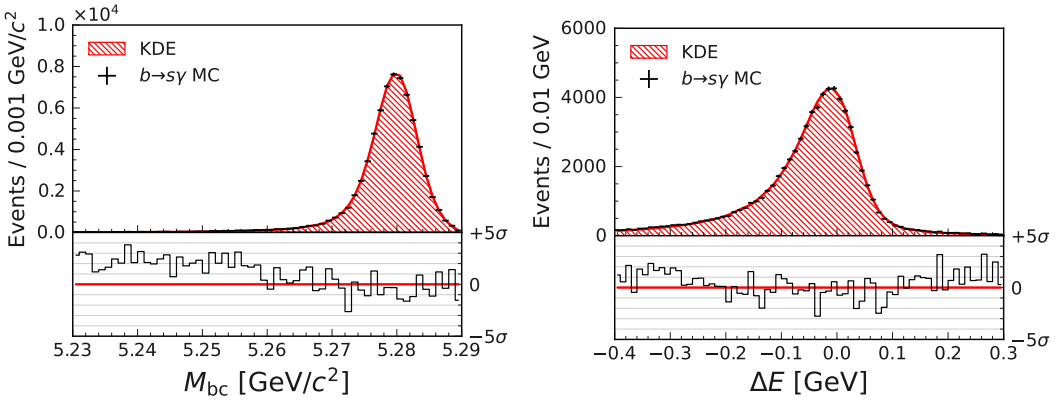
Signal candidates are peaking at $M_{\text{bc}} = m_B$ and $\Delta E \sim 0$ in the 2D distribution, with the correlated tail. We fit the KDE to truth-matched B^0 (B^+) candidates in MR1 using signal (control) MC sample. More details are described in Appendix C.

The signal B_{sig} candidates in MR1 dominantly decay via $K^*(892)$ resonance, and the signal distribution in $b \to s\gamma$ MC sample agrees with this KDE as shown in Figs. 5.1(a) and 5.1(c). Note that discrepancies between the models and the distributions are negligibly small and only visible with the extensive statistics of $b \to s\gamma$ MC sample, $O(100)$ times bigger than the data. Moreover, the discrepancy out of the signal-enhanced region can be ignored due to the statistical uncertainties from the dominant backgrounds. Indeed, the extracted F_{sig} does not significantly deviate from the truth in the statistics of $1\,\text{ab}^{-1}$, as shown in Section 5.1.5.

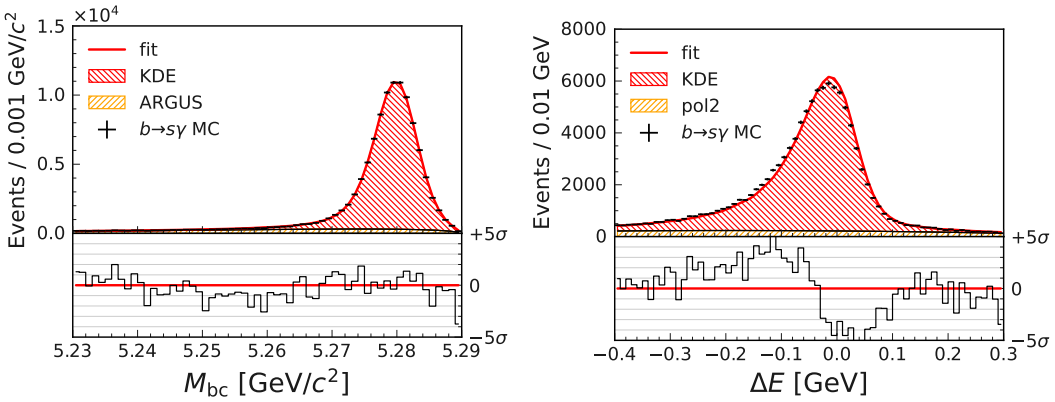
On the other hand, we need an additional ARGUS×pol2 model to explain the $M_{K\pi}$ tail of $K^*(892)^0$ for the non-MR1 signal:

$$\begin{aligned} P_{\text{sig}}(M_{\text{bc}}, \Delta E; F_{\text{peak}}, \chi_{\text{sig}}, m_{\text{sig}}, r_{\text{sig}}) = F_{\text{peak}} P_{\text{sig}}^{\text{KDE}}(M_{\text{bc}}, \Delta E) \\ +(1 - F_{\text{peak}}) f_{\text{ARGUS}}(M_{\text{bc}}; \chi_{\text{sig}}) f_{\text{pol2}}(\Delta E; m_{\text{sig}}, r_{\text{sig}}). \end{aligned} \tag{5.14}$$

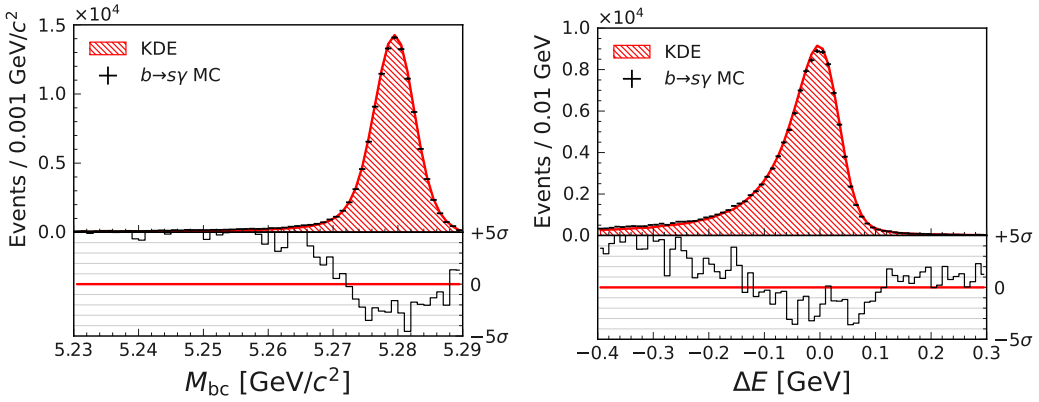
We determine the four free parameters, F_{peak}, χ_{sig}, m_{sig}, and r_{sig}, by fitting to the truth-matched signal candidates in the $b \to s\gamma$ sample, as shown in Fig. 5.1(b). The fit results are summarized in Tab. 5.1. Again a discrepancy between the model and the distribution is only visible with the high statistics of $b \to s\gamma$ sample and is not considered in this analysis.



(a) MR1



(b) non-MR1



(c) control

Figure 5.1: Comparison of the signal model (KDE for MR1 and control, "fit" for non-MR1) and 20M events of $b \to s\gamma$ mixed (charged) MC signals, corresponding to 56 ab^{-1} (53 ab^{-1}). The "fit" is identical to the sum of KDE and ARGUS or pol2. The lower subplots show the pulls between the model and the histogram.

Table 5.2: The $B\overline{B}$ background shape parameters determined in the fit to $B\overline{B}$ MC sample. The fit uncertainties are estimated using the Hesse matrix.

shape parameters	MR1	non-MR1	control
$F_{\text{feed-down}}$	0.751 ± 0.022	0.292 ± 0.016	0.824 ± 0.015
$\chi_{B\bar{B}}$	14.6 ± 0.6	11.89 ± 0.28	11.4 ± 1.0
$m_{B\bar{B}}$	-2.2 ± 0.5	-4.56 ± 0.11	-4.28 ± 0.35
$r_{B\bar{B}}$	0.9 ± 0.5	-0.240 ± 0.017	-0.18 ± 0.07

5.1.3 Distribution of $B\overline{B}$ background

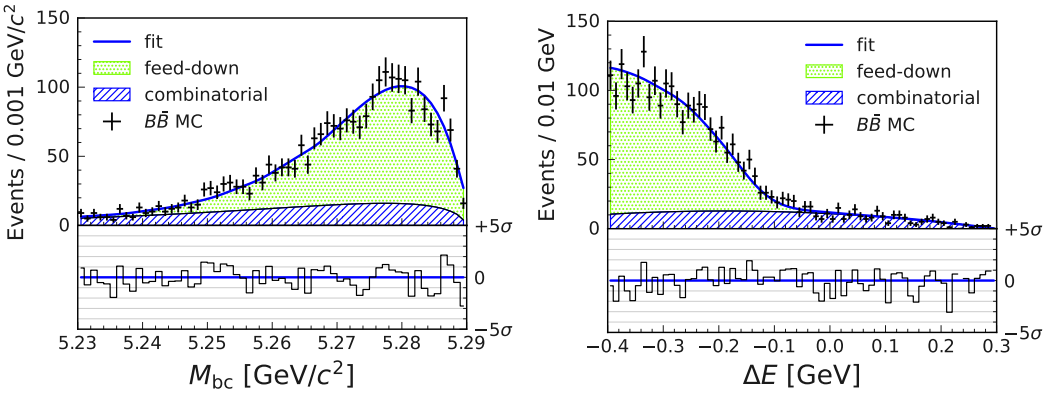
Some of the $B\overline{B}$ background candidates come from a partial reconstruction of B, referred to as $B\overline{B}$ feed-down. The dominant component for $B^0 \to K^{*0}\gamma$ signal is from $B^+ \to K_1(1270)^+(\to K^0_S\pi^0\pi^+)\gamma$, where we miss one π^+. We fit the KDE to the B^0 candidates in MR1 using $B^+ \to K_1(1270)^+(\to K^0_S\pi^0\pi^+)\gamma$ MC samples, which agrees with $B\overline{B}$ feed-down distributions in $B\overline{B}$ samples. For the control channel, we take the distribution from $3\,\text{ab}^{-1}$ $B\overline{B}$ samples to account for the minor contribution around $\Delta E = 0\,\text{GeV}$ from non-radiative decay. More details are described in Appendix C. The other $B\overline{B}$ background is a combinatorial reconstruction of both B mesons, which can be modeled with ARGUS×pol2. Combining the $B\overline{B}$ feed-down and the $B\overline{B}$ combinatorial components, we can model the $B\overline{B}$ background PDF as

$$\begin{aligned} P_{B\bar{B}}(M_{\text{bc}}, \Delta E; F_{\text{feed-down}}, \chi_{B\bar{B}}, m_{B\bar{B}}, r_{B\bar{B}}) = F_{\text{feed-down}} P^{\text{KDE}}_{B\bar{B}\text{ feed-down}}(M_{\text{bc}}, \Delta E) \\ +(1 - F_{\text{feed-down}}) f_{\text{ARGUS}}(M_{\text{bc}}; \chi_{B\bar{B}}) f_{\text{pol2}}(\Delta E; m_{B\bar{B}}, r_{B\bar{B}}). \end{aligned} \tag{5.15}$$

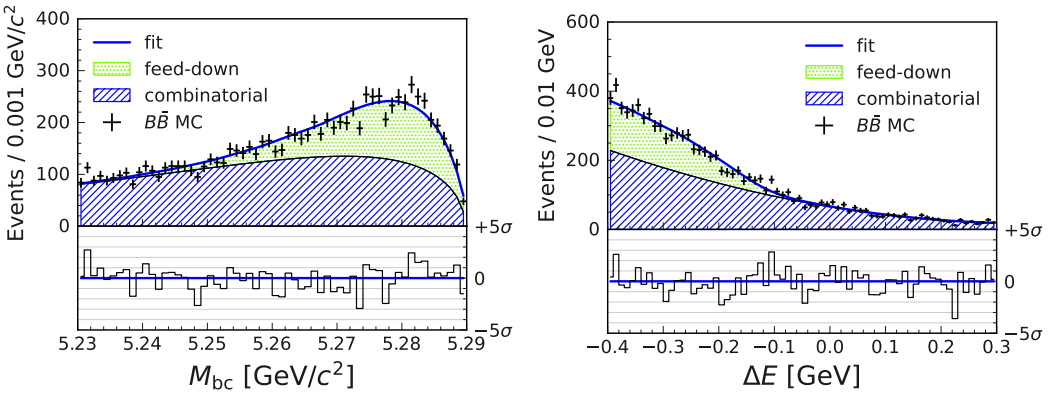
The fraction of $B\overline{B}$ feed-down and the shape parameters of $B\overline{B}$ combinatorial are determined through the fit to the $3\,\text{ab}^{-1}$ $B\overline{B}$ MC sample. The fit results are summarized in Tab. 5.2, and Fig. 5.2 shows the model and MC distribution comparison. We confirm the agreement between the model and the distribution.

5.1.4 Distribution of continuum background

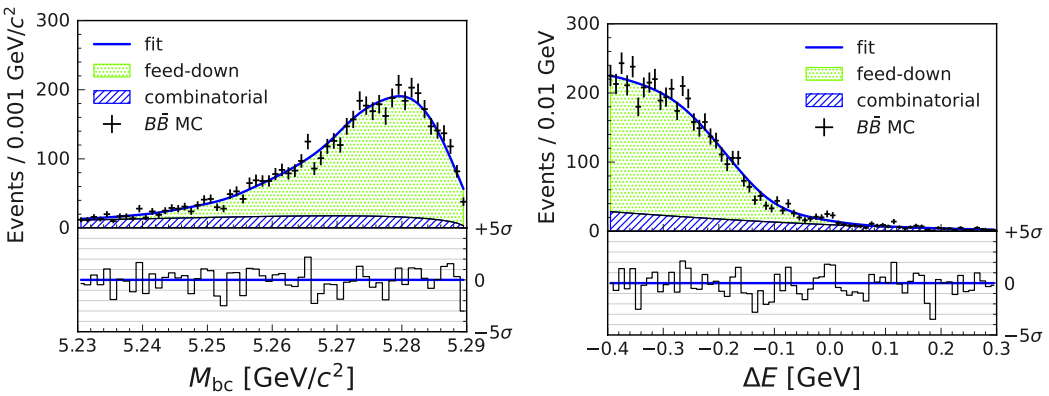
We first check the independence of M_{bc} and ΔE in $q\overline{q}$ MC sample with USP test [57,58], where the p-value against the null hypothesis, i.e., no correlation, is tested by resampling from the histogram PDF. The result in MR1 is shown in Fig. 5.3. The obtained p-values are 0.38, 0.73, and 0.73 for MR1, non-MR1, and control channels, respectively. Therefore, we can ignore the correlation and factorize the PDF.



(a) MR1



(b) non-MR1



(c) control

Figure 5.2: The results of $B\bar{B}$ background shape fit to 3 ab^{-1} of $B\bar{B}$ MC sample. The "fit" is identical to the sum of KDE ("feed-down") and ARGUS or pol2 ("combinatorial"). The lower subplots show the pulls between the model and the histogram.

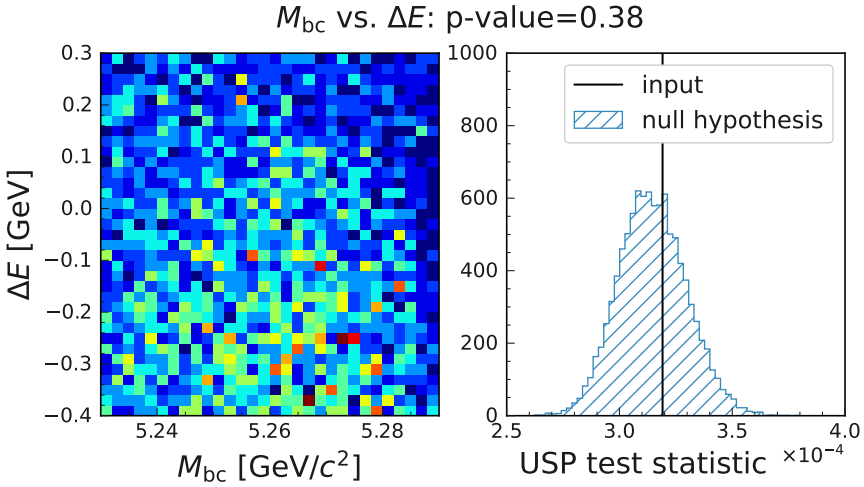


Figure 5.3: The $M_{\rm bc}$–ΔE two-dimensional histogram in $q\overline{q}$ MC for MR1 channel, and the distribution of USP test statistics.

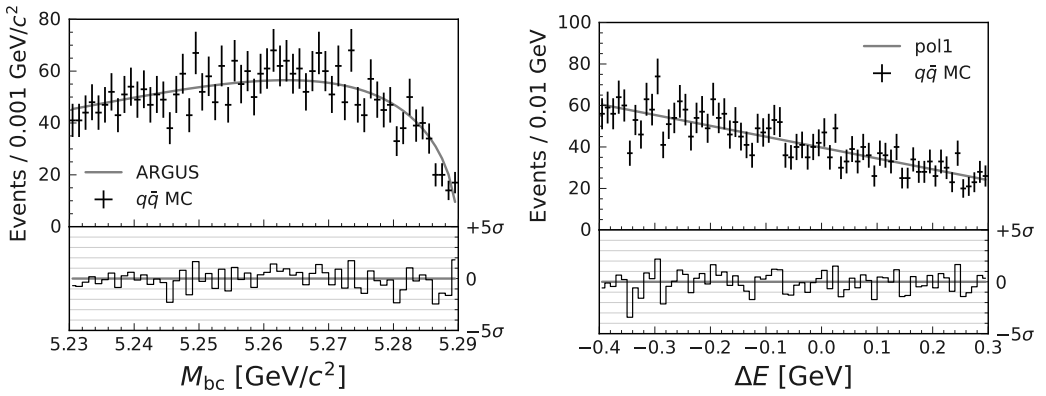
We model $q\overline{q}$ background distribution in MC with ARGUS×pol1 in MR1 and control channels, ARGUS×pol2 for non-MR1 channel. Figure 5.4 shows the distribution of 1 ab^{-1} $q\overline{q}$ MC sample and the fitted model. We let the shape parameters $\chi_{q\overline{q}}$ and $m_{q\overline{q}}$ free in the signal extraction fit.

5.1.5 Fit results

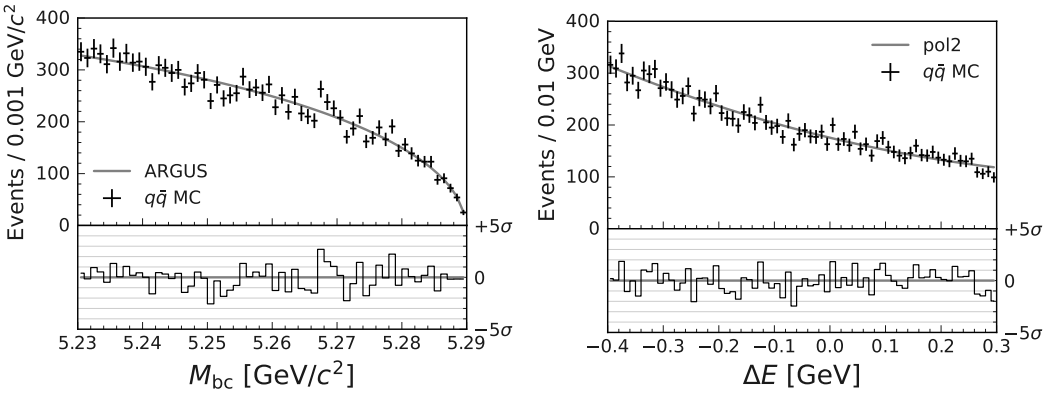
We extract signal and background fractions and $q\overline{q}$ shape parameters from the fit to the $M_{\rm bc}$–ΔE distributions using the PDF in Eq. (5.5) separately for three channels: MR1, non-MR1, and control. The other parameters, i.e., shape parameters of the non-MR1 signal and the MR1/non-MR1/control $B\overline{B}$ background, are fixed to the values obtained in Sections 5.1.2 and 5.1.3.

Figures 5.5 and 5.6 show $M_{\rm bc}$ and ΔE distributions and fit PDFs for 362 fb^{-1} data and 1 ab^{-1} MC, respectively. The fit results in data and MC are summarized in Tab. 5.3, together with the values expected from the MC truth. The yields in the signal-enhanced region are also evaluated as shown in Tab. 5.4. We obtain $N_{\rm sig} = 385 \pm 24$ for MR1 and $N_{\rm sig} = 171 \pm 23$ for non-MR1 in the data.

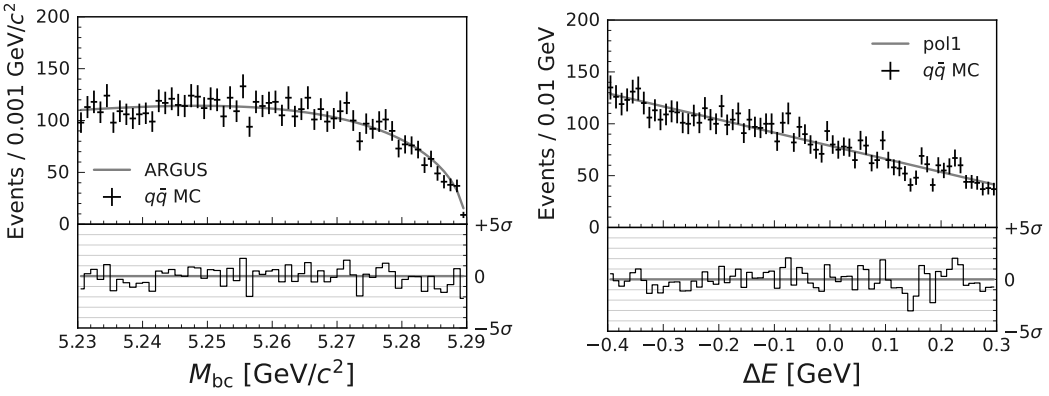
We first validate the fit procedure in MC. We confirm the overall agreement between the distribution and the fit model in Fig. 5.6. The floated $q\overline{q}$ shape parameters are consistent with the expected value determined in the fit to $q\overline{q}$-only sample. The extracted $F_{B\bar{B}}$ is 1–2σ larger than the expected value counted in MC. This could be due to the ne-



(a) MR1

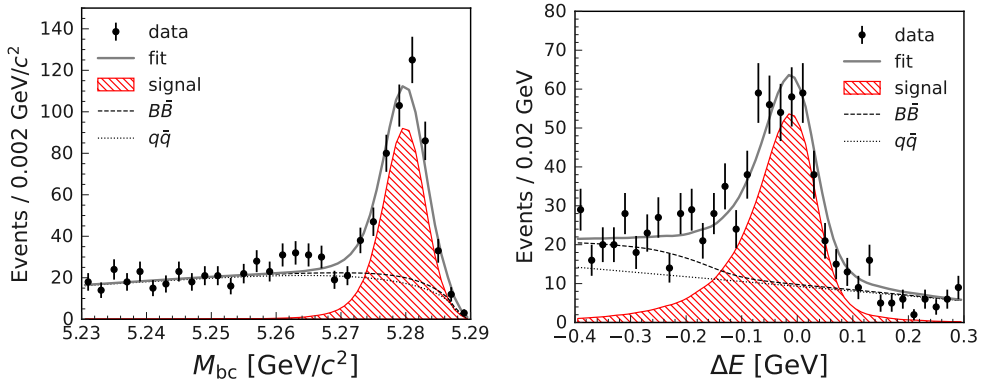


(b) non-MR1

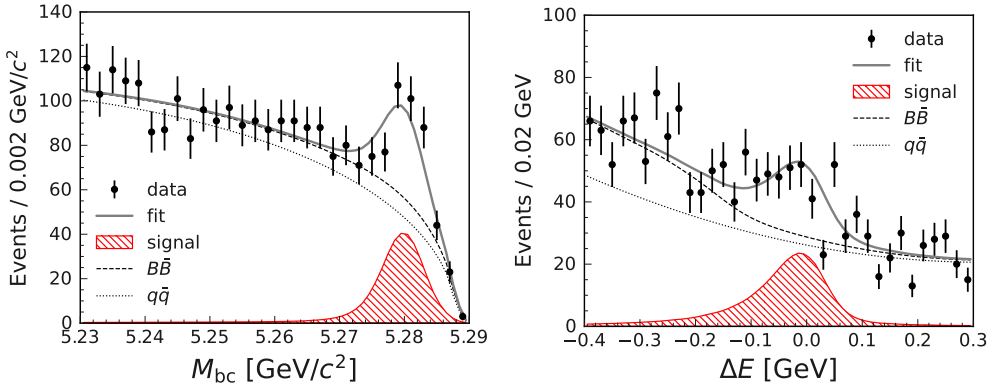


(c) control

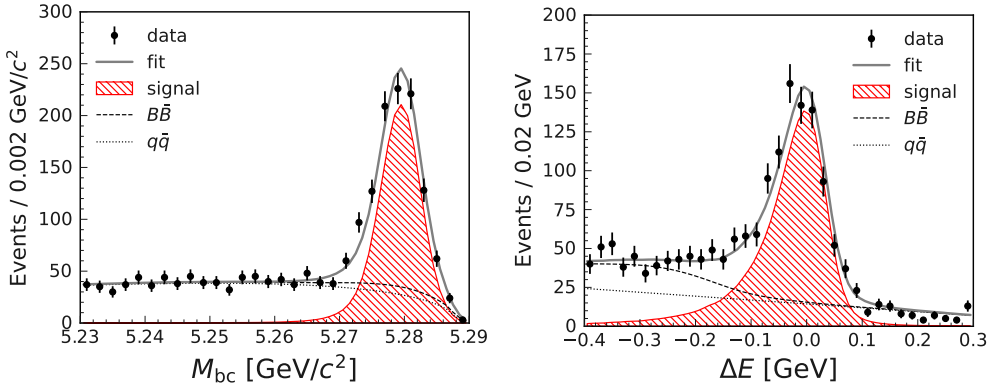
Figure 5.4: The results of $q\overline{q}$ background shape fit in 1 ab^{-1} of $q\overline{q}$ MC. The lower subplots show the pulls between the model and the histogram.



(a) MR1

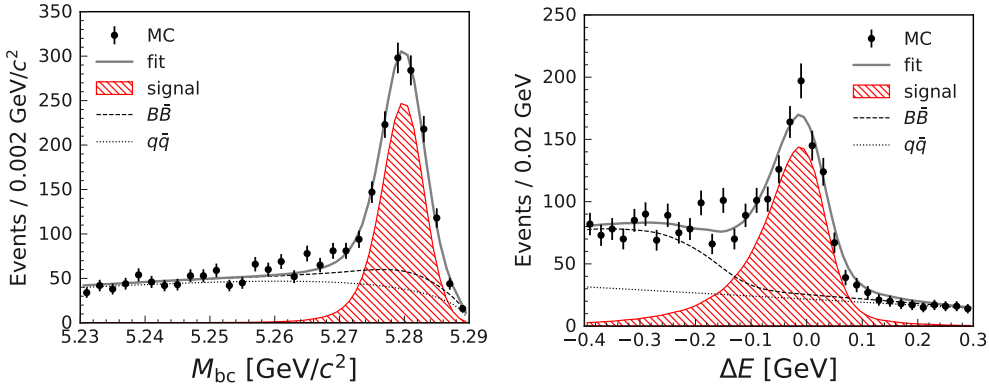


(b) non-MR1

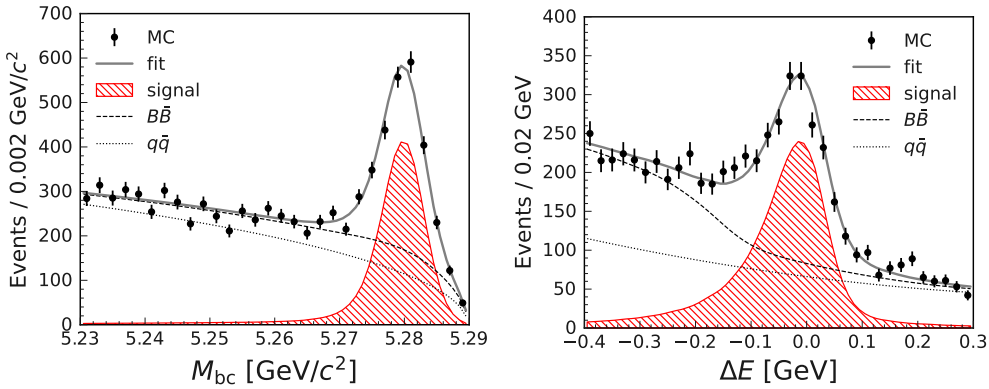


(c) control

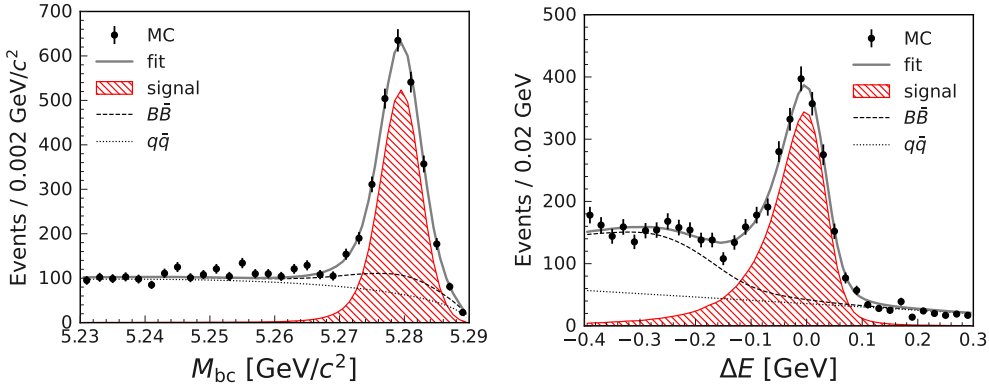
Figure 5.5: The results of signal extraction fit to 362 fb^{-1} data. $M_{\rm bc}$ (ΔE) distribution is shown in ΔE ($M_{\rm bc}$) signal-enhanced region.



(a) MR1



(b) non-MR1



(c) control

Figure 5.6: The results of signal extraction in 1 ab^{-1} of generic MC. $M_{\rm bc}$ (ΔE) distribution is shown in ΔE ($M_{\rm bc}$) signal-enhanced region.

Table 5.3: The results of signal extraction fit in 362 fb^{-1} data and 1 ab^{-1} MC. The expected values for fractions are calculated using the generated number of events in each component in 1 ab^{-1} MC. The expected values for $q\overline{q}$ shape parameters are obtained in the fit to 1 ab^{-1} $q\overline{q}$-only MC sample.

(a) MR1

fit parameters	data	MC 1 ab^{-1}	expected
F_{sig}	0.243 ± 0.014	0.235 ± 0.008	0.240
$F_{B\bar{B}}$	0.10 ± 0.04	0.258 ± 0.022	0.233
$\chi_{q\bar{q}}$	10.5 ± 0.8	$9.5^{+0.5}_{-0.6}$	$10.00^{+0.30}_{-0.31}$
$m_{q\bar{q}}$	$-1.78^{+0.27}_{-0.26}$	$-1.54^{+0.18}_{-0.17}$	-1.77 ± 0.12

(b) non-MR1

fit parameters	data	MC 1 ab^{-1}	expected
F_{sig}	0.036 ± 0.005	0.118 ± 0.004	0.114
$F_{B\bar{B}}$	$0.097^{+0.027}_{-0.028}$	$0.207^{+0.018}_{-0.017}$	0.182
$\chi_{q\bar{q}}$	$4.7^{+0.8}_{-1.0}$	$1.7^{+1.2}_{-4.7}$	$3.4^{+0.4}_{-0.5}$
$m_{q\bar{q}}$	$-1.90^{+0.15}_{-0.14}$	$-1.93^{+0.10}_{-0.09}$	-2.03 ± 0.06
$r_{q\bar{q}}$	$-0.128^{+0.027}_{-0.025}$	$-0.081^{+0.020}_{-0.019}$	$-0.084^{+0.016}_{-0.015}$

(c) control

fit parameters	data	MC 1 ab^{-1}	expected
F_{sig}	0.263 ± 0.010	0.233 ± 0.005	0.237
$F_{B\bar{B}}$	$0.136^{+0.024}_{-0.025}$	$0.262^{+0.013}_{-0.014}$	0.226
$\chi_{q\bar{q}}$	8.3 ± 0.07	6.9 ± 0.5	$7.95^{+0.27}_{-0.28}$
$m_{q\bar{q}}$	$-2.28^{+0.17}_{-0.16}$	$-1.98^{+0.12}_{-0.11}$	-2.12 ± 0.08

Table 5.4: The estimated yields in the signal-enhanced region ($5.27 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2$ and $-0.2 < \Delta E < 0.1\,{\rm GeV}$) in $362\,{\rm fb}^{-1}$ data and MC (scaled from $1\,{\rm ab}^{-1}$). The uncertainties of the yields are propagated from the fractions only, not from the $q\overline{q}$ shape parameters. The expected yields are the generated number of events in each component in $1\,{\rm ab}^{-1}$ MC, which is also scaled to $362\,{\rm fb}^{-1}$. The statistical uncertainties in expected yields are shown in parenthesis, which needs to be considered when compared with the data.

(a) MR1

yields	data	MC (scaled)	expected
$N_{\rm sig}$	385 ± 24	373 ± 13	369(±12)
$N_{B\bar{B}}$	20 ± 8	54 ± 5	55(±4)
$N_{q\bar{q}}$	149 ± 7	124 ± 4	127(±7)

(b) non-MR1

yields	data	MC (scaled)	expected
$N_{\rm sig}$	171 ± 23	631 ± 20	614(±15)
$N_{B\bar{B}}$	69 ± 19	158 ± 14	157(±8)
$N_{q\bar{q}}$	426 ± 13	388 ± 9	401(±12)

(c) control

yields	data	MC (scaled)	expected
$N_{\rm sig}$	843 ± 34	757 ± 18	751(±16)
$N_{B\bar{B}}$	55 ± 10	113 ± 6	88(±6)
$N_{q\bar{q}}$	234 ± 7	209 ± 4	236(±9)

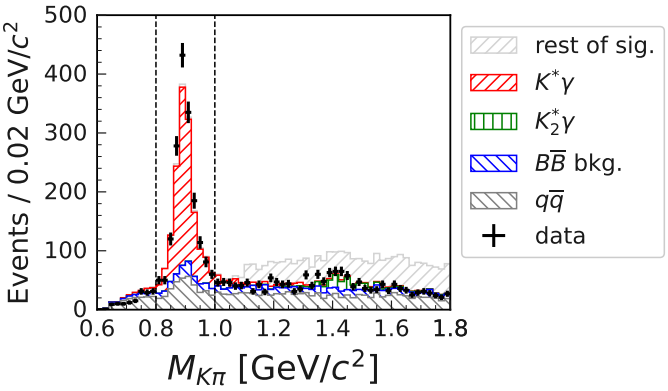


Figure 5.7: The comparison of $M_{K\pi}$ distribution between 352 fb^{-1} data and 1 ab^{-1} generic MC sample (scaled). The black dashed lines represent the edges of MR1. The rest of the signal in light gray is dominated by the X_s contribution.

glected correlation in $q\overline{q}$. We take the possible impact on the *CP*-violation parameters into the systematic uncertainty in Section 5.5.4.

We then discuss the fit result in the data. In Fig. 5.5, the fit model agrees with the distribution within the uncertainty. Overall, the estimated signal and background yields in the signal region also agree with MC expectations. However, we obtain smaller fractions for the $B\overline{B}$ background and non-MR1 signal. This discrepancy could be attributed to the data-MC difference in the $M_{K\pi}$ spectrum as shown in Fig. 5.7, possibly caused by the imperfect simulation of $X_s \to K\pi$ branching fraction through hadronization in PYTHIA. We consider the possible effect as a source of systematic uncertainty in Section 5.5.4.

5.1.6 Correction of Punzi effect

The event-by-event fractions can be calculated using $M_{\rm bc}$ and ΔE as

$$f_i(M_{\rm bc}, \Delta E) = \frac{F_i' P_i(M_{\rm bc}, \Delta E)}{\sum_j F_j' P_j(M_{\rm bc}, \Delta E)}, \tag{5.16}$$

where $i, j \in \{\text{sig}, B\overline{B}\text{ feed-down}, B\overline{B}\text{ combinatorial}, q\overline{q}\}$ and

$$\begin{aligned} F'_{\rm sig} &= F_{\rm sig}, \\ F'_{B\bar{B}\text{ feed-down}} &= (1 - F_{\rm sig}) F_{B\bar{B}} F_{\text{feed-down}}, \\ F'_{B\bar{B}\text{ combinatorial}} &= (1 - F_{\rm sig}) F_{B\bar{B}} (1 - F_{\text{feed-down}}), \\ F'_{q\bar{q}} &= (1 - F_{\rm sig})(1 - F_{B\bar{B}}). \end{aligned} \tag{5.17}$$

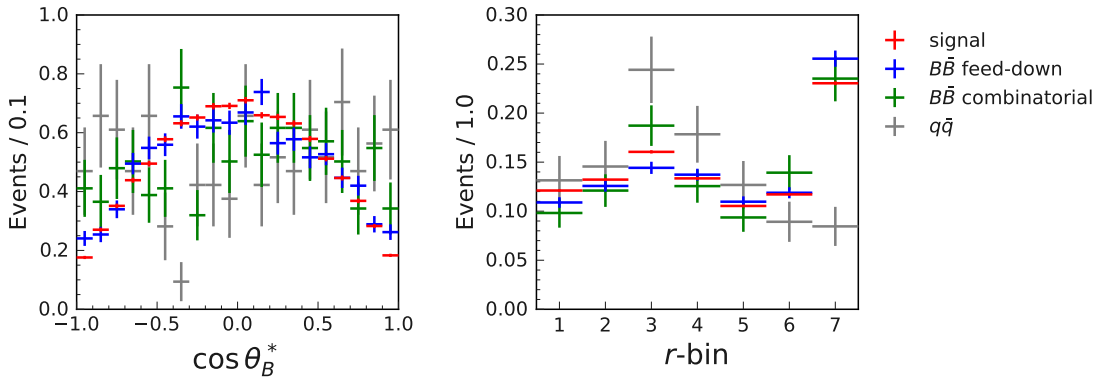


Figure 5.8: The distributions of $\cos\theta_B^*$ and r-bin in the signal region with the vertex quality requirement (TD) for each component.

Here, we evaluate the fractions separately for $B\overline{B}$ feed-down and $B\overline{B}$ combinatorial, as we expect different Δt resolution as described in Section 5.2.4. However, as reported by G. Punzi, the event-by-event fractions may bias the analysis if conditional variables distribute differently for components [59]. Among them, a significant difference is observed in the $\cos\theta_B^*$ distribution, which is used in the resolution function in Section 5.2.1(1), and the r-bin histogram, on which w and Δw in Section 4.4 depend. The component-wise comparison of these two variables is shown in Fig. 5.8 and for the others in Fig. 5.9. We can see differences among components in Fig. 5.8 but in Fig. 5.9. To avoid this possible bias, we multiply the probabilities $P_i(M_{\rm bc}, \Delta E)$ by the probabilities of these two variables, assuming the independence:

$$f_i(M_{\rm bc}, \Delta E, \cos\theta_B^*, r\text{-bin}) = \frac{F_i' P_i(M_{\rm bc}, \Delta E) P_i(\cos\theta_B^*) P_i(r\text{-bin})}{\sum_j F_j' P_j(M_{\rm bc}, \Delta E) P_j(\cos\theta_B^*) P_j(r\text{-bin})}. \tag{5.18}$$

To model the PDFs of these two variables with enough statistics, we use the $b \to s\gamma$ sample instead of the $B\overline{B}$ sample and merge the channels with similar distributions.

For the $\cos\theta_B^*$ model, we use B^0 candidates from signal MC for the signal component; B^0 and B^+ candidates combined from $b \to s\gamma$ sample for $B\overline{B}$ feed-down and combinatorial component. We fit these samples with the 2nd-order polynomial, floating the pole, p, and the value there, q. The fitted parameters are summarized in Tab. 5.5, and their curves and distributions are shown in Fig. 5.10. We model the $q\overline{q}$ background with the flat distribution. The obtained four models for signal, $B\overline{B}$ feed-down, $B\overline{B}$ combinatorial, and $q\overline{q}$ are commonly used in all three channels (MR1, non-MR1, and control). The models obtained in MC are directly applied to the data.

We model the r-bin by filling the histogram for each channel and component with the

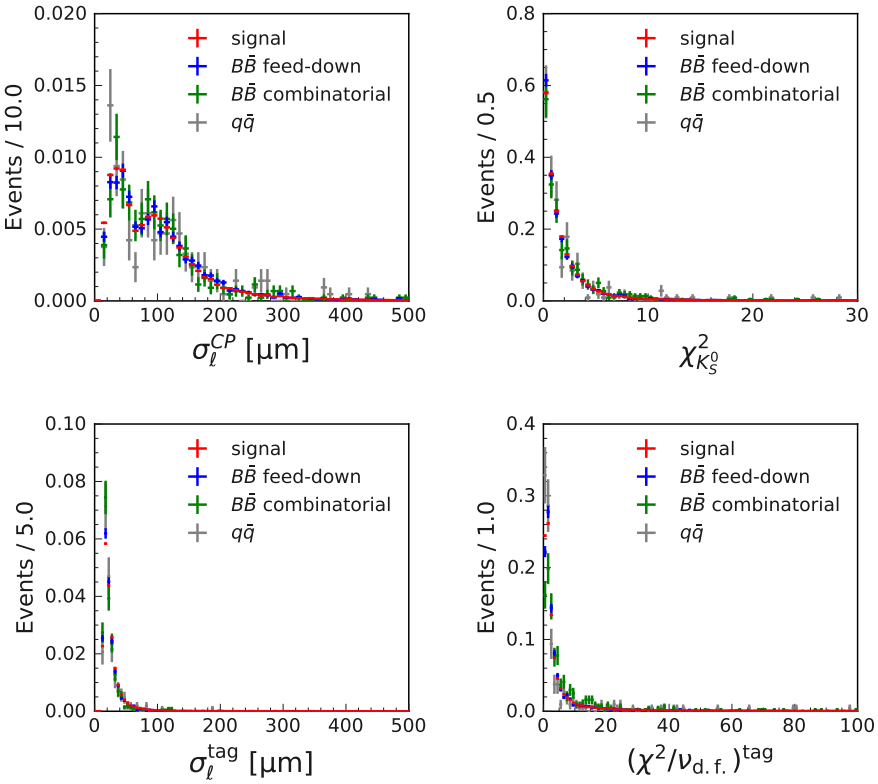
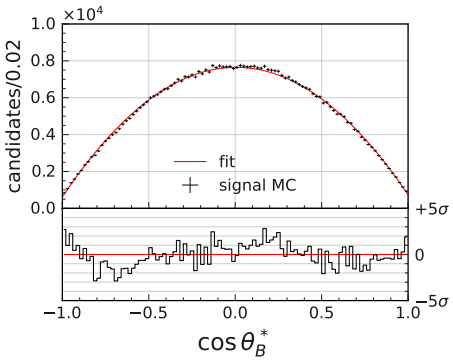
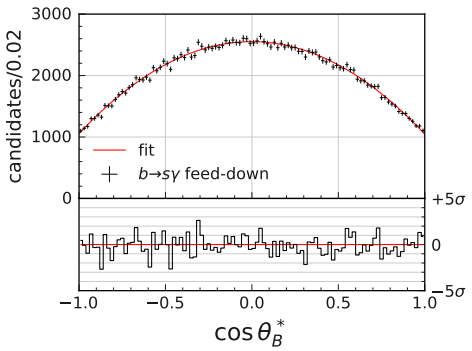


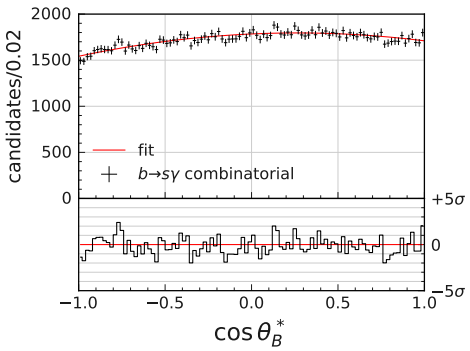
Figure 5.9: The distributions of the other conditional variables, $\chi^2_{K^0_S}$, σ_ℓ^{CP}, $(\chi^2/\nu_{\rm d.f.})^{\rm tag}$, and $\sigma_\ell^{\rm tag}$, in the signal region with the vertex quality requirement (TD) for each component.



(a) signal



(b) $B\bar{B}$ feed-down

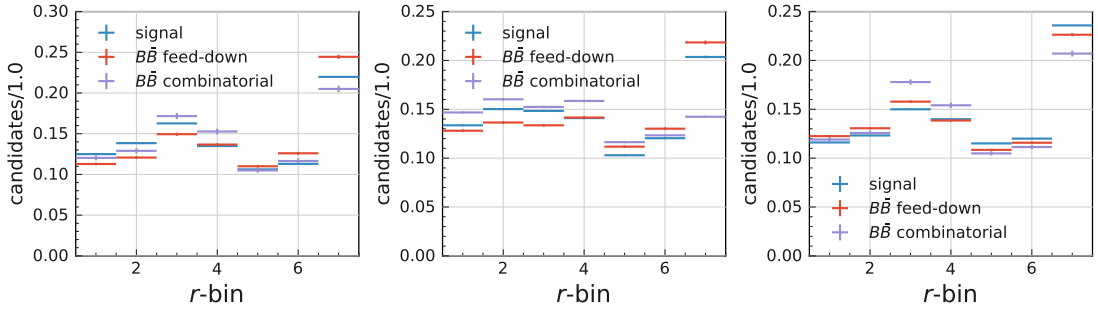


(c) $B\bar{B}$ combinatorial

Figure 5.10: The distributions of $\cos\theta_B^*$ and fitted model. The lower subplots show the pulls between the model and the histogram.

Table 5.5: The fitted parameters for $\cos\theta_B^*$. The fit uncertainties are estimated using the Hesse matrix.

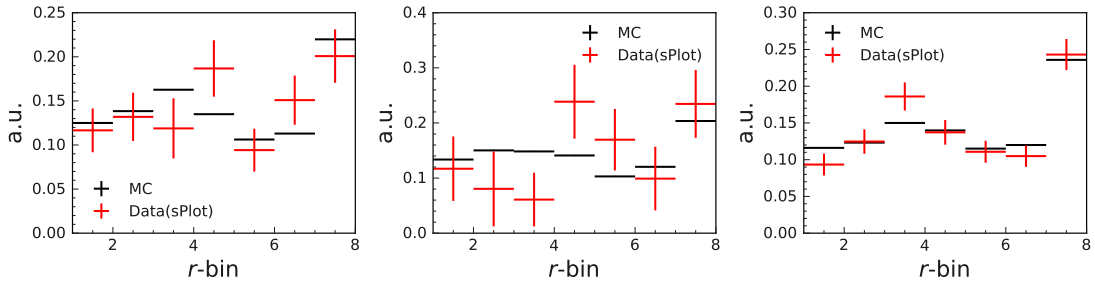
fit parameters	signal	$B\overline{B}$ feed-down	$B\overline{B}$ combinatorial
p	0.0042 ± 0.0006	0.0016 ± 0.0034	0.45 ± 0.06
q	0.7180 ± 0.0005	0.6214 ± 0.0016	0.5280 ± 0.0017



(a) MR1 (b) non-MR1 (c) control

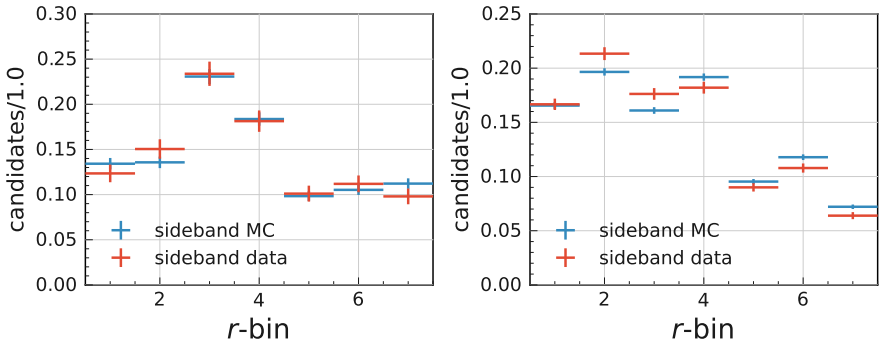
Figure 5.11: The distributions of r-bin for signal, $B\overline{B}$ feed-down, and $B\overline{B}$ combinatorial.

corresponding data. Figure 5.11 shows the nine histograms for signal, $B\overline{B}$ feed-down, and $B\overline{B}$ combinatorial component in MR1, non-MR1, and control channels. We use the signal, $b \to s\gamma$, and control MC samples for signal components in MR1, non-MR1, and control channels, respectively, and the $b \to s\gamma$ MC sample for $B\overline{B}$ feed-down and $B\overline{B}$ combinatorial in all three channels. Figure 5.12 shows that the signal MC agrees with ${}_s\mathcal{P}lot$ [60] data within their uncertainties, where the contribution of the background is subtracted from the distribution. We apply the histogram obtained in MC to both data and MC. For the $q\overline{q}$ modeling, we use $M_{\rm bc}$–ΔE sideband. As we do not observe differences between B^0 and B^+ candidates, we combine them indiscriminately to increase the statistics. The data–MC comparison in $M_{\rm bc}$–ΔE sideband is shown in Fig. 5.13. The difference between MR1 and non-MR1 comes from the different continuum suppression. We confirm that the data and MC agree in general. We use the histogram obtained in the data sideband (MC sideband) to model the $q\overline{q}$ component in the data (MC).



(a) MR1 (b) non-MR1 (c) control

Figure 5.12: The distributions of r-bin in signal MC and ${}_s\mathcal{P}lot$ data.



(a) MR1 and control (b) non-MR1

Figure 5.13: The data–MC comparison of r-bin histograms in sideband.

5.2 Proper-time difference modeling

We model Δt distribution based on our understanding of the vertex reconstruction in signal events, $B\overline{B}$ background events, and $q\overline{q}$ events, each.

5.2.1 Resolution of signal events

The measured proper-time difference Δt for signal events includes three sources of resolution:

1. kinematic approximation (R_{k}) corrects the bias from small B^0 momentum in the CM frame and statistically accounts for the relation between measured vertex distance and the proper-time difference,

2. B_{sig} decay vertex resolution (R_{rec}) accounts for the smearing of the decay vertex position by the finite detector resolution,

3. B_{tag} decay vertex resolution (R_{tag}) consists of the detector resolution (R_{asc}) and the bias from non-primary decay vertices mainly due to the finite lifetime of charmed intermediate states (R_{np}).

(1) Kinematic approximation

When we measure the proper-time difference with the decay vertex distance, we implicitly ignore B momentum in the $\Upsilon(4S)$ rest frame. We correct the bias from the difference in the $B\overline{B}$ momentum along $\Upsilon(4S)$ boost and consider the smearing from the difference in the proper time of $B\overline{B}$. The effect of this kinematical approximation referred to as R_{k} can be incorporated by requiring the following relation during the T integral of Eq. (1.19) and additionally integrate over $\Delta\tau$ from $-\infty$ to $+\infty$:

$$\Delta t_{\mathrm{true}} = a\Delta\tau + c_{\mathrm{k}}T, \tag{5.19}$$

where

$$a \coloneqq \frac{E^*_B}{m_{B^0}} \quad \text{and} \quad c_{\mathrm{k}} \coloneqq \frac{p^*_B \cos\theta^*_B}{\beta m_{B^0}}. \tag{5.20}$$

Here, $\cos\theta_B^*$ is the angle between the boost of $\Upsilon(4S)$ and the momentum of $B_{\rm sig}$ in the CM frame. The $R_{\rm k}$-convoluted PDF can be written as

$$
\begin{aligned}
P \otimes R_{\rm k}(\Delta t_{\rm true}, q) = \frac{1}{4a\tau_{B^0}} e^{-\frac{|\Delta t_{\rm true}|}{(a\pm c_{\rm k})\tau_{B^0}}} \Bigg[1 - q\Delta w + \frac{q(1-2w)}{1+(c_{\rm k}\tau_{B^0}\Delta m_d/a)^2} \\
\times \left\{ \left(A_{CP} - S_{CP}\frac{c_{\rm k}\tau_{B^0}\Delta m_d}{a} \right) \cos\frac{\Delta m_d \Delta t_{\rm true}}{a \pm c_{\rm k}} \right. \\
\left. + \left(S_{CP} + A_{CP}\frac{c_{\rm k}\tau_{B^0}\Delta m_d}{a} \right) \sin\frac{\Delta m_d \Delta t_{\rm true}}{a \pm c_{\rm k}} \right\} \Bigg],
\end{aligned}
\tag{5.21}
$$

including the effect of w and Δw, where the $\pm$ sign corresponds to the sign of $\Delta t_{\rm true}$. The derivation of this form is described elsewhere [61].

(2) Signal-side resolution model

We model the signal-side resolution function $R_{\rm rec}$ with a double Gaussian,

$$
\begin{aligned}
R_{\rm rec}\left(\delta\ell^{CP}; \chi^2_{K^0_S}, \sigma_\ell^{CP}\right) = (1 - f_{\rm tail}^{CP}(\chi^2_{K^0_S}))G\left(\delta\ell^{CP}; \mu = 0, \sigma = s_{\rm main}^{CP}(\chi^2_{K^0_S})\sigma_\ell^{CP}\right) \\
+ f_{\rm tail}^{CP}(\chi^2_{K^0_S})G\left(\delta\ell^{CP}; \mu = 0, \sigma = s_{\rm tail}^{CP}\sigma_\ell^{CP}\right),
\end{aligned}
\tag{5.22}
$$

with

$$
f_{\rm tail}^{CP}(\chi^2_{K^0_S}) = f_{\rm tail}^{CP,0} + f_{\rm tail}^{CP,1}\chi^2_{K^0_S} \text{ and} \tag{5.23}
$$

$$
s_{\rm main}^{CP}(\chi^2_{K^0_S}) = s_{\rm main}^{CP,0} + s_{\rm main}^{CP,1}\chi^2_{K^0_S}, \tag{5.24}
$$

where $\delta\ell^{CP}$ is the residual between reconstructed and true $B_{\rm sig}$ decay vertices along the $\Upsilon(4S)$ boost, and $f_{\rm tail}^{CP,0(1)}$, $s_{\rm main}^{CP,0(1)}$ and $s_{\rm tail}^{CP}$ are free parameters. The χ^2 dependence is employed as a means to account for underestimated uncertainty. When we reconstruct a track using hits that do not actually belong to the track, the uncertainties associated with those hits may not adequately explain the deviation from the true track. This leads to an underestimation of the vertex uncertainty and results in an inflated fit χ^2. For the $R_{\rm rec}$, we use $\chi^2_{K^0_S}$ for the resolution model since $K^0_S \to \pi^+\pi^-$ are the only tracks from $B_{\rm sig}$. We studied the $\chi^2_{K^0_S}$ dependence by fitting $\delta\ell^{CP}$ with double Gaussian in the 20 slices of $\chi^2_{K^0_S}$. Figure D.1 in Appendix D shows fit results in all the $\chi^2_{K^0_S}$ slices. The fit results of three shape parameters $f_{\rm tail}$, $s_{\rm main}$, and $s_{\rm tail}$ are shown in Fig. 5.14. We observe clear linear relation in $s_{\rm main}$ and weaker relation in $f_{\rm tail}$ and $s_{\rm tail}$. We tried to introduce $\chi^2_{K^0_S}$ dependence in all three parameters. However, the $\chi^2_{K^0_S}$ dependence of $s_{\rm tail}$ is unstable

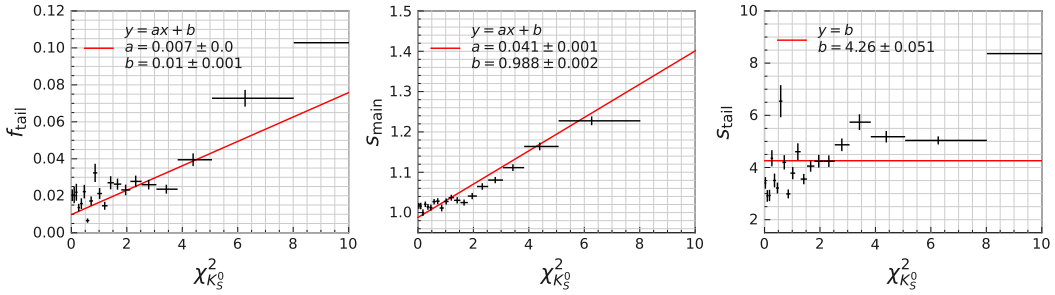


Figure 5.14: The shape parameters of fitted double Gaussian for each $\chi^2_{K^0_S}$ slices. The red line shows the result of the least-squares fit with a linear function.

Table 5.6: The results of $\delta\ell^{CP}$ fit in signal MC.

parameter	fit result
$f_{\text{tail}}^{CP,0}$	0.0075 ± 0.0005
$f_{\text{tail}}^{CP,1}$	0.0131 ± 0.0004
$s_{\text{main}}^{CP,0}$	1.0007 ± 0.0020
$s_{\text{main}}^{CP,1}$	0.0303 ± 0.0010
s_{tail}^{CP}	4.03 ± 0.06
f_{outlier}	0.00051 ± 0.00006
s_{outlier} [μm]	$3.01^{+0.21}_{-0.20} \times 10^3$

and less significant than the other two parameters. Moreover, we do not observe an improvement in the model–distribution agreement even with the $\chi^2_{K^0_S}$ dependence of s_{tail}. Therefore, we introduce $\chi^2_{K^0_S}$ dependence only in f_{tail} and s_{main} and not in s_{tail}.

For the signal MC fit, we add another Gaussian to account for the broad outlier. Our fit PDF is

$$\begin{aligned} P\left(\delta\ell^{CP}; \chi^2_{K^0_S}, \sigma_\ell^{CP}\right) = (1 - f_{\text{outlier}}) R_{\text{rec}}\left(\delta\ell^{CP}; \chi^2_{K^0_S}, \sigma_\ell^{CP}\right) \\ + f_{\text{outlier}} G\left(\delta\ell^{CP}; \mu = 0, \sigma = s_{\text{outlier}}\right). \end{aligned} \tag{5.25}$$

The $\delta\ell^{CP}$ distribution and fit curve are shown in Fig. 5.15. The resulting fit parameters are summarized in Tab. 5.6. We also confirm that the model describes the distribution well in all $\chi^2_{K^0_S}$ regions, as shown in Fig. D.2 in Appendix D.

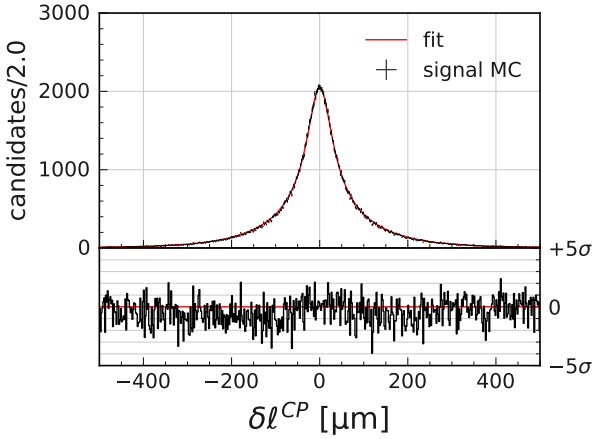


Figure 5.15: $\delta\ell^{CP}$ distribution and fitted $R_{\rm rec}$ in 2M events of signal MC, corresponding to $> 400\,\mathrm{ab}^{-1}$. The lower subplot shows the pulls between the model and the histogram.

(3) Tag-side resolution model

In principle, $B_{\rm tag}$ vertex reconstruction is independent of $B_{\rm sig}$ reconstruction. Thus, we use the same resolution model as $B^0 \to K^0_S\,K^0_S\,K^0_S$ analysis [47]:

$$R_{\rm tag} = R_{\rm asc} \otimes R_{\rm np} \tag{5.26}$$

with

$$\begin{aligned} &R_{\rm asc}\left(\delta\ell^{\rm tag};(\chi^2/\nu_{\rm d.f.})^{\rm tag},\sigma_\ell^{\rm tag}\right) \\ &\quad= (1-f_{\rm tail}^{\rm tag})G\left(\delta\ell^{\rm tag};\mu=0,\sigma=(s_{\rm main}^{\rm tag,0}+s_{\rm main}^{\rm tag,1}(\chi^2/\nu_{\rm d.f.})^{\rm tag})\sigma_\ell^{\rm tag}\right) \\ &\qquad+ f_{\rm tail}^{\rm tag}G\left(\delta\ell^{\rm tag};\mu=0,\sigma=(s_{\rm tail}^{\rm tag,0}+s_{\rm tail}^{\rm tag,1}(\chi^2/\nu_{\rm d.f.})^{\rm tag})\sigma_\ell^{\rm tag}\right) \end{aligned} \tag{5.27}$$

and

$$\begin{aligned} R_{\rm np}\left(\delta\ell^{\rm tag};(\chi^2/\nu_{\rm d.f.})^{\rm tag},\sigma_\ell^{\rm tag}\right) &= f_\delta\delta(\delta\ell^{\rm tag}) + (1-f_\delta)f_{\rm p}E_{\rm p}(\delta\ell^{\rm tag};\tau=\tau_{\rm np}\sigma_\ell^{\rm tag}) \\ &\quad+(1-f_\delta)(1-f_{\rm p})E_{\rm n}(\delta\ell^{\rm tag};\tau=\tau_{\rm np}\sigma_\ell^{\rm tag}), \end{aligned} \tag{5.28}$$

where $E_{\rm p}$ and $E_{\rm p}$ are one-sided exponential functions,

$$E_{\rm p}(x;\tau) = \begin{cases} \exp(-x/\tau)/\tau & (x>0) \\ 0 & (x\le 0) \end{cases} \quad \text{and} \tag{5.29}$$

$$E_{\rm n}(x;\tau) = \begin{cases} 0 & (x>0) \\ \exp(+x/\tau)/\tau & (x\le 0) \end{cases}, \tag{5.30}$$

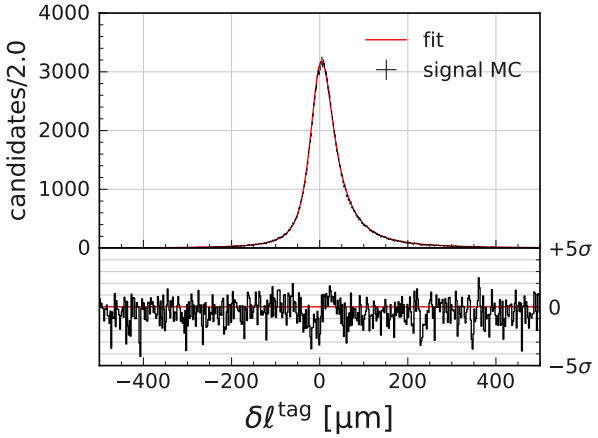


Figure 5.16: $\delta\ell^{\text{tag}}$ distribution and fitted R_{tag} in 2M events of signal MC, corresponding to $> 400\,\text{ab}^{-1}$. The lower subplot shows the pulls between the model and the histogram. The asymmetric distribution reflects the asymmetric R_{np}.

and f_δ, f_{p}, and τ_{np} are expressed by seven free parameters, $f_\delta^{0,1}$, $f_{\text{p}}^{0,1}$, and $\tau^{0,1,\text{max}}$:

$$f_\delta = \begin{cases} 0 & (f_\delta^0 + f_\delta^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} < 0) \\ f_\delta^0 + f_\delta^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} & (0 \le f_\delta^0 + f_\delta^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} \le 1)\,, \\ 1 & (f_\delta^0 + f_\delta^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} > 1) \end{cases} \tag{5.31}$$

$$f_{\text{p}} = \begin{cases} 0 & (f_{\text{p}}^0 + f_{\text{p}}^1\sigma_\ell^{\text{tag}} < 0) \\ f_{\text{p}}^0 + f_{\text{p}}^1\sigma_\ell^{\text{tag}} & (0 \le f_{\text{p}}^0 + f_{\text{p}}^1\sigma_\ell^{\text{tag}} \le 1)\,, \text{ and} \\ 1 & (f_{\text{p}}^0 + f_{\text{p}}^1\sigma_\ell^{\text{tag}} > 1) \end{cases} \tag{5.32}$$

$$\tau_{\text{np}} = \begin{cases} \tau^0 + \tau^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} & (\tau^0 + \tau^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} \le \tau^{\text{max}}) \\ \tau^{\text{max}} & (\tau^0 + \tau^1(\chi^2/\nu_{\text{d.f.}})^{\text{tag}} > \tau^{\text{max}}) \end{cases}. \tag{5.33}$$

The fitted $\delta\ell^{\text{tag}}$ distribution in signal MC is shown in Fig. 5.16. The fitted 12 parameters are summarized in Tab. 5.7. The model correctly describes the $(\chi^2/\nu_{\text{d.f.}})^{\text{tag}}$ dependence, as we see no significant deviation in the slice distributions shown in Fig. D.3 in Appendix D.

(4) Lifetime fit

We perform Δt fit to extract B^0 lifetime τ_{B^0}, referred to as *lifetime fit*, with the PDF of

$$E_{\text{f}}(\Delta\tau) = \frac{1}{2\tau_{B^0}} e^{-\frac{|\Delta\tau|}{\tau_{B^0}}}, \tag{5.34}$$

Table 5.7: The results of $\delta\ell^{\rm tag}$ fit in signal MC.

	parameter	fit result
$R_{\rm asc}$	$f_{\rm tail}^{\rm tag}$	$0.00265^{+0.00018}_{-0.00016}$
	$s_{\rm main}^{\rm tag,0}$	$1.0131^{+0.0041}_{-0.0042}$
	$s_{\rm main}^{\rm tag,1}$	$0.0718^{+0.0010}_{-0.0014}$
	$s_{\rm tail}^{\rm tag,0}$	15.8 ± 0.7
	$s_{\rm tail}^{\rm tag,1}$	$1.85^{+0.15}_{-0.14}$
$R_{\rm np}$	f_δ^0	0.737 ± 0.008
	f_δ^1	-0.1622 ± 0.0029
	$f_{\rm p}^0$	0.8677 ± 0.0026
	$f_{\rm p}^1$ [$\mu{\rm m}^{-1}$]	-0.00158 ± 0.00006
	τ^0	$1.027^{+0.013}_{-0.012}$
	τ^1	0.1925 ± 0.0025
	$\tau^{\rm max}$	$5.23^{+0.05}_{-0.06}$

in order to test the resolution functions and their parameters listed in Tabs. 5.6 and 5.7. Using the MC truth information, we can examine the resolution functions and their parameters with several configurations as follows:

1. $\Delta\tau$ fit with $E_{\rm f}$ to check the bias in the reconstruction and selection in Chapter 4,
2. $\Delta t_{\rm true}$ fit with $E_{\rm f} \otimes R_{\rm k}$ to check the bias from (1) $R_{\rm k}$,
3. $(\ell^{CP} - \ell^{\rm tag}_{\rm true})/\beta\gamma c$ fit with $E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm rec} + G_{\rm outlier}$ to check the bias from (2) $R_{\rm sig}$,
4. $(\ell^{CP}_{\rm true} - \ell^{\rm tag})/\beta\gamma c$ fit with $E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm tag}$ to check the bias from (3) $R_{\rm tag}$, and
5. Δt fit with $E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm rec} \otimes R_{\rm tag} + G_{\rm outlier}$ to check the overall bias in Δt modeling.

The fitted lifetimes are summarized in Tab. 5.8. Figure 5.17 shows the Δt distribution with the overall fit model. We confirm that all the biases listed above are negligible.

Table 5.8: The results of lifetime fit in signal MC.

proper-time difference	model	fit result [ps]
MC input	(no fit)	1.519
$\Delta\tau$	$E_{\rm f}$	1.5185 ± 0.0028
$\Delta t^{\rm true} = (\ell^{CP}_{\rm true} - \ell^{\rm tag}_{\rm true})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k}$	1.5187 ± 0.0028
$(\ell^{CP} - \ell^{\rm tag}_{\rm true})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm rec} + G_{\rm outlier}$	1.5186 ± 0.0035
$(\ell^{CP}_{\rm true} - \ell^{\rm tag})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm tag}$	1.5188 ± 0.0031
$\Delta t = (\ell^{CP} - \ell^{\rm tag})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm rec} \otimes R_{\rm tag} + G_{\rm outlier}$	1.519 ± 0.004

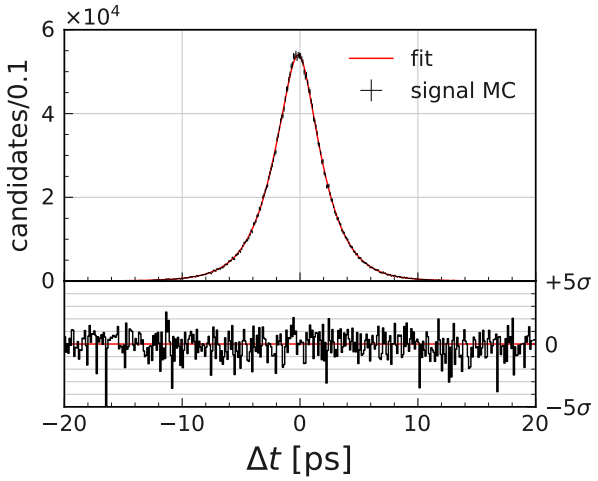


Figure 5.17: Δt distribution and fitted model in 2M events of signal MC, corresponding to $> 400\,\mathrm{ab}^{-1}$. The lower subplot shows the pulls between the model and the histogram. The asymmetric distribution reflects the asymmetric $R_{\rm np}$.

5.2.2 Resolution calibration using fast simulation

We develop a fast simulator that can reproduce the χ^2 dependence of the resolution function by modeling the helix pulls with double Gaussian (see Appendix E). Together with data/MC correction of the double Gaussian shape parameters (see Appendix F), the fast simulator enables the calibration of the resolution functions in the following way:

1. Generate two sets of fast simulations from the MC samples used for the resolution function modeling: the *original* fast simulation with the helix pull models obtained in the signal MC and the *calibrated* fast simulation with the models corrected by data/MC correction.

Table 5.9: The results of $\delta\ell^{CP}$ and $\delta\ell^{\mathrm{tag}}$ fit in two fast simulations, one from original signal MC and the other calibrated with cosmic data. The fit uncertainties are estimated using the Hesse matrix. The $\delta\ell^{\mathrm{tag}}$ fit in the calibrated fast simulation results in an above EDM threshold.

	parameter	original fast sim	calibrated fast sim
R_{rec}	$f_{\mathrm{tail}}^{CP,0}$	0.0119 ± 0.0011	0.0393 ± 0.0023
	$f_{\mathrm{tail}}^{CP,1}$	0.0109 ± 0.0008	0.0327 ± 0.0018
	$s_{\mathrm{main}}^{CP,0}$	1.0154 ± 0.0025	0.9962 ± 0.0029
	$s_{\mathrm{main}}^{CP,1}$	0.0336 ± 0.0012	0.0287 ± 0.0018
	s_{tail}^{CP}	2.95 ± 0.08	2.207 ± 0.029
$\delta\ell^{CP}$ outlier	f_{outlier}	0.00067 ± 0.00007	0.00118 ± 0.00008
	s_{outlier} [μm]	$2.37 \pm 0.14 \times 10^3$	$1.96 \pm 0.09 \times 10^3$
R_{asc}	$f_{\mathrm{tail}}^{\mathrm{tag}}$	0.00277 ± 0.00018	0.00270 ± 0.00018
	$s_{\mathrm{main}}^{\mathrm{tag},0}$	0.969 ± 0.004	0.9846 ± 0.0034
	$s_{\mathrm{main}}^{\mathrm{tag},1}$	0.0825 ± 0.0014	0.0812 ± 0.0012
	$s_{\mathrm{tail}}^{\mathrm{tag},0}$	14.6 ± 0.6	14.4 ± 0.7
	$s_{\mathrm{tail}}^{\mathrm{tag},1}$	1.94 ± 0.14	1.96 ± 0.15
R_{np}	f_{δ}^{0}	0.728 ± 0.007	0.731 ± 0.007
	f_{δ}^{1}	-0.1494 ± 0.0021	-0.1509 ± 0.0023
	f_{p}^{0}	0.8823 ± 0.0026	0.8847 ± 0.0023
	f_{p}^{1} [$\mu\mathrm{m}^{-1}$]	-0.00171 ± 0.00007	-0.00162 ± 0.00007
	τ^{0}	1.000 ± 0.012	0.996 ± 0.012
	τ^{1}	0.2414 ± 0.0027	0.2446 ± 0.0027
	τ^{max}	5.73 ± 0.06	5.76 ± 0.06

2. Fit R_{rec} and R_{tag} to $\delta\ell^{CP}$ and $\delta\ell^{\mathrm{tag}}$ in the two sets of fast simulations. The original and calibrated fast simulations give the resolution functions for MC and data, respectively.

The obtained resolution function parameters are summarized in Tab. 5.9.

We reanalyze the MC with the resolution function obtained in the fast simulation without data/MC correction to validate the resolution functions obtained in this procedure.

Figure 5.18 shows the distributions of $\delta\ell^{CP}$ and $\delta\ell^{\mathrm{tag}}$. By comparing the curve and the distribution, we confirm that the resolution functions obtained in the fast simulation sufficiently explain the signal MC.

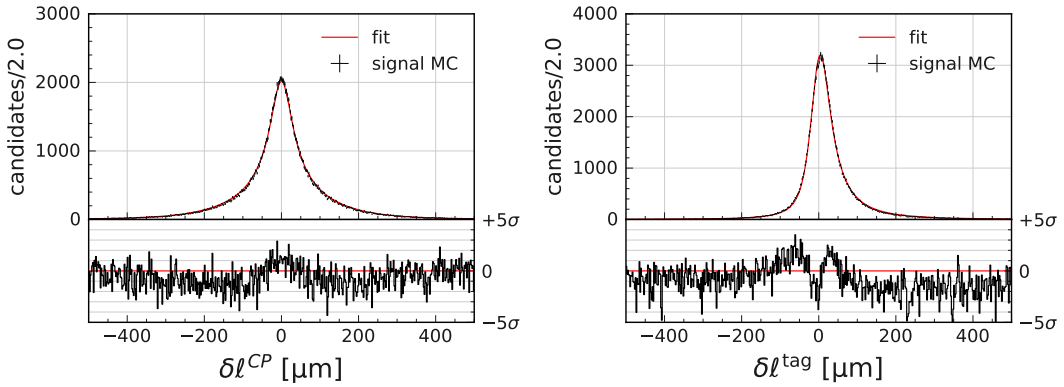


Figure 5.18: Comparison of $R_{\rm rec}$ and $R_{\rm tag}$ obtained in the fast simulation with $\delta\ell^{CP}$ and $\delta\ell^{\rm tag}$ distributions in 2M events of signal MC, corresponding to $> 400\,{\rm ab}^{-1}$. The lower subplot shows the pulls between the model and the histogram.

Table 5.10: The results of $\delta\ell^{\rm tag}$ fit in signal MC with resolution functions obtained in fast simulation. The MC input, $\Delta\tau$ fit, and the $\Delta t_{\rm true}$ fit are the same as Tab. 5.8.

proper-time difference	model	fit result [ps]
$(\ell^{CP} - \ell^{\rm tag}_{\rm true})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm rec} + G_{\rm outlier}$	1.5292 ± 0.0035
$(\ell^{CP}_{\rm true} - \ell^{\rm tag})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm tag}$	1.5086 ± 0.0031
$\Delta t = (\ell^{CP} - \ell^{\rm tag})/\beta\gamma c$	$E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm rec} \otimes R_{\rm tag} + G_{\rm outlier}$	1.520 ± 0.004

We also repeat the lifetime fit as summarized in Tab. 5.10. The resolution function and signal MC distribution are compared in Fig. 5.19. We observe some bias in the lifetime due to the small mismodeling of the resolution. We consider this deviation into systematic uncertainty in Section 5.5.5.

We confirm that the resolution functions obtained in the fast simulation sufficiently work in signal MC. Similarly, we expect that the calibrated resolution functions work for data.

5.2.3 Resolution of signal events in non-MR1 channel

We check the resolution of the non-MR1 signal to see if the resolution functions of MR1 obtained in $B^0 \to K^*(892)^0\gamma$ can apply to the resonance other than $K^*(892)^0$. Figure 5.20 compares the residual and the resolution function in signal MC. We confirm that the resolution function obtained in the MR1 signal sample can successfully explain the resolution of the non-MR1 signal. The results of lifetime fit are also summarized

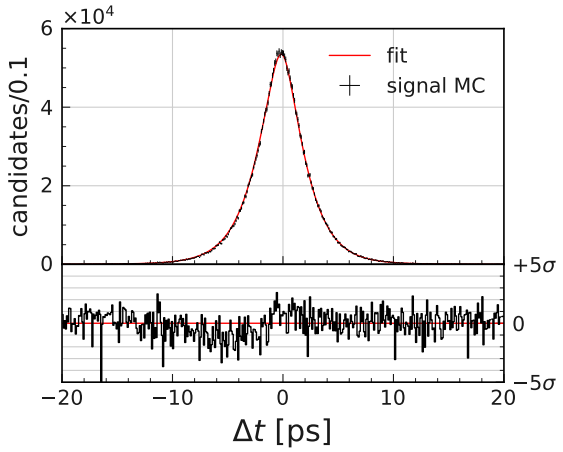


Figure 5.19: Comparison of Δt PDF obtained in fast simulation and Δt distribution in 2M events of signal MC, corresponding to $> 400\,\mathrm{ab}^{-1}$. The lower subplot shows the pulls between the model and the histogram.

Table 5.11: The results of lifetime fit for non-MR1 signal in $b \to s\gamma$ MC sample.

proper-time difference	model	fit result [ps]
MC input	(no fit)	1.519
$\Delta\tau$	E_f	1.510 ± 0.006
$\Delta t^\mathrm{true} = (\ell^{CP}_\mathrm{true} - \ell^\mathrm{tag}_\mathrm{true})/\beta\gamma c$	$E_\mathrm{f} \otimes R_\mathrm{k}$	1.510 ± 0.006
$(\ell^{CP} - \ell^\mathrm{tag}_\mathrm{true})/\beta\gamma c$	$E_\mathrm{f} \otimes R_\mathrm{k} \otimes R_\mathrm{rec} + G_\mathrm{outlier}$	1.526 ± 0.008
$(\ell^{CP}_\mathrm{true} - \ell^\mathrm{tag})/\beta\gamma c$	$E_\mathrm{f} \otimes R_\mathrm{k} \otimes R_\mathrm{tag}$	1.502 ± 0.007
$\Delta t = (\ell^{CP} - \ell^\mathrm{tag})/\beta\gamma c$	$E_\mathrm{f} \otimes R_\mathrm{k} \otimes R_\mathrm{rec} \otimes R_\mathrm{tag} + G_\mathrm{outlier}$	1.512 ± 0.009

in Tab. 5.11. We observe a slight bias in the true proper-time difference fit, i.e., from the reconstruction and selection. Otherwise, the behavior is similar to the MR1 signal. Therefore, we use the same resolution function as in MR1.

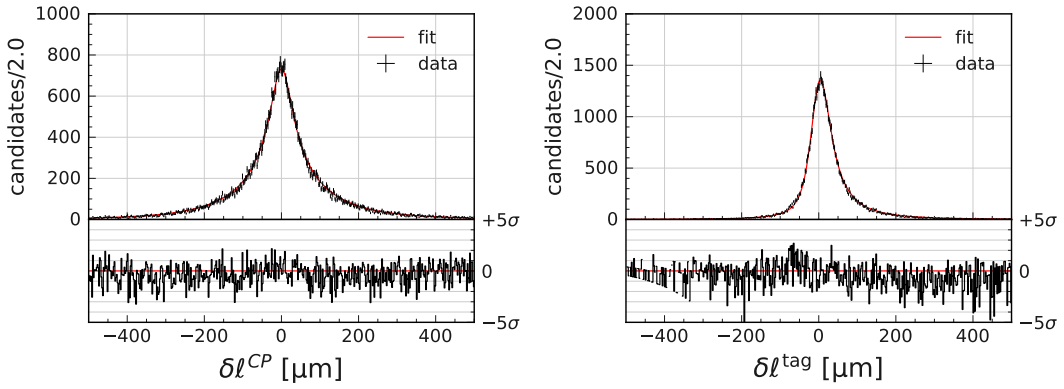


Figure 5.20: Comparison of $R_{\rm rec}$ and $R_{\rm tag}$ for MR1 obtained in the fast simulation with $\delta\ell^{CP}$ and $\delta\ell^{\rm tag}$ distributions of non-MR1 signal in 20M events of $b \to s\gamma$ MC sample, corresponding to 56 ab^{-1}. The lower subplot shows the pulls between the model and the histogram.

5.2.4 Resolution of $B\bar{B}$ background events

We model the resolution function in the $B\bar{B}$ background, $R_{B\bar{B}}$, by changing the parameters of $R_{\rm sig}$. Figure 5.21 illustrates the vertex reconstruction in the $B\bar{B}$ background events.

Since we can reconstruct K^0_S candidates with good purity, fake $B_{\rm sig}$ candidates in the $B\bar{B}$ background have the same resolution as the signal $B_{\rm sig}$ candidates. We confirm that the $R_{\rm rec}$ for the signal candidates agrees with the $\delta\ell^{CP}$ distribution of the $B\bar{B}$ background, as shown in Fig. 5.22.

On the other hand, fake $B_{\rm tag}$ candidates suffer from contamination by $B_{\rm sig}$ daughter

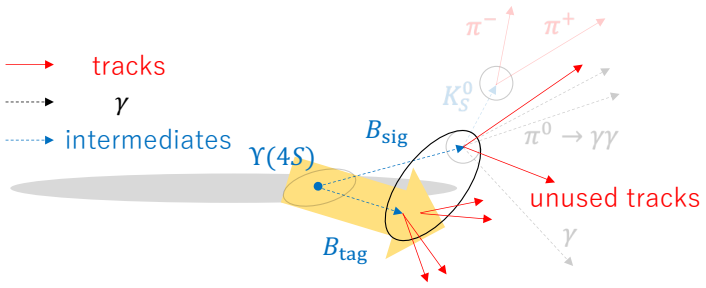


Figure 5.21: A schematics of vertex reconstruction in $B\bar{B}$ feed-down background. The fake $B_{\rm sig}$ vertex points to one of the B vertex thanks to the high purity of K^0_S. On the other hand, the fake $B_{\rm tag}$ vertex is heavily smeared by the mixed-up daughter tracks.

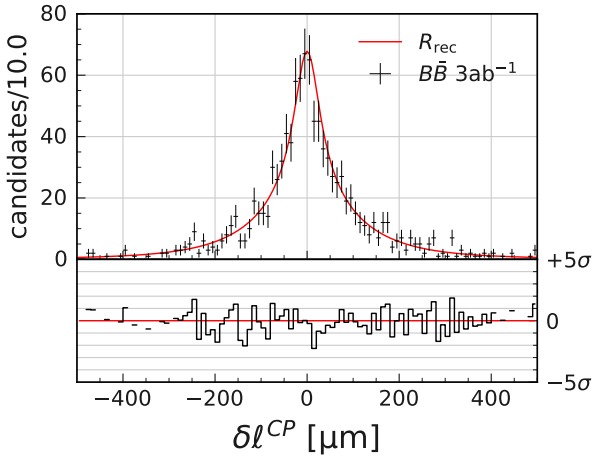


Figure 5.22: Background $\delta\ell^{CP}$ distribution in $B\overline{B}$ 3 ab^{-1} MC and $R_{\rm rec}$ for the signal candidates obtained in the original signal MC. The lower box shows the discrepancy between the distribution and the model.

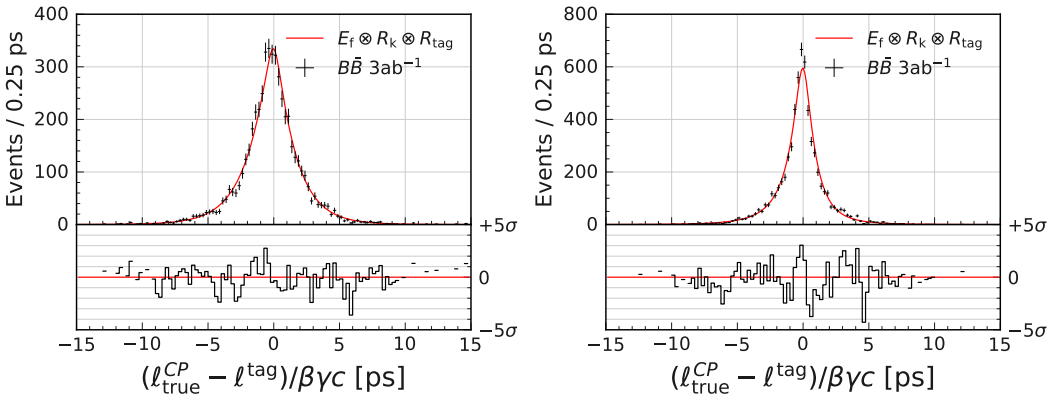
tracks. This biases the $B_{\rm tag}$ vertex closer to the $B_{\rm sig}$ vertex and shortens Δt. We model this effect with a lifetime shorter than the truth, $\tau_{B\bar{B}} < \tau_{B^0}$ in Eq. (5.2). We additionally change the parameters of $R_{\rm np}$ to account for the bias due to the contamination of $B_{\rm sig}$ daughters. These two sets of parameters are simultaneously fitted to $(\ell^{CP}_{\rm true} - \ell^{\rm tag})/\beta\gamma c$ using the $E_{\rm f} \otimes R_{\rm k} \otimes R_{\rm tag}$ PDF. Though the $B\overline{B}$ backgrounds contain a similar number of $B^0\overline{B}^0$ and B^+B^- events, we use m_{B^0} to calculate $R_{\rm k}$. The effect is negligible since the deviation is only 6×10^{-5} of m_{B^+}. We do not float $R_{\rm asc}$ since the detector resolution is expected to be the same as the signal sample in principle. We use 3 ab^{-1} $B\overline{B}$ MC samples for the fit, including the candidates outside of the $M_{\rm bc}$–ΔE signal-enhanced region to increase the statistics. We separate the $B\overline{B}$ feed-down and combinatorial as in Section 5.1.3 to account for the different sharpness of $(\ell^{CP}_{\rm true} - \ell^{\rm tag})/\beta\gamma c$ distribution among the two components, as shown in Fig. 5.23. This is related to the number of $B_{\rm sig}$ tracks used in the $B_{\rm tag}$ vertexing: a few for feed-down reconstruction and possibly many for combinatorial reconstruction since $B_{\rm sig}$ can decay without γ or π^0. The fit result is summarized in Tab. 5.12. We fix the $f^1_{\rm p}$ parameters to 0 in $B\overline{B}$ combinatorial fit since we obtain small but positive values in the fit. Figure 5.23 compares the fitted model and the distribution. We confirm that the fitted models agree well with the MC.

We check the combined model using the signal extraction fit result in Section 5.1.3. Figure 5.24 compares the model and Δt distribution of $B\overline{B}$ MC sample in $M_{\rm bc}$–ΔE signal-enhanced region. We confirm that the overall model successfully explains the $B\overline{B}$ background Δt in MC.

Table 5.12: The results of $(\ell^{CP}_{\text{true}} - \ell^{\text{tag}})/\beta\gamma c$ fit in $B\bar{B}$ sample ($3\,\text{ab}^{-1}$). The fit uncertainties are estimated using the Hesse matrix.

parameter		$B\bar{B}$ feed-down	$B\bar{B}$ combinatorial
lifetime	$\tau_{B\bar{B}}$	1.336 ± 0.020	0.63 ± 0.06
	f^0_δ	1.21 ± 0.04	0.98 ± 0.04
	f^1_δ	−0.213 ± 0.012	−0.134 ± 0.008
	f^0_{p}	0.766 ± 0.022	0.638 ± 0.018
R_{np}	f^1_{p} [μm^{-1}]	−0.0028 ± 0.0007	0 (fixed)
	τ^0	2.84 ± 0.12	4.9 ± 0.4
	τ^1	0.312 ± 0.010	0.116 ± 0.027
	τ^{max}	7.24 ± 0.18	8.2 ± 1.0

We also calibrated the resolution function parameters using fast simulations. The obtained parameters are summarized in Tab. 5.13.



(a) $B\bar{B}$ feed-down (b) $B\bar{B}$ combinatorial

Figure 5.23: $(\ell^{CP}_{\text{true}} - \ell^{\text{tag}})/\beta\gamma c$ distribution and fitted $E_{\text{f}} \otimes R_{\text{k}} \otimes R_{\text{tag}}$ in $B\bar{B}$ sample ($3\,\text{ab}^{-1}$). The candidates outside of the M_{bc}–ΔE signal-enhanced region are included to increase the statistics. The lower box shows the discrepancy between the distribution and fit model.

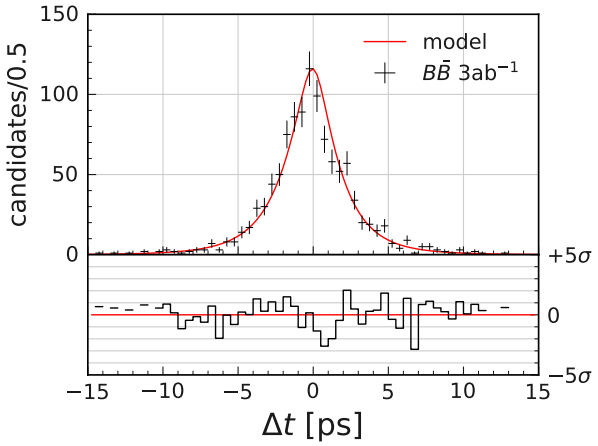


Figure 5.24: The $B\bar{B}$ background Δt model and the background Δt distribution in $3\,\text{ab}^{-1}$ $B\bar{B}$ MC samples. The lower subplot shows the pulls between the model and the histogram.

Table 5.13: The results of $(\ell^{CP}_{\text{true}} - \ell^{\text{tag}})/\beta\gamma$ fit in two fast-simulated $B\overline{B}$ samples, one from the original signal MC and the other calibrated with cosmic data. The fit uncertainties are estimated using the Hesse matrix.

(a) $B\overline{B}$ feed-down

parameter		original fast sim	calibrated fast sim
lifetime	$\tau_{B\bar{B}}$	1.317 ± 0.025	1.300 ± 0.028
	f^0_δ	1.232 ± 0.034	1.24 ± 0.06
	f^1_δ	-0.268 ± 0.011	-0.246 ± 0.025
	f^0_{p}	0.728 ± 0.020	0.715 ± 0.023
R_{np}	f^1_{p} [μm^{-1}]	-0.00205 ± 0.00020	-0.00151 ± 0.00025
	τ^0	1.29 ± 0.23	3.12 ± 0.28
	τ^1	0.693 ± 0.024	0.415 ± 0.033
	τ^{max}	9.24 ± 0.26	9.7 ± 0.7

(b) $B\overline{B}$ combinatorial

parameter		original fast sim	calibrated fast sim
lifetime	$\tau_{B\bar{B}}$	0.667 ± 0.023	0.489 ± 0.028
	f^0_δ	1.035 ± 0.013	0.924 ± 0.020
	f^1_δ	-0.120 ± 0.006	-0.145 ± 0.006
	f^0_{p}	0.672 ± 0.016	0.622 ± 0.013
R_{np}	f^1_{p} [μm^{-1}]	0 (fixed)	
	τ^0	5.75 ± 0.29	5.47 ± 0.18
	τ^1	0.147 ± 0.019	0.204 ± 0.019
	τ^{max}	9.2 ± 0.4	9.7 ± 0.4

5.2.5 Resolution of continuum background events

We model Δt PDF for $q\overline{q}$ background using the $q\overline{q}$-rich $M_{\rm bc}$–ΔE sideband shown in Fig. 4.11. We have developed an empirical PDF based on the double Gaussian model used in the $B^0 \to K^0_S K^0_S K^0_S$ analysis [47] to better explain the MC samples:

$$\begin{aligned} P_{q\overline{q}}\Big(\Delta t; X^2, \sigma_{\Delta t}\Big) = (1-f_2)(1-f_3)G\Big(\Delta t; \mu = 0, \sigma = (s^0_1 + s^1_1 X^2)\sigma_{\Delta t}\Big) \\ + f_2(1-f_3)G\Big(\Delta t; \mu = 0, \sigma = (s^0_2 + s^1_2 X^2)\sigma_{\Delta t}\Big) \\ + f_3 G(\Delta t; \mu = 0, \sigma = s_3\sigma_{\Delta t}), \end{aligned} \tag{5.35}$$

where X^2 is the average reduced χ^2 and $\sigma_{\Delta t}$ is the Δt uncertainty,

$$X^2 = \frac{\chi^2_{K^0_S}/1.6 + (\chi^2/\nu_{\rm d.f.})^{\rm tag}}{2}, \tag{5.36}$$

$$\sigma_{\Delta t} = \frac{\sqrt{(\sigma^{CP}_\ell)^2 + (\sigma^{\rm tag}_\ell)^2}}{\beta\gamma}. \tag{5.37}$$

The factor of 1.6 in X^2 accounts for the effective number of degrees of freedom for the K^0_S vertex in the B^0 vertex fit. The seven free parameters, namely f_2, f_3, s^0_1, s^1_1, s^0_2, s^1_2, and s_3, are determined in the fit to the sideband.

To better constrain the fit parameters, we remove the continuum suppression (see Section 4.3) and combine MR1, non-MR1, and control channels. The $q\overline{q}$ purity in the acquired sideband sample is 98.5% in MC as shown in Fig. 5.25. We confirm in Fig. 5.26 that Δt distribution and its width depending on the X^2 in the sideband agree with $q\overline{q}$ samples in the signal-enhanced region, either in MR1, non-MR1, or control channel.

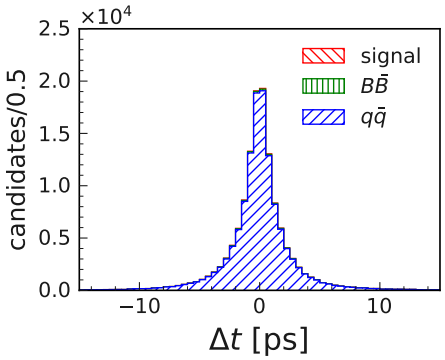


Figure 5.25: Δt distribution breakdown of generic MC in $M_{\rm bc}$–ΔE sideband.

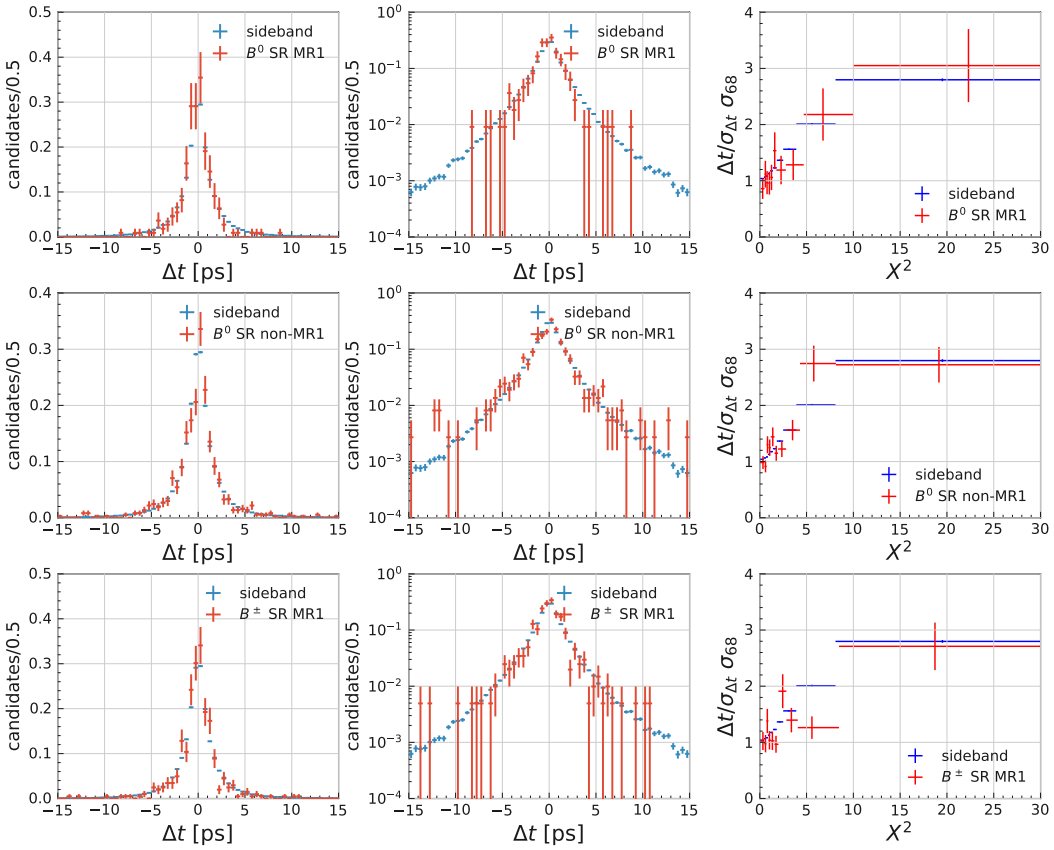


Figure 5.26: Δt distribution in linear/log scale and X^2 dependence of σ_{68} (half-width between 16% and 84% quantiles) of $\Delta t/\sigma_{\Delta t}$ distribution are compared between signal region $q\overline{q}$ and sideband in generic MC. Top for MR1 B^0 candidates, middle for non-MR1 B^0 candidates, and bottom for B^+ candidates. The Δt distribution in the sideband agrees with that of $q\overline{q}$ in the signal region. The sideband also represents the X^2 dependence of $q\overline{q}$ in the signal region.

Having defined the fit model and the sideband, we analyze the sideband in the data. Also in data, by comparing Δt and X^2 distribution in Fig. 5.27 and the X^2 dependence of Δt width in Fig. 5.28 (together with the MC), we have confirmed that the Δt resolution model remains consistent even with the continuum suppression and in different channels. Tab. 5.14 summarizes the fit parameters, and Fig. 5.29 shows the Δt distribution and the fit model, together with the results in MC. We confirm that the resolution function models for MC and data agree with the distribution in MC and data, respectively.

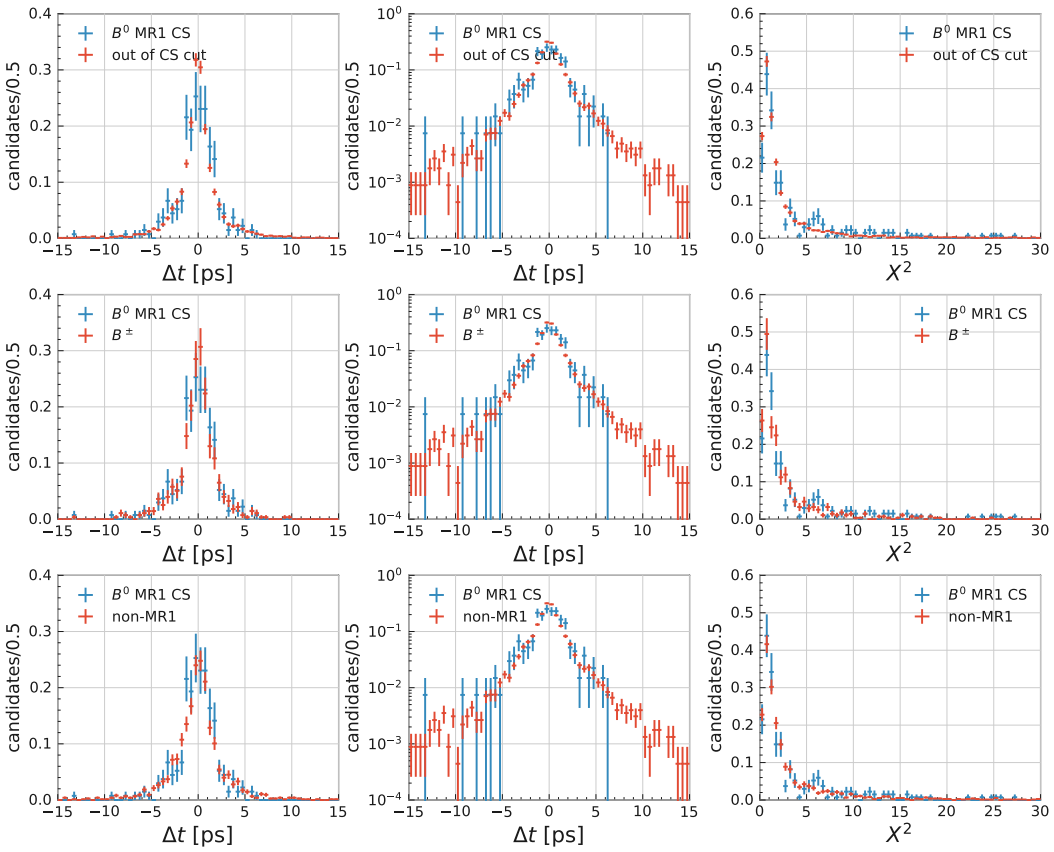


Figure 5.27: Δt (in linear and log scale) and X^2 distribution in sideband data are compared for in/out of the continuum suppression, B^0/B^+, and MR1/non-MR1. The Δt and X^2 distributions agree in general. The slightly sharper Δt distribution for B^0 out of the continuum suppression selection or B^+ can be explained by the slightly smaller X^2.

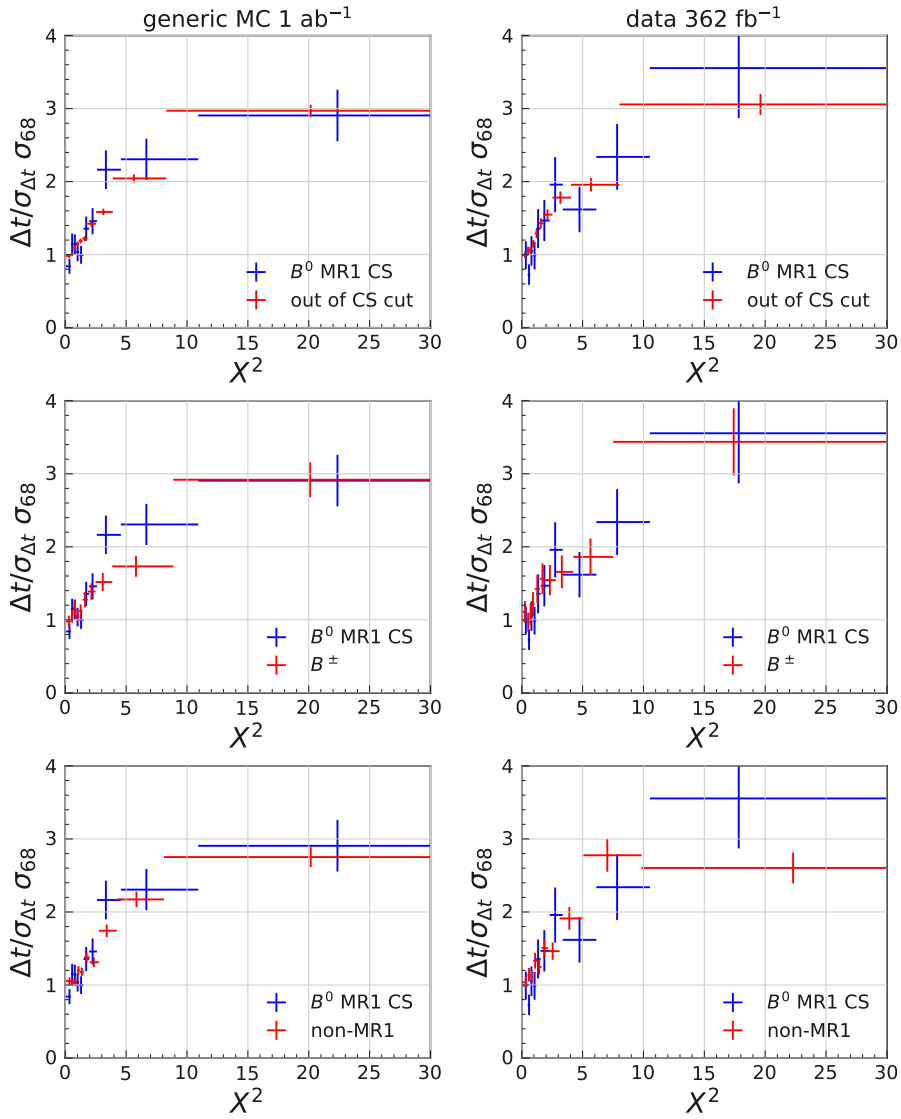


Figure 5.28: X^2 dependence of σ_{68} (half-width between 16% and 84% quantiles) of $\Delta t/\sigma_{\Delta t}$ distribution in ΔE-$M_{\rm bc}$ sideband are compared. Top for in/out of the continuum suppression cut, middle for B^0/B^+, and bottom for MR1/non-MR1. Left for generic MC and right for data.

Table 5.14: The results of Δt fit in ΔE-$M_{\rm bc}$ sideband, in MC and data.

fit parameter	MC sideband	data sideband
f_2	0.208 ± 0.006	$0.239^{+0.013}_{-0.012}$
f_3	0.0080 ± 0.0004	$0.0097^{+0.0009}_{-0.0008}$
s_1^0	0.916 ± 0.005	0.930 ± 0.009
s_1^1	0.0843 ± 0.0019	$0.0825^{+0.0036}_{-0.0035}$
s_2^0	1.598 ± 0.027	$1.61^{+0.05}_{-0.04}$
s_2^1	$0.552^{+0.014}_{-0.013}$	$0.478^{+0.021}_{-0.020}$
s_3	23.0 ± 0.7	$17.9^{+0.9}_{-0.8}$

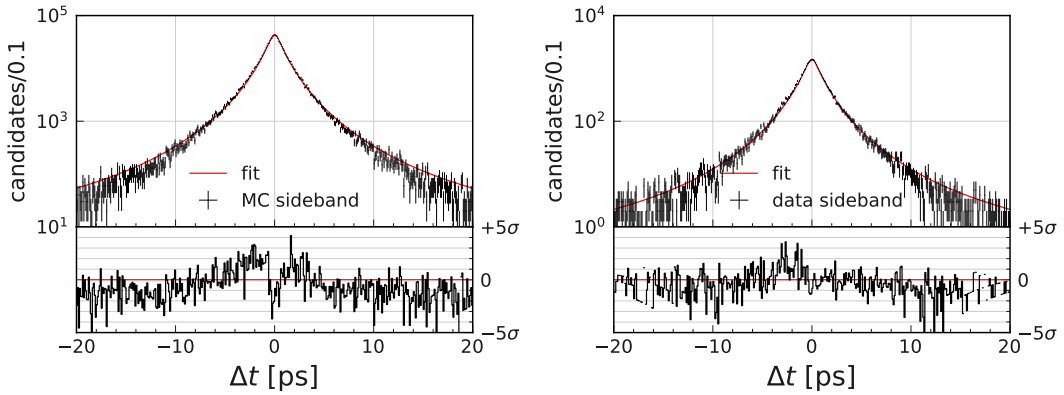


Figure 5.29: Δt distribution and fitted model in sideband region of 1 ab^{-1} generic MC (left) and 362 fb^{-1} data (right). The resolution function models agree with the distribution both in MC and data.

5.3 Validation

With the event-by-event fractions obtained in Section 5.1 and the resolution functions obtained in Section 5.2, we can now fit the S_{CP} and A_{CP} with the PDFs in Eqs. (5.1) and (5.3). Note that the resolution model is fixed except for the event-by-event conditional variables. We also fix the lifetime and mixing parameters to the PDG values [46],

$$\tau_{B^0} = 1.519 \pm 0.004\,\text{ps}, \tag{5.38}$$

$$\Delta m_d = 0.5065 \pm 0.0019\,\text{ps}^{-1}, \tag{5.39}$$

where these uncertainties are accounted as a source of systematic uncertainty in Section 5.5.5. Therefore, S_{CP} and A_{CP} are the only free parameters in the fit. In this section, we describe various validations for this fit.

5.3.1 Linearity test in toy Monte Carlo

We have developed a toy Monte Carlo generator according to the PDF used in our fit model. We test the linearity of *CP* fit using this toy MC by varying either A_{CP} or S_{CP} from -1 to 1 in steps of 0.2 while fixing the other to 0. For this purpose, we only perform *CP* fit, where the signal and background fractions are fixed at the generator value in the fit.

The pull shape parameters of fit results to 1,000 pseudo-experiments with 362 fb^{-1} each are summarized in Fig. 5.30(a). We observe mean bias for A_{CP} and S_{CP} and uncertainty overestimation in S_{CP}. This is due to the low statistics, as these effects are gone in the results with 4 ab^{-1} in Fig. 5.30(b). We do not observe any significant deviation from the expectation and confirm the healthiness of the fitter.

5.3.2 Linearity test in full simulation

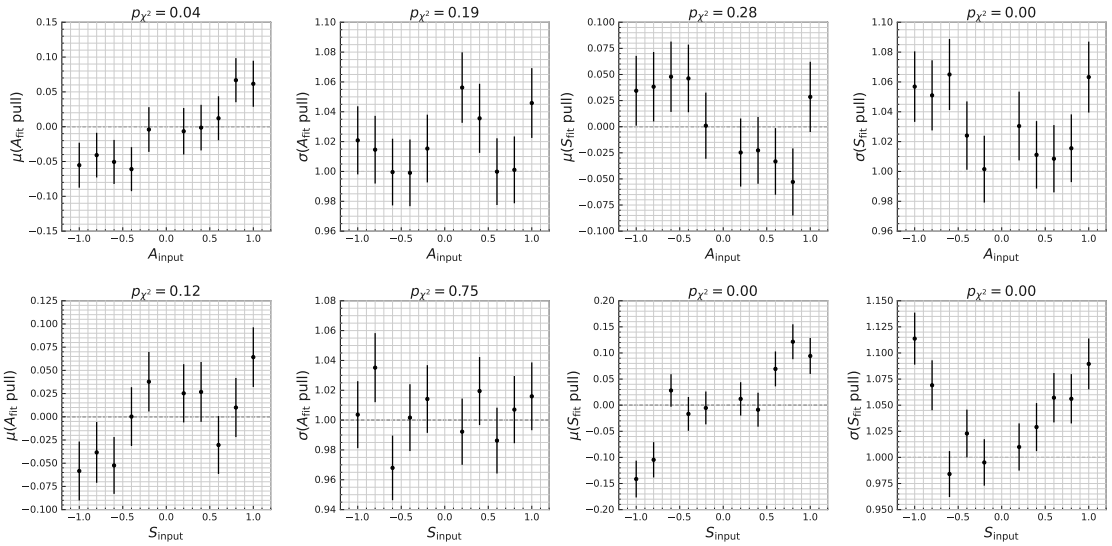
To test the entire model including the resolution function, we generate fully-simulated signal samples with various *CP*-violation inputs and perform *CP* fit in each sample.

We modify the `SVP_CP` model in EvtGen [38] according to Eqs. (1.4) and (1.5) and vary either A_{CP} or S_{CP} in the same way as in Section 5.3.1. We generate 50,000 events at each configuration and reconstruct and select the signal events as described in Chapter 4. We confirm the correct *CP* asymmetry generation with the modified `SVP_CP` model by checking the linearity of *CP*-violation parameters fitted to $\Delta\tau$ and q_{true} in the generator truth using the PDF in Eq. (1.20), as shown in Fig. 5.31.

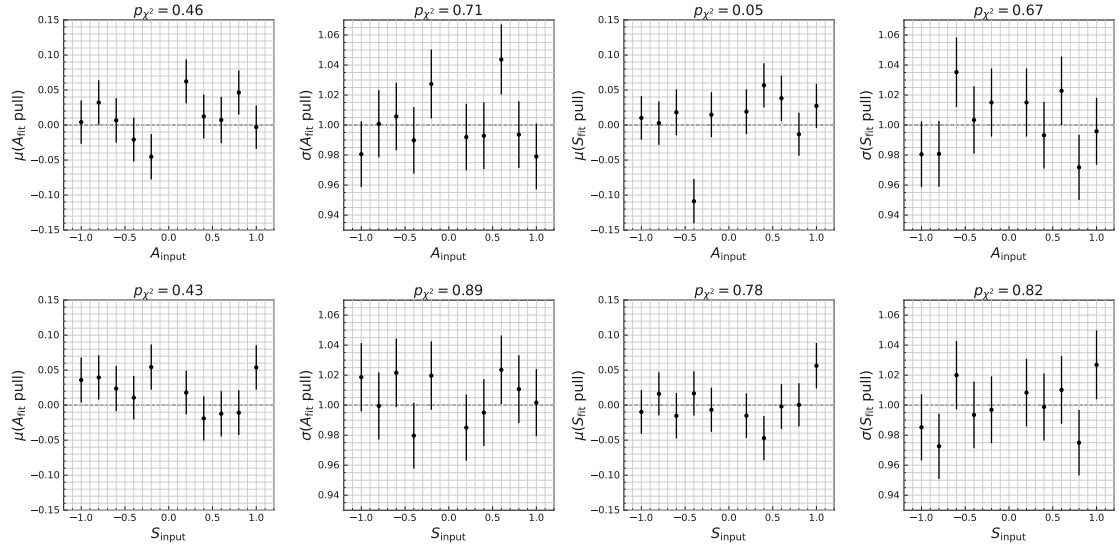
We test the linearity of *CP*-violation parameters fitted to reconstructed Δt and q with only the signal component in Eqs. (5.1) and (5.3). Figure 5.32 summarizes the fit results. We do not see any discrepancy and confirm that the modeling of Δt and q does not bring significant bias.

5.3.3 Resolution in the control sample

To validate the resolution model in data, we use $B^+ \to K^0_S\pi^+\gamma$ sample to check the residual between the B^+ vertex with and without π^+ tracks, $\ell^{CP} - \ell^{CP}_{\pi^+}$. Since the right π^+ track directly comes from the B^+ decay vertex, we can estimate the decay vertex with much better resolution than in the original vertex reconstruction using only K^0_S trajectory. Hence, we can utilize $\ell^{CP} - \ell^{CP}_{\pi^+}$ as an alternative for the residual from the true decay vertex, $\delta\ell^{CP}$, without using truth information. We compare $\ell^{CP} - \ell^{CP}_{\pi^+}$ and $\delta\ell^{CP}$ in control MC. The distribution of $\ell^{CP} - \ell^{CP}_{\pi^+}$ shows a sharper peak at 0, which is expected from the



(a) 362 fb^{-1}



(b) 4 ab^{-1}

Figure 5.30: The mean μ and the standard deviation σ of pull distribution for various *CP*-violation inputs. We fit 1,000 toy MC samples in each input. The titles show the p-values of χ^2.

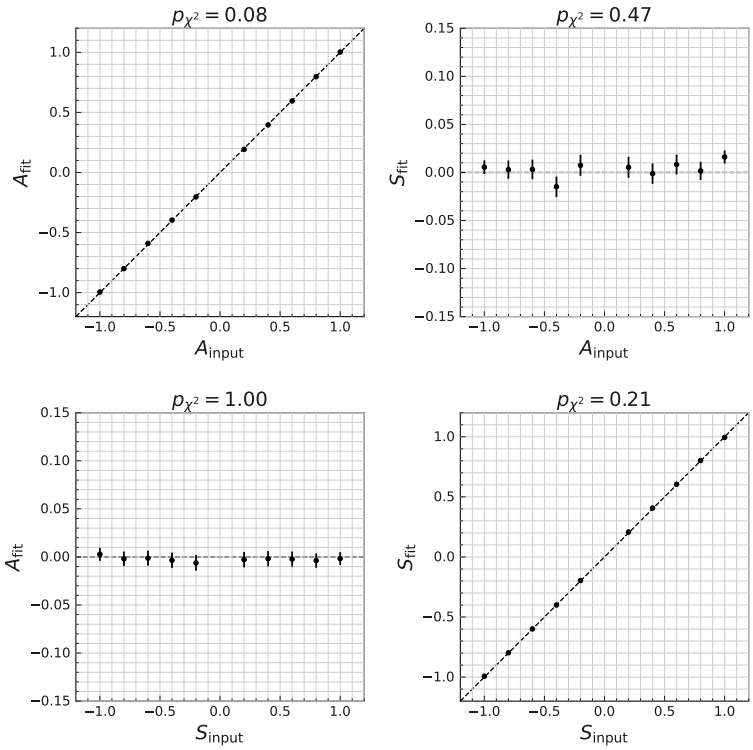


Figure 5.31: Comparison of generated and fitted *CP* asymmetries. We use the modified `SVP_CP` model in the generation, and fit A_{CP} and S_{CP} to $\Delta\tau$ and q_{true} with Eq. (1.20). The titles show the p-values of χ^2.

correlation between the two vertex reconstructions. We confirm the excellent agreement in the tail part of the distribution. Thus, we can test the resolution in data.

Figure 5.34 compares ${}_s\mathcal{P}lot$ of the residual and the resolution model obtained in the calibrated fast simulation. We confirm that the resolution model agrees with the residual distribution in the tail part.

5.3.4 Lifetime fit in data

We perform a lifetime fit in the data. Figure 5.35 shows ${}_s\mathcal{P}lot$ of Δt and the fitted curve. We observe no significant deviation between the data and the model. The results are summarized in Tab. 5.15. The obtained lifetimes are consistent with the PDG values [46] within the uncertainties.

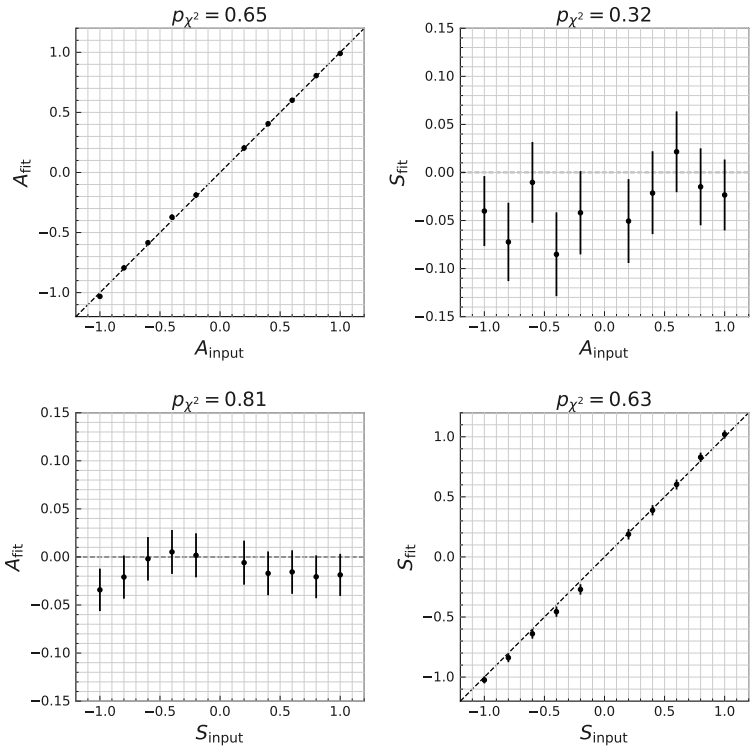


Figure 5.32: Comparison of generated and fitted *CP* asymmetries. We use the modified `SVP_CP` model in the generation, and fit A_{CP} and S_{CP} to reconstructed Δt and q with the signal model in Eqs. (5.1) and (5.3). The titles show the p-values of χ^2.

5.3.5 Calibration of bias from non-primary decay vertices

We use the incomplete D decay table in the MC generation. For example, some of the $O(0.1\%)$ semi-leptonic modes in the D^+ decay table are missing, e.g., $K_1(1270)^0 e^+\nu$ or $\pi^+\pi^- e^+\nu$. This data-MC difference brings mismodeling of the bias from non-primary decay vertices, i.e., the width of E_{p} and E_{n} in R_{np}. To calibrate this possible difference, we introduce scale factor s_{np} for τ_{np} as

$$R_{\text{np}}\left(\delta\ell^{\text{tag}}; (\chi^2/\nu_{\text{d.f.}})^{\text{tag}}, \sigma_\ell^{\text{tag}}, s_{\text{np}}\right) = f_\delta \delta(\delta\ell^{\text{tag}}) + (1-f_\delta) f_{\text{p}} E_{\text{p}}(\delta\ell^{\text{tag}}; \tau = s_{\text{np}}\tau_{\text{np}}\sigma_\ell^{\text{tag}}) + (1-f_\delta)(1-f_{\text{p}}) E_{\text{n}}(\delta\ell^{\text{tag}}; \tau = s_{\text{np}}\tau_{\text{np}}\sigma_\ell^{\text{tag}}). \tag{5.40}$$

We fit Δt distribution floating B-meson lifetime and s_{np}. We separately fit B^0 and B^+ to consider the difference of branching fractions decaying into charged and neutral D. We merge the MR1 and non-MR1 signals for this fit. The fitted s_{np}'s are summarized in

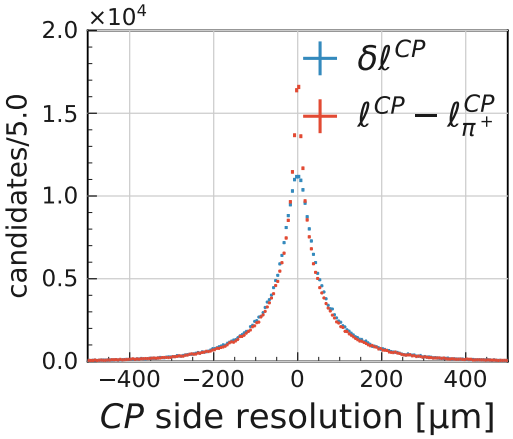


Figure 5.33: Comparison between $\ell^{CP} - \ell^{CP}_{\pi^+}$ and $\delta\ell^{CP}$ in control MC.

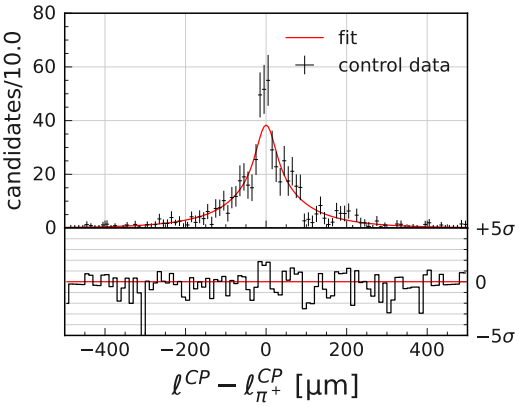
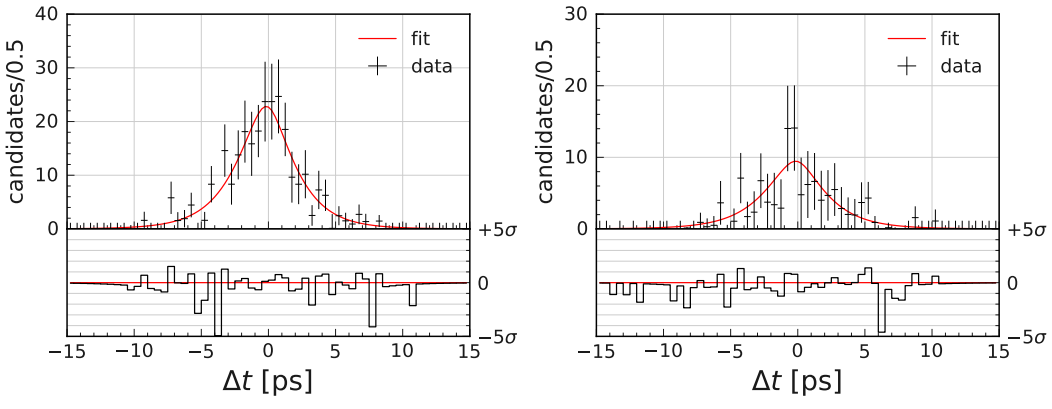


Figure 5.34: ${}_s\mathcal{P}lot$ of the residual between the B^+ vertex with and without π^+ tracks, together with the resolution model obtained in the calibrated fast simulation.

Tab. 5.16. The results in both signal and control samples are consistent with 1, i.e., no correction. We fix s_{np} to 1 and consider the fitted uncertainty as a source of systematic uncertainty in Section 5.5.5.

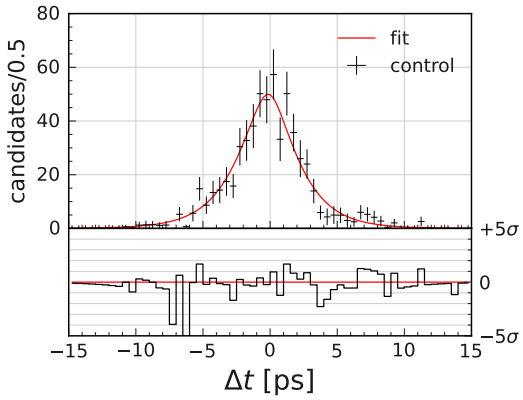
5.3.6 *CP* fit in the control sample

We perform the *CP* fit with the control sample ($B^+ \to K^0_S\pi^+\gamma$), where S_{CP} and A_{CP} are expected to be 0. We use 778 events in the TD category and 379 events in the TI



(a) MR1

(b) non-MR1



(c) control

Figure 5.35: ${}_s\mathcal{P}lot$ of Δt in 362 fb^{-1} data together with the fitted curve.

category. We obtain

$$S_{CP}(K^0_S\pi^+\gamma) = 0.11 \pm 0.14, \text{and} \tag{5.41}$$

$$A_{CP}(K^0_S\pi^+\gamma) = 0.03 \pm 0.08. \tag{5.42}$$

The fit result is consistent with the null CP asymmetry. Figure 5.36 compares the ${}_s\mathcal{P}lot$ and fitted curve, and we confirm that the model agrees with the data well.

Table 5.15: The results of lifetime fit in 362 fb^{-1} data.

sample	fitted lifetime [ps]	PDG value
MR1	1.55 ± 0.14	1.519 ± 0.004
non-MR1	1.58 ± 0.24	
control	1.65 ± 0.09	1.638 ± 0.004

Table 5.16: The lifetime and s_{np} fit in 362 fb^{-1} data.

sample	fitted lifetime [ps]	s_{np}
MR1 + non-MR1	1.56 ± 0.12	1.00 ± 0.24
control	1.65 ± 0.10	1.04 ± 0.22

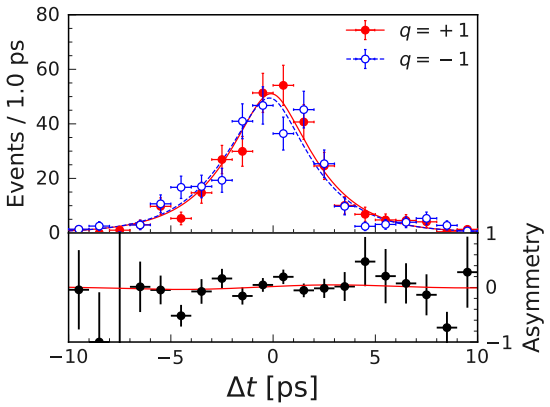
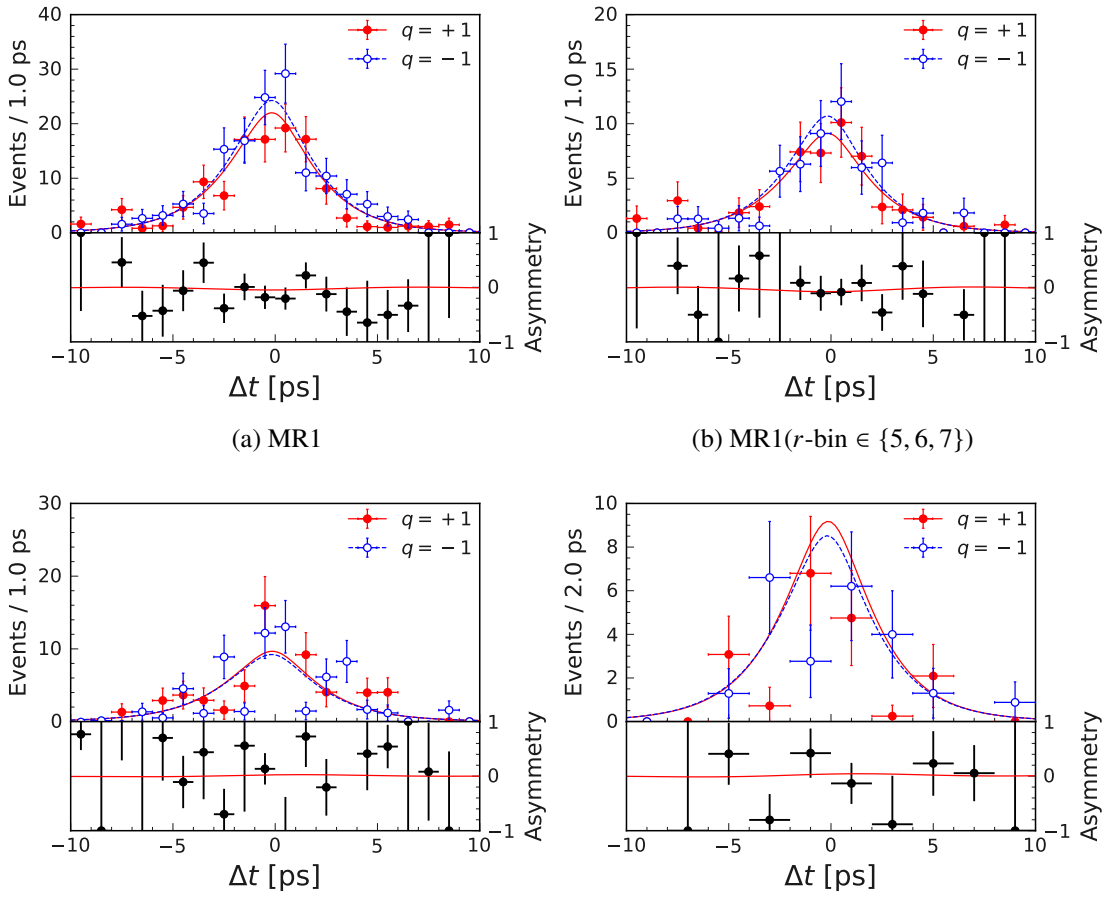


Figure 5.36: ${}_s\mathcal{P}lot$ of Δt for $q = \pm 1$ in 362 fb^{-1} control data, together with the fitted curve. The lower plot shows the asymmetry, defined as $(N(q = +1) - N(q = -1))/(N(q = +1) + N(q = -1))$.

5.4 Results

We perform the *CP* fit in $B^0 \to K^0_S \pi^0 \gamma$ decays separately for MR1 and non-MR1 channels. We use 369 (449) events in the TD category and 179 (220) events in the TI category in MR1 (non-MR1) channel. The results are summarized in Tab. 5.17. Figure 5.37 compares the ${}_s\mathcal{P}lot$ and fitted curve, and also shows the *CP* asymmetry. We also show the plots of the events with high r-bin to confirm that we do not see *CP* asymmetry even with the small dilution.



(a) MR1

(b) MR1(r-bin $\in \{5, 6, 7\}$)

(c) non-MR1

(d) non-MR1(r-bin $\in \{5, 6, 7\}$)

Figure 5.37: ${}_s\mathcal{P}lot$ of Δt for $q = \pm 1$ and its asymmetry in 362 fb^{-1} data, together with the fitted curve. The lower plot shows the asymmetry, defined as $(N(q = +1) - N(q = -1))/(N(q = +1) + N(q = -1))$.

Table 5.17: The *CP*-violation parameters fitted to 362 fb^{-1} data.

channel	S_{CP}	A_{CP}
MR1	$0.00^{+0.27}_{-0.26}$	-0.10 ± 0.13
non-MR1	$0.04^{+0.45}_{-0.44}$	0.06 ± 0.25

5.5 Systematic uncertainties

The systematic uncertainties are evaluated with the following three methods:

- *Refitting*
  We propagate the statistical fluctuation of the fit model by refitting the data 1,000 times. The systematic uncertainty is then determined as the standard deviation of the fitted *CP*-violation parameters.
  Especially for the case of fixed parameters in the fit model, we account for their statistical fluctuations by adding random Gaussian deviations. The standard deviation of the Gaussian is obtained from the uncertainty of the corresponding parameter. In case the parameters are determined with MC, we conservatively double the deviation (2σ) instead of 1σ.
- *Toy MC*
  We evaluate the possible model bias by generating pseudo-experiments with an alternative model. We conservatively select the larger value of the shift of the mean (as the single-sided uncertainty) and the uncertainty of the mean (as the double-sided uncertainty). We generated 10,000 pseudo-experiments with 4 ab^{-1} integrated luminosity.[1]
- *Testing*
  For the other cases, we analyze the data or MC with a test configuration and take the difference from the nominal fit result as their systematic uncertainties.

The systematic uncertainties in this measurement are summarized in Tab. 5.18. In the following sections, we explain the evaluation of each component.

[1]Before obtaining the results from the data, we choose $S_{CP} = -0.2$ and $A_{CP} = 0$ as input to be close to the world average [46] in MR1. We use the evaluation with this parameter setting since the obtained *CP*-violation parameters are smaller than these inputs, and the evaluation is conservative enough.

Table 5.18: The summary of systematic uncertainties.

	MR1		non-MR1	
source	δS_{CP}	δA_{CP}	δS_{CP}	δA_{CP}
reconstruction of decay	± 0.017	± 0.015	± 0.083	± 0.047
flavor tagging	± 0.005	$^{+0.009}_{-0.012}$	± 0.008	$^{+0.009}_{-0.013}$
decay vertex measurement	± 0.021	± 0.009	± 0.023	± 0.036
event-by-event fractions	± 0.003	$^{+0.003}_{-0.004}$	± 0.032	± 0.013
proper-time difference modeling	± 0.014	± 0.009	± 0.032	± 0.013
$B\bar{B}$ background asymmetry	$^{+0.010}_{-0.021}$	± 0.022	$^{+0.023}_{-0.015}$	$^{+0.033}_{-0.032}$
tag-side interference	< 0.001	$+0.002$	-0.001	-0.001
total	$^{+0.033}_{-0.037}$	$^{+0.031}_{-0.032}$	$^{+0.100}_{-0.098}$	$^{+0.070}_{-0.071}$

Table 5.19: The systematic uncertainties from the reconstruction of decay.

	MR1		non-MR1	
source	δS_{CP}	δA_{CP}	δS_{CP}	δA_{CP}
p scale for $\pi^{\pm}$	± 0.0017	± 0.0016	± 0.0024	± 0.0019
E scale for γ	± 0.0170	± 0.0148	± 0.0825	± 0.0467
total	± 0.0171	± 0.0149	± 0.0825	± 0.0467

5.5.1 Reconstruction

During the event reconstruction, we apply MC/data scaling factors to the track momentum and the ECL cluster energy, as noted in Sections 4.2.1 to 4.2.3. We evaluate the systematic uncertainties by testing the reconstruction of the data with the scaling factor shifted up and down by the measurement uncertainty. We conservatively use the larger deviation of the upper and lower shift as the systematic uncertainty. The results are summarized in Tab. 5.19.

5.5.2 Flavor tagging

We introduce the performance of the flavor tagger in the PDF in Section 4.4. We evaluate the uncertainty from w and Δw by refitting. We evaluate the effect of ignored μ in Eq. (4.9) by toy MC with the effect of μ. The results are summarized in Tab. 5.20.

Table 5.20: The systematic uncertainties related to the flavor tagger.

	MR1		non-MR1	
source	δS_{CP}	δA_{CP}	δS_{CP}	δA_{CP}
w and Δw	± 0.0045	± 0.0076	± 0.0070	± 0.0085
μ	± 0.0008	-0.0084	± 0.0009	-0.0088
total	± 0.0046	$^{+0.0076}_{-0.0113}$	± 0.0070	$^{+0.0085}_{-0.0122}$

5.5.3 Decay vertex measurement

We model the measurement of the B decay vertices with the resolution functions. However, some possible biases are left unmodeled. We consider the following four sources:

1. *detector misalignment*
   We evaluate the effect of the possible detector misalignment by testing with the full simulation in the four scenarios of misalignment configuration. We generated 1M signal events with $S_{CP} = A_{CP} = 0$, and the detector response is fully simulated in each scenario. We calculate the differences from the nominal alignment configuration sample in four scenarios and then add them in quadrature to obtain the contribution to the systematic uncertainty.

2. *IP measurement*
   In the B-decay vertex fit, we constrain the $\Upsilon(4S)$ decay vertex to the IP measured in $e^+e^- \to \mu^+\mu^-$ events (see Section 4.5). We quote the uncertainty from this measurement evaluated for $\sin 2\phi_1$ measurement in $B^0 \to J/\psi K^0_S$ decay [62] as the systematic uncertainty.

3. *helix uncertainty correction*
   The helix uncertainties used in the vertex fit are corrected during the reconstruction (see Sections 4.2.1 and 4.5). We evaluate the effect by testing the reconstruction of the data without correction. To avoid overestimation by the statistical fluctuation from the in and out of the events, we use the same events for both cases. Instead, we repeat the same evaluation in signal MC allowing the in and out of the events to address the effect ignored in the above procedure with much higher statistics. The difference is small, so we take the fit uncertainty. We add the uncertainties evaluated in data and signal MC in quadrature as the systematic uncertainty.

4. *vertex quality selection*

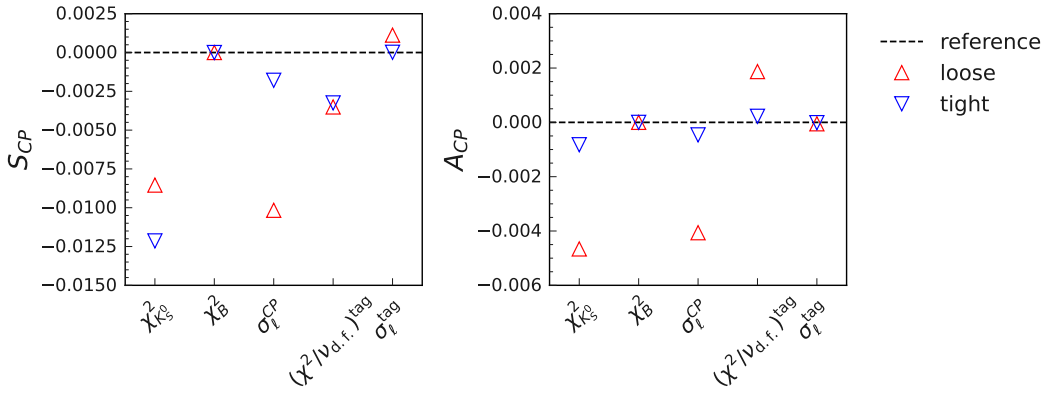


Figure 5.38: The deviation of S_{CP} and A_{CP} in the control sample with loosened and tightened vertex quality selections from nominal thresholds.

We only use events with good vertex quality in the TD fit (see Section 4.5). The possible data–MC difference in the correlation between vertex quality and Δt leads to the mismodeling of Δt in data. We evaluate the effect by testing the *CP* fit in the data in ten configurations with selection thresholds tightened or loosened by 20%:

- $\chi^2_{K^0_S} < 30$ changed to $\chi^2_{K^0_S} < 24$ or 36,
- $\chi^2_B < 100$ changed to $\chi^2_B < 80$ or 120,
- $\sigma_\ell^{CP} < 500$ µm changed to $\sigma_\ell^{CP} < 400$ µm or 600 µm,
- $(\chi^2/\nu_{\mathrm{d.f.}})^{\mathrm{tag}} < 100$ changed to $(\chi^2/\nu_{\mathrm{d.f.}})^{\mathrm{tag}} < 80$ or 120, and
- $\sigma_\ell^{\mathrm{tag}} < 500$ µm changed to $\sigma_\ell^{\mathrm{tag}} < 400$ µm or 600 µm.

We use the control channel to increase the statistics and use the same evaluation both for MR1 and non-MR1. The differences from the nominal thresholds are shown in Fig. 5.38. We take the larger deviation of tighter and looser thresholds and assign the sum in quadrature as the uncertainty.

The results are summarized in Tab. 5.21.

5.5.4 Event-by-event fractions

We perform the signal extraction fit to estimate the event-wise probability of each component in Section 5.1. The mismeasurement of this probability directly leads to the mismodeling of Δt and q. We consider the following sources:

Table 5.21: The systematic uncertainties from the measurement of Δt.

	MR1		non-MR1	
source	δS_{CP}	δA_{CP}	δS_{CP}	δA_{CP}
detector misalignment	±0.0049	±0.0012	±0.0048	±0.0012
IP measurement	±0.0024	±0.0020	±0.0024	±0.0020
helix uncertainty correction	±0.0118	±0.0065	±0.0158	±0.0351
vertex quality selection	±0.0163	±0.0065	±0.0163	±0.0065
total	±0.0209	±0.0095	±0.0233	±0.0357

1. *limited statistics for KDE model*
   Our KDE models are affected by the limited statistics of the fitted MC sample (see Appendix C). To consider the statistical fluctuation of the model, we refit KDE to 1,000 bootstrap samples, where the same number of events as the original dataset are resampled allowing the overlap (the same events can be selected). The effect is evaluated by refitting with the obtained 1,000 KDEs.

2. *bias from KDE*
   The $M_{\rm bc}$–ΔE distribution estimated with KDE can be biased and consequently affect the signal extraction fit. To assess this potential bias, we utilize toy MC where ΔE and $M_{\rm bc}$ are generated without employing KDE, but rather by bootstrapping the MC sample used for fitting KDE.

3. *pre-fit parameters for signal and $B\overline{B}$ background*
   We fit the signal and $B\overline{B}$ background before the signal extraction fit and fix the shape parameters in Sections 5.1.2 and 5.1.3. We assess this uncertainty by refitting.

4. *signal extraction fit*
   The fractions and $q\overline{q}$ shape parameters determined in the signal extraction fit has statistical uncertainties (see Section 5.1.5). We assess this uncertainty by refitting.

5. $\cos\theta^*_B$ *PDF model*
   We model $\cos\theta^*_B$ distribution for each component in MC in Section 5.1.6. We assess the uncertainty by refitting.

6. *r-bin PDF model*
   We model r-bin by the histograms obtained in MC or sideband data in Section 5.1.6.

We assess the uncertainty by refitting, where we independently randomize the histogram entries with Poisson distribution for each r-bin and rebuild the PDF. To account for the possible data–MC difference, we divide the entry by 4 for the histogram obtained in MC to double the Poisson fluctuation.

7. *bias from $M_{\rm bc}$–ΔE correlation in $q\overline{q}$*
We observe a 1–2σ overestimation of $F_{B\bar{B}}$ in MC (see Section 5.1.5). One possible cause is the $M_{\rm bc}$–ΔE correlation in $q\overline{q}$, which is confirmed to be small and not modeled in Section 5.1.4. However, if only 2% of the $q\overline{q}$ background is correlated like $B\overline{B}$ feed-down, we can explain the observed bias. We evaluate this effect by testing the *CP* fit with varied fractions by regarding the portion of $B\overline{B}$ background as $q\overline{q}$ background. The ratio is determined from MC using the following equation:

$$\left(\left[F_{B\bar{B}}\right]_{\rm fit} - \left[F_{B\bar{B}}\right]_{\rm true}\right) / \left[F_{B\bar{B}}\right]_{\rm true}, \tag{5.43}$$

which was found to be less than 20% in both the MR1 and non-MR1 signals, as well as in the control channels. We conservatively take twice the ratio, 40%, to cover possible data–MC discrepancy in the $M_{\rm bc}$–ΔE correlation.

8. *bias from X_s mass spectrum modeling in MC*
The $M_{K\pi}$ spectrum in MC is deviated from real data, and the yield and distribution differ from those in the data. This may bring the data–MC discrepancy in the KDE–ARGUS×pol2 decomposition of the non-MR1 signal candidates. We evaluate this effect by testing the *CP* fit with varied fractions by regarding the signal ARGUS×pol2 component as combinatorial $B\overline{B}$ background.

9. *Punzi effect*
We do not consider the conditional variables $\chi^2_{K^0_S}$, σ_ℓ^{CP}, $(\chi^2/\nu_{\rm d.f.})^{\rm tag}$, and $\sigma_\ell^{\rm tag}$ into the Punzi effect correction, as the differences among the components are minor (see Section 5.1.6). We evaluate the possible bias by toy MC where these four conditional variables are randomly sampled for each component from the corresponding distributions shown in Fig. 5.9.

The results are summarized in Tab. 5.22.

5.5.5 Proper-time difference modeling

The Δt model used for the fit (see Section 5.2) may deviate from the truth and cause systematic bias. We consider the following sources:

Table 5.22: The systematic uncertainties from the probability estimation for each component.

	MR1		non-MR1	
source	δS_{CP}	δA_{CP}	δS_{CP}	δA_{CP}
limited statistics for KDE model	± 0.0017	± 0.0012	± 0.0060	± 0.0035
bias from KDE model	± 0.0008	-0.0007	± 0.0008	± 0.0004
pre-fit parameters for signal and $B\overline{B}$ bkg.	± 0.0002	± 0.0006	± 0.0033	± 0.0019
signal extraction fit	± 0.0011	± 0.0011	± 0.0220	± 0.0111
$\cos\theta_B^*$ model	± 0.0003	± 0.0001	± 0.0004	± 0.0002
r-bin model	± 0.0022	± 0.0029	± 0.0074	± 0.0047
ignored $M_{\rm bc}$–ΔE correlation in $q\overline{q}$	± 0.0003	± 0.0005	± 0.0064	± 0.0011
X_s mass spectrum modeling in MC	-	-	± 0.0191	± 0.0032
Punzi effect	± 0.0008	± 0.0004	± 0.0008	± 0.0004
total	± 0.0032	$^{+0.0034}_{-0.0035}$	± 0.0315	± 0.0132

1. Δt *resolution model for signal component*
   We evaluate the possible bias from the signal Δt resolution modeling in Section 5.2.1 by comparing the *CP* fit in signal MC with and without the resolution.

   - *CP*-violation parameters without the resolution are estimated by fitting the PDF in Eq. (1.20) to $\Delta\tau$ and $q_{\rm true}$.
   - *CP*-violation parameters with the resolution are estimated through the fit to the Δt and $q_{\rm true}$.

   The deviation between the two fit results is large enough to include the bias from possible mismodeling of all the resolution functions in MC, namely $R_{\rm k}$, $R_{\rm rec}$, $R_{\rm asc}\otimes R_{\rm np}$, including the mismodeling from fast simulation.
   We also variate the input *CP*-violation parameter one by one to cover the possible dependence, whereas the other is fixed to 0 for simplicity. The results are shown in Fig. 5.39. The dependencies on the input *CP*-violation parameters are modeled with a linear function determined by the fit. We add the effects from input S_{CP} and A_{CP} in quadrature and take them as the systematic uncertainty,

$$\delta S = \pm\sqrt{(0.0096\cdot A - 0.0079)^2 + (-0.0121\cdot S + 0.0058)^2}, \tag{5.44}$$

$$\delta A = \pm\sqrt{(0.0013\cdot A - 0.0022)^2 + (0.0043\cdot S - 0.0081)^2}. \tag{5.45}$$

   Similarly, we evaluate this source of systematic uncertainty for non-MR1 using

$b \to s\gamma$ MC sample and obtain

$$\delta S = \pm\sqrt{(-0.0140 \cdot A + 0.0052)^2 + (0.0468 \cdot S - 0.0257)^2}, \tag{5.46}$$

$$\delta A = \pm\sqrt{(-0.0026 \cdot A + 0.0072)^2 + (0.0167 \cdot S + 0.0094)^2}. \tag{5.47}$$

2. Δt *resolution function for signal*
   We also consider the fluctuation of resolution function parameters from the limited statistics of the fast-simulated MC (see Section 5.2.2) by refitting.

3. Δt *resolution function for* $B\overline{B}$ *background*
   We model the $B\overline{B}$ background Δt with the resolution functions determined in the fast simulation (see Section 5.2.4). We assess the uncertainty by refitting.

4. *resolution function calibration*
   To correct the possible data–MC difference of the wrong hit assignment in the track, we use the calibrated fast simulation to determine the resolution function for signal and $B\overline{B}$ background (see Section 5.2.2). We evaluate the possible deviation of this method by testing the *CP* fit with the resolution function determined in the original fast simulation.

5. Δt *resolution function for* $q\overline{q}$ *background*
   We model $q\overline{q}$ background Δt distribution using the $M_{\rm bc}$–ΔE sideband data in Section 5.2.5. We assess the uncertainty by refitting.

6. $s_{\rm np}$
   We confirm that the $s_{\rm np}$ is consistent with 1 in data (see Section 5.3.5). We assess the uncertainty by refitting.[2]

7. *physics constants*
   We fix τ_{B^0} and Δm_d to the PDG value [46] in Eqs. (5.38) and (5.39). We assess the uncertainty by refitting.

The results are summarized in Tab. 5.23.

5.5.6 $B\overline{B}$ background asymmetry

The fit model in Eqs. (5.1) and (5.3) does not include the time-dependent *CP* asymmetry of $B\overline{B}$ background. To evaluate this impact, we generate toy MC with the *CP* asymmetry

[2]For better handling of adding the deviations, we take the logarithm and re-evaluate the uncertainty as $\log s_{\rm np} = 0 \pm 0.23$.

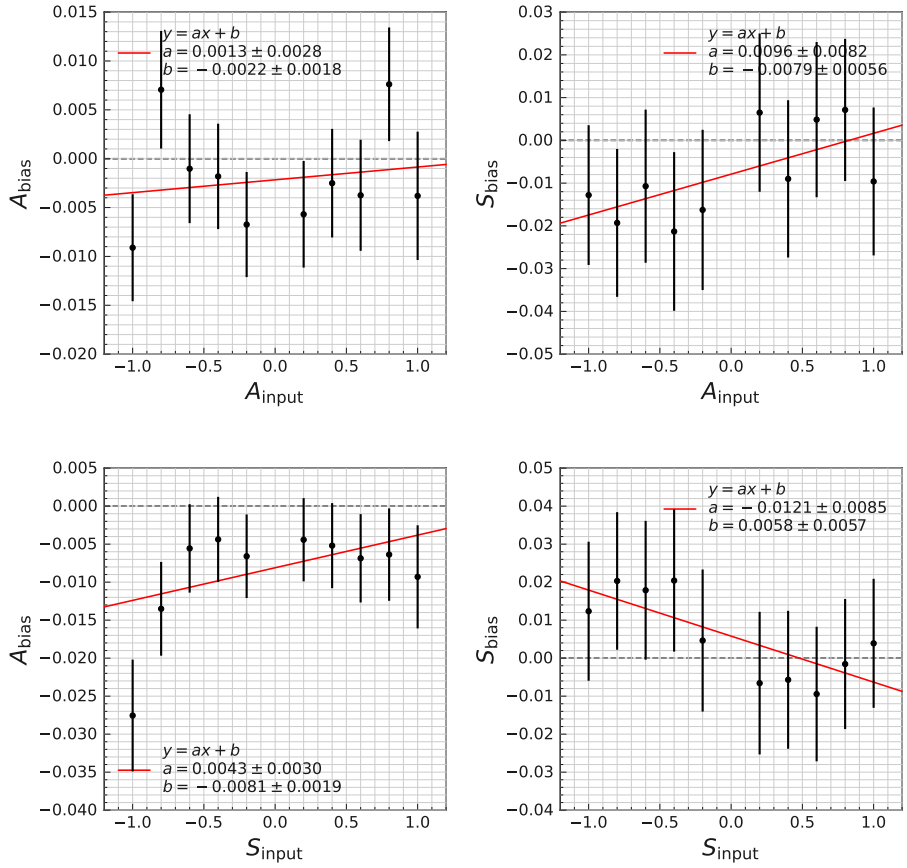


Figure 5.39: The deviation of S_{CP} and A_{CP} in signal MC with CP fit using $\Delta\tau$ and Δt. We use q_{true} for both configurations. The errors are estimated from the difference between the uncertainties of the two configurations in quadrature.

in $B\overline{B}$ background according to the following configuration for the MR1 sample (the numbers for the non-MR1 sample are also shown in parenthesis):

- The B^+B^- events, which amounts to 50% of the whole $B\overline{B}$ background, do not show time-dependent CP asymmetry.
- Among the rest of the $B\overline{B}$ background, i.e., $B^0\overline{B}^0$ events, $b \to s\gamma$ backgrounds are dominant. They consist of $K^0_S\eta\gamma$ events, which amounts to 5.4 (36.7) events per 1 ab^{-1}, and other $b \to s\gamma$ events, which amounts to 52% (25%) of the $B^0\overline{B}^0$ events. Since we have not measured the CP asymmetries for $b \to s\gamma$ in various resonances with enough precision, we conservatively assign $A_{CP} = \pm 1$ or $S_{CP} = \pm 1$ and take the largest positive/negative deviations as the positive/negative uncertainty.
- The miss-reconstructed signal events corresponding to 0.060% of the signal com-

Table 5.23: The systematic uncertainties from the modeling of Δt.

	MR1		non-MR1	
source	δS_{CP}	δA_{CP}	δS_{CP}	δA_{CP}
resolution model for signal	±0.0106	±0.0084	±0.0246	±0.0125
resolution function for signal	±0.0011	±0.0002	±0.0024	±0.0006
resolution function for $B\overline{B}$ background	±0.0005	±0.0002	±0.0017	±0.0009
resolution function calibration	±0.0012	±0.0024	±0.0178	±0.0007
resolution function for $q\overline{q}$ background	±0.0011	±0.0004	±0.0024	±0.0010
$s_{\rm np}$	±0.0084	±0.0015	±0.0081	±0.0020
physics parameters	±0.0002	±0.0002	±0.0034	±0.0006
total	±0.0137	±0.0089	±0.0318	±0.0127

ponent enters in $B^0\overline{B}^0$ events with the same S_{CP} and A_{CP} as the signal.

- The *CP* asymmetry in the rest of the $B^0\overline{B}^0$ events depends on the *CP* egienvalue ξ_{CP}: 75% (13%) with $\xi_{CP} = +1$, 25% (51%) with $\xi_{CP} = -1$, 0% (27%) of non *CP*-eigenstate, i.e., with no *CP* asymmetry, and 0% (9%) of a mixture of *CP*-eigenstates. We conservatively assign ξ_{CP} of the last categories as the same as $b \to s\gamma$ background: +1 for $S_{CP} < 0$, −1 for $S_{CP} > 0$, and 0 for $S_{CP} = 0$. We assign *CP* asymmetry of $S_{CP} = -\xi_{CP} \sin 2\phi_1$ and $A_{CP} = 0$ with $\sin 2\phi_1 = 0.699$ [46].

These fractions are determined by counting the number of events in the 3 ab^{-1} $B\overline{B}$ background MC samples. In case of the miss-reconstructed signal, we use signal MC sample. For the evaluation of $K^0_S\eta\gamma$ yield, we generate a dedicated MC sample, where the mass spectrum of $X_s \to K^0_S\eta$ is modeled not by Kagan–Neubert model [42] but by a single peak at 1.45 GeV/c^2 with the width of 0.30 GeV/c^2 determined from the measured distribution in Belle [63].

5.5.7 Tag-side interference

The B^0–$\overline{B}^0$ interference in the $B_{\rm tag}$ decay cause *CP* asymmetry called tag-side interference [64]. In Eq. (1.17), we assume that $f_{\rm tag}$ is perfectly flavor-specific without non-diagonal components. However, in the case of hadronic flavor-specific decays like $\overline{B}^0 \to D^+\pi^-$, non-diagonal transitions like $B^0 \to D^+\pi^-$ are also allowed with the doubly Cabbibo suppression. This causes interference in $B_{\rm tag}$ and brings bias on the measured *CP*-violation parameters. The evaluation is detailed in Appendix G.

Chapter 6

Discussion

By combining the fit result in Section 5.4 and the systematic uncertainty in Section 5.5, we obtain the following results: for the exclusive $B^0 \to K^*(892)^0\gamma$ decays (MR1 channel),

$$S_{CP}(K^{*0}\gamma) = 0.00^{+0.27}_{-0.26}(\text{stat})^{+0.03}_{-0.04}(\text{syst}) \text{ and}$$
$$A_{CP}(K^{*0}\gamma) = -0.10 \pm 0.13(\text{stat}) \pm 0.03(\text{syst}),$$

and for the inclusive $B^0 \to K^0_S\pi^0\gamma$ decays (non-MR1 channel),

$$S_{CP}(K^0_S\pi^0\gamma) = +0.04^{+0.45}_{-0.44}(\text{stat}) \pm 0.10(\text{syst}) \text{ and}$$
$$A_{CP}(K^0_S\pi^0\gamma) = +0.06 \pm 0.25(\text{stat}) \pm 0.07(\text{syst}).$$

In this chapter, we discuss the impact of our measurement and the prospects.

6.1 Comparison with the previous measurements

We compare the obtained $S_{CP}(K^{*0}\gamma)$ and $A_{CP}(K^{*0}\gamma)$, the previous results explained in Section 1.5, and the world average in Tab. 6.1(a). The obtained result is consistent with the preceding measurements within the uncertainties. Figure 6.1 shows the detailed comparison of $S_{CP}(K^{*0}\gamma)$, including the theoretical expectations. As a single measurement, this analysis gives the most precise measurement to date, even with the smaller integrated luminosity. The world average is updated to

$$S_{CP}(K^{*0}\gamma) = -0.09 \pm 0.17, \tag{6.1}$$

Table 6.1: The comparison of the obtained *CP*-violation parameters and the previous measurements.

(a) exclusive $B^0 \to K^*(892)^0\gamma$

	data size	S_{CP}	A_{CP}
this analysis	$388 \times 10^6 B\overline{B}$	$0.00^{+0.27}_{-0.26}$	-0.10 ± 0.13
world average [46]		-0.15 ± 0.22	-0.04 ± 0.16
Belle (2006) [23]	$535 \times 10^6 B\overline{B}$	$-0.32^{+0.36}_{-0.33}$	-0.20 ± 0.25
BaBar (2008) [24]	$467 \times 10^6 B\overline{B}$	-0.03 ± 0.29	$+0.14 \pm 0.16$

(b) inclusive $B^0 \to K^0_S\pi^0\gamma$

	data size	S_{CP}	A_{CP}
this analysis	$388 \times 10^6 B\overline{B}$	$+0.04^{+0.46}_{-0.45}$	$+0.06 \pm 0.26$
Belle (2006) [23]	$535 \times 10^6 B\overline{B}$	$+0.50 \pm 0.68$	-0.20 ± 0.39

with the inclusion of the result of this analysis. All the measured and the averaged S_{CP} are consistent with the SM expectations, -0.023 ± 0.016 [15] and -0.035 ± 0.017 [16], i.e., we have found no signature of BSM physics in this analysis.

We also compare the obtained $S_{CP}(K^0_S\pi^0\gamma)$ and $A_{CP}(K^0_S\pi^0\gamma)$ and the Belle result reported in 2006 [23] in Tab. 6.1(b), where the similar $M_{K\pi}$ region is analyzed. This analysis gives a consistent result with the preceding measurement within the uncertainties. We also summarize the comparison of $S_{CP}(K^{*0}\gamma)$ in Fig. 6.2, including the theoretical expectation. Since the SM expectation [12] and the measurements still have large uncertainties, we cannot discuss the effect of BSM physics. Providing $M_{K\pi}$ dependence with much higher statistics in the future will be a good test of the QCD calculation.

6.1.1 Improvements from the Belle experiment

As a single measurement, this analysis gives the most precise measurement to date. Compared to the Belle experiment, the integrated luminosity is around half. The 3.3 times higher (TD) signal efficiency achieves such high precision. Due to the enlargement of the VXD detector, K^0_S vertex acceptance is improved by a factor of 1.6. As for the other sources, improvements in K^0_S reconstruction criteria are important, where we use BDT-based classifer, while the cut-based selection in Belle.

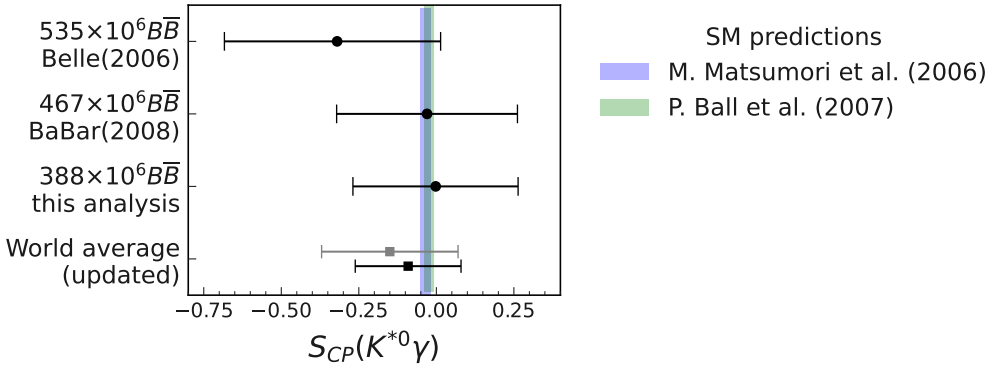


Figure 6.1: The measurements of $S_{CP}(K^{*0}\gamma)$ and its theoretical expectations (blue and green bands) in the SM. The world averages before and after the update are also shown in gray and black, respectively.

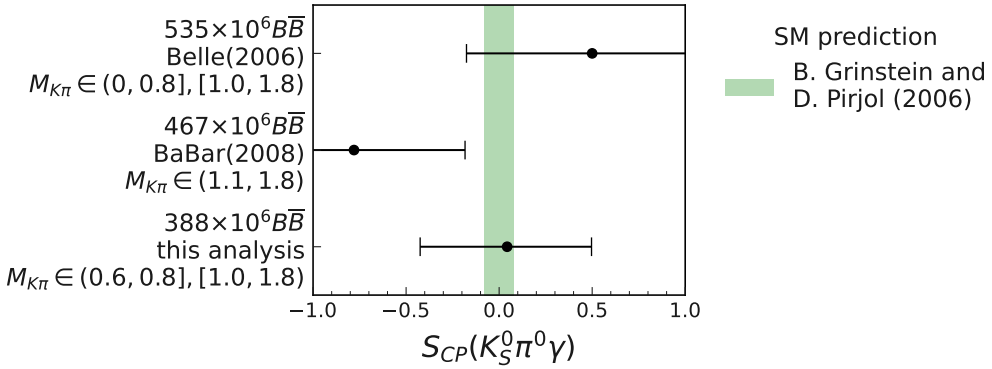


Figure 6.2: The measurements of $S_{CP}(K^0_S\pi^0\gamma)$ and its theoretical expectation (green band) in the SM.

6.2 Constraints on the BSM physics

As we discuss in Section 1.4, we can constrain $C_7'^{\rm BSM}$ with the obtained $S_{CP}(K^{*0}\gamma)$:

$$S_{CP}(K^{*0}\gamma) = 0.00^{+0.27}_{-0.26}(\text{stat})^{+0.03}_{-0.04}(\text{syst}). \tag{6.2}$$

To visualize the constraint, we use the flavio package [65]. We assume Gaussian distribution for each constraint on the observables and model parameters and run toy Monte Carlo. The obtained contour for the real and imaginary part of $C_7'^{\rm BSM}$ is shown in Fig. 6.3. We show the constraint solely from this measurement and the global constraint [66] where we additionally consider

- world average of the previous $S_{CP}(K^{*0}\gamma)$ measurements [8],

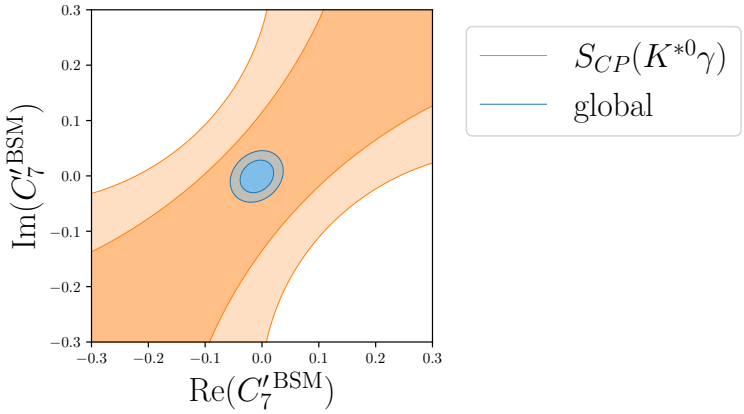


Figure 6.3: Constraints on the real and imaginary part of $C_7'^{\mathrm{BSM}}$ from this measurement only ($S_{CP}(K^{*0}\gamma)$) and in combination with other measurements (global). The darker region corresponds to 1σ, and the lighter 1–2σ.

- branching ratios of $B^0 \to K^{*0}\gamma$, $B^+ \to K^{*+}\gamma$, $B \to X_s\gamma$, and $B_s^0 \to \phi\gamma$ [8],
- the coefficient of sinh-term in the time-dependent CP violation of $B_s \to \phi\gamma$ decays [67],

$$\mathcal{A}^{\Delta} = -\frac{2\xi_{CP}\,\mathrm{Re}\left[\left(\frac{q}{p}\right)_{B_s}\left(\mathcal{M}_R^*\overline{\mathcal{M}_L} + \mathcal{M}_L^*\overline{\mathcal{M}_R}\right)\right]}{\left|\mathcal{M}_R\right|^2 + \left|\mathcal{M}_L\right|^2 + \left|\overline{\mathcal{M}_R}\right|^2 + \left|\overline{\mathcal{M}_L}\right|^2}, \tag{6.3}$$

  and
- angular analysis of $B^0 \to K^{*0}e^+e^-$ [68].

6.3 Prospect of the measurement at Belle II

6.3.1 Improvements anticipated shortly

Some of the tools used in this measurement are being further improved. We evaluate the gain of these modifications and consider them in the prospects.

(1) Modification of tag-side vertex fitting

The current tag-side vertex fitting ignores the effect from the displaced vertex of charm mesons. We consider this bias in the resolution function R_{np}. To reduce bias, we iteratively eliminate the track with the highest χ^2 contribution in the vertex fitting process.

By this modification, we can narrow $R_{\rm np}$ and achieve a smaller fit χ^2. From the study using signal MC, we expect a 10% yield increase in the TD category. This is consistent with the efficiency of the $(\chi^2/\nu_{\rm d.f.})^{\rm tag}$ selection in Tab. 4.3.

(2) Modification of flavor tagging

The current flavor tagging algorithm is based on a combination of category-based BDT classifiers. We develop a new flavor tagging algorithm based on the graph neural network architecture. The effective tagging efficiency,

$$\varepsilon_{\rm eff} = \sum_i \varepsilon_i \cdot (1 - 2w_i)^2, \tag{6.4}$$

is evaluated to be 20% larger than the current one from a study using the $B^0 \to J/\psi\, K^{*0}$ data. Since ε_i and hence $\varepsilon_{\rm eff}$ is dependent on the decay channel, further studies are ongoing to confirm this improvement in the $B^0 \to K^0_S\pi^0\gamma$ channel.

6.3.2 Prospects of statistical and systematic uncertainties

The statistic uncertainty is anticipated to scale with the data size by $1/\sqrt{\int L{\rm d}t}$. In addition, the systematic uncertainties from the measurements are also expected to scale in parallel:

- MC/data correction factors for track momentum and photon energy,
- w and Δw in flavor tagging,
- signal extraction fit, and
- r-bin and Δt modeling for $q\overline{q}$ background.

Another dominant systematic uncertainty from vertex quality selection is evaluated as the difference between the two configurations and includes the fluctuation from the limited data size. Assuming the domination of statistic uncertainty, we also scale this uncertainty by $1/\sqrt{\int L{\rm d}t}$. Considering also the 10% increased statistics from the improvement in tag-side vertex fitting, we visualize the prospects of the uncertainty on $S_{CP}(K^{*0}\gamma)$ in Fig. 6.4. The systematic uncertainty is expected to be dominant after accumulating $30\,{\rm ab}^{-1}$. We need to understand our measurements further and decrease the systematic uncertainty before that. With the target integrated luminosity of $50\,{\rm ab}^{-1}$, the uncertainty is expected to be 0.037. Other main contributions are also expected to decrease in the future but are not considered in the prospect:

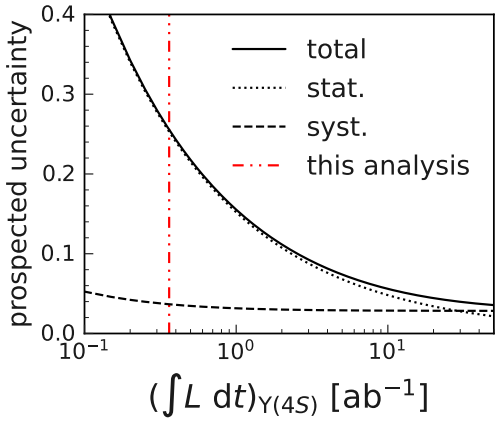


Figure 6.4: The prospects of the S_{CP} uncertainty vs. the data size.

- The systematic uncertainty from the helix uncertainty correction is evaluated by removing the correction. We can evaluate it more accurately by repeating the event reconstruction considering the uncertainty of the correction factors.
- For the $B\bar{B}$ background asymmetry, we conservatively consider $A_{CP} = \pm 1$ and $S_{CP} = \pm 1$ for independent $b \to s\gamma$ decays such as $B^0 \to K^0_S \eta\gamma$. This will become smaller if the future measurement gives better constraints on the size of the CP violation in these decay channels.
- The X_s mass spectrum and $X_s \to K\pi$ branching fractions will be tuned based on the future measurement, e.g., using $B^0 \to K^0_S \pi^0 \ell^+ \ell^-$.

6.3.3 Prospects of constraints on BSM physics

Assuming that the central value is unchanged from this measurement, we can visualize the future constraints on $C_7'^{\mathrm{BSM}}$ in Fig. 6.5. Note that we do not consider the prospects of the other measurements used for the global constraint.

We can limit the LRSM parameters from Eqs. (1.13) and (1.14) as

$$M_{W_2} > \frac{g_R}{g_L} \sqrt{\mathrm{Im}\left[V^R_{ts} e^{i(2\phi_1+\omega)}\right]} \cdot 9.42\,\mathrm{TeV}/c^2. \tag{6.5}$$

In the MSSM, we can limit $\left(\delta^d_{RL}\right)_{23}$ from Eqs. (1.13) and (1.15) as

$$-\mathrm{Im}\left[\left(\delta^d_{RL}\right)_{23} e^{-2i\phi_1}\right] < \frac{m_{\tilde{q}}}{896\,\mathrm{TeV}/c^2}. \tag{6.6}$$

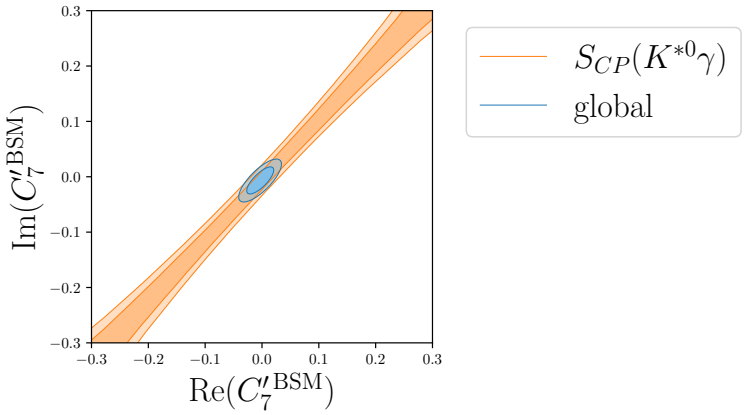


Figure 6.5: Constraints on the real and imaginary part of $C_7'^{\mathrm{BSM}}$ from the expected S_{CP} results with the integrated luminosity of $50\,\mathrm{ab}^{-1}$ ($S_{CP}(K^{*0}\gamma)$) and in combination with other measurements (global). No improvement on the other measurements is assumed. The darker region corresponds to 1σ, and the lighter 1–2σ.

Chapter 7

Conclusion

$B^0 \to K^0_S \pi^0 \gamma$ decay is mediated by $b \to s\gamma$ radiative penguin transition. The mixing-induced *CP* violation is suppressed by the left-handedness of the weak current within the SM and provides a probe for BSM physics. We measure the time-dependent *CP* asymmetries in $B^0 \to K^0_S \pi^0 \gamma$ decay using $(387.5 \pm 5.8) \times 10^6$ $B\bar{B}$ pairs collected near the $\Upsilon(4S)$ resonance with SuperKEKB and Belle II from 2019 to 2022. For the measurement, we reconstruct one neutral B meson in the $B^0 \to K^0_S \pi^0 \gamma$ decay channel, identify the flavor of the accompanying B meson from its decay products, and measure the proper-time difference between the two B mesons. We obtain the following results: for the $K^0_S \pi^0$ invariant mass around $K^*(892)^0$ resonance (0.8 to 1.0 GeV/c^2),

$$\begin{aligned} S_{CP}(K^{*0}\gamma) &= 0.00^{+0.27}_{-0.26}(\text{stat})^{+0.03}_{-0.04}(\text{syst}) \text{ and} \\ A_{CP}(K^{*0}\gamma) &= -0.10 \pm 0.13(\text{stat}) \pm 0.03(\text{syst}), \end{aligned}$$

and for the rest of the mass regions up to 1.8 GeV/c^2,

$$\begin{aligned} S_{CP}(K^0_S \pi^0 \gamma) &= +0.04^{+0.45}_{-0.44}(\text{stat}) \pm 0.10(\text{syst}) \text{ and} \\ A_{CP}(K^0_S \pi^0 \gamma) &= +0.06 \pm 0.25(\text{stat}) \pm 0.07(\text{syst}). \end{aligned}$$

We achieved the most precise measurement to date as a single measurement, even with the less integrated luminosity. The obtained $S_{CP}(K^{*0}\gamma)$ is consistent with the SM expectation, -0.023 ± 0.016 [15] or -0.035 ± 0.017 [16], and no signature of BSM physics is found.

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor Prof. Yutaka Ushiroda, who has provided invaluable support in his roles as a university professor, experiment manager, predecessor in the same analysis channel in Belle, and senior researcher.

I appreciate all the Japanese TDCPV group members: Dr. Kazutaka Sumisawa for carefully tracking my analysis in the weekly report and discussing the plan; Prof. Akimasa Ishikawa for helpful comments based on a deep understanding of the theory including the latest developments in the field; Dr. Koji Hara for Helpful comments on the analysis methodology, including the statistical treatment; Prof. Kenkichii Miyabayashi for his educational comments on physics, analysis, and detectors; Prof. Takeo Higuchi for constantly reviewing and managing the pace of the analysis also as a convener of the TDCPV working group; Dr. Yosuke Yusa for reviewing the analysis; Dr. Yo Sato for helping with technical issues in the analysis software as one of the core developers; Dr. Yu Nakazawa for incorporating new technologies into the analysis; Dr. Kota Nakgairi for the development and handover of the modification work for K_S^0 reconstruction software.

I also appreciate all the TDCPV working group members, especially Dr. Thibaud Humair, for following and reviewing the analysis.

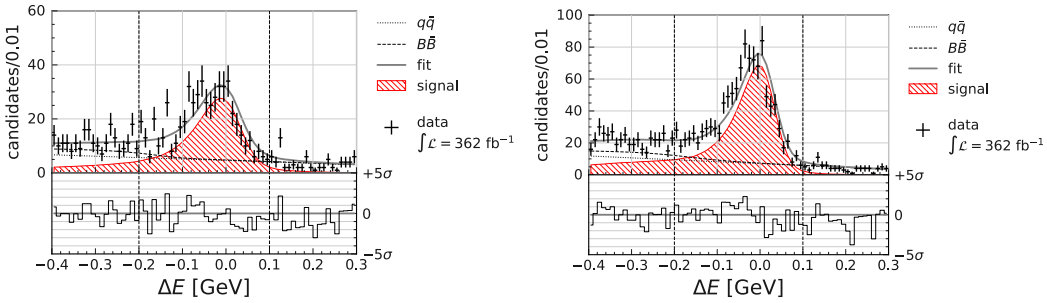
I want to give a shout-out to all the UTokyo/Sokendai B-factory group members for their support, both in the analysis work and in my daily life. Special thanks to Mr. Hikaru Tanigawa and Mr. Hanwook Bae, my great lab buddies. Our discussions on analysis and physics, along with our casual chats and shared meals, have enriched my journey as a researcher.

I also want to take this opportunity to thank my family for their continuous support.

Appendix A

Scaling of hard photon energy

In the experimental data, we observe a slight discrepancy in the right tail of ΔE peak, as shown in Fig. A.1. One possible reason is the lack of MC/data correction for E_γ larger than 1.95 GeV, where most of the prompt γ from B meson lies. To correct possible data–MC discrepancy of E_γ for those γ's, we perform a one-dimensional signal extraction fit to $B^0 \to K^+\pi^-\gamma$ decays. We analyze 10M events of $B \to K^*(892)^0\gamma$ MC samples for study.



(a) $B^0 \to K^0_S\pi^0\gamma$ in MR1

(b) $B^+ \to K^+\pi^-\gamma$ in MR1

Figure A.1: The ΔE distributions in 362 fb^{-1} data without correction and fitted model. The distribution is shown for $M_{\rm bc}$ signal-enhanced region. The vertical dashed lines represent the border of ΔE signal-enhanced region. Note that the signal and background definition is inconsistent with the current one.

A.1 Strategy

The basic concept is the same as the E_γ correction for lower energy: apply MC/data scale factor s_γ on the measured photon energy-momentum in the Lab frame,

$$E_\gamma^{\mathrm{Lab}} \to s_\gamma E_\gamma^{\mathrm{Lab}}, \tag{1.1}$$

$$\mathbf{p}_\gamma^{\mathrm{Lab}} \to s_\gamma \mathbf{p}_\gamma^{\mathrm{Lab}}. \tag{1.2}$$

We apply the same scale factor for E_γ^{Lab} and $\mathbf{p}_\gamma^{\mathrm{Lab}}$ since $m_\gamma = 0$. However, modeling E_γ^{Lab} distribution is not straightforward. We move to the $\Upsilon(4S)$ CM frame:

$$E_\gamma^{\mathrm{cms}} = \gamma E_\gamma^{\mathrm{Lab}} - \gamma \boldsymbol{\beta} \cdot \mathbf{p}_\gamma^{\mathrm{Lab}}. \tag{1.3}$$

By applying the scale factor, we get scaled E_γ^{cms}:

$$E_\gamma^{\mathrm{cms}} \to \gamma s_\gamma E_\gamma^{\mathrm{Lab}} - \gamma \boldsymbol{\beta} \cdot s_\gamma \mathbf{p}_\gamma^{\mathrm{Lab}} = s_\gamma E_\gamma^{\mathrm{cms}}. \tag{1.4}$$

We can still extract the scale factor s_γ using E_γ^{cms}. Moreover, we can use B-candidates' ΔE for scale factor determination. Recalling the ΔE definition,

$$\Delta E = E_\gamma^{\mathrm{cms}} + E_{K^+}^{\mathrm{cms}} + E_{\pi^-}^{\mathrm{cms}} - E_{\mathrm{beam}}/2, \tag{1.5}$$

we can derive ΔE after applying the scale factor as

$$\Delta E \to s_\gamma E_\gamma^{\mathrm{cms}} + E_{K^+}^{\mathrm{cms}} + E_{\pi^-}^{\mathrm{cms}} - E_{\mathrm{beam}}/2 = \Delta E|_{s_\gamma=1} + (s_\gamma - 1) E_\gamma^{\mathrm{cms}}, \tag{1.6}$$

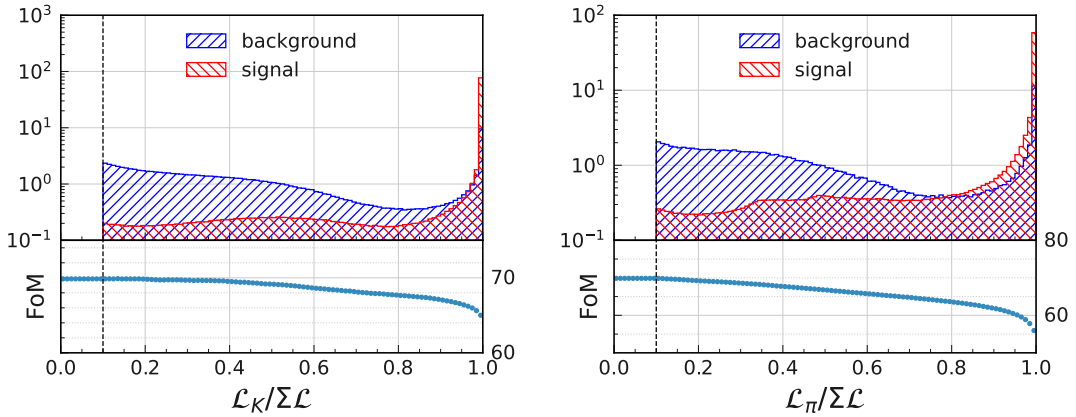
the shift proportional to E_γ^{cms}. Thus, we can determine the scale factor s_γ through ΔE signal extraction fit with shift proportional to the event-by-event E_γ^{cms}.

A.2 Event reconstruction and selection

We use the same γ reconstruction and selection as the signal channel in Section 4.2.3. We perform the same selection optimization procedure as the signal channel for the other selections.

For the K^+ and π^-, we select tracks with at least 20 CDC hits which satisfies $\Delta d < 0.5\,\mathrm{cm}$ and $|\Delta z| < 2\,\mathrm{cm}$. We also require the likelihood ratio of particle identification (PID) for K (π), $\mathcal{L}_K/\sum\mathcal{L}$ ($\mathcal{L}_\pi/\sum\mathcal{L}$), to be larger than 0.1 for K^+ (π^-). The selection on the PID is tight enough, as shown in Figs. A.2(a) and A.2(b).

We require the invariant mass of $K^+\pi^-$ system to be $0.8 < M_{K\pi} < 1.0\,\mathrm{GeV}/c^2$ as $K_S^0\pi^0\gamma$ in MR1. We also require that the reconstructed B^0 must satisfy $5.2 < M_{\mathrm{bc}} <$



(a) $\mathcal{L}_K/\sum\mathcal{L}$ of K^+ (b) $\mathcal{L}_\pi/\sum\mathcal{L}$ of π^-

Figure A.2: The $\mathcal{L}_K/\sum\mathcal{L}$ ($\mathcal{L}_\pi/\sum\mathcal{L}$) distribution of K^+ (π^-) in $B^0 \to K^+\pi^-\gamma$ candidates. The red distribution shows signal candidates, and the blue shows the fake candidates. The blue dotted line in the lower plot shows the FoM of B^0 candidates at each threshold value. The vertical dashed line shows the determined threshold.

5.29 GeV/c^2, $|\Delta E| < 0.5$ GeV. We trained a dedicated BDT classifier $C_{q\bar{q}}(K^+\pi^-\gamma)$ for continuum suppression. We use the set of input variables for the non-MR1 signal channel so as not to bring any severe deformation in background ΔE distributions. We optimize the $C_{q\bar{q}}(K^+\pi^-\gamma)$ threshold to be 0.67, as shown in Fig. A.3. For the events with multiple B^0 candidates, we select a single candidate randomly not to bring any deformation in background ΔE and $M_{\rm bc}$ distributions.

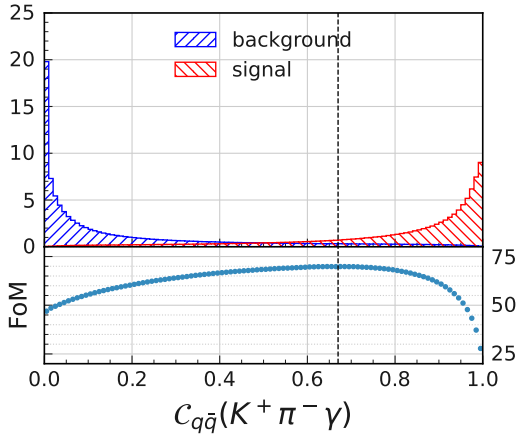


Figure A.3: The $C_{q\bar{q}}(K^+\pi^-\gamma)$ distribution of signal (fake) B^0 candidates are shown in red (blue). The blue dotted line in the lower plot shows the FoM at each threshold value. The vertical dashed line shows the threshold at the highest FoM.

A.3 Fit models

We separate ΔE distributions into the signal, $B\overline{B}$ background, and $q\overline{q}$ background components.

The signal distributions are modeled with Johnson function [69],

$$f_{\text{Jhonson}}\,(x=\Delta E;\mu,\sigma,\gamma,\lambda) = \frac{\lambda}{\sigma\sqrt{2\pi}}\frac{1}{\sqrt{1+\left(\frac{x-\mu}{\sigma}\right)^2}}\,\exp\left[-\frac{1}{2}\left(\gamma+\lambda\sinh^{-1}\left(\frac{x-\mu}{\sigma}\right)\right)^2\right]. \tag{1.7}$$

The four shape parameters are fixed to the signal MC fit,

$$\mu = 0.02141 \pm 0.00010, \tag{1.8}$$
$$\sigma = 0.04532 \pm 0.00010, \tag{1.9}$$
$$\gamma = 0.8696 \pm 0.0024, \tag{1.10}$$
$$\lambda = 1.1447 \pm 0.0019. \tag{1.11}$$

The distribution of ΔE in $B^0 \to K^*(892)^0\gamma$ MC and fitted model are shown in Fig. A.4. The discrepancy is only significant with the extensive statistics of the MC sample.

One-third of $B\overline{B}$ background comes from X_s resonance, but its yield and distribution shape is unreliable due to the Kagan–Neubert X_s mass spectrum model. Figure A.5(a) shows the distribution for X_s-resonant and the other $B\overline{B}$ background. We model $B\overline{B}$ backgrounds other than X_s resonance using spline fit and ignore the contribution from

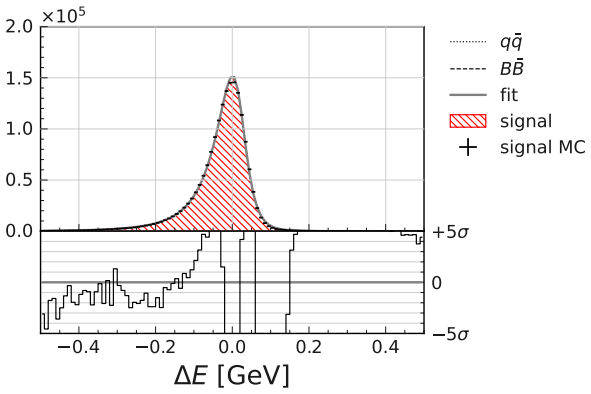
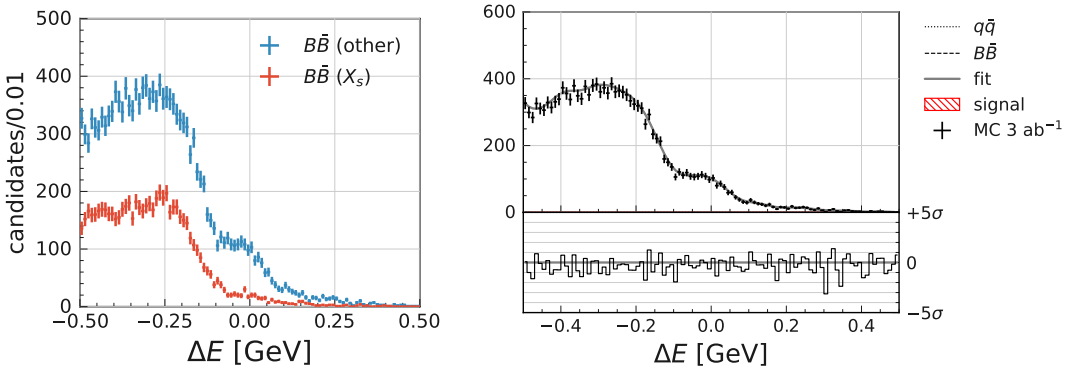


Figure A.4: The ΔE distribution of signal candidates in $B^0 \to K^*(892)^0\gamma$ MC sample.



(a) comparison between X_s and other

(b) spline fit for $B\bar{B}$ background other than X_s

Figure A.5: The ΔE distributions of $B\bar{B}$ background in 3 ab^{-1} MC.

X_s resonance. The comparison with MC distribution and the spline curve is shown in Figure A.5(b). We can confirm that the spline curve smooths the histogram without bias.

$q\bar{q}$ background is modeled with a linear function with only one parameter: slope, $m_{q\bar{q}}$. We leave floating the shape parameter in the signal extraction fit. The 1 ab^{-1} MC distribution and fitted model are shown in Fig. A.6. We do not observe any bias with this simple modeling.

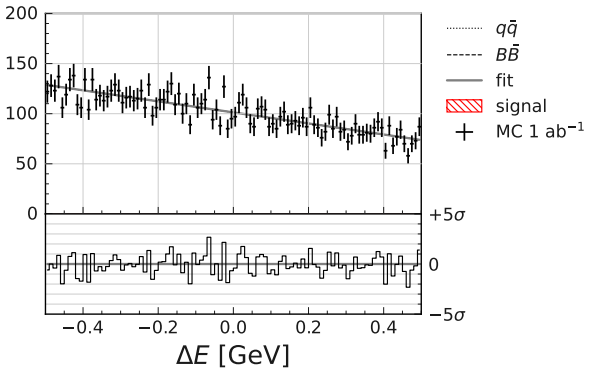


Figure A.6: The ΔE distributions of $q\overline{q}$ background in 1 ab^{-1} MC.

A.4 Signal extraction fit

We perform an unbinned maximum likelihood fit to only ΔE with PDF analogous to Eq. (5.5), floating fractions, $q\overline{q}$ shape, and E_γ scale factor.

We apply shift from E_γ scale factors for signal and $B\overline{B}$ background. We do not need to shift the $q\overline{q}$ shape because we float the shape in the fit. We fit with a narrower range to allow a shift of the spline curve, $-0.4 < \Delta E < 0.4$ GeV. The result in a dress rehearsal on 1 ab^{-1} MC is shown in Fig. A.7. The resulting fit parameters are summarized in Tab. A.1. We observe significant overestimation in $F_{B\bar{B}}$ and $m_{q\bar{q}}$, which means the background breakdown is not correctly working. However, we do not observe significant bias in the signal fraction and E_γ scale. Since we aim to obtain this E_γ scale, we proceed with this fit model.

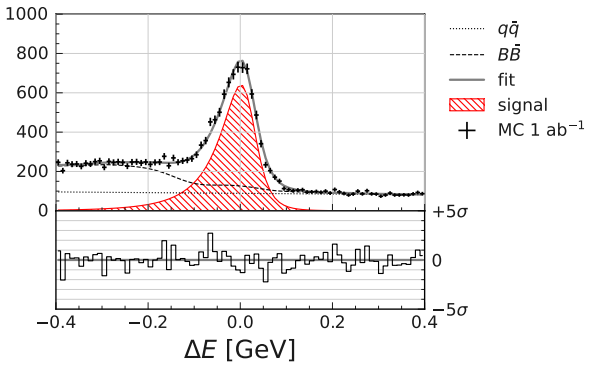


Figure A.7: The ΔE distributions in 1 ab^{-1} generic MC and fitted model.

Table A.1: The results of signal extraction fits of $B^0 \to K^+\pi^-\gamma$ candidates in 1 ab^{-1} generic MC.

parameter	fit result	expectations
$(s_\gamma - 1)[\%]$	-0.032 ± 0.028	0
F_{sig}	0.393 ± 0.006	0.397
$F_{B\bar{B}}$	$0.380^{+0.035}_{-0.034}$	0.283
$m_{q\bar{q}}$	$-0.25^{+0.27}_{-0.24}$	-0.544 ± 0.034

A.5 Results

The fit results on data are shown in Fig. A.8 and Tab. A.2. The result of signal extraction fit with scaling factor is shown in Section 5.1.5. As shown in Fig. 5.5, the right tail of ΔE peak better agrees with the data.

Table A.2: The results of signal extraction fits of $B^0 \to K^+\pi^-\gamma$ candidates in 362 fb^{-1} data.

parameter	fit result
$(s_\gamma - 1)[\%]$	$+0.22 \pm 0.05$
F_{sig}	0.435 ± 0.010
$F_{B\bar{B}}$	0.18 ± 0.06
$m_{q\bar{q}}$	$-0.49^{+0.38}_{-0.32}$

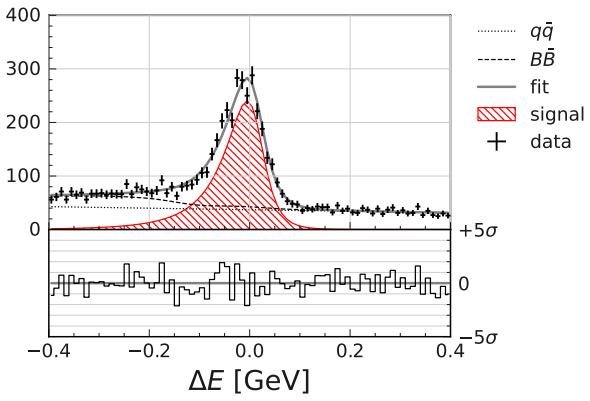


Figure A.8: The ΔE distributions in 362 fb^{-1} data and fitted model.

Appendix B

Control sample $B^+ \to K^0_S \pi^+ \gamma$

B.1 Selections for the control sample

We reconstruct the control sample using π^+ instead of π^0 in the signal mode. We reconstruct the π^+ from a track in the CDC acceptance, which satisfies $|\Delta d| < 0.5\,\mathrm{cm}$, $|\Delta z| < 3\,\mathrm{cm}$. We also require the likelihood ratio of particle identification $\mathcal{L}_\pi / \sum \mathcal{L}$ to be larger than 0.01. We optimize the cut on $\mathcal{L}_\pi / \sum \mathcal{L}$ of π^+ and confirm that the preselection is tight enough, as shown in Fig. B.1. We use the same γ and K^0_s selection criteria as the signal decay.

The primary source of multiplicity comes from the π^+ tracks. However, any π^+ candidates are acceptable for the control sample since we smear the π^+ track in the vertex fit. Thus, we select one candidate randomly when multiple B^+ candidates exist. Similarly, the truth-matching of B^+ does not require the truth-matching of π^+.

We use the same continuum suppression and apply the same $C_{q\bar{q}}$ selection criteria as the signal mode.

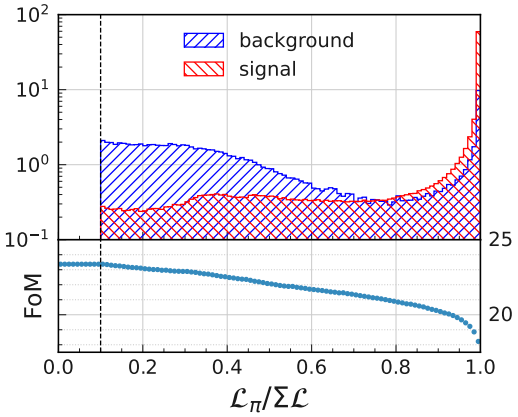


Figure B.1: The comparison of $\mathcal{L}_\pi/\sum\mathcal{L}$ between π^+ from signal B^+ decay in 10M control MC (signal, red) and fake π^+ in 1 ab^{-1} generic MC (background, blue). The entries for each component are normalized to 100. The vertical black dashed line represents the threshold. The FoM of B^+ candidates at each threshold is also shown in the lower subplot.

B.2 Vertex reconstruction and resolution of the control sample

The vertex reconstruction of the control sample is similar to the signal mode except for the $\pi^\pm$ track, as illustrated in Fig. B.2. To reconstruct the $B_{\rm sig}$ vertex with the same resolution as the signal, we ignore the position information of $\pi^\pm$ track in the vertex fit. We confirm that the signal $R_{\rm rec}$ model determined in fast-simulated signal MC successfully explains $\delta\ell^{CP}$, as shown in Fig. B.3. The $B_{\rm tag}$ decay vertex reconstruction is entirely the same as the signal. The required vertex quality for TD events is also aligned with the signal.

The tag side resolutions for the signal, $B\overline{B}$ feed-down, and $B\overline{B}$ combinatorial backgrounds are modeled in the same procedure as the signal. We summarized the resolution parameters in Tab. B.1. The obtained resolution function models are compared with the distributions in Figs. B.4 and B.5. The observed discrepancies are only significant with the extensive statistics of the MC samples.

The $q\overline{q}$ model is by construction combined with the signal sample.

We perform B^+ lifetime fit using resolution parameters obtained in fast simulation. The results are summarized in Tab. B.2. The obtained model and signal MC distribution are compared in Fig. B.6. We observe a similar behavior as the signal mode. The lifetime

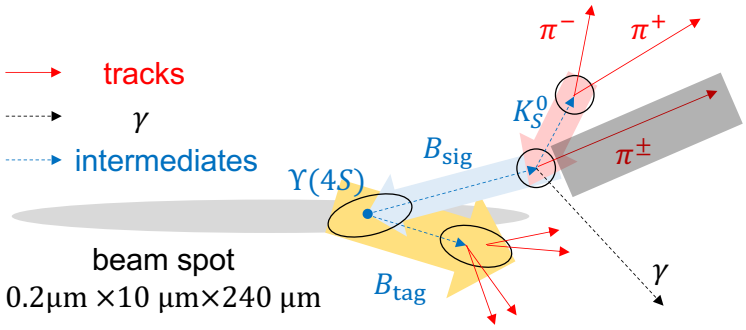


Figure B.2: The schematic explanation for $B\bar{B}$ decay vertex reconstruction of the control sample.

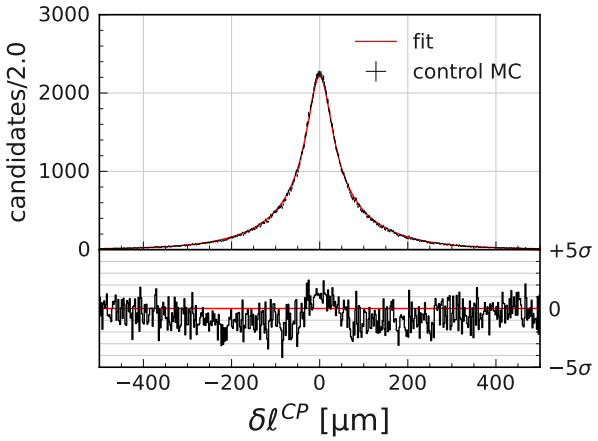


Figure B.3: $\delta\ell^{CP}$ distribution in control MC and $R_{\rm rec}$ obtained in fast-simulated signal MC. The lower box shows the discrepancy between the distribution and fit model.

bias in the event reconstruction and selection is slight, while the small mismodeling of the resolutions brings some bias. However, this level of bias is acceptable for the validation purpose.

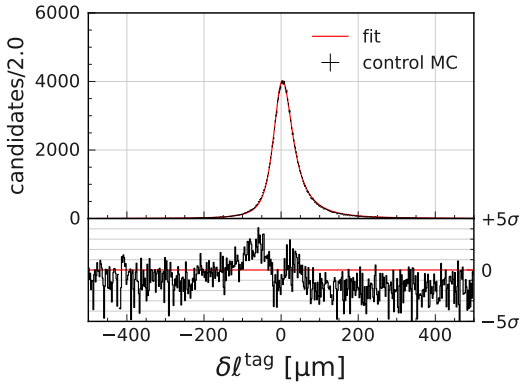


Figure B.4: $\delta\ell^{\text{tag}}$ distribution in control MC and R_{tag} obtained in the original fast simulation. The lower box shows the discrepancy between the distribution and fit model.

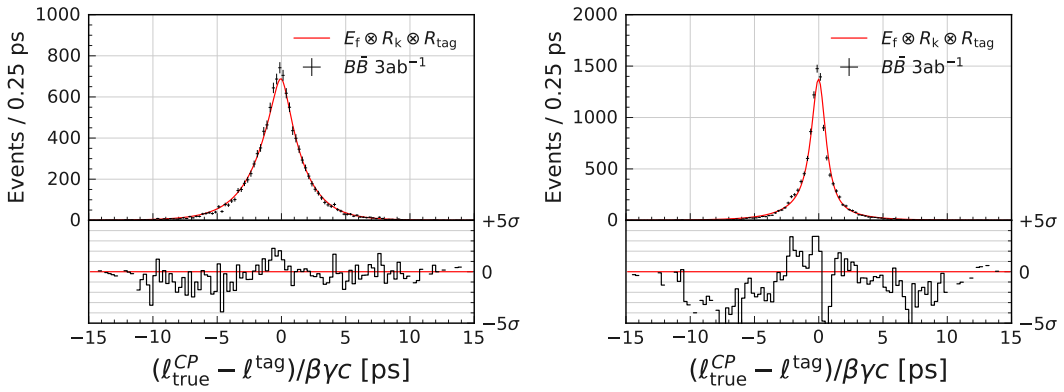


Figure B.5: $(\ell^{CP}_{\text{true}} - \ell^{\text{tag}})/\beta\gamma c$ distribution in $B\overline{B}$ sample ($3\,\text{ab}^{-1}$) and $E_{\text{f}} \otimes R_{\text{k}} \otimes R_{\text{tag}}$ obtained in the original fast simulation. The candidates outside the M_{bc}–ΔE signal-enhanced region are included to increase the statistics. The lower box shows the discrepancy between the distribution and fit model.

Table B.1: The results of $\delta\ell^{\text{tag}}$ fit in the original fast-simulated $b \to s\gamma$ sample and the calibrated fast-simulated $b \to s\gamma$ sample. The fit uncertainty is estimated from the Hesse matrix.

(a) signal

parameter		original fast sim	calibrated fast sim
R_{asc}	$f_{\text{tail}}^{\text{tag}}$	0.00077 ± 0.00008	0.00080 ± 0.00007
	$s_{\text{main}}^{\text{tag},0}$	0.986 ± 0.004	1.000 ± 0.004
	$s_{\text{main}}^{\text{tag},1}$	0.0747 ± 0.0018	0.0728 ± 0.0017
	$s_{\text{tail}}^{\text{tag},0}$	27.4 ± 2.2	26.9 ± 2.0
	$s_{\text{tail}}^{\text{tag},1}$	2.8 ± 0.5	2.6 ± 0.4
R_{np}	f_{δ}^{0}	0.739 ± 0.007	0.763 ± 0.011
	f_{δ}^{1}	-0.1490 ± 0.0029	-0.157 ± 0.005
	f_{p}^{0}	0.8769 ± 0.0028	0.8827 ± 0.0029
	f_{p}^{1} [μm^{-1}]	-0.00154 ± 0.00008	-0.00161 ± 0.00009
	τ^{0}	0.941 ± 0.010	0.954 ± 0.013
	τ^{1}	0.2066 ± 0.0028	0.2109 ± 0.0034
	τ^{max}	5.24 ± 0.10	5.24 ± 0.08

(b) $B\bar{B}$ feed-down

parameter		original fast sim	calibrated fast sim
lifetime	$\tau_{B\bar{B}}$	1.40 ± 0.02	1.396 ± 0.021
R_{np}	f_{δ}^{0}	1.103 ± 0.020	1.90 ± 0.05
	f_{δ}^{1}	-0.186 ± 0.011	-0.624 ± 0.021
	f_{p}^{0}	0.783 ± 0.027	0.722 ± 0.026
	f_{p}^{1} [μm^{-1}]	-0.0018 ± 0.0007	-0.0003 ± 0.0008
	τ^{0}	2.51 ± 0.30	1.52 ± 0.033
	τ^{1}	0.40 ± 0.04	0.54 ± 0.04
	τ^{max}	7.9 ± 0.4	7.7 ± 0.4

(c) $B\bar{B}$ combinatorial

parameter		original fast sim	calibrated fast sim
lifetime	$\tau_{B\bar{B}}$	0.400 ± 0.024	0.411 ± 0.019
R_{np}	f_{δ}^{0}	0.756 ± 0.032	0.775 ± 0.014
	f_{δ}^{1}	-0.067 ± 0.007	-0.0798 ± 0.0026
	f_{p}^{0}	0.607 ± 0.009	0.612 ± 0.009
	f_{p}^{1} [μm^{-1}]	0	0
	τ^{0}	5.84 ± 0.12	5.59 ± 0.11
	τ^{1}	0.081 ± 0.010	0.107 ± 0.011
	τ^{max}	8.8 ± 0.4	8.9 ± 0.4

Table B.2: The results of lifetime fit in control MC.

proper-time difference	model	fit result [ps]
MC input	(no fit)	1.638
$\Delta\tau^{\text{true}} = \tau^{\text{true}}_{B_{\text{sig}}} - \tau^{\text{true}}_{B_{\text{tag}}}$	E_{f}	1.6341 ± 0.0028
$\Delta t^{\text{true}} = t^{\text{true}}_{B_{\text{sig}}} - t^{\text{true}}_{B_{\text{tag}}}$	$E_{\text{f}} \otimes R_{\text{k}}$	1.6349 ± 0.0028
$t_{B_{\text{sig}}} - t^{\text{true}}_{B_{\text{tag}}}$	$E_{\text{f}} \otimes R_{\text{k}} \otimes R_{\text{rec}} + G_{\text{outlier}}$	1.651 ± 0.004
$t^{\text{true}}_{B_{\text{sig}}} - t_{B_{\text{tag}}}$	$E_{\text{f}} \otimes R_{\text{k}} \otimes R_{\text{tag}}$	1.6241 ± 0.0030
$\Delta t = t_{B_{\text{sig}}} - t_{B_{\text{tag}}}$	$E_{\text{f}} \otimes R_{\text{k}} \otimes R_{\text{rec}} \otimes R_{\text{tag}} + G_{\text{outlier}}$	1.641 ± 0.004

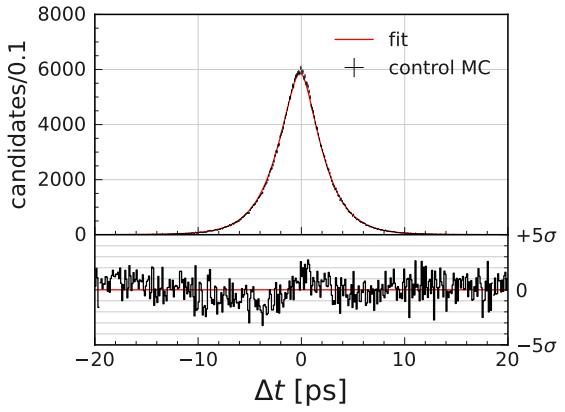


Figure B.6: Δt distribution in control MC and model obtained in fast simulation. The lower subplot shows the pull of the discrepancy.

Appendix C

Gaussian kernel density estimation

We use Gaussian kernel density estimation (KDE) in Scikit-learn package [70] to construct correlated two-dimensional distribution from the MC samples, as shown in Fig. C.1. The KDE smoothes the empirical distribution by assigning one Gaussian for each data point. In the case of multi-dimension, we need to apply the scaling factor to data so that the standard dedication of the Gaussian is common to all the dimensions. The scale factors tuned by hand are summarized in Tab. C.1. The common standard deviation, or bandwidth, h is determined according to Scott's factor [71]:

$$h = n^{-1/(d+4)}, \tag{3.1}$$

where n for sample size and d for dimension. Figures C.2 and C.3 shows the one-deimensional projections for signal and $B\bar{B}$ feed-down background. The agreement between the KDE and the distribution is good in general, while we observe slight over-smoothing of the steep histogram, such as around the peak. We take the possible impact into the systematic uncertainty described in Section 5.5.4.

			$M_{\rm bc}$ [GeV/c^2]	ΔE [GeV]
candidate	component	scale	0.01	0.14
B^0	signal	0.2	0.002	0.028
	feed-down bkg.	1	0.01	0.14
B^+	signal	0.2	0.002	0.028
	feed-down bkg.	1.2	0.012	0.168

Table C.1: Scale factors for $M_{\rm bc}$ and ΔE in each component.

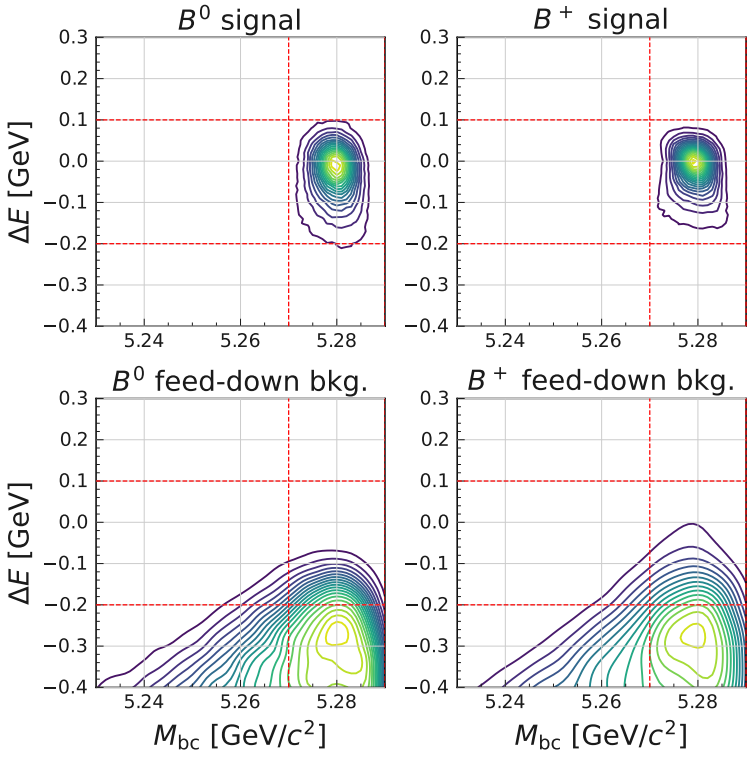


Figure C.1: The 2D contour plot for KDE curves. The upper half is for signal in MR1, the lower half is for $B\bar{B}$ feed-down in MR1, the left-hand side is for B^0, and the right-hand side is for $B^{\pm}$.

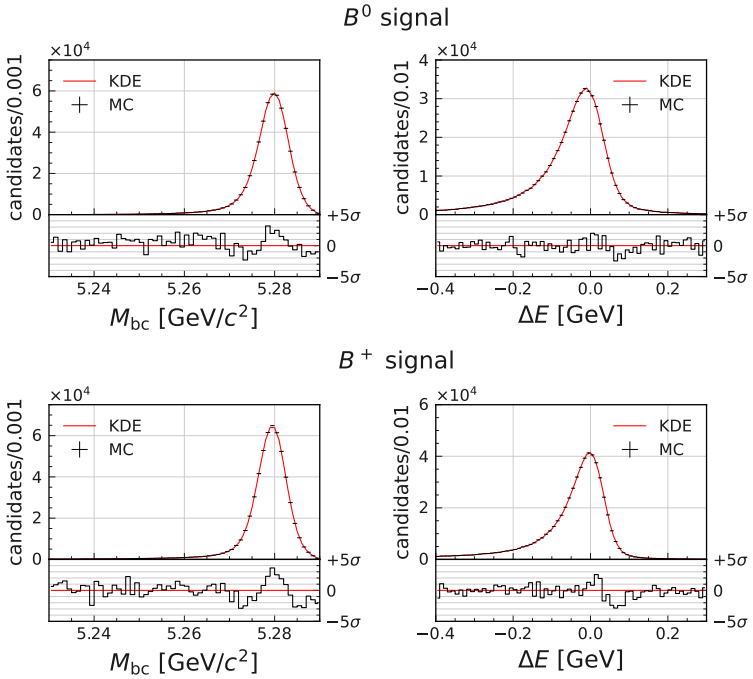


Figure C.2: The comparison of MC histogram and KDE curve in 1D projection for correct reconstruction. The upper half is for B^0, and the lower half is for $B^{\pm}$.

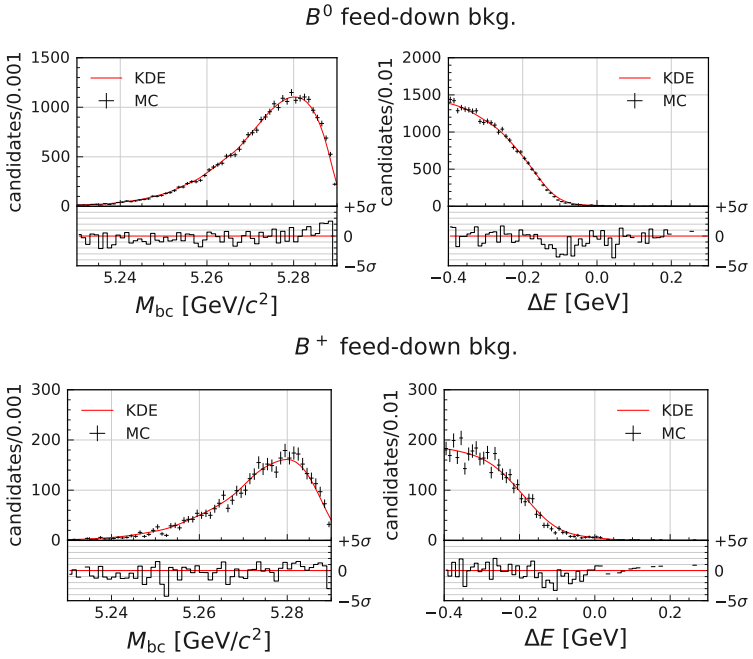


Figure C.3: The comparison of MC histogram and KDE curve in 1D projection for feed-down reconstruction. The upper half is for B^0, and the lower half is for $B^{\pm}$.

Appendix D

Resolution in χ^2 slice

We show the vertex position residual distribution in the slice of χ^2. We confirm that $R_{\rm rec}$ and $R_{\rm asc} \otimes R_{\rm np}$ agree well with the distribution even in the χ^2 slice.

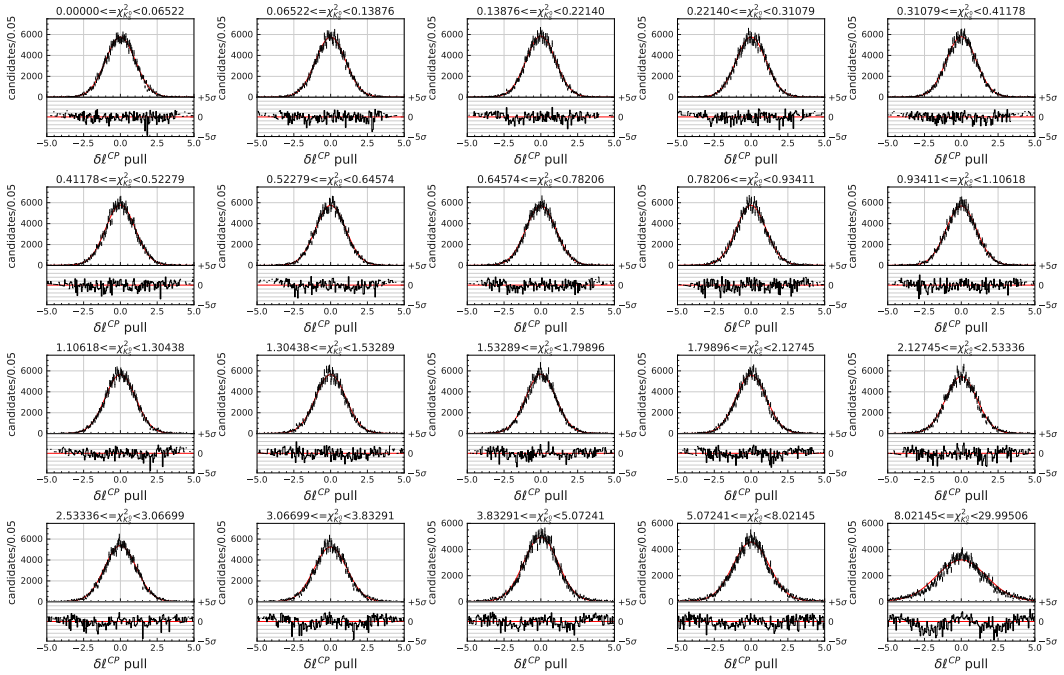


Figure D.1: $\delta\ell^{CP}$ pull distribution and fitted double Gaussian in signal MC are shown for each $\chi^2_{K^0_S}$ slice. The lower box shows the discrepancy between the distribution and fit model.

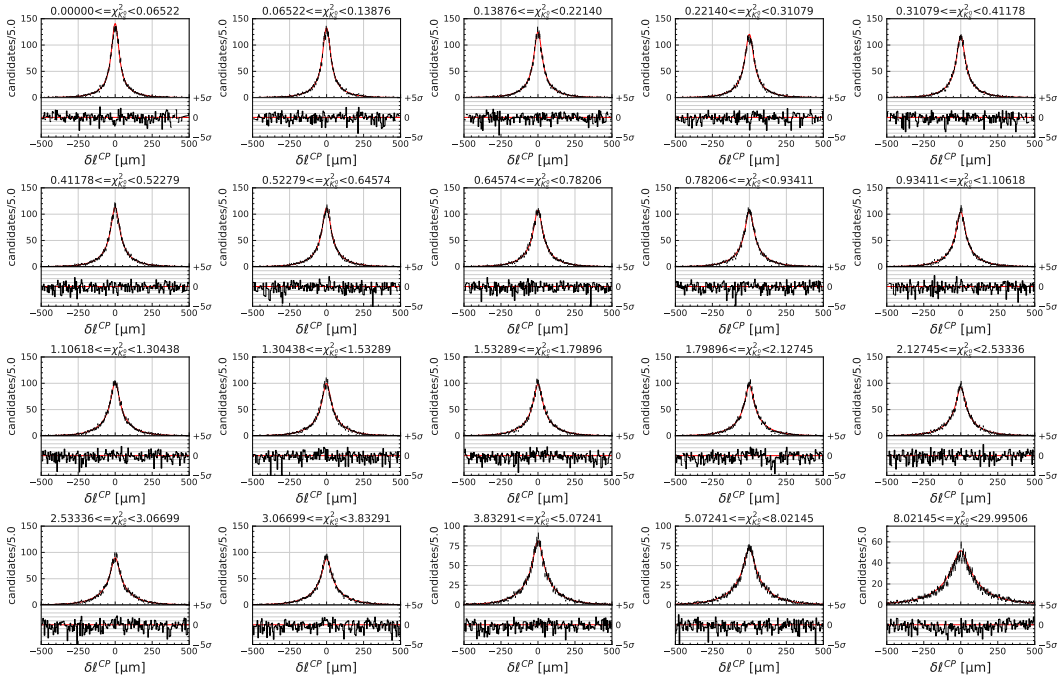


Figure D.2: $\delta\ell^{CP}$ distribution and fitted R_{rec} in signal MC are shown for each $\chi^2_{K^0_S}$ slice. The lower box shows the discrepancy between the distribution and fit model.

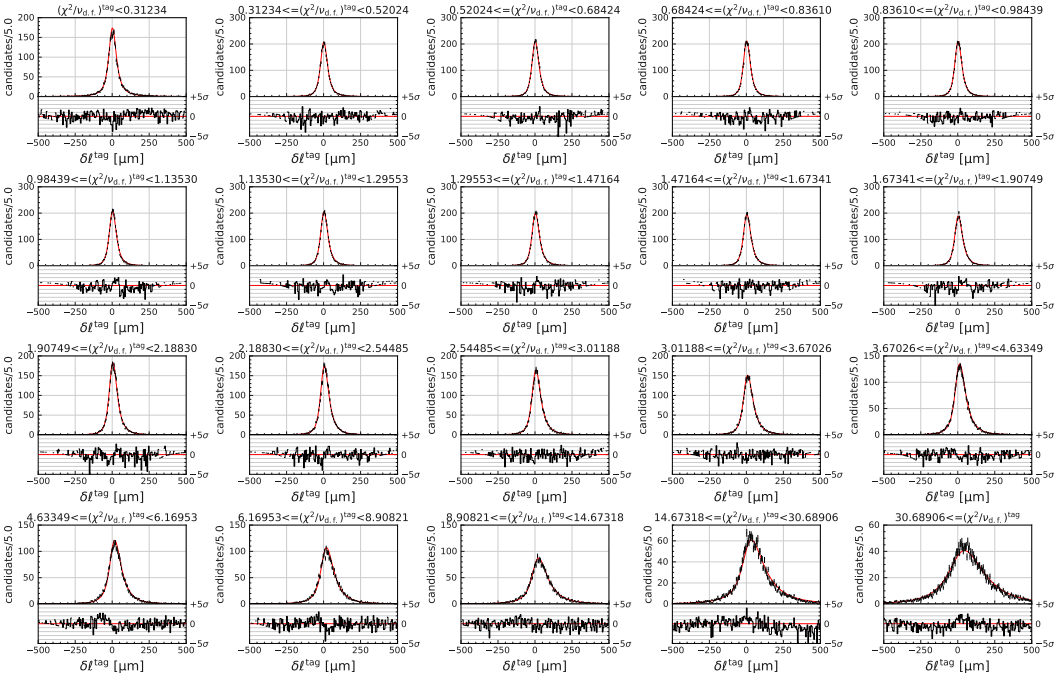


Figure D.3: $\delta\ell^{\mathrm{tag}}$ distribution and fitted R_{tag} in signal MC are shown for each $(\chi^2/\nu_{\mathrm{d.f.}})^{\mathrm{tag}}$ slice. The lower box shows the discrepancy between the distribution and fit model.

Appendix E

Fast simulation

To study the relation between the helix and the resolution function parameter, we developed a fast simulation where the helix pull is randomly sampled from two Gaussian distributions (main + tail). The tail component is key to reproducing the χ^2 dependence of the resolution function. Without the tail component, which means when the uncertainty of the vertex fit is correctly estimated, no χ^2 dependency is observed in the resolution function.

In this helix study, we use only signals in the signal-enhanced region. Since $B^0 \to K^*(892)^0(\to K^0_S\pi^0)\gamma$ decay has only K^0_S daughter tracks, we need to take the helices not at the IP POCA (point of closest approach) but at the true decay vertex POCA. We also randomize the helix from the true helix at the true decay vertex.

The correlations between helix pulls are shown in Fig. E.1. We observe a strong correlation between d_0 and ϕ_0, z_0 and $\tan\lambda$. We apply two-dimensional rotation by $\mp 45^\circ$ as

$$\begin{pmatrix} (d_0, \phi_0)_{\text{wide}} \\ (d_0, \phi_0)_{\text{narrow}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} d_0 \\ \phi_0 \end{pmatrix}, \tag{5.1}$$

$$\begin{pmatrix} (z_0, \tan\lambda)_{\text{wide}} \\ (z_0, \tan\lambda)_{\text{narrow}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} z_0 \\ \tan\lambda \end{pmatrix}. \tag{5.2}$$

$(d_0, \phi_0)_{\text{narrow}}$ and $(z_0, \tan\lambda)_{\text{narrow}}$ represent the sharpness of correlation. This correlation is smeared by multiple scattering between the decay vertex and the first (i.e., innermost) hit. In particular, the tracks originating inside the beampipe have smeared correlation. Thus, we separately model K^0_S decays outside the beam pipe and K^0_S decays inside the beampipe, referred to as outer- and inner-K^0_S, respectively. The results of double Gaussian fit to $(d_0, \phi_0)_{\text{narrow}}$ and $(z_0, \tan\lambda)_{\text{narrow}}$ for outer- and inner-K^0_S are shown in Figs. E.2

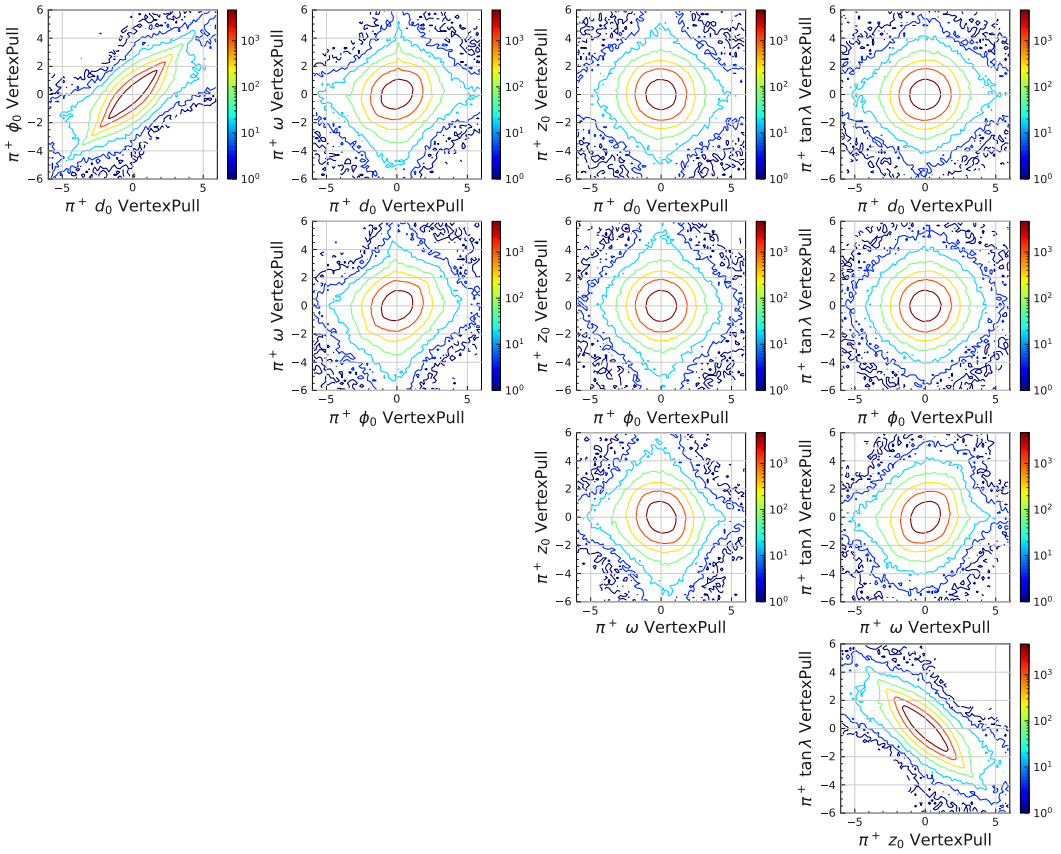


Figure E.1: Two-dimensional helix pull distributions of π^+ tracks from K^0_S in signal MC.

and E.3. We fit the distribution in a wider range to model the broader distribution of inner-K^0_S. The resulting fit parameters are summarised in Tab. E.1. The result of the inner-K^0_S fit is also used to model tag-side tracks since they originate from IP.

In principle, the wrong hit assignment leads to a significant deviation in all five helices. Thus, we force $(d_0, \phi_0)_{\text{wide}}$, $(z_0, \tan\lambda)_{\text{wide}}$, and ω to be in main or tail simultaneously. The result of simultaneous fit is shown for three helices in Fig. E.4. The resulting fit parameters are summarised in Tab. E.2. This modeling is not entirely accurate due to one-dimensional wrong hits such as u/v-side SVD hits. This causes the narrower and larger tail component in the fit. By adding the dilution to allow main-tail cross-term, $\sim$ 10% of the main component goes to the cross-term, but we ignore this effect for simplicity.

We run the fast simulation based on these fit parameters. The rotated helices before and after fast simulation show good agreement, as shown in Figs. E.5 to E.7. Slightly higher peaks come from the narrower tail component in the model. The helix pulls

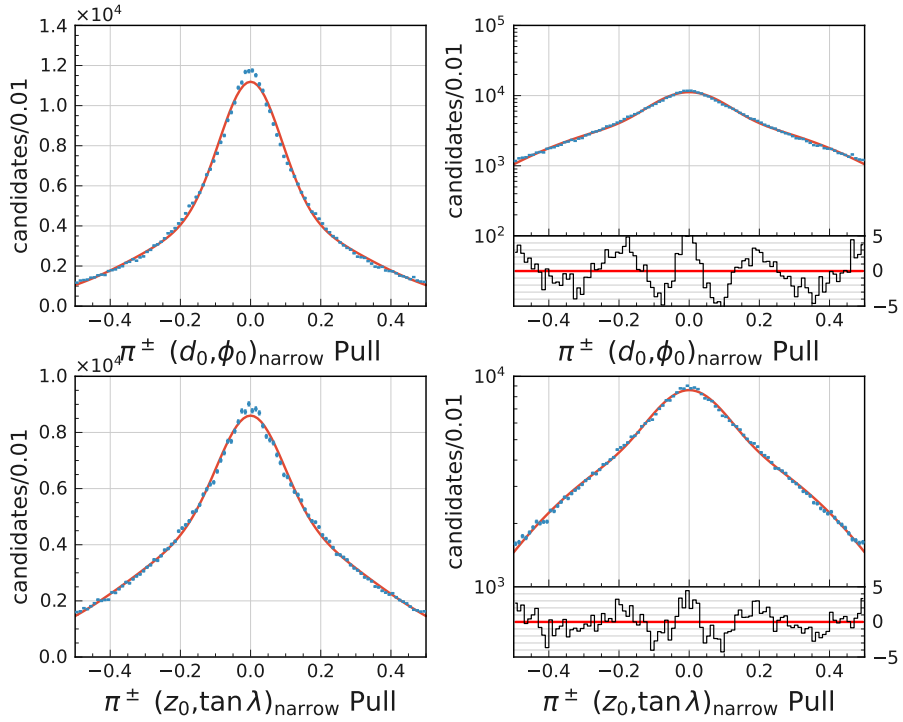


Figure E.2: $(d_0, \phi_0)_{\text{narrow}}$ (upper) and $(z_0, \tan\lambda)_{\text{narrow}}$ (lower) distribution and their fit results of outer-K_S^0 daughters in signal MC. Left for linear scale and right for log scale.

correlation in fast simulation is shown in Fig. E.8. The strong correlation between d_0 and ϕ_0, z_0 and $\tan\lambda$ is reproduced. Since main-tail cross terms in $(d_0, \phi_0)_{\text{wide}}$, $(z_0, \tan\lambda)_{\text{wide}}$, and ω are restricted in our model, the two-dimensional plot shows perfect circle contour.

We fit the R_{rec} and R_{tag} in this fast simulation sample. For the fit, only fully fast-simulated samples are used, i.e., we require helix randomization of both K_S^0 daughter $\pi^\pm$ for $\delta\ell^{CP}$ fit and all the tracks used for tag-side vertex fit for $\delta\ell^{\text{tag}}$. The fit parameters are summarized in Tab. 5.9. The obtained models and distribution in the original signal MC are compared in Fig. 5.18. The standard deviation of R_{rec} depending on $\chi^2_{K_S^0}$ is shown in Fig. E.9, together with $\chi^2_{K_S^0}$ distribution before and after fast simulation.

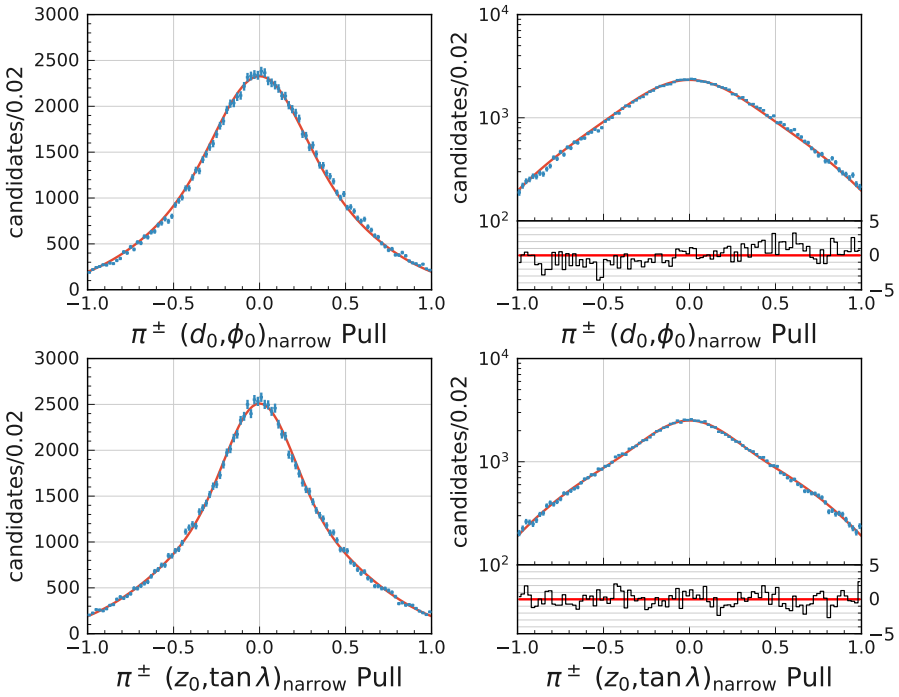


Figure E.3: $(d_0, \phi_0)_{\text{narrow}}$ (upper) and $(z_0, \tan\lambda)_{\text{narrow}}$ (lower) distribution and their fit results of inner-K_S^0 daughters in signal MC. Left for linear scale and right for log scale.

Table E.1: The results of $(d_0, \phi_0)_{\text{narrow}}$ and $(z_0, \tan\lambda)_{\text{narrow}}$ fit in signal MC.

		parameter	fit result
outer-K_S^0	$(d_0, \phi_0)_{\text{narrow}}$	f_{main}	0.3018 ± 0.0034
		s_{main}	0.0864 ± 0.0007
		s_{tail}	0.2920 ± 0.0016
	$(z_0, \tan\lambda)_{\text{narrow}}$	f_{main}	0.180 ± 0.004
		s_{main}	0.0915 ± 0.0013
		s_{tail}	$0.3227^{+0.0021}_{-0.0020}$
inner-K_S^0	$(d_0, \phi_0)_{\text{narrow}}$	f_{main}	$0.259^{+0.021}_{-0.019}$
		s_{main}	0.232 ± 0.008
		s_{tail}	0.512 ± 0.008
	$(z_0, \tan\lambda)_{\text{narrow}}$	f_{main}	0.237 ± 0.011
		s_{main}	0.189 ± 0.005
		s_{tail}	0.503 ± 0.005

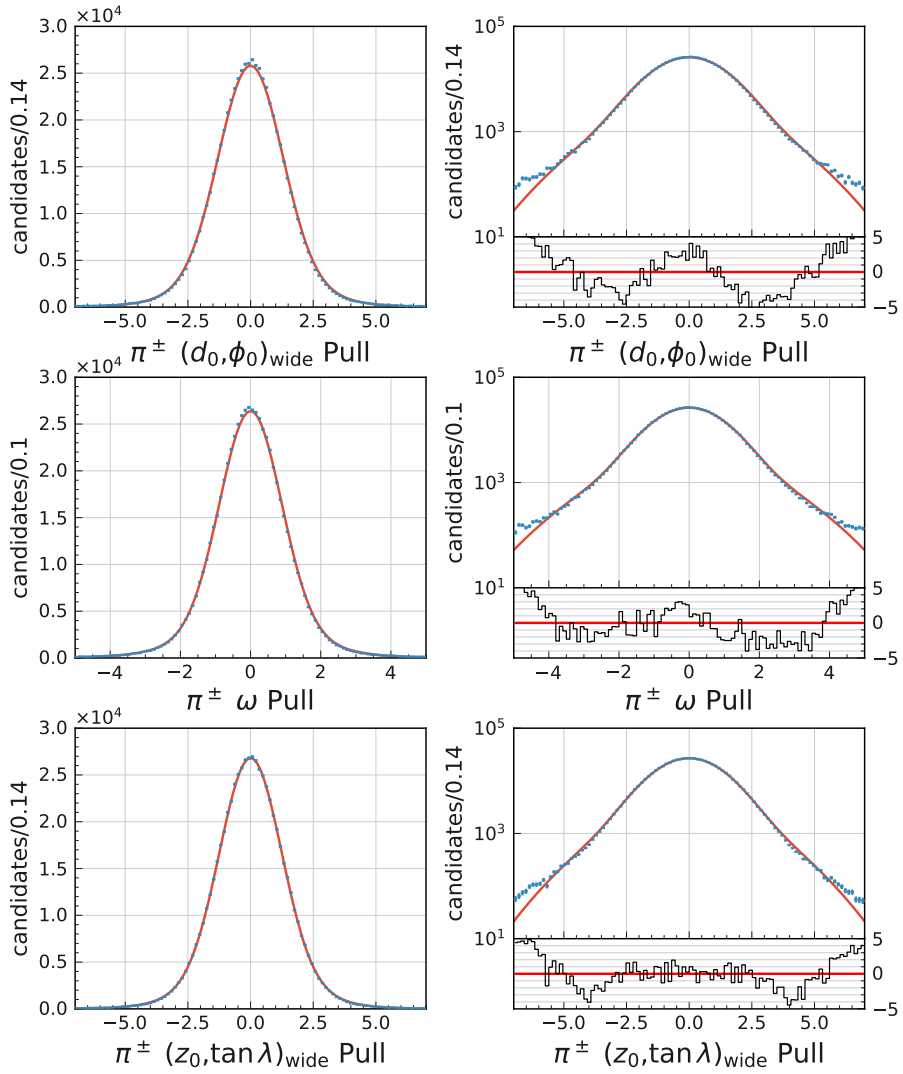


Figure E.4: $(d_0, \phi_0)_{\text{wide}}$, ω, and $(z_0, \tan\lambda)_{\text{wide}}$ (upper, middle, and lower) distribution and the fit result of K_S^0 daughters in signal MC. Left for linear scale and right for log scale.

Table E.2: The results of $(d_0, \phi_0)_{\text{narrow}}$ and $(z_0, \tan\lambda)_{\text{narrow}}$ fit in signal MC.

	parameter	fit result
$(d_0, \phi_0)_{\text{wide}}$	f_{main}	$0.8117^{+0.0018}_{-0.0019}$
	s_{main}	1.2650 ± 0.0019
	s_{tail}	2.331 ± 0.008
ω	s_{main}	0.8744 ± 0.0015
	s_{tail}	1.781 ± 0.007
$(z_0, \tan\lambda)_{\text{wide}}$	s_{main}	1.2184 ± 0.0018
	s_{tail}	2.226 ± 0.007

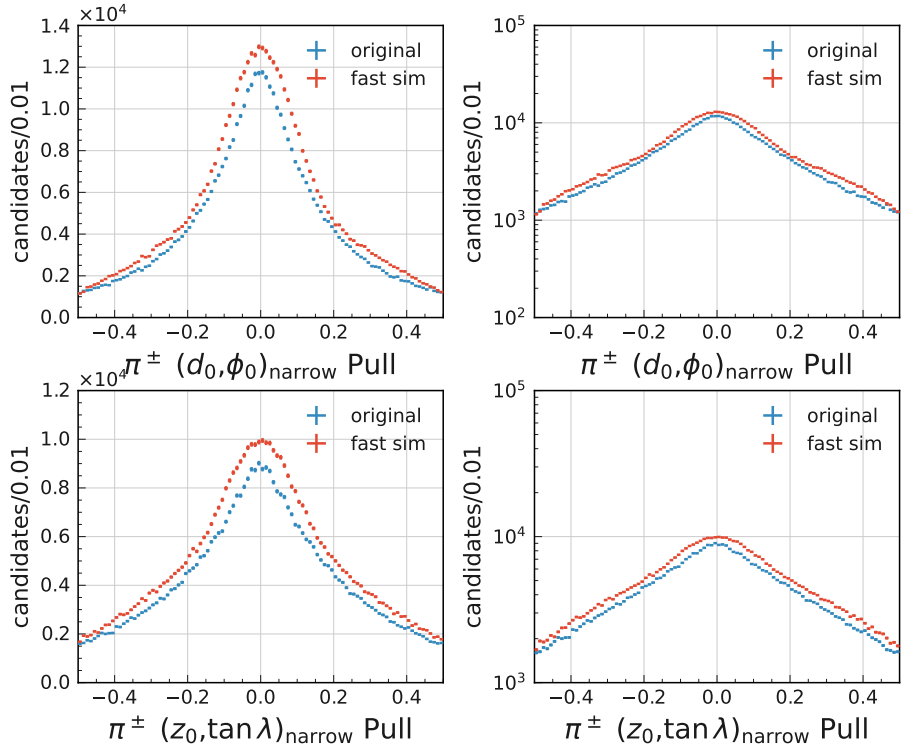


Figure E.5: $(d_0, \phi_0)_{\text{narrow}}$ (upper) and $(z_0, \tan\lambda)_{\text{narrow}}$ (lower) distributions of outer-K^0_S daughters in original signal MC and fast simulation. Left for linear scale and right for log scale.

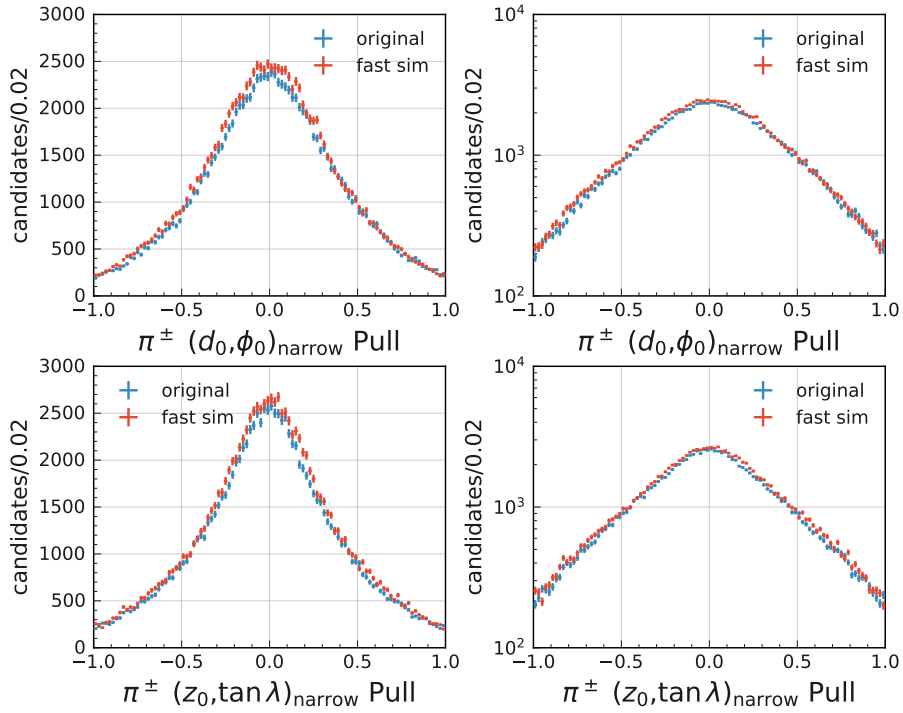


Figure E.6: $(d_0, \phi_0)_{\text{narrow}}$ (upper) and $(z_0, \tan\lambda)_{\text{narrow}}$ (lower) distributions of inner-K_S^0 daughters in original signal MC and fast simulation. Left for linear scale and right for log scale.

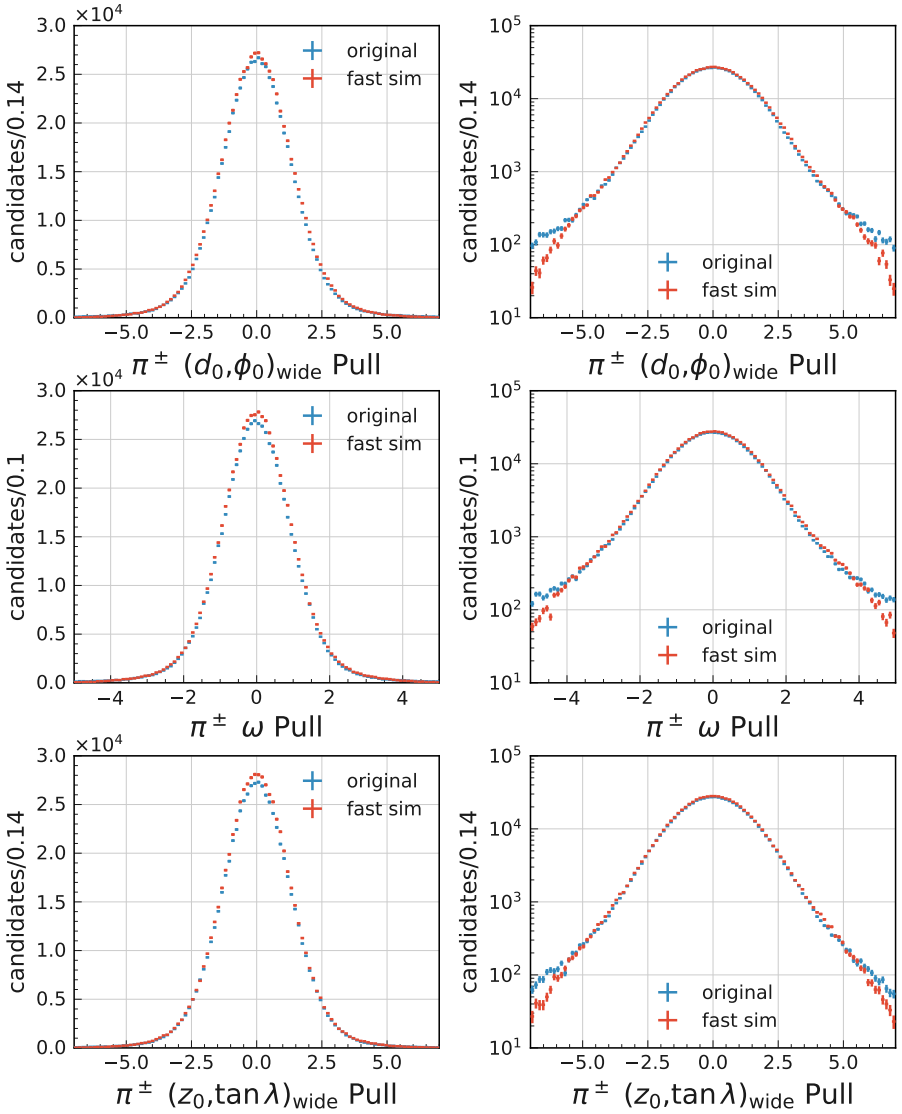


Figure E.7: $(d_0, \phi_0)_{\text{wide}}$, ω, and $(z_0, \tan\lambda)_{\text{wide}}$ (upper, middle, and lower) distributions of K_S^0 daughters in original signal MC and fast simulation. Left for linear scale and right for log scale.

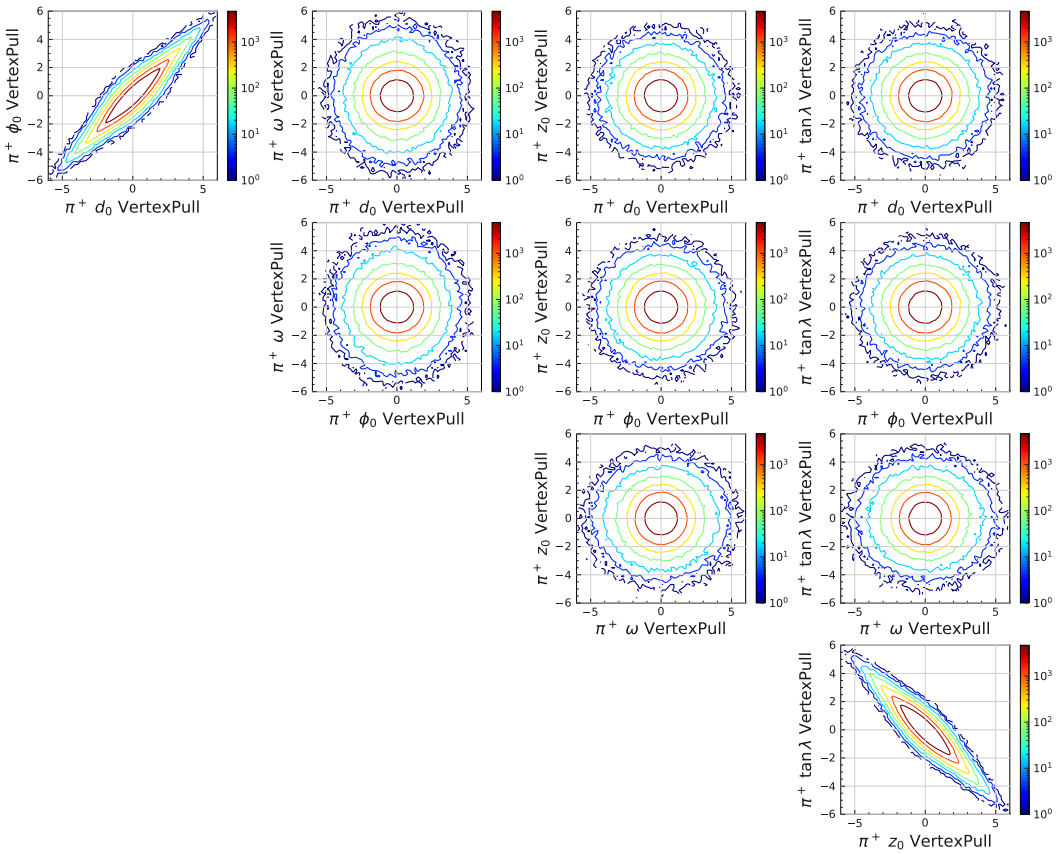


Figure E.8: Two-dimensional helix pull distributions of π^+ tracks from K_S^0 in fast simulation.

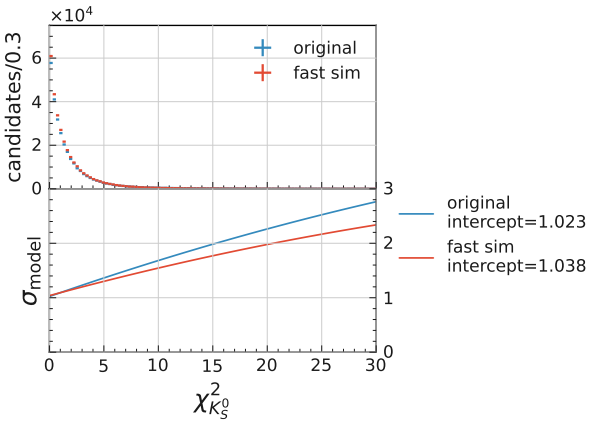


Figure E.9: The standard deviation of R_{rec} depending on $\chi^2_{K^0_S}$ and $\chi^2_{K^0_S}$ distribution before and after fast simulation.

Appendix F

Helix pull calibration

We can calibrate the difference of resolution function parameters between data and MC using cosmic MC and data, assuming that the wrong hit assignment causes the tail component in helix pulls and χ^2 dependence of resolution. We analyze cosmic data taken during beam operation, where we expect the same background condition as the physics run. We prepare two sets of reconstruction. One is for calibrating the effect of wrong PXD and SVD hits, where we require both upper and lower tracks to satisfy

- At least one PXD hits detected.
- $|z_0| > 0.2$ cm to reject peak around IP. We think these tracks come from beam collision events.
- -2 cm $< z_0 < 4$ cm to reject tracks hitting Ti, which is not in the simulated geometry.
- $|d_0| < 1.0$ cm to have material effect from beampipe.

The other is for calibrating the effect of wrong SVD hits only. We do not reconstruct PXD hits for this purpose. We require both tracks to satisfy

- At least three SVD hits detected.
- $|z_0| > 0.2$ cm to reject peak around IP. We think these tracks come from beam collision events.
- -8 cm $< z_0 < 12$ cm
- $|d_0| < 2.5$ cm
- $|p| < 3$ GeV/c

We reconstruct cosmic MC with the same method and criteria.

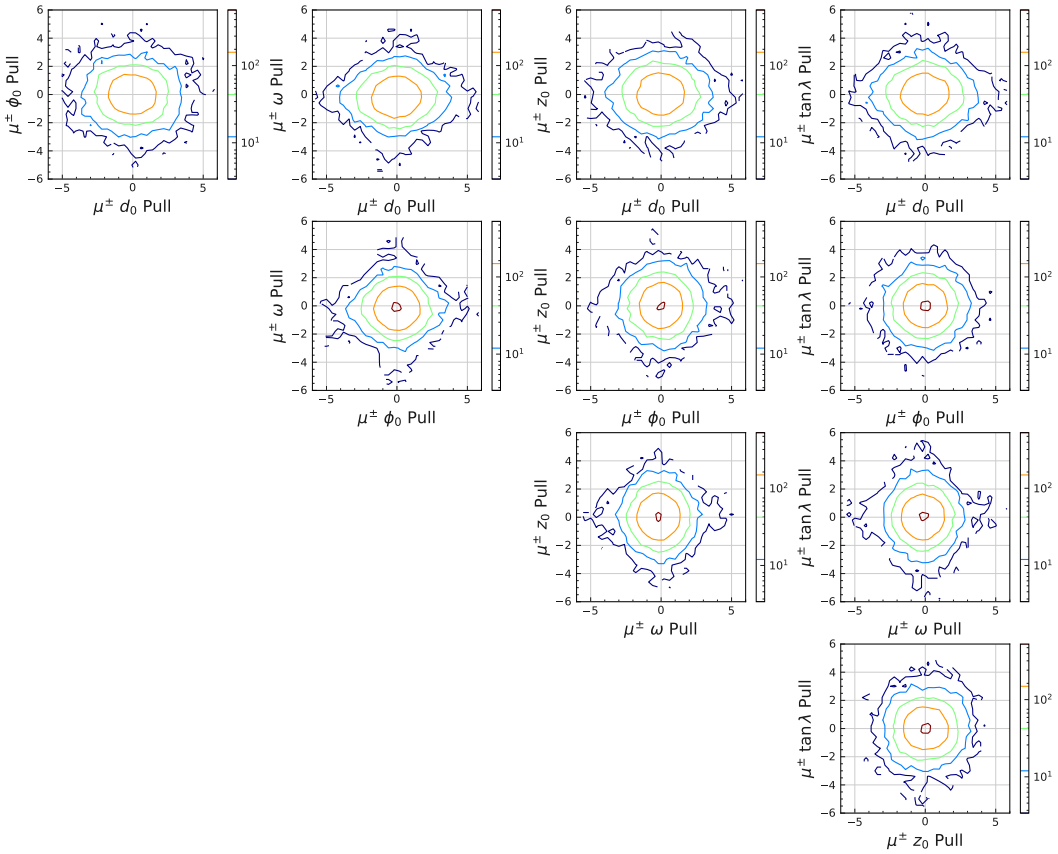


Figure F.1: Two-dimensional helix pull distributions of $\mu^\pm$ tracks in the SVD-reconstructed cosmic data.

We can extract pull information by comparing upper and lower tracks,

$$\text{Pull}(h) = \frac{h_{\text{upper}} - h_{\text{lower}}}{\sqrt{\sigma^2_{h_{\text{upper}}} + \sigma^2_{h_{\text{lower}}}}}, \tag{6.1}$$

where $h_{\text{upper(lower)}}$ for upper (lower) helix and $\sigma_{h_{\text{upper(lower)}}}$ for their errors. The upper cosmic ray is first reconstructed as the track from IP. Then, the charge and momentum of the track, i.e., d_0, ω, and $\tan\lambda$, are flipped. Due to this operation, the strong correlations in the cosmic pulls are smeared as shown in Fig. F.1, i.e., the correlation shows convolution of the one in signal MC and the one with flipped along with d_0 or $\tan\lambda$. We model $(d_0, \phi_0)_{\text{wide}}$ $((z_0, \tan\lambda)_{\text{wide}})$ and $(d_0, \phi_0)_{\text{narrow}}$ $((z_0, \tan\lambda)_{\text{narrow}})$ with the same distribution, convolution of wide and narrow double Gaussian.

Since the narrow width is buried in the wide width, we use the same shape parameters as signal MC. The cosmic ray includes both inner- and outer-K^0_S-like components as we

Table F.1: The results of simultaneous fit of five helices in PXD-reconstructed cosmic MC and data, and obtained correction factors.

	parameter	PXD MC	PXD Data	correction
$(d_0, \phi_0)_{\text{wide}}$	f_{main}	0.9306 ± 0.0022	$0.878^{+0.015}_{-0.016}$	$-0.052^{+0.015}_{-0.016}$
	s_{main}	0.8454 ± 0.0028	0.825 ± 0.015	$(-2.4 \pm 1.8)\%$
	s_{tail}	$2.044^{+0.025}_{-0.024}$	$1.73^{+0.08}_{-0.07}$	$(-15.3^{+3.9}_{-3.6})\%$
ω	μ	0.006 ± 0.0004	-0.274 ± 0.019	N/A
	s_{main}	1.015 ± 0.004	0.975 ± 0.019	$(-4.0 \pm 1.9)\%$
	s_{tail}	$1.785^{+0.029}_{-0.028}$	$1.85^{+0.12}_{-0.10}$	$(+3.5^{+6.8}_{-6.0})\%$
$(z_0, \tan\lambda)_{\text{wide}}$	s_{main}	0.9037 ± 0.0028	0.890 ± 0.015	$(-1.5 \pm 1.7)\%$
	s_{tail}	1.949 ± 0.022	$1.79^{+0.08}_{-0.07}$	$(+8.3^{+4.2}_{-3.8})\%$

analyze cosmic rays regardless of their passing through materials or not. We require $|d_0| < 2.5$ cm in the above selection. Therefore, we use inner-K^0_S-like components, assuming that most tracks pass through PXD or beampipe.

We fit seven parameters to describe three double Gaussian for wide helices and mean μ of ω as we observe shifts in data. The result of simultaneous fit is summarized in Tabs. F.1 and F.2. The obtained model is compared with the distribution in Figs. F.2 and F.3. We obtain correction factors from data–MC comparison. We apply corrections to the helix uncertainty to account for the underestimated uncertainty in PXD hits. Unlike signal MC, the pull width shows dependency on the uncertainty even after this correction. To be ignorant of this possible tail contribution, we correct the fraction with a bias factor. For the width, we simply correct with scales. We do not apply the correction if the correction factor is null-consistent within the uncertainty.

Table F.2: The results of simultaneous fit of five helices in SVD-reconstructed cosmic MC and data, and obtained correction factors.

	parameter	SVD MC	SVD Data	correction
$(d_0, \phi_0)_{\text{wide}}$	f_{main}	0.865 ± 0.005	0.841 ± 0.004	-0.0239 ± 0.0068
	s_{main}	0.908 ± 0.005	0.933 ± 0.006	$(+2.81 \pm 0.87)\%$
	s_{tail}	2.046 ± 0.028	$2.304^{+0.028}_{-0.027}$	$(+12.6 \pm 2.0)\%$
ω	μ	0.009 ± 0.0006	-0.168 ± 0.0006	N/A
	s_{main}	0.940 ± 0.006	0.936 ± 0.005	$(-0.48 \pm 0.84)\%$
	s_{tail}	$1.555^{+0.026}_{-0.025}$	1.596 ± 0.023	$(+2.6 \pm 2.2)\%$
$(z_0, \tan\lambda)_{\text{wide}}$	s_{main}	0.852 ± 0.005	0.839 ± 0.005	$(-1.58 \pm 0.79)\%$
	s_{tail}	1.709 ± 0.022	$1.950^{+0.022}_{-0.021}$	$(+14.1 \pm 1.9)\%$

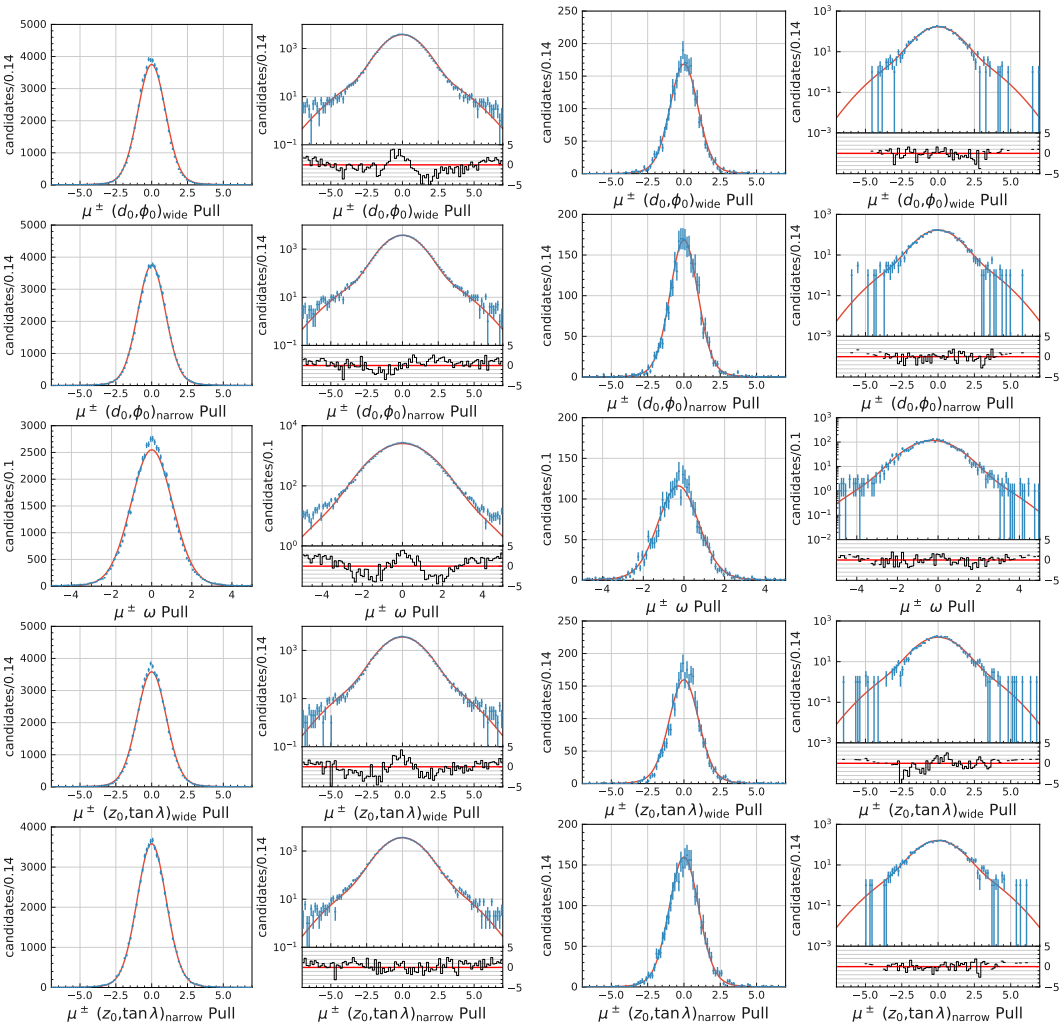


Figure F.2: $(d_0, \phi_0)_{\text{wide}}$, $(d_0, \phi_0)_{\text{narrow}}$, ω, $(z_0, \tan\lambda)_{\text{wide}}$, and $(z_0, \tan\lambda)_{\text{narrow}}$ (from top to bottom) distributions and fitted model of PXD-reconstructed cosmic ray data (right two columns) and MC (left two columns). Left for linear scale and right for log scale.

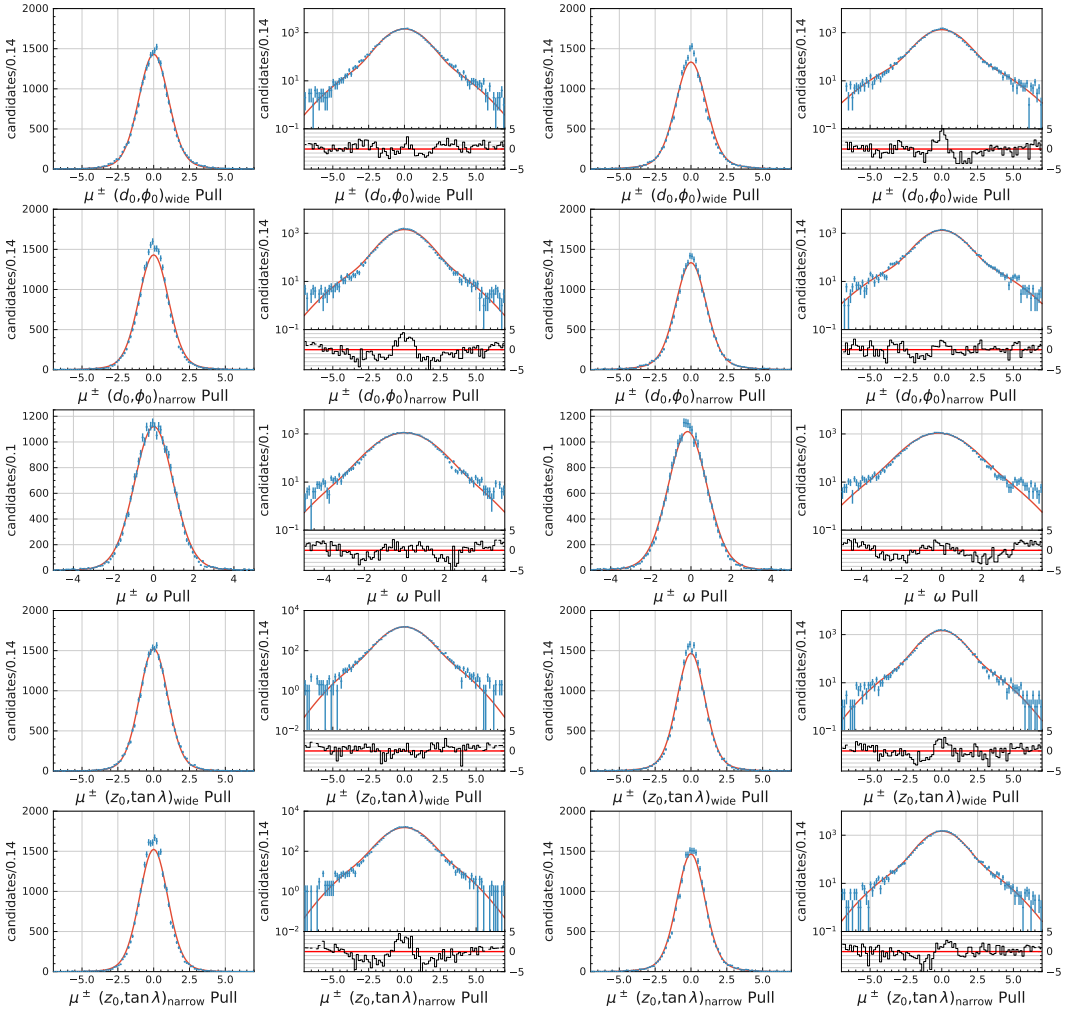


Figure F.3: $(d_0, \phi_0)_{\text{wide}}$, $(d_0, \phi_0)_{\text{narrow}}$, ω, $(z_0, \tan\lambda)_{\text{wide}}$, and $(z_0, \tan\lambda)_{\text{narrow}}$ (from top to bottom) distributions and fitted model of SVD-reconstructed cosmic ray data (right two columns) and MC (left two columns). Left for linear scale and right for log scale.

Appendix G

Tag-side interference

The B^0–$\overline{B}^0$ interference in the $B_{\rm tag}$ decay also cause CP asymmetry called tag-side interference [64]. In Eq.1.20, we model the time-evolution of $B_{\rm tag}$ by decay and B^0–$\overline{B}^0$ mixing. However, the hadronic flavor-specific decay like $\overline{B}^0 \to D^+\pi^-$ interferes with the doubly Cabbibo suppressed decay $B^0 \to D^+\pi^-$. This could bring bias on the measured CP-violation parameters. We can calculate the tag-side interfered CP-violation parameters depending on the flavor,

$$S'_{CP}(q) = \xi_{CP}\,\mathrm{Im}[\lambda] - q\cdot r'(|\lambda|^2-1)\sin(2\phi_1+\phi_3-q\delta')/R(q) \text{ and} \tag{7.1}$$

$$A'_{CP}(q) = \frac{|\lambda|^2-1}{2} - 2q\cdot r'\,\mathrm{Im}[\lambda]\cos(2\phi_1+\phi_3-q\delta')/R(q), \text{ with} \tag{7.2}$$

$$R(q) = \frac{|\lambda|^2+1}{2} - 2q\cdot r'\,\mathrm{Re}[\lambda]\cos(2\phi_1+\phi_3-q\delta'). \tag{7.3}$$

λ is related to the bare CP-violation parameters through Eq. (1.3). The CKM angle ϕ_1 and ϕ_3 are taken from HFLAV [8] as

$$\phi_1 = (22.2\pm0.7)^\circ \quad \text{and} \quad \phi_3 = (65.9^{+3.5}_{-3.3})^\circ. \tag{7.4}$$

The two parameters r' and δ' are transformed from the measured S_+ and S_- in Belle using $B^0 \to D^{*-}\ell^+\nu$ decay [72],

$$S_+ = +0.0096\pm0.0073 \quad \text{and} \quad S_- = -0.0067\pm0.0073, \tag{7.5}$$

through the relation

$$S_+ = 2r'\sin(2\phi_1+\phi_3+\delta') \text{ and} \tag{7.6}$$

$$S_- = 2r'\sin(2\phi_1+\phi_3-\delta'). \tag{7.7}$$

We obtain

$$r' = 0.0118 \quad \text{and} \quad \delta' = 274^\circ. \tag{7.8}$$

For the evaluation, we conservatively assume that all the B_{tag} decays in the hadronic flavor-specific way. We can fairly assume that the number of $q = +1$ and $q = -1$ events are the same, as the flavor tagging asymmetry μ is 0-consistent in hadronic B decay data. $K^0_S\pi^0\gamma$ is a mixture of CP-even ($\xi_{CP} = +1$) and odd ($\xi_{CP} = -1$) states. We calculate the ratio of CP-even and odd states from the measured S_{CP} as

$$\frac{N_{\text{even}} - N_{\text{odd}}}{N_{\text{even}} + N_{\text{odd}}} = \frac{S_{CP}}{\sin 2\phi_1}, \tag{7.9}$$

approximating the bare S_{CP} with measured S_{CP}. We take the difference between the bare (calculated) and tag-side interfered (observed) CP-violation parameters as the single-sided uncertainty.

Bibliography

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

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

[3] CKMfitter Group, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. C **41** (2005) 1, `arXiv:hep-ph/0406184`, ρ and η are calculated from $\overline{\rho}$ and $\overline{\eta}$ using Eq. (17).

[4] A. B. Carter and A. I. Sanda, CP *Nonconservation in Cascade Decays of B Mesons*, Phys. Rev. Lett. **45** (1980) 952.

[5] A. B. Carter and A. I. Sanda, CP *violation in B-meson decays*, Phys. Rev. D **23** (1981) 1567.

[6] I. I. Bigi and A. I. Sanada, *Notes on the observability of CP violations in B decays*, Nucl. Phys. B **193** (1981) 85.

[7] T. Inami and C. S. Lim, *Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $K_L \to \mu\bar{\mu}$, $K^+ \to \pi^+\nu\overline{\nu}$ and $K^0 \leftrightarrow \overline{K}^0$*, Prog. Theor. Phys. **65** (1981) 297, [Erratum: Prog.Theor.Phys. 65, 1772 (1981)].

[8] HFLAV, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. **C81** (2021) 226, `arXiv:1909.12524`, updated results and plots available at `https://hflav.web.cern.ch/`.

[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** (1997) 185, `arXiv:9704272`.

[10] D. Atwood *et al.*, *Mixing-induced CP violation in $B \to P_1 P_2 \gamma$ in search of clean new physics signals*, Phys. Rev. D **71** (2005) 076003.

[11] B. Grinstein *et al.*, *Photon polarization in $B \to X\gamma$ in the standard model*, Phys. Rev. D **71** (2005) 011504.

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

[13] A. Khodjamirian *et al.*, *QCD estimate of the long-distance effect in $B \to K^* \gamma$*, Physics Letters B **402** (1997) 167.

[14] P. Ball and R. Zwicky, *Time-dependent CP asymmetry in $B \to K^* \gamma$ as a (quasi)null test of the Standard Model*, Physics Letters B **642** (2006) 478.

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

[16] M. Matsumori and A. I. Sanda, *Mixing-induced CP asymmetry in $B \to K^* \gamma$ decays with perturbative QCD approach*, Phys. Rev. D **73** (2006) 114022.

[17] D. Bečirević *et al.*, *Future prospects for the determination of the Wilson coefficient $C'_{7\gamma}$*, Journal of High Energy Physics **2012** (2012) 90.

[18] J. C. Pati and A. Salam, *Lepton number as the fourth "color"*, Phys. Rev. D **10** (1974) 275.

[19] F.-S. Yu, E. Kou, and C.-D. Lü, *Photon polarization in the $b \to s\gamma$ processes in the left-right symmetric model*, Journal of High Energy Physics **2013** (2013) 102.

[20] ATLAS, G. Aad *et al.*, *Search for heavy resonances decaying into a Z or W boson and a Higgs boson in final states with leptons and b-jets in* 139 fb^{-1} *of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **06** (2023) 016, arXiv:2207.00230.

[21] S. Dimopoulos and H. Georgi, *Softly broken supersymmetry and SU(5)*, Nucl. Phys. B **193** (1981) 150.

[22] H. Eberl *et al.*, *Imprint of SUSY in radiative B-meson decays*, Phys. Rev. D **104** (2021) 075025.

[23] Belle, Y. Ushiroda *et al.*, *Time-Dependent CP Asymmetries in $B^0 \to K^0_S\pi^0\gamma$ transitions*, Phys. Rev. **D74** (2006) 111104, arXiv:hep-ex/0608017.

[24] BaBar, B. Aubert *et al.*, *Measurement of Time-Dependent CP Asymmetry in $B^0 \to K^0_S\pi^0\gamma$ Decays*, Phys. Rev. **D78** (2008) 071102, arXiv:0807.3103.

[25] Belle II Collaboration, T. Abe, *Belle II Technical Design Report*, arXiv:1011.0352.

[26] SuperKEKB, K. Akai, K. Furukawa, and H. Koiso, *SuperKEKB Collider*, Nucl. Instrum. Meth. **A907** (2018) 188, arXiv:1809.01958, Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn).

[27] https://www.kek.jp/ja/imagearchive/images/20180320_superkekb_002.png.

[28] P. M. Lewis *et al.*, *First measurements of beam backgrounds at SuperKEKB*, Nucl. Instrum. Meth. **A914** (2019) 69.

[29] Belle II DEPFET, PXD, and SVD Collaborations, T. Bilka *et al.*, *Belle II Vertex Detector Performance*, PoS **Vertex2019** (2020) 001.

[30] Belle-II PXD, Q. Liu *et al.*, *Operational Experience and Performance of the Belle II Pixel Detector*, JPS Conf. Proc. **34** (2021) 010002.

[31] K. Adamczyk *et al.*, *The design, construction, operation and performance of the Belle II silicon vertex detector*, J. Instrum. **17** (2022) P11042.

[32] Belle-II DEPFET, PXD, B. Wang *et al.*, *Operational experience of the Belle II pixel detector*, Nucl. Instrum. Meth. **A1032** (2022) 166631.

[33] Older revision is publically available in https://confluence.desy.de/download/attachments/35009238/BellellTopview415R12.pdf.

[34] V. Bertacchi *et al.*, *Track finding at Belle II*, Computer Physics Communications **259** (2021) 107610.

[35] R. Itoh *et al.*, *Data flow and high level trigger of belle II DAQ system*, in *2012 18th IEEE-NPSS Real Time Conference*, pp. 1–4, 2012. doi: 10.1109/RTC.2012.6418174.

[36] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, Comput. Softw. Big Sci. **1** (2017) 2.

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

[38] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[39] S. Jadach, B. F. L. Ward, and Z. Wąs, *The Precision Monte Carlo event generator K K for two fermion final states in e+ e- collisions*, Comput. Phys. Commun. **130** (2000) 260, arXiv:hep-ph/9912214.

[40] T. Sjöstrand *et al.*, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[41] S. Jadach, J. H. Kuhn, and Z. Wąs, *TAUOLA: A library of Monte Carlo programs to simulate decays of polarized tau leptons*, Comput. Phys. Commun. **64** (1990) 275.

[42] A. L. Kagan and M. Neubert, *QCD anatomy of $B \to X_s\gamma$ decays*, The European Physical Journal C - Particles and Fields **7** (1999) 5.

[43] GEANT4, S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[44] Belle-II Framework Software Group, T. Kuhr *et al.*, *The Belle II Core Software*, Comput. Softw. Big Sci. **3** (2019) 1, arXiv:1809.04299.

[45] R. Storn and K. Price, *Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces*, J. Global Optim. **11** (1997) 341.

[46] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, Prog. Theor. Exp. Phys. **2020** (2020) 071102, 083C01.

[47] H. Tanigawa, *Measurement of time-dependent CP asymmetry in $B^0 \to K^0_S K^0_S K^0_S$ decays at the Belle II experiment*, 2022. [Doctoral dissertation, The University of Tokyo]. To be uploaded to UTokyo Repository. https://repository.dl.itc.u-tokyo.ac.jp/.

[48] A. Khotanzad and Y. H. Hong, *Invariant image recognition by zernike moments*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990) 489.

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

[50] BaBar, Belle, A. J. Bevan *et al.*, *The Physics of the B Factories*, Eur. Phys. J. **C74** (2014) 3026, arXiv:1406.6311.

[51] S. Brandt *et al.*, *The principal axis of jets — an attempt to analyse high-energy collisions as two-body processes*, Physics Letters **12** (1964) 57.

[52] E. Farhi, *Quantum Chromodynamics Test for Jets*, Phys. Rev. Lett. **39** (1977) 1587.

[53] Belle Collaboration, S. H. Lee and other, *Evidence for $B^0 \to \pi^0\pi^0$*, Phys. Rev. Lett. **91** (2003) 261801.

[54] D. M. Asner *et al.*, *Search for exclusive charmless hadronic B decays*, Phys. Rev. **D53** (1996) 1039.

[55] The Belle II Collaboration, F. Abudinén *et al.*, *Measurement of the B^0 lifetime and flavor-oscillation frequency using hadronic decays reconstructed in 2019–2021 Belle II data*, Phys. Rev. D **107** (2023) L091102.

[56] Belle II, J.-F. Krohn *et al.*, *Global decay chain vertex fitting at Belle II*, Nucl. Instrum. Meth. **A976** (2020) 164269, arXiv:1901.11198.

[57] T. B. Berrett and R. J. Samworth, *USP: an independence test that improves on Pearson's chi-squared and the G-test*, Proc. R. Soc. A **407** (2021) .

[58] H. Dembinski *et al.*, *scikit-hep/resample*, https://github.com/scikit-hep/resample.

[59] G. Punzi, *Comments on likelihood fits with variable resolution*, eConf **C030908** (2003) WELT002, arXiv:physics/0401045.

[60] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[61] H. Tajima *et al.*, *Proper-time resolution function for measurement of time evolution of B mesons at the KEK B-Factory*, Nucl. Instrum. Meth. **A533** (2004) 370.

[62] Belle II Collaboration, I. Adachi *et al.*, *Measurement of decay-time-dependent CP violation in $B^0 \to J/\psi K^0_S$ decays using 2019-2021 Belle II data*, arXiv:2302.12898.

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

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

[65] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, 2018.

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

[67] LHCb Collaboration, R. Aaij *et al.*, *Measurement of CP-Violating and Mixing-Induced Observables in $B^0_s \to \phi\gamma$ Decays*, Phys. Rev. Lett. **123** (2019) 081802.

[68] 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*, Journal of High Energy Physics **2020** (2020) 81.

[69] N. L. Johnson, *Systems of Frequency Curves Generated by Methods of Translation*, Biometrika **36** (1949) 149.

[70] F. Pedregosa *et al.*, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** (2011) 2825.

[71] D. W. Scott, *On optimal and data-based histograms*, Biometrika **66** (1979) 605, arXiv:https://academic.oup.com/biomet/article-pdf/66/3/605/632347/66-3-605.pdf.

[72] Y. Yusa and K. Miyabayashi, *Measurement of time-dependent CP violation in $B^0 \to (c\overline{c})K^0$ decays with 772M $B\overline{B}$*, Belle Note 1149 (2011).