Doctoral Dissertation
博士論文

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
(Belle II実験における $B^0 \to K^0_S K^0_S K^0_S$ 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 2022
令和4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

Hikaru Tanigawa
谷川　輝

Abstract

$B^0 \to K^0_S K^0_S K^0_S$ decay is mediated by $b \to sq\overline{q}$ penguin transition within the Standard Model. Comparison of *CP*-violating parameters between the penguin-dominated decays and $b \to c\bar{c}s$ decays such as $B^0 \to J/\psi K^0_S$ provides a probe for new physics which contributes to the $b \to s$ transition. We report a measurement of time-dependent *CP* violation in $B^0 \to K^0_S K^0_S K^0_S$ decays, using a data set containing 198×10^6 $B\overline{B}$ pairs collected at the Belle II experiment from 2019 to 2021. We obtain

$$S = -1.86^{+0.91}_{-0.46}(\text{stat}) \pm 0.09(\text{syst}), \text{ and}$$
$$A = -0.22^{+0.30}_{-0.27}(\text{stat}) \pm 0.04(\text{syst}).$$

The result is consistent with the SM expectation $(S, A) = (-0.70 \pm 0.02, 0)$ based on the measurements in $B^0 \to (c\bar{c})K^0$ and the previous measurements at Belle, BaBar, and Belle II. We have established the analysis procedure anticipating the high-precision measurement using $50\,\text{ab}^{-1}$ data to be taken at the Belle II experiment.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Yutaka Ushiroda, for offering me the opportunity to participate in the Belle II experiment and for his guidance.

I would like to thank all members of TDCPV working group, especially Dr. Thibaud Humair and Dr. Varghese Babu for their valuable advice and review to the analysis.

I am also grateful for the internal review committee members, Dr. Hans-Günther Moser, Prof. Livio Lanceri, and Dr. Kookhyun Kang, for their careful review and insightful suggestions.

I would like to thank the Belle II and SuperKEKB colleagues including the Belle secretaries for their efforts to realize the stable operation of the experiment.

I appreciate all the support on academic and technical aspects from Japanese TCPV group members, Dr. Koji Hara, Prof. Kenkichi Miyabayashi, Dr. Akimasa Ishikawa, Dr. Takeo Higuchi, Dr. Yosuke Yusa, Dr. Yo Sato, and Dr. Kota Nakagiri. I would like to especially thank Dr. Kazutaka Sumisawa, without whom I could have never completed the analysis.

Thanks should also go to UTokyo/Sokendai B-factory group members for their educational and moral support. The face-to-face discussions with Mr. Yuma Uematsu and Mr. Hanwook Bae gave me great inspirations and refreshment during the difficult situation due to the pandemic.

Personal contribution

The measurement reported in this thesis is based on the joint effort of more than 1000 members in the Belle II collaboration and the SuperKEKB group. The construction and operation of the experimental apparatus, data management by distributed computing system, data calibration, and development of basic analysis frameworks including vertex fitter, flavor tagger and multivariate analysis are done in collaboration. Main contributions by the author are the development of the event reconstruction algorithm and selection criteria described in Chapter 4, and the development of *CP*-asymmetry estimation procedure described in Chapter 5.

The author has eagerly studied vertex reconstruction and the response functions representing the vertex resolution. The author has revealed that uncertainties of fundamental track parameters are underestimated and developed a correction to the uncertainties as described in Appendix F. The vertex reconstruction adopts novel constraining techniques developed by collaborators, which fully exploit the merit of much smaller beam collision spot than the previous experiment. The tag-side resolution model is directly applicable and useful for future time-dependent *CP* violation analyses, not restricting the decay channel. To validate the response functions, the author has developed a data-driven technique described in Appendix B, which will be more useful with more statistics, and extended the study to the control channel, $B^+ \to K^0_S K^0_S K^+$, where a special technique is used to emulate the vertex resolution of the signal channel. Based on the vertex resolution study, the author takes a conservative approach to restrict the use of decay time information to carefully selected events, while recycling the filtered events without using the decay time to restore the sensitivity. The author has shown that the sensitivity can be fully restored in near future as described in Appendix E.

The author pursued better sensitivity by including events contaminated by $q\overline{q}$ background in the *CP* fit rather than applying tight continuum suppression. Accordingly, the signal fraction is extended to take into account the $q\overline{q}$ contamination.

The author has sophisticated the analysis by resolving possible small biases in the measurement, anticipating a high-precision measurement with a larger data set to be taken at the Belle II experiment. The improvement includes the use of r bin and $\cos\theta^*_B$ in signal fraction estimation to avoid the Punzi effect and the removal of a possible peaking background component that was neglected in previous analyses.

Outline

In Chapter 1 we review the mechanism of time-dependent *CP* violation in B-meson system and manifest the sensitivity of $B^0 \to K^0_S K^0_S K^0_S$ decays to physics beyond the Standard Model. In Chapter 2 we describe the experimental apparatus with which we have recorded our data set, the SuperKEKB collider and the Belle II detector. In Chapter 3 we give an overview of our measurement of time-dependent *CP*-violating parameters of $B^0 \to K^0_S K^0_S K^0_S$ decays. In Chapter 4 we explain the reconstruction algorithm and selection criteria for the $B^0 \to K^0_S K^0_S K^0_S$ decays. In Chapter 5 we describe the method to extract the *CP*-violating parameters from the reconstructed and selected data, various validation tests of the method, the result of our measurement, and systematic uncertainties in it. In Chapter 6 we discuss on the statistical significance of the measurement, improvement from Belle, and future prospect. We conclude the thesis in Chapter 7.

Contents

List of Figures

List of Tables

Chapter 1

Physics motivation

The Standard Model (SM) describes fundamental interactions between elementary particles. While it is a greatly successful theory to describe most of observations, there remain unsolved phenomena and unnaturalness such as the matter-antimatter asymmetry problem, dark matter, dark energy, quantum gravity, hierarchy problem, and strong *CP* problem. The matter-antimatter asymmetry problem indicates the existence of an additional source of violation of *CP* symmetry other than the quark mixing matrix in the SM. C and P refer to charge-conjugate and parity transformations, respectively.

We review the mechanism of *CP* violation in B-meson system within the SM and that the precise measurement of *CP* violation can probe the additional *CP*-violating phase arising from physics beyond the SM (new physics, NP) [2].

1.1 CKM matrix and Unitarity triangle

The Standard Model (SM) of elementary particles is a chiral gauge field theory with $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry. The interactions are described by the gauge symmetry. $SU(3)_C$ describes the strong interaction. Because of the nonzero vacuum expectation value of the Higgs field, $SU(2)_L \times U(1)_Y$ spontaneously breaks to $U(1)_{EM}$. Consequently, the $SU(2)_L$ and $U(1)_Y$ gauge bosons are mixed into photon and weak bosons, which respectively mediate the electromagnetic and weak interaction.

The SM contains three generations of up-type and down-type quarks with the quantum numbers shown in Tab. 1.1. The left-handed fields form doublets of the $SU(2)_L$ group and the right-handed fields are singlets. The quark fields bear different $U(1)_Y$ hypercharges Y. The coupling between the quarks and the Higgs field is called the Yukawa interaction. Because the Higgs field has a nonzero vacuum expectation value v, the Yukawa interaction gives rise to the quark mass terms:

$$\mathcal{L}_m = -v\ \overline{u}'_{Lj} Y^u_{jk} u'_{Rk} - v\overline{d}'_{Lj} Y^d_{jk} d'_{Rk} + \text{h.c.} \tag{1.1}$$

Here, $u'_{L,Rj}$ and $d'_{L,Rj}$ are weak eigenstates of the quark fields of jth generation, Y^u and

Table 1.1: Quantum numbers of the quark fields. $u_{L(R)}$ and $d_{L(R)}$ represent left-handed (right-handed) up-type and down-type quark fields, respectively. The index $j = 1, 2, 3$ denotes the generation. T_3 is the third component of weak isospin, Y is hypercharge, Q is electric charge of $U(1)_{EM}$ symmetry.

	T_3	Y	Q
$\begin{pmatrix} u_{Lj} \\ d_{Lj} \end{pmatrix}$	$+\frac{1}{2}$ $-\frac{1}{2}$	$+\frac{1}{3}$	$+\frac{2}{3}$ $-\frac{1}{3}$
u_{Rj}	0	$+\frac{4}{3}$	$+\frac{2}{3}$
d_{Rj}	0	$-\frac{2}{3}$	$-\frac{1}{3}$

Y^d are 3×3 complex Yukawa matrices, and h.c. stands for Hermitian conjugate. The terms are summed over the repeated indices. The Yukawa matrices can be diagonalized by unitary rotations of the quark fields in the generation space, $S^{u(d)}_{L(R)}$:

$$u'_{L(R)j} = S^u_{L(R)jk} u_{L(R)k}, \tag{1.2}$$

$$d'_{L(R)j} = S^d_{L(R)jk} d_{L(R)k}, \tag{1.3}$$

$$\begin{aligned} \mathcal{L}_m &= -v\overline{u}_{Lj} S^{u\dagger}_{Ljl} Y^u_{lm} S^u_{Rmk} u_{Rk} - v\overline{d}_{Lj} S^{d\dagger}_{Ljl} Y^d_{lm} S^d_{Rmk} d'_{Rk} + \text{h.c.} \\ &= -\overline{u}_{Lj} M^u_{jk} u_{Rk} - \overline{d}_{Lj} M^d_{jk} d_{Rk} + \text{h.c.} \\ &= -\sum_{q=u,d,s,c,b,t} m_q (\overline{q}_L q_R + \overline{q}_R q_L) \\ &= -\sum_{q=u,d,s,c,b,t} m_q \overline{q} q \end{aligned} \tag{1.4}$$

where $M^u = S^{u\dagger} Y^u S^u = \text{diag}(m_u, m_c, m_t)$ and $M^d = S^{d\dagger} Y^d S^d = \text{diag}(m_d, m_s, m_b)$ are the diagonalized mass matrices, $u_{L(R)1,2,3} = u_{L(R)}, c_{L(R)}, t_{L(R)}$ and $d_{L(R)1,2,3} = d_{L(R)}, s_{L(R)}, b_{L(R)}$ are the mass eigenstates of quarks, and $q = q_L + q_R$ is the Dirac field of quarks.

Using the weak eigenstates, the charged current of weak interaction is expressed as

$$\mathcal{L}_W = \frac{g_2}{\sqrt{2}} \left[\overline{u}'_{Lj} \gamma^\mu d'_{Lj} W^+_\mu + \overline{d}'_{Lj} \gamma^\mu d'_{Lj} W^-_\mu \right], \tag{1.5}$$

where $W^\pm_\mu$ is the W-boson field, γ^μ is gamma matrices, and g_2 is the coupling constant of $SU(2)_L$. This is also transformed by the unitary rotation of the left-handed quark fields $S^{u,d}_L$:

$$\begin{aligned} \mathcal{L}_W &= \frac{g_2}{\sqrt{2}} \left[\overline{u}_{Lj} S^{u\dagger}_{Ljl} \gamma^\mu S^d_{Llk} d_{Lk} W^+_\mu + \overline{d}_{Lk} S^{d\dagger}_{Ljl} \gamma^\mu S^u_{lk} d_{Lj} W^-_\mu \right] \\ &= \frac{g_2}{\sqrt{2}} \left[\overline{u}_{Lj} V_{jk} \gamma^\mu d_{Lk} W^+_\mu + \overline{d}_{Lk} V^*_{jk} \gamma^\mu d_{Lj} W^-_\mu \right] \end{aligned} \tag{1.6}$$

where $V = S^{u\dagger} S^d$ is a 3×3 unitary matrix in the generation space called the Cabibbo-Kobayashi-Maskawa (CKM) matrix. We redefine the weak doublets of quarks as

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s' \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b' \end{pmatrix}_L, \tag{1.7}$$

where the weak eigenstates of the down-type quarks relate to the mass eigenstates via the CKM matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.8}$$

A 3×3 unitary matrix is characterized by three rotation angles and six complex phases. Since the mass terms and kinetic terms of quarks are invariant under the phase rotation of the individual quark Dirac fields, five phases of the CKM matrix are absorbed by rotating the relative phases between the six quark fields. Thus, four parameters are left free: three angles $\theta_{ij}(ij = 12, 13, 23)$ and one complex phase δ. Defining $s_{ij} \equiv \sin\theta_{ij}, c_{ij} \equiv \cos\theta_{ij}$, we write the matrix as

$$\begin{aligned} V &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}. \end{aligned} \tag{1.9}$$

δ is the only source of *CP* violation in the SM. The Wolfenstein parametrization is often used to explicitly show the hierarchy between the mixing angles $1 \gg s_{12} \gg s_{23} \gg s_{13}$ [3][4]:

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \tag{1.10}$$

$$s_{23} = A\lambda^2 = \lambda \left| \frac{V_{cb}}{V_{us}} \right|, \tag{1.11}$$

$$s_{13}e^{i\delta} = V_{ub}^* = A\lambda^3(\rho + i\eta) = \frac{A\lambda^3(\overline{\rho} + i\overline{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1 - \lambda^2}(1 - A^2\lambda^4(\overline{\rho} + i\overline{\eta}))}, \tag{1.12}$$

$$V = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4), \tag{1.13}$$

where, $\lambda \simeq 0.22$ and $A, \rho, \eta = O(1)$.

The unitarity of the CKM matrix requires $\sum_i V_{ij}V_{ik}^* = \delta_{jk}$ and $\sum_j V_{ij}V_{kj}^* = \delta_{ik}$. Among the unitarity conditions, the following equation is suitable for measurement since all terms are of equal order, $O(A\lambda^3)$:

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

Divided by $V_{cd}V_{cb}^*$, the unitarity condition is expressed as the closure of the triangle shown

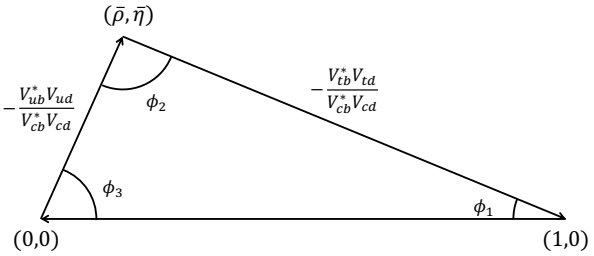


Figure 1.1: Unitarity triangle

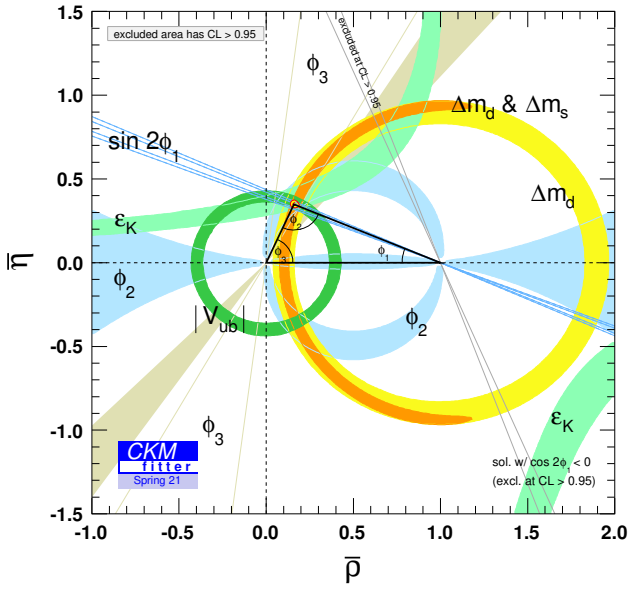


Figure 1.2: Constraints of the unitarity triangle [5]

in Fig. 1.1, called the "unitarity triangle". Its interior angles are given by

$$\begin{aligned}\phi_1 &= \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right)\\ \phi_2 &= \alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right)\\ \phi_3 &= \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right).\end{aligned} \tag{1.15}$$

The unitarity has been intensively tested by various independent measurements of the side lengths and angles as shown in Fig. 1.2.

Measurements of *CP* violation in B-meson system are sensitive to the interior angles. ϕ_1 is most precisely determined through the time-dependent *CP* asymmetry measurement of $B^0 \to J/\psi K_S^0$ as explained later. ϕ_2 is determined by measuring the *CP* asymmetry of $b \to u\overline{u}d$ decays. The isospin analyses of $B \to \pi\pi,\ \rho\rho,\ \rho\pi$ individually determine ϕ_2. ϕ_3 is extracted from the interference of $b \to c\overline{u}s$ and $b \to u\overline{c}s$ transitions in $B \to DK$ decays. Since the decay modes include no penguin contribution, they provide theoretically clean measurement of ϕ_3.

1.2 Time-dependent *CP* violation

For a final state f reachable from both B^0 and $\overline{B}^0$, there are two paths in which B^0 decays to f: either directly ($B^0 \to f$) or via mixing to $\overline{B}^0$ ($B^0 \to \overline{B}^0 \to f$). Because the $B^0 - \overline{B}^0$ mixing involves a *CP* violating phase, the interference of the two paths leads to time-dependent *CP* violation. [6]

1.2.1 *CP* violation induced by $B^0 - \overline{B}^0$ mixing

We consider the time evolution of B-meson system, which is initially a superposition of B^0 and $\overline{B}^0$:

$$|\psi(0)\rangle = a(0)\,|B^0\rangle + b(0)\,|\overline{B}^0\rangle\,, \tag{1.16}$$

$$|\psi(t)\rangle = a(t)\,|B^0\rangle + b(t)\,|\overline{B}^0\rangle + \sum_i c_i(t)\,|f_i\rangle\,, \tag{1.17}$$

where f_i denotes any final state to which B meson decays. Aiming to obtain the expressions of only $a(t)$ and $b(t)$, we restrict $|\psi(t)\rangle$ to the subspace of $|B^0\rangle$ and $|\overline{B}^0\rangle$, and do not argue the evolution of decay final states. Then the time evolution of $|\psi(t)\rangle$ is described by the following Schrödinger equation:

$$\begin{aligned} i\frac{\partial}{\partial t}\psi(t) &= \boldsymbol{H}\psi(t),\\ \boldsymbol{H} &= \begin{pmatrix}\langle B^0|\mathcal{H}|B^0\rangle & \langle B^0|\mathcal{H}|\overline{B}^0\rangle\\ \langle \overline{B}^0|\mathcal{H}|B^0\rangle & \langle \overline{B}^0|\mathcal{H}|\overline{B}^0\rangle\end{pmatrix}\\ &= \boldsymbol{M} - i\boldsymbol{\Gamma},\end{aligned} \tag{1.18}$$

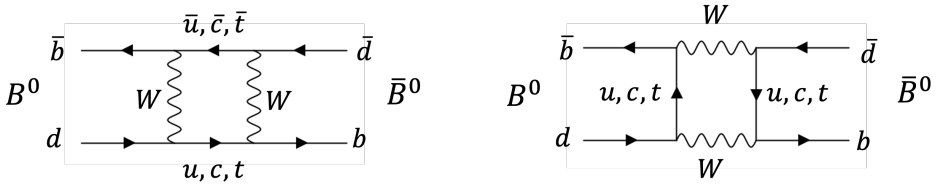


Figure 1.3: Box diagrams of $B^0 - \overline{B}^0$ mixing

where $\psi(t) = (a(t), b(t))^T$, $\mathcal{H}$ is the effective Hamiltonian of weak interaction, $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ are Hermitian matrices, and $\boldsymbol{H}$ is a non-Hermitian matrix. $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ correspond to dispersive (via off-shell intermediate states) and absorptive (via on-shell intermediate states) parts, respectively. The $B^0 - \overline{B}^0$ transition occurs via the box diagrams shown in Fig. 1.3. Assuming CPT invariance $(CPT)\mathcal{H}(CPT)^{-1} = \mathcal{H}$, where T is the time-reversal transformation, we obtain $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$.

The mass eigenstates of the system is given by diagonalizing $\boldsymbol{H}$:

$$|B_{1,2}\rangle \equiv p\,|B^0\rangle \pm q\,|\overline{B}^0\rangle\,, \tag{1.19}$$

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}, \tag{1.20}$$

$$|p|^2 + |q|^2 = 1. \tag{1.21}$$

$$\tag{1.22}$$

The eigenvalues correspond to the masses and decay widths of $B_{1,2}$:

$$m_{1,2} - \frac{i}{2}\Gamma_{1,2} \;=\; M_{11} - \frac{i}{2}\Gamma_{11} \pm \frac{q}{p}\Big(M_{12} - \frac{i}{2}\Gamma_{12}\Big). \tag{1.23}$$

We define the mean and difference of them:

$$\begin{aligned} m &\equiv (m_1 + m_2)/2, \\ \Gamma &\equiv (\Gamma_1 + \Gamma_2)/2, \\ \Delta m_d - \frac{i}{2}\Delta\Gamma &\equiv \left(m_2 - \frac{i}{2}\Gamma_2\right) - \left(m_1 - \frac{i}{2}\Gamma_1\right) \\ &= 2\sqrt{M_{12} - \frac{i}{2}\Gamma_{12}}\sqrt{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}. \end{aligned} \tag{1.24}$$

Eq. (1.18) is solved as:

$$\begin{aligned} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} &= \begin{pmatrix} \frac{1}{2p} & \frac{1}{2p} \\ \frac{1}{2q} & -\frac{1}{2q} \end{pmatrix} \begin{pmatrix} e^{-im_1 t - \frac{1}{2}\Gamma_1 t} & 0 \\ 0 & e^{-im_2 t - \frac{1}{2}\Gamma_2 t} \end{pmatrix} \begin{pmatrix} p & q \\ q & -q \end{pmatrix} \begin{pmatrix} a(0) \\ b(0) \end{pmatrix} \\ &= \begin{pmatrix} g_+(t) & -\frac{q}{p}g_-(t) \\ -\frac{p}{q}g_-(t) & g_+(t) \end{pmatrix} \begin{pmatrix} a(0) \\ b(0) \end{pmatrix}, \end{aligned} \tag{1.25}$$

where we define

$$g_\pm(t) \equiv \frac{1}{2}\left(e^{-im_2 t - \frac{1}{2}\Gamma_2 t} \pm e^{-im_1 t - \frac{1}{2}\Gamma_1 t}\right). \tag{1.26}$$

Now we consider the decay of initially pure B^0 or $\overline{B}^0$ meson to a common CP eigenstate f with its eigenvalue η_f. Using Eq. (1.25), we can write the states after a proper time t as

$$|B^0(t)\rangle = g_+(t)\,|B^0\rangle - \frac{q}{p} g_-(t)\,|\overline{B}^0\rangle\,, \tag{1.27}$$

$$|\overline{B}^0(t)\rangle = g_+(t)\,|\overline{B}^0\rangle - \frac{p}{q} g_-(t)\,|B^0\rangle\,. \tag{1.28}$$

Defining

$$x \equiv \Delta m_d/\Gamma, y \equiv \Delta\Gamma/(2\Gamma), \tag{1.29}$$

$$\mathcal{A} \equiv \langle f|\mathcal{H}|B^0\rangle\,, \overline{\mathcal{A}} \equiv \langle f|\mathcal{H}|\overline{B}^0\rangle\,, \tag{1.30}$$

$$\lambda_f \equiv \frac{q}{p}\frac{\overline{\mathcal{A}}}{\mathcal{A}}, \tag{1.31}$$

we find the decay rates at t to be

$$\Gamma(B^0 \to f;t) \propto \left|\langle f|\mathcal{H}|B^0(t)\rangle\right|^2 \tag{1.32}$$

$$= e^{-\Gamma t}\Bigg[\left(|\mathcal{A}|^2 + \left|\frac{q}{p}\overline{\mathcal{A}}\right|^2\right)\cosh(y\Gamma t) + \left(|\mathcal{A}|^2 - \left|\frac{q}{p}\overline{\mathcal{A}}\right|^2\right)\cos(x\Gamma t)$$

$$+ 2\mathrm{Re}\left(\frac{q}{p}\mathcal{A}^*\overline{\mathcal{A}}\right)\sinh(y\Gamma t) - 2\mathrm{Im}\left(\frac{q}{p}\mathcal{A}^*\overline{\mathcal{A}}\right)\sin(x\Gamma t)\Bigg], \tag{1.33}$$

$$= e^{-\Gamma t}|\mathcal{A}|^2\Bigg[\left(1+|\lambda_f|^2\right)\cosh(y\Gamma t) + \left(1-|\lambda_f|^2\right)\cos(x\Gamma t)$$

$$+ 2\mathrm{Re}(\lambda_f)\sinh(y\Gamma t) - 2\mathrm{Im}(\lambda_f)\sin(x\Gamma t)\Bigg], \tag{1.34}$$

$$\Gamma(\overline{B}^0 \to f;t) \propto \left|\langle f|\mathcal{H}|\overline{B}^0(t)\rangle\right|^2 \tag{1.35}$$

$$= e^{-\Gamma t}\Bigg[\left(\left|\frac{p}{q}\mathcal{A}\right|^2 + |\overline{\mathcal{A}}|^2\right)\cosh(y\Gamma t) - \left(\left|\frac{p}{q}\mathcal{A}\right|^2 - |\overline{\mathcal{A}}|^2\right)\cos(x\Gamma t)$$

$$+ 2\mathrm{Re}\left(\frac{p}{q}\mathcal{A}\overline{\mathcal{A}}^*\right)\sinh(y\Gamma t) - 2\mathrm{Im}\left(\frac{p}{q}\mathcal{A}\overline{\mathcal{A}}^*\right)\sin(x\Gamma t)\Bigg], \tag{1.36}$$

$$= e^{-\Gamma t}\left|\frac{p}{q}\right|^2|\mathcal{A}|^2\Bigg[\left(1+|\lambda_f|^2\right)\cosh(y\Gamma t) - \left(1-|\lambda_f|^2\right)\cos(x\Gamma t)$$

$$+ 2\mathrm{Re}(\lambda_f)\sinh(y\Gamma t) + 2\mathrm{Im}(\lambda_f)\sin(x\Gamma t)\Bigg]. \tag{1.37}$$

For neutral B meson, $\Delta\Gamma \ll \Gamma$ and $|q/p| = 1$ hold to a good approximation. Then the asymmetry of the decay rates is written in a simple form:

$$asym(t) = \frac{\Gamma(\overline{B}^0 \to f;t) - \Gamma(B^0 \to f;t)}{\Gamma(\overline{B}^0 \to f;t) + \Gamma(B^0 \to f;t)} \tag{1.38}$$

$$= S\sin\Delta m_d t + A\cos\Delta m_d t, \tag{1.39}$$

where we define the CP asymmetries as

$$S \equiv \frac{2\mathrm{Im}(\lambda_f)}{|\lambda_f|^2+1}, \quad A \equiv \frac{|\lambda_f|^2-1}{|\lambda_f|^2+1}. \tag{1.40}$$

S is called mixing-induced CP asymmetry and occurs via the interference between decays with and without net $B^0 - \overline{B}^0$ mixing. A is called direct CP asymmetry and occurs only when $|\mathcal{A}| \neq |\overline{\mathcal{A}}|$. This requires that $B^0 \to f$ proceeds in more than two different diagrams whose CP-violating phases (weak phases) and CP-invariant phases (strong phases) are both different between the diagrams. A and S are bounded in a unit circle by definition:

$$S^2 + A^2 \leq 1. \tag{1.41}$$

B^0 is suitable for CP violation measurement because the lifetime and mixing period are on the same order: $\tau_{B^0} = 1.519 \pm 0.004\,\mathrm{ps}$ and $\Delta m_d = 0.5065 \pm 0.0019\,\mathrm{ps}^{-1}$ [7].

1.2.2 CP violation measurement at Belle II

At the Belle II experiment, $B^0\overline{B}^0$ pair is coherently produced via $\Upsilon(4S)$ vector resonance, $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$. Because the mass of $\Upsilon(4S)$ is only slightly above the $B\overline{B}$ production threshold, it decays into B^+B^- or $B^0\overline{B}^0$ with the branching fraction of approximately 50% each. As the $B\overline{B}$ pair preserve the charge-conjugate eigenvalue $C = -1$, they cannot be B^0B^0 or $\overline{B}^0\overline{B}^0$ at the same time while proceeding with $B^0 - \overline{B}^0$ oscillation. At the moment when one B (B_{tag}) decays into a final state that tells its flavor (for example semileptonic decays), we can identify the flavor of the other B (B_{CP}) to be the opposite to that of B_{tag} based on the quantum entanglement between them.

Let us consider the explicit expression for time evolution of B_{CP} and B_{tag} after proper times t_{CP} and t_{tag} respectively from the production. They are in a $C = -1$ state at the production time $t_{CP} = t_{\mathrm{tag}} = 0$:

$$|B_{CP}B_{\mathrm{tag}}(t_{CP} = t_{\mathrm{tag}} = 0)\rangle = \frac{1}{\sqrt{2}}\left(|B^0_{CP}\overline{B}^0_{\mathrm{tag}}\rangle - |\overline{B}^0_{CP}B^0_{\mathrm{tag}}\rangle\right). \tag{1.42}$$

Using Eq. (1.25), $\Delta\Gamma = 0, t_+ \equiv t_{CP} + t_{\mathrm{tag}}$, and $\Delta t \equiv t_{CP} - t_{\mathrm{tag}}$, we obtain

$$\begin{aligned}
|B_{CP}B_{\mathrm{tag}}(t_{CP}, t_{\mathrm{tag}})\rangle = \frac{1}{\sqrt{2}}\Bigg\{ & \frac{p}{q}\left[g_-(t_{CP})g_+(t_{\mathrm{tag}}) - g_+(t_{CP})g_-(t_{\mathrm{tag}})\right]|B^0_{CP}B^0_{\mathrm{tag}}\rangle \\
& + \left[g_+(t_{CP})g_+(t_{\mathrm{tag}}) - g_-(t_{CP})g_-(t_{\mathrm{tag}})\right]|B^0_{CP}\overline{B}^0_{\mathrm{tag}}\rangle \\
& + \left[g_-(t_{CP})g_-(t_{\mathrm{tag}}) - g_+(t_{CP})g_+(t_{\mathrm{tag}})\right]|\overline{B}^0_{CP}B^0_{\mathrm{tag}}\rangle \\
& + \frac{q}{p}\left[g_+(t_{CP})g_-(t_{\mathrm{tag}}) - g_-(t_{CP})g_+(t_{\mathrm{tag}})\right]|\overline{B}^0_{CP}\overline{B}^0_{\mathrm{tag}}\rangle\Bigg\}
\end{aligned} \tag{1.43}$$

$$\begin{aligned}
= \frac{1}{\sqrt{2}}e^{-imt_+}e^{-\frac{1}{2}\Gamma t_+}\Bigg\{ & i\sin\frac{\Delta m_d\Delta t}{2}\left[\frac{p}{q}|B^0_{CP}B^0_{\mathrm{tag}}\rangle - \frac{q}{p}|\overline{B}^0_{CP}\overline{B}^0_{\mathrm{tag}}\rangle\right] \\
& + \cos\frac{\Delta m_d\Delta t}{2}\left[|B^0_{CP}\overline{B}^0_{\mathrm{tag}}\rangle - |\overline{B}^0_{CP}B^0_{\mathrm{tag}}\rangle\right]\Bigg\}.
\end{aligned} \tag{1.44}$$

Eq. (1.44) confirms that the $B\overline{B}$ pair cannot be B^0B^0 or $\overline{B}^0\overline{B}^0$ at the same time ($\Delta t = 0$). We consider the probability that B_{CP} decays into a CP eigenstate f_{CP} and $B_{\rm tag}$ into a flavor specific state $f_{\rm tag}$ or its conjugate $\overline{f}_{\rm tag}$. We define $\mathcal{A} \equiv \langle f_{CP}|B^0_{CP}\rangle\,, \overline{\mathcal{A}} \equiv \langle f_{CP}|\overline{B}^0_{CP}\rangle\,, a \equiv \langle f_{\rm tag}|B^0_{\rm tag}\rangle\,, \overline{a} \equiv \langle \overline{f}_{\rm tag}|\overline{B}^0_{\rm tag}\rangle$ and assume $\langle f_{\rm tag}|\overline{B}^0_{\rm tag}\rangle = \langle \overline{f}_{\rm tag}|B^0_{\rm tag}\rangle = 0$ so that the tag-side final state tells the $B_{\rm tag}$ flavor. Then we obtain

$$|\langle f_{CP} f_{\rm tag}|B_{CP}B_{\rm tag}\rangle|^2 = \frac{1}{4}e^{-\Gamma t_+}\left|\frac{p}{q}\mathcal{A}a\right|^2 \times \left[(1+|\lambda_f|^2) + (|\lambda_f|^2-1)\cos\Delta m_d\Delta t + 2{\rm Im}(\lambda_f)\sin\Delta m_d\Delta t\right] \tag{1.45}$$

$$\left|\langle f_{CP} \overline{f}_{\rm tag}|B_{CP}B_{\rm tag}\rangle\right|^2 = \frac{1}{4}e^{-\Gamma t_+}\left|\mathcal{A}\overline{a}\right|^2 \times \left[(1+|\lambda_f|^2) + (|\lambda_f|^2-1)\cos\Delta m_d\Delta t + 2{\rm Im}(\lambda_f)\sin\Delta m_d\Delta t\right]. \tag{1.46}$$

Using Eq. (1.40), $|q/p| = 1$, and $|a| = |\overline{a}|$, and integrating over $t_{CP}, t_{\rm tag} \geq 0$, we obtain the decay probability for $B^0_{\rm tag}$ and $\overline{B}^0_{\rm tag}$ events as

$$\Gamma(f_{\rm tag};\Delta t) \propto \frac{1}{4\tau_{B^0}}e^{-\frac{|\Delta t|}{\tau_{B^0}}}\left[1 + S\sin(\Delta m_d\Delta t) + A\cos(\Delta m_d\Delta t)\right] \tag{1.47}$$

$$\Gamma(\overline{f}_{\rm tag};\Delta t) \propto \frac{1}{4\tau_{B^0}}e^{-\frac{|\Delta t|}{\tau_{B^0}}}\left[1 - S\sin(\Delta m_d\Delta t) - A\cos(\Delta m_d\Delta t)\right]. \tag{1.48}$$

In reality we use $b \to c\overline{u}d$ process such as $\overline{B}^0 \to D^+\pi^-$ as one of signatures to tag the $\overline{B}^0$ flavor. These final states are also reachable from B^0 via $\overline{b} \to \overline{u}c\overline{d}$ process though it is doubly CKM suppressed compared to $b \to c\overline{u}d$. It violates the assumption of $\langle f_{\rm tag}|\overline{B}^0_{\rm tag}\rangle = \langle \overline{f}_{\rm tag}|B^0_{\rm tag}\rangle = 0$ and give rise to additional interference terms that changes the ostensible CP asymmetries at permille level (tag-side interference) [8].

1.3 Measurement of $\sin 2\phi_1$

1.3.1 Measurement of $\sin 2\phi_1$ using $b \to c\bar{c}s$ and $b \to sq\overline{q}$ decays

Let us consider more concrete expressions of S and A for particular decays. Because the box diagrams involving top quarks in the loop dominate the $B^0 - \overline{B}^0$ mixing due to its heavy mass, we can approximate $M_{12} \propto (V_{tb}V^*_{td})^2$. $|\Gamma_{12}| \ll |M_{12}|$ also holds because Γ_{12} represents the transition via on-shell intermediate states and does not include top quarks. Then we approximate the phase factor due to $B^0 - \overline{B}^0$ mixing in Eq. (1.20) as

$$\frac{q}{p} \simeq \sqrt{\frac{M^*_{12}}{M_{12}}} \simeq \frac{V^*_{tb}V_{td}}{V_{tb}V^*_{td}} \simeq e^{-2i\phi_1} \tag{1.49}$$

When the decay $B^0 \to f$ proceeds via a single dominant diagram with a negligibly small weak phase, we obtain $\lambda_f = \eta_f e^{-2i\phi_1}$ and Eq. (1.40) leads to

$$S = -\eta_f \sin 2\phi_1,\ A = 0. \tag{1.50}$$

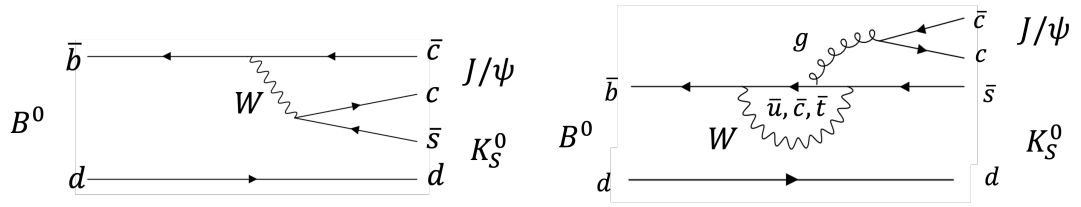


Figure 1.4: Tree (left) and penguin (right) diagrams contributing to $B^0 \to J/\psi K_S^0$

Considering a subdominant contribution,

$$\mathcal{A} = r_1 e^{i\theta_1} e^{i\delta_1} + r_2 e^{i\theta_2} e^{i\delta_2}, \tag{1.51}$$

we find the *CP* asymmetries deviate from Eq. (1.50):

$$S = -\eta_f \left(\sin 2\phi_1 + 2r \sin\theta_2 \cos\delta_{21} \cos 2\phi_1\right) + O(r^2), \tag{1.52}$$

$$A = 2r \sin\theta_2 \sin\delta_{21} + O(r^2), \tag{1.53}$$

where $r_{1,2}, \theta_{1,2}$, and $\delta_{1,2}$ are magnitude, weak phase, and strong phase of each contribution, $\delta_{21} \equiv \delta_2 - \delta_1$, $r \equiv r_2/r_1$ and we assume $r_2/r_1 \ll 1$ and $\theta_1 \sim 0$.

Being dominated by a tree-level diagram shown in Fig. 1.4, $b \to c\bar{c}s$ transitions such as $B^0 \to J/\psi K_S^0, J/\psi K_L^0$, and $\psi(2S)K_S^0$ are suitable for the precise measurement of ϕ_1. The subdominant contribution from penguin diagram shown in Fig. 1.4 is loop suppressed and doubly CKM suppressed to less than one percent.

On the other hand, penguin-dominated decays $b \to sq\bar{q}(q = u, d, s)$ such as $B^0 \to \phi K_S^0$, $\eta' K_S^0$, and $K_S^0 K_S^0 K_S^0$ shown in Fig. 1.5 work as a probe for new physics. The branching fraction of these decays are suppressed by the loop in the diagram to be $\mathcal{B}(B^0 \to \phi K^0) = (7.3 \pm 0.7) \times 10^{-5}$, $\mathcal{B}(B^0 \to \eta' K^0) = (6.6 \pm 0.4) \times 10^{-5}$, and $\mathcal{B}(B^0 \to K_S^0 K_S^0 K_S^0) = (6.0 \pm 0.5) \times 10^{-6}$ compared to tree-level decays such as $\mathcal{B}(B^0 \to J/\psi K^0) = (8.91 \pm 0.21) \times 10^{-4}$ [7]. Within the SM, they also exhibit the same *CP* asymmetries as $b \to c\bar{c}s$ up to small correction. However, potential interference with a NP contribution would change the *CP* asymmetries as in Eq. (1.52). Thanks to loop suppression in the SM penguin amplitude, they are sensitive to new physics. If we significantly observe either a deviation of S from the measurements in $c\bar{c}s$ decays, a deviation of S between $b \to sq\bar{q}$ decays, or a nonzero A, it is an evidence of new physics.

As can be seen in Eq. (1.52), the presence of new physics does not necessarily change the *CP* asymmetries but it depends on the phase difference from the SM contribution. Therefore it is important to collect as many decay modes as possible to increase the sensitivity.

Figure 1.6 summarizes the latest measurements of S and A in $b \to sq\bar{q}$ decays. While such a signature of new physics is not yet observed, most of the measurements are still statistically limited and leave large room for improvement.

Among the $b \to sq\bar{q}$ decays, the ones involving $q = u$ are polluted by $b \to u$ tree process. Theoretical prediction of *CP*-violating parameters for these decays accompany

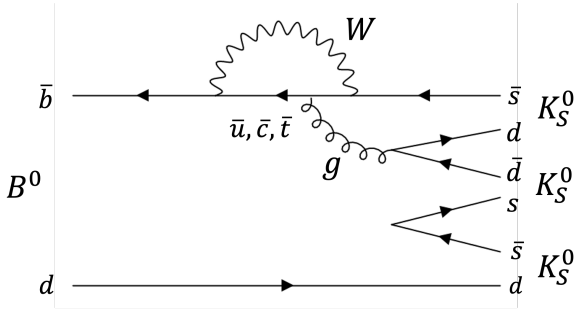


Figure 1.5: Penguin diagram contributing to $B^0 \to K_S^0 K_S^0 K_S^0$

additional uncertainties due to the pollution. On contrary, $b \to sq\overline{q}(q = d, s)$ decays like $B^0 \to K_S^0 K_S^0 K_S^0$ are theoretically clean without the tree pollution so the CP asymmetry is precisely predicted within the SM. For example the difference of $S \equiv \sin 2\phi_1^{eff}$ in $B^0 \to K^+K^-K_S^0, K_S^0 K_S^0 K_S^0$ from $\sin 2\phi_1$ is predicted to be [9]

$$\Delta \sin 2\phi_1^{eff}{}_{K^+K^-K_S^0} = 0.025^{+0.084\ +0.004}_{-0.017\ -0.015}, \tag{1.54}$$

$$\Delta \sin 2\phi_1^{eff}{}_{K_S^0 K_S^0 K_S^0} = 0.024^{+0.000\ +0.007}_{-0.000\ -0.018}. \tag{1.55}$$

Here, the uncertainty arising from that of $\phi_3 = (58.6 \pm 7)^\circ$ is factored out and shown separately as the second uncertainty. ϕ_3 is sensitive to the CKM elements in the penguin diagram. While $B^0 \to K^+K^-K_S^0$ involves a relatively large theoretical uncertainty due to the color-allowed tree process, the uncertainty on $B^0 \to K_S^0 K_S^0 K_S^0$ is negligibly small. The remaining small uncertainty will be reduced in future by improvement in ϕ_3 measurement.

Experimentally, the CP violation measurement of $B^0 \to K_S^0 K_S^0 K_S^0$ is difficult because only neutral particles emerge from the B^0 decay vertex. The Belle II experiment has a unique sensitivity to the decay because its collision energy is moderate enough to let the K_S^0's decay inside its vertex detector. Thanks to the large volume of upgraded vertex detector, the reconstruction efficiency of $B^0 \to K_S^0 K_S^0 K_S^0$ decays are expected to improve significantly.

1.3.2 SUSY phenomenology in $b \to sq\overline{q}$

We review a supersymmetric extension of SM (SUSY) conserving R parity as an example of new physics model. In SUSY, $b \to sq\overline{q}$ acquires an additional contribution from the diagram shown in Fig. 1.7, where the gluon-quark loop is replaced with the gluino-squark loop. The mass matrix for squark is not flavor diagonal in general, which allows flavor violating transitions between sbottom and sstrange in the diagram. In the framework of mass insertion approximation, where the off-diagonal element of the squark mass matrix is small compared to the average squark mass, the flavor-violating coupling is described by mass insertion parameters $(\delta^d_{AB})_{23}$, where $A, B = L, R$ indicates the helicity of corresponding quarks [11]. There are four types of mass insertion parameters:

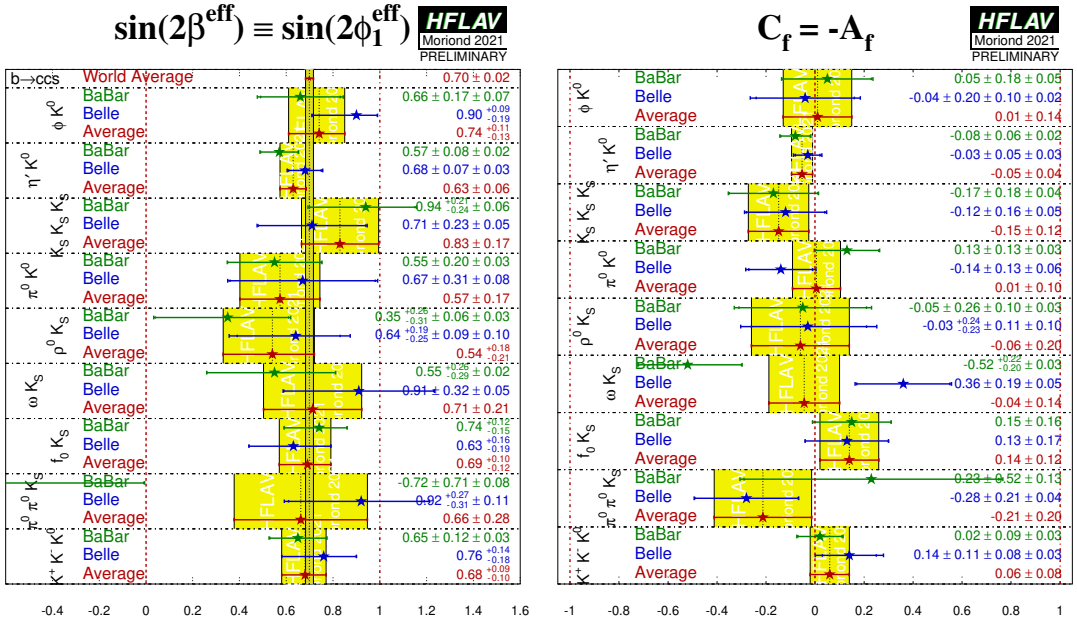


Figure 1.6: Measurements of CP asymmetries, S (left) and $C(=-A)$ (right) in $b \to sq\overline{q}$ decays [10]

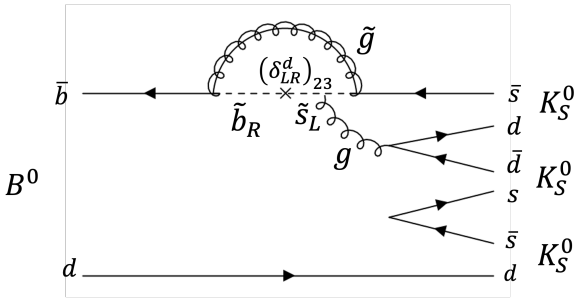


Figure 1.7: SUSY diagram contributing to $B^0 \to K^0_S K^0_S K^0_S$

$AB = LL, LR, RL, RR$. As the mass insertion parameters are complex in general, it introduces an additional source of CP violation.

Among the four types of mass insertion, $(\delta^d_{LR})_{23}$ and $(\delta^d_{RL})_{23}$ are particularly interesting for $b \to sq\overline{q}$. Because they require chirality flip on the gluino line, their contributions are enhanced by the gluino mass compared to RR and LL contributions where chirality is flipped on the quark line. Due to the enhancement, $b \to s$ transitions such as $b \to sq\overline{q}$, $b \to s\ell\ell$, and $b \to s\gamma$ are sensitive to $(\delta^d_{LR})_{23}$ and $(\delta^d_{RL})_{23}$ while $B_s - \overline{B}_s$ mixing cannot effectively constrain the parameters [12]. The enhancement also implies that the SUSY contribution does not decouple, or it is not suppressed even if the SUSY particles are heavy. As of today direct searches at the LHC have excluded gluino and squark masses up to around 2 TeV but indirect searches using $b \to s$ transition could shed light to them [13].

Though the experimental input are not up-to-date, the deviation in the *CP* asymmetries of $B^0 \to \phi K^0_S, \pi^0 K^0_S, \eta' K^0_S$, and ωK^0_S due to SUSY was predicted to be potentially as large as $O(0.1)$ [12]. Interplay with *CP* violation measurements of $B_s \to \phi\phi, B_s \to \phi\eta'$, and $B^0 \to K^0_S K^0_S$ can also provide probes to the SUSY [14].

1.3.3 Status of measurements of *CP* violation in $B^0 \to K^0_S K^0_S K^0_S$

Belle, BaBar, and Belle II have measured the *CP* violation in $B^0 \to K^0_S K^0_S K^0_S$ to obtain the following results:

$$\left.\begin{aligned} S &= -0.71 \pm 0.23 \pm 0.05 \\ A &= 0.12 \pm 0.16 \pm 0.05 \end{aligned}\right\} \text{Belle } (711\,\text{fb}^{-1}) \text{ [15]} \tag{1.56}$$

$$\left.\begin{aligned} S &= -0.94^{+0.24}_{-0.21} \pm 0.06 \\ A &= 0.17 \pm 0.18 \pm 0.04 \end{aligned}\right\} \text{BaBar } (426\,\text{fb}^{-1}) \text{ [16]} \tag{1.57}$$

$$\left.\begin{aligned} S &= -0.82 \pm 0.85 \pm 0.07 \\ A &= -0.21 \pm 0.28 \pm 0.06 \end{aligned}\right\} \text{Belle II } (63\,\text{fb}^{-1}) \text{ [17]}, \tag{1.58}$$

where the first and second uncertainties are statistical and systematic respectively and the size of the data set used in each analysis is specified in the parentheses. By integrating the data set of $50\,\text{ab}^{-1}$ at Belle II, the uncertainty on S is expected to be reduced to 0.04 by the previous work [17]. As the theoretical uncertainties of $\Delta \sin 2\phi_1^{\mathit{eff}}{}_{K^0_S K^0_S K^0_S}$ and the measurement uncertainty of $\sin 2\phi_1$ in $B^0 \to J/\psi K^0_S$ expected at $50\,\text{ab}^{-1}$ (0.006) are negligible, the potential deviation of $O(0.1)$ in $\Delta \sin 2\phi_1^{eff}$ will be detectable [18]. The sensitivity will become even higher by combining the measurements of S and A as they probe the additional *CP*-violating phase complementarily in terms of strong phase difference.

1.3.4 Target of this thesis

In this thesis, we present a measurement of *CP*-violating parameters in $B^0 \to K^0_S K^0_S K^0_S$ decays by using a data set corresponding to $189.3\,\text{fb}^{-1}$ collected at Belle II. We aim to explore the NP in $b \to s$ transition by the measurement and to refine the analysis procedure anticipating a high-precision measurement with the unprecedentedly large data set of $50\,\text{ab}^{-1}$ to be taken at Belle II.

Chapter 2

The Belle II experiment

Belle II is a B factory experiment operating at the SuperKEKB asymmetric-energy e^+e^- collider at the High Energy Accelerator Research Organization (KEK), in Tsukuba, Japan. The e^+e^- collision data are recorded by the Belle II detector mainly at $\Upsilon(4S)$ resonance, which subdominantly decays into $B\overline{B}$ pair. The main physics targets are precision measurements in flavor physics using B, D, and τ and searches for dark-sector particles.

The Belle II experiment aims to accumulate $50\,\mathrm{ab}^{-1}$ of integrated luminosity, which is 50 times as large as the data set recorded by the Belle experiment at the KEKB collider.

2.1 SuperKEKB

Figure 2.1 shows the main components of SuperKEKB. It consists of a linear accelerator, a positron damping ring, and the two main storage rings: the 7 GeV electron ring called the High Energy Ring (HER) and the 4 GeV positron ring called the Low Energy Ring (LER). The beams collide at an interaction point (IP) where the Belle II detector is located. The center-of-mass energy is set to the $\Upsilon(4S)$ resonance, 10.58 GeV, to selectively produce $B\overline{B}$ pairs. The center-of-mass system is boosted due to the asymmetric beam energy by the boost factor of $\beta\gamma = 0.287$. The boost factor has been reduced by a factor of 2/3 from KEKB to reduce the Touschek effect and maintain the beam lifetime for LER.

The SuperKEKB aims to achieve the instantaneous luminosity of $6 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and to accumulate $50\,\mathrm{ab}^{-1}$. The thirty-fold higher luminosity than KEKB comes from increasing the beam currents and adopting a so-called nano-beam scheme. The nano-beam scheme is characterized by a large beam crossing angle of 83 mrad and extremely small vertical beta functions of 0.3 mm at the IP. The large crossing angle allows to effectively increase the luminosity inversely proportional to the beta function by avoiding the deterioration of luminosity due to an hourglass effect.

The higher beam currents, smaller beam sizes, and higher luminosity all leads to severer beam background rate. The Belle II detector is required to record the collision

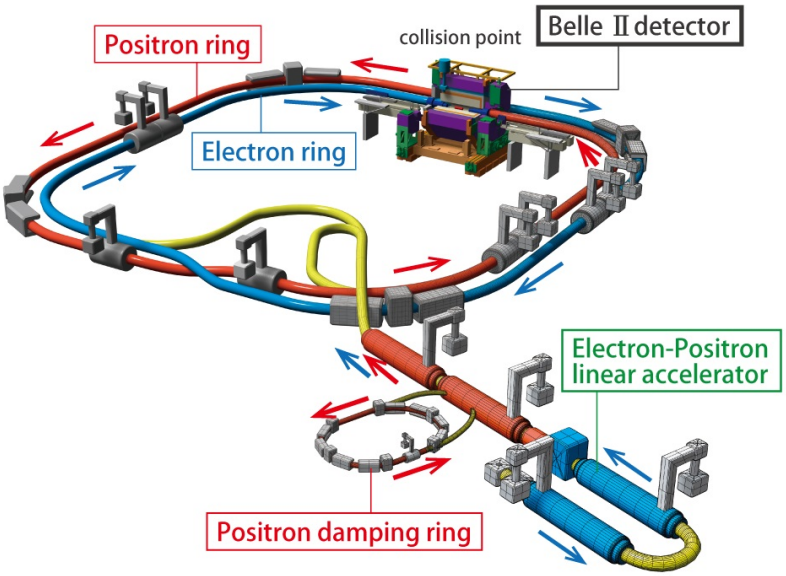


Figure 2.1: SuperKEKB

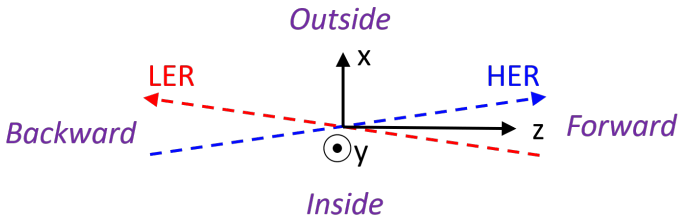


Figure 2.2: Coordinate system seen from the top

data at 30 times higher event rate and to tolerate the severe background. From the perspective of time-dependent *CP* violation analyses, the smaller boost factor than KEKB is a challenge since it deteriorates the decay time resolution. It is also notable that the IP size has been significantly reduced, which allows a new vertex reconstruction technique.

2.1.1 Coordinate system

We define (x, y, z) coordinates using a right-handed Cartesian system whose origin is at the nominal position of the IP. As shown in Fig. 2.2, the z axis is defined as the median line of the HER and LER, and the z direction points toward the HER direction. The y axis points toward the vertically upper direction and the x axis toward the outside of the main ring. We also define the distance from the z axis and the polar angle by $r = \sqrt{x^2 + y^2}$ and $\theta = \arccos\left(z/\sqrt{x^2 + y^2 + z^2}\right)$, respectively.

2.2 Belle II detector

The Belle II detector is a 4π general-purpose detector surrounding the IP. It consists of the following subdetectors:

- Vertex Detector (VXD)
- Central Drift Chamber (CDC)
- Time of Propagation counter (TOP)
- Aerogel Ring Imaging Cherenkov counter (ARICH)
- Electromagnetic Calorimeter (ECL)
- K_L^0 and μ detector (KLM)

Figure 2.3 shows the cross section of the detector seen from the top. To effectively catch the particles boosted forward, it has an asymmetric coverage in polar angle, $17° < \theta < 150°$. The subdetectors other than the KLM is in a superconducting solenoid magnet, which generates a 1.5 T magnetic field along the z direction.

The Belle II detector mainly records $e^+e^- \to q\overline{q}(q = u, d, s, c), B\overline{B}$, and $\tau^+\tau^-$ events and other signatures of interest for dark-sector searches. Charged particles in the final states, $e^\pm, \mu^\pm, \pi^\pm, K^\pm, p,$ and $\bar{p}$, are reconstructed as trajectories (tracks) in the tracking system, the VXD and CDC. A track in the magnet field draws a helix, which is parameterized by the following five *helix parameters*,

- d_0: the signed distance of the point of closest approach (POCA) to the z axis;
- ϕ_0: the angle defined by the x axis and the track transverse momentum at the POCA;
- ω: track curvature signed with the particle charge;
- z_0: the z coordinate of the POCA; and
- $\tan\lambda$: the tangent of the dip angle.

The definition of the helix parameters on x-y plane is illustrated in Fig. 2.4. Neutral particles such as photons and K_L^0's are reconstructed as the clusters of energy detected in the ECL and KLM that are not associated to any tracks.

2.2.1 VXD

Figure 2.5 shows the innermost subdetector, the VXD. It consists of six layers of cylindrical arrays of silicon sensors based on two different technologies. The inner two layers (Layer 1,2) are called the Pixel Detector (PXD) and uses DEPFET (Depleted P-channel Field Effect Transistor) silicon pixel sensors. The outer four layers (Layer 3–6) are called the Silicon Vertex Detector (SVD) using DSSDs (Double-sided Silicon Strip Detectors). The geometry of the sensors are summarized in Tab. 2.1.

The six VXD layers are aligned in a windmill structure shown in Fig. 2.6 at the radii of 14 mm, 22 mm, 39 mm, 80 mm, 104 mm, and 135 mm. There are currently only two

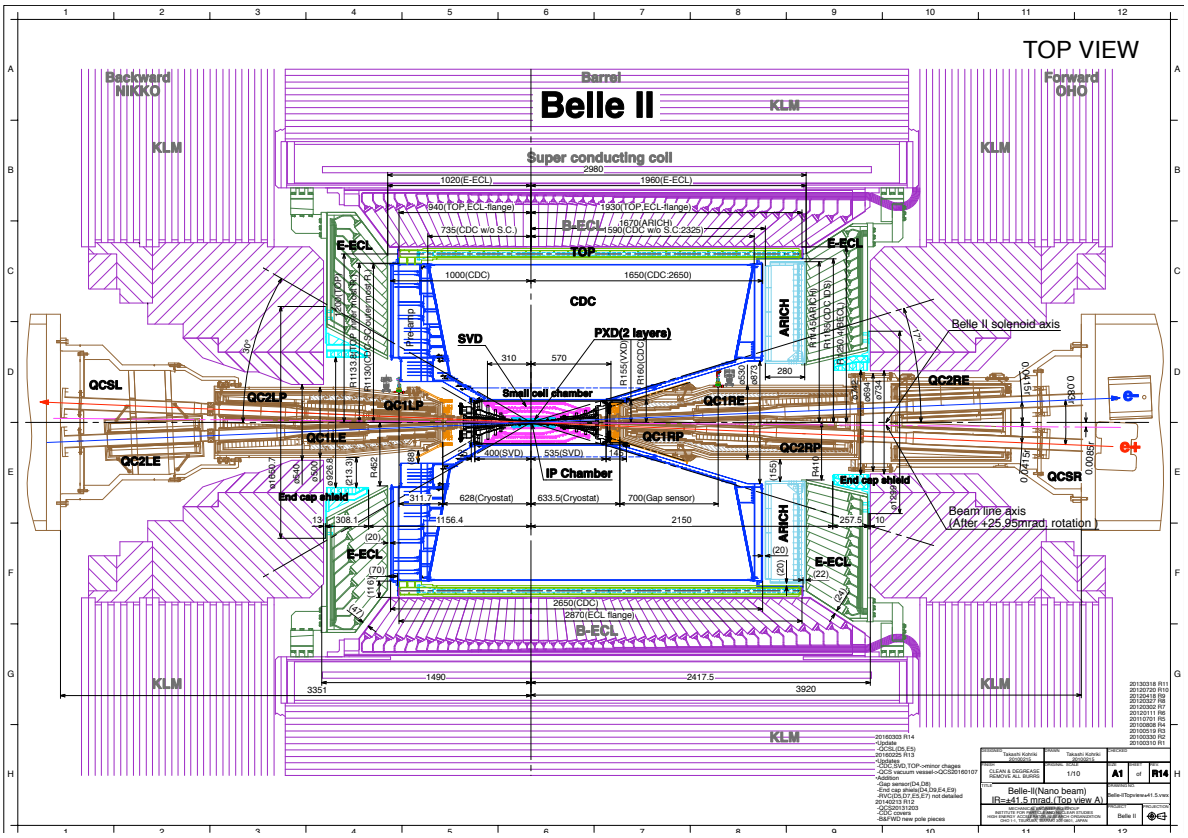


Figure 2.3: Cross-sectional top view of the Belle II detector

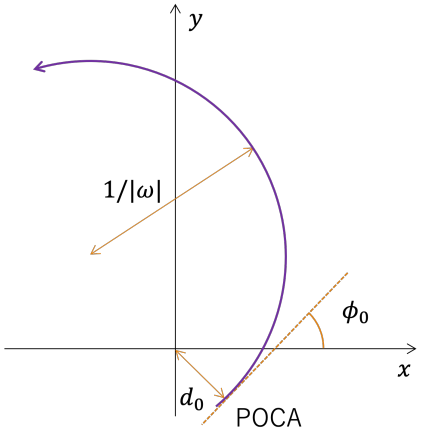


Figure 2.4: Definition of helix parameters, d_0, ϕ_0, and ω, on x-y plane. The arc is the trajectory of a charged particle.

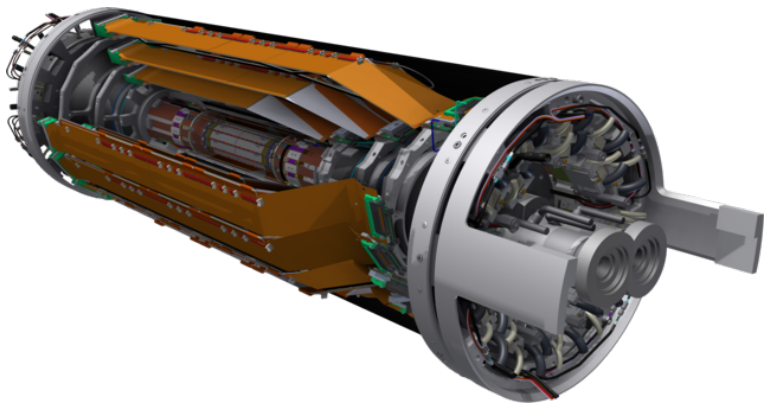


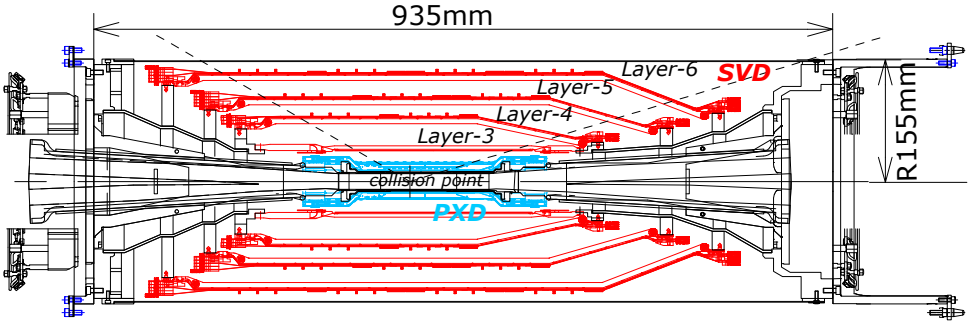
Figure 2.5: 3D drawing of the VXD

Layer-2 modules installed. In the long shutdown period starting in summer 2022, the current PXD is going to be replaced with a new PXD fully equipped with 12 Layer-2 modules.

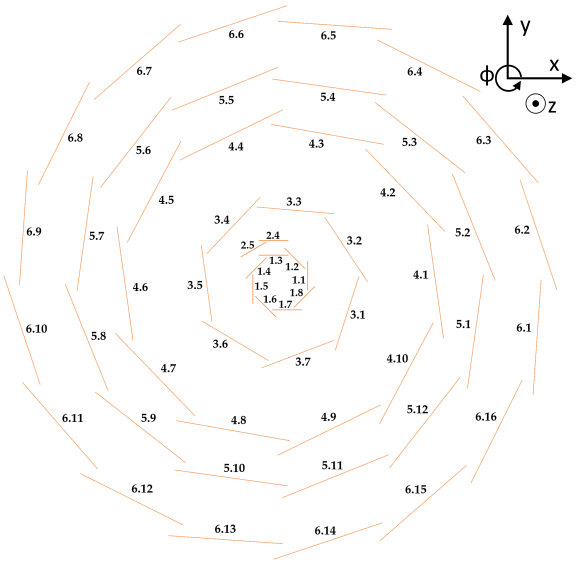
The VXD plays an crucial role in measuring the position of charged particles, and thus the measurement of B^0 decay vertex position necessary for time-dependent CP violation analyses. Layer 1 is closer to the IP compared to its predecessor, the Belle SVD2, which was 20 mm apart [19]. It is achieved by reducing the radius and thickness of the beam pipe. Being closer to the IP is advantageous for better vertex position resolution while hit rate increases. To cope with the high hit rate, the pixelated sensor is adopted for the inner layers. To reduce the data size below a bandwidth requirement, the PXD data are read out only within the region of interest which we determine by extrapolating tracks from the SVD onto the PXD. Using the PXD, the resolution of d_0 and z_0 is measured to be $13.64 \pm 0.08\,\mu$m and $14.92 \pm 0.07\,\mu$m for tracks which come from Bhabha scattering and fulfill $|\theta - \pi/2| < 0.5, p_T > 1\,\mathrm{GeV}/c$ and $p\beta(\sin\theta)^{3/2} > 2\,\mathrm{GeV}/c$ [20]. Here, $p_{(T)}$ is the transverse momentum of the particle, θ is the polar angle, and β is the velocity divided by the speed of light.

The SVD is essential for the precise measurement of K^0_S decay vertex position. We expect the improvement in the K^0_S reconstruction efficiency thanks to the larger coverage volume than the Belle SVD2. The z (r-ϕ) position resolution on the SVD sensors is measured to be around 25(16) μm on Layer 3 and 34(17) μm on Layer 4–6 for tracks of perpendicular incidence [21].

Other important roles of the SVD are the tracking of low-momentum tracks that do not reach the CDC and the identification of the particle species (PID) for them by the measurement of energy loss per unit passage length dE/dx.



(a) z-r view



(b) x-y view

Figure 2.6: Cross sections of the VXD on z-r and x-y planes

Table 2.1: Geometry of the VXD sensors [22]

Component	Layer	Radius (mm)	Pitch (z) (μm)	Pitch (r-ϕ) (μm)	Thickness (μm)
PXD	1	14	55–60	50	75
	2	22	70–85	50	75
SVD	3	39	160	50	320
	4	80	240	50–75	300–320
	5	104	240	50–75	300–320
	6	135	240	50–75	300–320

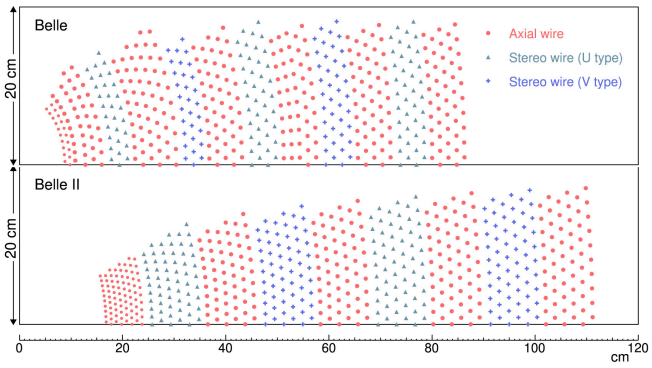


Figure 2.7: Layer configuration of the Belle and Belle II CDCs viewed on x-y plane [23]

2.2.2 CDC

The CDC is a drift chamber filled with the gas mixture of 50 % He and 50 % C_2H_6. It contains 14,336 sense wires arranged in 56 layers. The layers are grouped into nine parts called superlayers as shown in different colors in Fig. 2.7. There are five superlayers of axial wires along the z axis and four superlayers of stereo wires in skewed positions. Combination of the signals from axial and stereo wires provides 3-dimensional position information. Compared to the Belle, the outer radius of the CDC is extended from 880 mm to 1,130 mm, and a finer cell size is adopted in the innermost superlayer to cope with higher hit rate.

The CDC measures the trajectories and energy losses of charged particles for three purposes: measurement of the momenta of the tracks, trigger generation using the track information, and PID using dE/dx. The PID is important especially for low-momentum tracks that do not reach the TOP and ARICH.

Using cosmic rays, the transverse momentum resolution is measured to be $(0.127 \pm 0.001)p_T \oplus (0.321 \pm 0.003)\%$ (p_T in GeV/c) and impact parameter resolution to be around 120 μm for tracks with $p_T > 5$ GeV/c, as shown in Fig. 2.8.

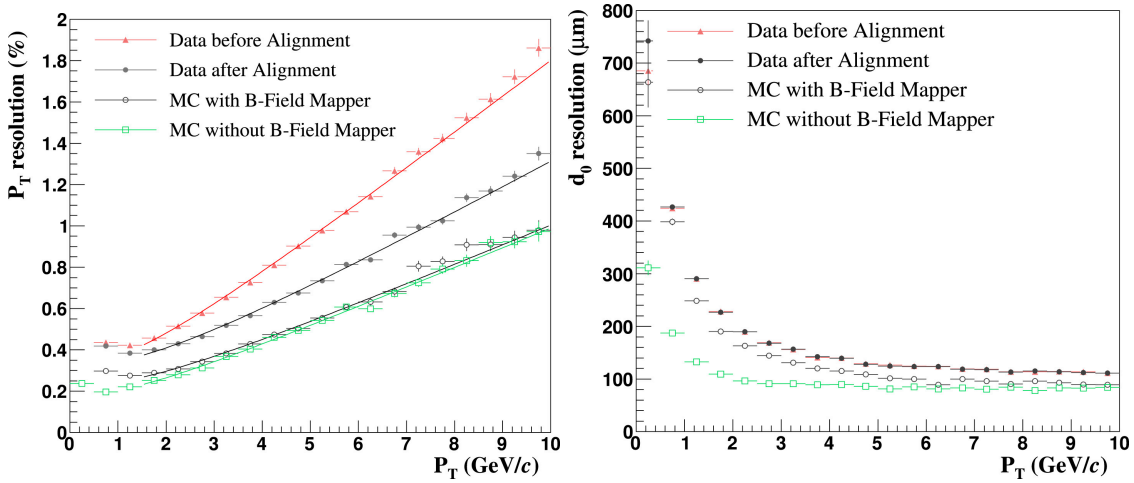


Figure 2.8: Transverse momentum and impact parameter resolution of the CDC [23]

2.2.3 TOP and ARICH

The TOP and ARICH are ring-imaging Cherenkov counters built to identify mainly $\pi^{\pm}$ and $K^{\pm}$.

The TOP covers the barrel part of the CDC with 16 modules. A TOP module consists of 2.6 m-long quartz radiator with a focusing mirror on one end and micro channel plate photomultipliers (MCP-PMTs) on the other end, as shown in Fig. 2.9. When a charged particle traverses the quartz radiator with a velocity exceeding the speed of light in the medium, it emits Cherenkov photons, which are reflected on the quartz surface and the mirror and detected by the MCP-PMTs. 20–40 photons are typically detected [24]. The MCP-PMTs have timing resolution better than 100 ps, enabling the TOP to also supply the event timing information to the trigger system. Provided with the information of track incident from the inner tracking detectors, the probability density function (PDF) of temporal and spatial hit pattern at the MCP-PMTs is calculated on different hypotheses of particle types. PDFs on pion and kaon hypotheses for a kaon track is shown as an example in Fig. 2.10. Then the likelihood for each hypothesis is calculated based on the PDF and measured hit pattern. Using only the TOP, the $K^{\pm}$ selection efficiency of 85% is achieved, with $\pi^{\pm}$ fake rate of 10% [24].

The ARICH covers the forward endcap region. It uses aerogel as the Cherenkov radiator and detects Cherenkov light by hybrid avalanche photon detectors (HAPDs) placed 160 mm apart from the aerogel tiles as shown in Fig. 2.11. The radiator consists of two layers of aerogel with different refractive indices to increase the number of radiated Cherenkov photons without blurring the Cherenkov ring. The ARICH provides a good K/π separation for tracks from 0.4 GeV/c to 4.0 GeV/c. The $K^{\pm}$ selection efficiency of $93.5 \pm 0.6\%$ with $\pi^{\pm}$ fake rate of $10.9 \pm 0.9\%$ is confirmed using $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decays [26].

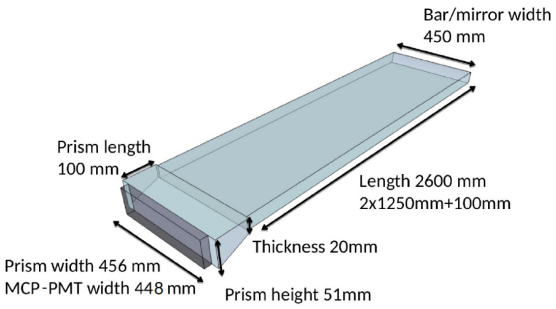


Figure 2.9: Schematic drawing of a TOP module [25]

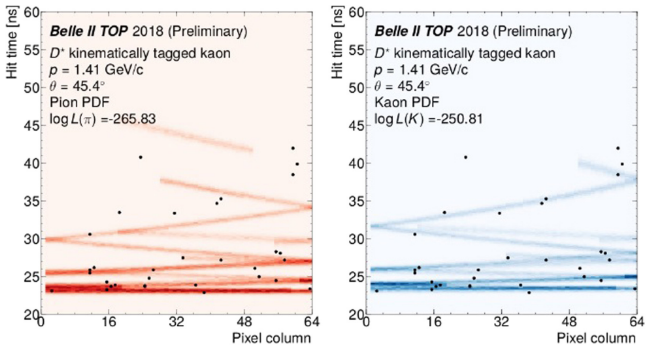


Figure 2.10: Hit pattern PDFs for a kaon track traversing the TOP based on pion and kaon hypotheses, overlaid with the actual hit pattern [25]

2.2.4 ECL

The ECL is a scintillation counter consisting of array of CsI(Tl) crystals. The scintillation light is read out by PIN-photodiodes attached to the crystals. The size of each crystal is about $6 \times 6\,\mathrm{cm}^2$ in cross section and 30 cm ($16.1X_0$) in length. The ECL is used for the following purposes:

- detect photons and measure their energy and angular coordinates,
- identify electrons from other hadrons using E/p, which is defined as deposited energy divided by track momentum,
- provide information for trigger generation, and
- measure luminosity from the rate of $e^+e^- \to e^+e^-, \gamma\gamma$ events.

Most of the detector components, CsI(Tl) crystals, preamplifiers, and support structures, are reused from Belle, where the energy resolution was evaluated to be 4% (1.6%) at 100 MeV (8 GeV) and angluar resolution to be 13 mrad (3 mrad) at low (high) energies [27]. Because of the relatively long decay time of scintillations in Cs(Tl), large pile-up noise is expected at the higher background level of Belle II. To deal with the pile-up noise, readout electronics has been renewed to enable wave form analysis.

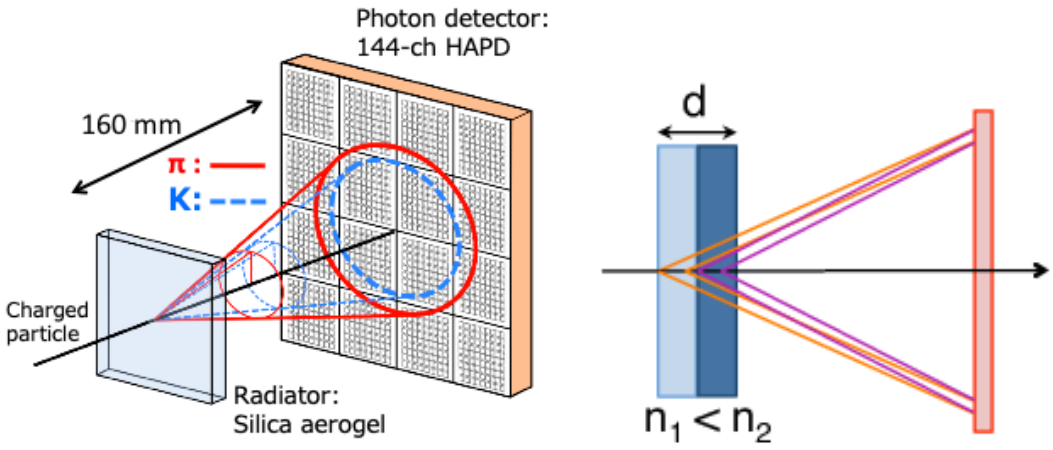


Figure 2.11: Working principle of ARICH counter [26]

2.2.5 KLM

The KLM is used to detect and identify K_L^0 mesons and muons. It consists of alternating layers of 4.7 cm-thick iron plates and detectors: resistive plate chamber (RPC) or plastic scintillator. It surrounds the solenoid and all other subdetectors. Muons and K_L^0's are detected through the interactions with the KLM material. The iron plates also serve as the return yoke for the magnetic field.

The RPC has been inherited from the Belle KLM. Due to its long dead time, it is used only in the barrel part where the background hit rate is moderate. In the endcap part and the two innermost layers in the barrel, scintillator strips with wavelength shifting fibers embedded and silicon photomultipliers are used.

Muons are identified as hits in the KLM associated with the CDC tracks, while K_L^0's are detected as isolated hits in the KLM.

2.2.6 Trigger and Data Acquisition System

The level-1 trigger system continuously processes signals from four subdetectors in parallel to detect physics events of interest. The occurrence of the physics events are mainly detected using the CDC and ECL, while the TOP and KLM assist to estimate precise event timings and to identify muons, respectively. The signals from the subdetectors are combined in Global Reconstruction Logic (GRL) and are examined in Global Decision Logic (GDL) to judge if the corresponding event is to be recorded. On detecting such an event, the system sends a trigger to the data acquisition (DAQ) system to start data readout from all subdetectors.

The expected event rates from main physics processes are listed in Tab. 2.2. Here, the instantaneous luminosity of $8 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ is assumed. The trigger system is required by the DAQ to reduce the level-1 trigger rate below 30 kHz. As the physics processes

Table 2.2: Estimated cross sections and trigger rates from various physics processes at the $\Upsilon(4S)$ resonance. Bhabha and $\gamma\gamma$ rates are prescaled by a factor of 1/100.

Process	σ [nb]	Rate [Hz]
$\Upsilon(4S)$	1.2	960
$q\bar{q}$	2.8	2,200
$\mu^+\mu^-$	0.8	640
$\tau^+\tau^-$	0.8	640
Bhabha	44	350
$\gamma\gamma$	2.4	19
Two-photon	12	10,000
Total	67	$\sim 15,000$

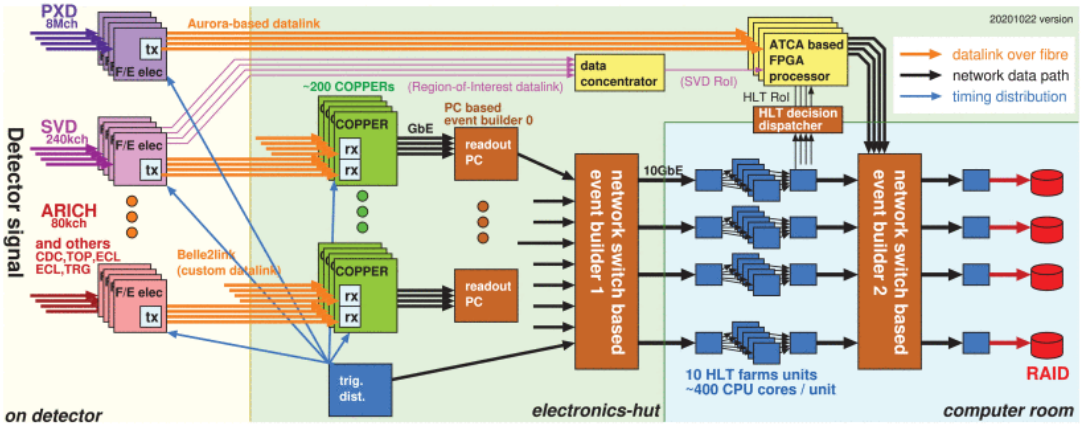


Figure 2.12: Schematic view of DAQ system [30]

amount to 15 kHz, fake trigger rate issued by background particles should be suppressed to 15 kHz by the GDL. The whole trigger decision process should be completed within 5 μs latency to meet a requirement from the ASICs used for the frontend readout of the SVD data.

The schematic view of the DAQ system is shown in Fig. 2.12. The signals from each subdetectors are digitized by dedicated frontend electronics and sent to common readout modules called COPPER except for the PXD. Then data from different subdetectors are combined in the network switches and sent to the computing nodes called high level triggers (HLTs). In the HLT, the data are reconstructed by the same software used for offline analysis called Belle II Analysis Software Framework (BASF2) [28][29]. Using the result of full event reconstruction, the event rate is suppressed to 1/5. The PXD data size is also reduced to 1/10 by selecting only the region of interest where charged tracks are extrapolated from the SVD. The HLT suppresses the maximum data rate to storage to 1.8 GB/s.

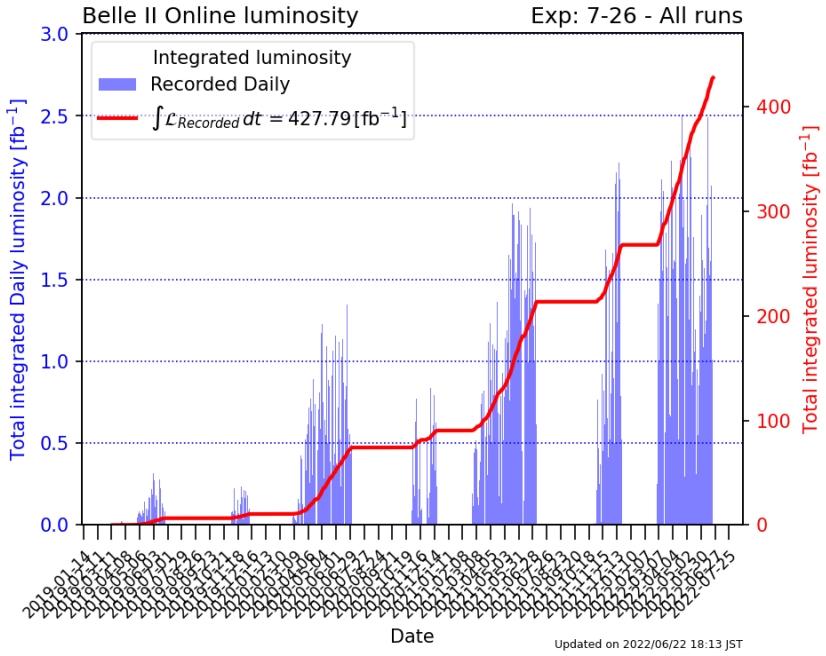


Figure 2.13: History of integrated luminosity in the Belle II experiment. The histogram shows the daily recorded luminosity and the line shows its integral.

2.3 Experiment status

The Belle II experiment has accumulated $428\,\mathrm{fb}^{-1}$ since the start of physics data taking in 2019 until summer 2022 as shown in Fig. 2.13. SuperKEKB has achieved the instantaneous luminosity of $4.7 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In the following analysis we use a data set corresponding to $190\,\mathrm{fb}^{-1}$ taken at the $\Upsilon(4S)$ resonance until summer 2021, which contains $(198.0 \pm 3.0) \times 10^6$ $B\overline{B}$ pairs.

Figure 2.14 shows the long-term projection of instantaneous and integrated luminosity. From the summer 2022, the experiment plans a long shutdown for about 1.5 years to replace the current PXD with a new one equipped with the full Layer 2. Another long shutdown is planned around 2027 for modification of the final beam focusing system around the interaction region, which is required to achieve the target luminosity.

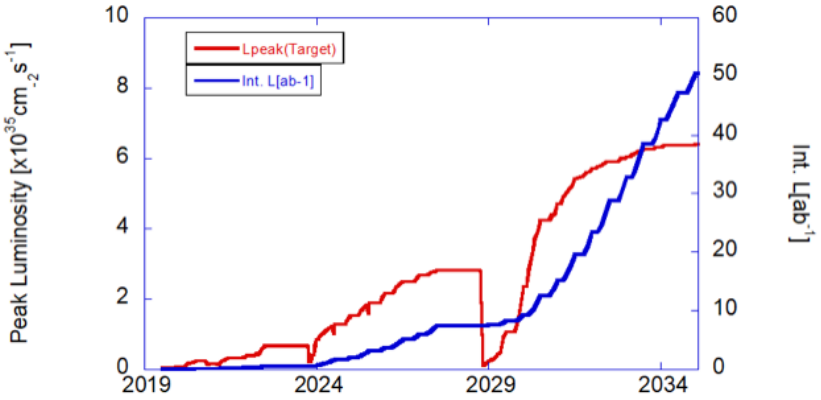


Figure 2.14: Projection of instantaneous and integrated luminosity in the Belle II experiment. The red (blue) lines show the instantaneous (integrated) luminosity as a function of data-taking year.

Chapter 3

Analysis Strategy

As discussed in Sec. 1.2.2, to measure the time-dependent CP asymmetry in $B^0 \to K^0_S K^0_S K^0_S$ using $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ events, we need to:

1. reconstruct one B (B_{CP}) decaying into the target final state $K^0_S K^0_S K^0_S$,
2. identify the flavor of the other B ($B_{\rm tag}$) from its decay products, and
3. measure the difference of proper decay time Δt between the pair of B's.

Figure 3.1 illustrates the above procedure. The ideal PDF of Δt and the $B_{\rm tag}$ flavor q ($= +1$ for $B^0_{\rm tag}$, -1 for $\overline{B}^0_{\rm tag}$) is given by

$$P^{\rm TD}_{\rm sig}(\Delta t, q) = \frac{1}{4\tau_{B^0}} e^{-\frac{|\Delta t|}{\tau_{B^0}}} \left[1 + q(S\sin(\Delta m_d \Delta t) + A\cos(\Delta m_d \Delta t))\right]. \tag{3.1}$$

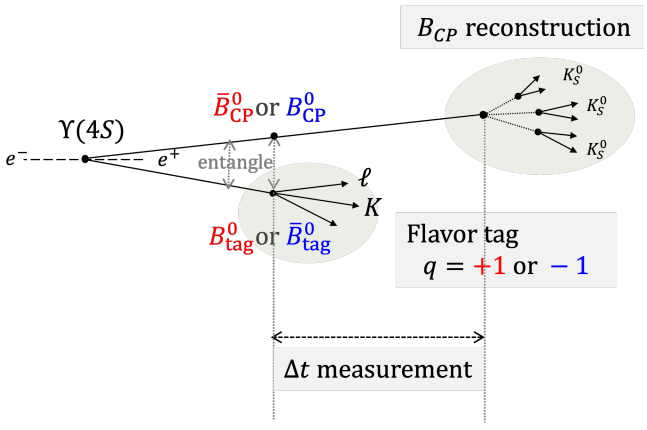


Figure 3.1: Overview of time-dependent CP violation measurement

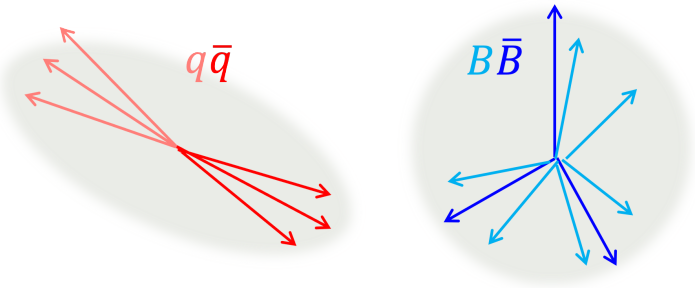


Figure 3.2: Illustration of event topology for $q\bar{q}$ (left) and $B\overline{B}$ events (right)

3.1 Reconstruction

We reconstruct the $B^0 \to K^0_S K^0_S K^0_S$ decays where all three K^0_S's decay into $\pi^+\pi^-$. K^0_S candidates are formed by combining two oppositely charged tracks. Having the lifetime of $c\tau = 2.7\,\text{cm}$, most of K^0_S's decay at significantly displaced position from the IP. That signature makes it easy to discriminate K^0_S from fake candidates formed by random combination of tracks originating from the IP. We select K^0_S candidates using a boosted-decision-tree (BDT) classifier and the $\pi^+\pi^-$ invariant mass [31].

Based on $\mathcal{B}(K^0_S \to \pi^+\pi^-) = 0.692$, we expect the $3(\pi^+\pi^-)$ final state to cover 33% of all $B^0 \to K^0_S K^0_S K^0_S$ decays. Although allowing one $K^0_S \to \pi^0\pi^0$ decay would potentially release 44% in addition to that, this final state suffers from worse purity due to fake π^0 candidates. For example, the latest BaBar analysis reconstructed the $2K^0_S(\pi^+\pi^-)K^0_S(\pi^0\pi^0)$ mode with 0.3 times smaller signal yield and 2.3 times more background than the $3K^0_S(\pi^+\pi^-)$ mode. [16] In addition, the Δt resolution of the $2K^0_S(\pi^+\pi^-)K^0_S(\pi^0\pi^0)$ mode is worse than the $3K^0_S(\pi^+\pi^-)$ mode. Therefore, we put the first priority on the $3(\pi^+\pi^-)$ final state and concentrate on it in this analysis.

B^0 candidates are reconstructed from the combination of three K^0_S candidates and selected using its invariant mass and beam-energy-constrained mass.

Background candidates arise from $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) events and $e^+e^- \to B\overline{B}$ events. Despite of similar cross sections, $q\bar{q}$ events dominate over $B\overline{B}$ because of relatively higher track momentum of them. As the light quark pair is largely boosted by the collision energy, the momenta of final-state particles in $q\bar{q}$ events tend to distribute in a jet-like shape in the center-of-mass frame as shown in Fig. 3.2. In contrary, $B\overline{B}$ event tend to have an isotropic momentum distribution. We discriminate the $q\bar{q}$ events based on a BDT classifier using event-topology variables (*continuum suppression*).

3.2 Flavor tag

We identify the B_{tag} flavor from the charge of its decay products. Leptons from semileptonic decays $b \to c\ell^-\overline{\nu}$ ($\ell = e, \mu$) and kaons from CKM-favored $b \to c \to s$ decays

Table 3.1: Flavor tagger performance parameters ε, w, μ, and Δw in % measured in different bins of r [1]. The first and second uncertainties are statistical and systematic, respectively.

r	ε	w	μ	Δw
0–0.1	$19.0 \pm 0.3 \pm 0.1$	$47.1 \pm 1.6 \pm 0.5$	$4.4 \pm 3.2 \pm 0.9$	$8.8 \pm 2.0 \pm 0.6$
0.1–0.25	$17.1 \pm 0.3 \pm 0.1$	$41.3 \pm 1.7 \pm 0.5$	$3.9 \pm 3.3 \pm 0.9$	$6.1 \pm 2.1 \pm 0.6$
0.25–0.5	$21.3 \pm 0.3 \pm 0.1$	$30.3 \pm 1.4 \pm 0.4$	$6.8 \pm 2.9 \pm 0.8$	$2.7 \pm 1.9 \pm 0.6$
0.5–0.625	$11.3 \pm 0.3 \pm 0.1$	$22.9 \pm 1.8 \pm 0.6$	$3.2 \pm 4.0 \pm 1.1$	$5.5 \pm 2.6 \pm 0.8$
0.625–0.75	$10.7 \pm 0.3 \pm 0.1$	$12.4 \pm 1.8 \pm 0.5$	$-0.5 \pm 4.1 \pm 1.1$	$0.7 \pm 2.9 \pm 0.7$
0.75–0.875	$8.2 \pm 0.2 \pm 0.1$	$9.4 \pm 1.9 \pm 0.5$	$10.8 \pm 4.3 \pm 1.1$	$7.7 \pm 3.2 \pm 0.9$
0.875–1	$12.4 \pm 0.2 \pm 0.1$	$2.3 \pm 1.3 \pm 0.4$	$-3.7 \pm 3.2 \pm 1.0$	$0.6 \pm 2.4 \pm 0.7$

are particularly discriminative because of the large branching fraction of these decays. Negatively charged leptons and kaons indicate $\overline{B}^0$ and vice versa.

We use a BDT-based software called *flavor tagger* in the BASF2 [1]. The flavor tagger estimates the flavor q and its credibility r ranging from 0 (ambiguous tag) to 1 (obvious tag) for each B_{tag} candidate. The performance of the flavor tagger is expressed by

$$\varepsilon \equiv \frac{\varepsilon_{B^0} + \varepsilon_{\overline{B}^0}}{2}, \tag{3.2}$$

$$w \equiv \frac{w_{B^0} + w_{\overline{B}^0}}{2}, \tag{3.3}$$

$$\mu \equiv \frac{\varepsilon_{B^0} - \varepsilon_{\overline{B}^0}}{\varepsilon_{B^0} + \varepsilon_{\overline{B}^0}} \text{ and,} \tag{3.4}$$

$$\Delta w \equiv w_{B^0} - w_{\overline{B}^0}, \tag{3.5}$$

where $\varepsilon_{B^0(\overline{B}^0)}$ and $w_{B^0(\overline{B}^0)}$ are the tagging efficiency of B^0 ($\overline{B}^0$) and the probability to misidentify B^0 ($\overline{B}^0$) as $\overline{B}^0$ (B^0) (wrong-tag fraction). μ and Δw represent the difference in tagging efficiency and wrong-tag fraction between B^0 and $\overline{B}^0$. These quantities are functions of r in general. Instead of using continuous r, we use a rough binning with the following bin edges and call the binned r as r bin:

$$(0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.875, 1). \tag{3.6}$$

The flavor tagger performance is calibrated by the measurement of the time-integrated mixing probability using $B^0 \to D^{(*)-}h^+$ ($h = \pi, \rho, a_1$) decays [1]. The calibration is based on a data set corresponding to $62.8\,\text{fb}^{-1}$, which was collected before summer 2020. Table 3.1 shows ε, w, μ, and Δw measured in each r bin. The statistical uncertainty for CP asymmetry is proportional to $1/\sqrt{\varepsilon_{\text{eff}}}$, where $\varepsilon_{\text{eff}} \equiv \sum_{rbin} \varepsilon(1-2w)^2$ is called effective tagging efficiency and is estimated to be $30.0 \pm 1.2(\text{stat}) \pm 0.4(\text{syst})\%$.

3.3 Δt measurement

The absolute value of the proper decay time difference Δt is typically so short around $\tau_{B^0} = 1.519\,\text{ps}$ that it is hard to directly measure. To overcome the difficulty, the center-of-mass system is boosted by the asymmetric beam energies with $\beta\gamma = 0.287$ at the Belle II experiment. As the $B\overline{B}$ pair is almost at rest with momentum of only $0.3\,\text{GeV}/c$ in the center-of-mass system, we approximate the boost factor of B mesons as that of $\Upsilon(4S)$ and measure Δt as the difference of decay vertex positions of $B\overline{B}$,

$$\Delta t = \frac{\ell_{CP} - \ell_{\text{tag}}}{\beta\gamma c}, \tag{3.7}$$

where ℓ_{CP} and ℓ_{tag} are the decay vertex position of B_{CP} and B_{tag}, which is projected to the boost direction. The approximation of the boost factor is corrected by a response function described in Sec. 5.3.1.

The decay vertex position is measured by kinematic fit using charged tracks in the final state. The vertex position resolution is typically tens of micrometers and thus can never be neglected compared to the average flight length of B^0, $130\,\mu\text{m}$. Moreover, the resolution varies from around $10\,\mu\text{m}$ to $100\,\mu\text{m}$ depending on the situation such as the number of K^0_S having VXD hits for B_{CP} vertex and the number of available tracks for B_{tag}. Therefore, precise understanding of the vertex resolution on the event-by-event basis is a key to the time-dependent CP violation measurement.

We classify the reconstructed events into two categories based on the quality of Δt measurement: *time-differential (TD) events* and *time-integrated (TI) events*. For TI events, where we cannot measure Δt with good quality, we do not use the Δt information but only the flavor information q. Therefore we integrate the PDF of Eq. (3.1) over Δt:

$$P^{\text{TI}}_{\text{sig}}(q) = \frac{1}{2}\left(1 + qA\frac{1}{1+x_d^2}\right), \tag{3.8}$$

where $x_d = \Delta m_d \tau_{B^0}$. As explicitly shown by Eq. (3.8), TI events are sensitive to only A through the dilution factor of $1/(1+x_d^2)$. The classification criteria are summarized in Tab. 4.3.

3.4 Control sample

We use $B^+ \to K^0_S K^0_S K^+$ decays as the control channel for the analysis of $B^0 \to K^0_S K^0_S K^0_S$. The branching fraction for the decay is $\mathcal{B}(B^+ \to K^0_S K^0_S K^+) = (1.05 \pm 0.04) \times 10^{-5}$ so the expected number of signals are twice larger than $B^0 \to K^0_S K^0_S K^0_S$, whose branching fraction is $\mathcal{B}(B^0 \to K^0_S K^0_S K^0_S) = (6.0 \pm 0.5) \times 10^{-6}$ [7].

Because of the similarity in decay kinematics and vertex resolution between $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$, we use the control sample to model the background shapes for B masses and the continuum suppression classifier, and to validate our knowledge of vertex resolution obtained from simulation.

Chapter 4

Event Reconstruction and Selection

4.1 Data and simulation samples

4.1.1 Data sample

We use the e^+e^- collision data taken at the $\Upsilon(4S)$ resonance at the Belle II experiment from 2019 to summer 2021, which corresponds to $189.3\,\text{fb}^{-1}$ and contains $(198.0\pm3.0)\times10^6$ $B\overline{B}$ pairs. As we are interested in hadronic $q\overline{q}$ and $B\overline{B}$ events with high track multiplicity, we select events with three or more "clean" tracks. The clean tracks are required to have high transverse momentum ($p_T > 0.2\,\text{GeV}/c$) and to originate from the IP ($|d_0| < 2\,\text{cm}$ and $|z_0| < 4\,\text{cm}$) to be discriminated from beam background tracks. We veto Bhabha scattering events that would otherwise increase the data size. Bhabha events are identified by fulfilling all of following conditions:

- two or more clean tracks that are back to back in the center-of-mass system (the opening angle larger than 2.88 radian),
- one or more tracks identified as electron ($p^* > 5\,\text{GeV}/c$ and $E/p > 0.8$),
- two tracks with $p^*/E_{\text{beam}} > 0.35$, and
- total energy of ECL clusters larger than 4 GeV,

where p^* and E_{beam} are the track momentum and beam energy in the center-of-mass system.

4.1.2 Simulation sample

We generate the Monte Carlo (MC) simulation samples of the following event types:

- $B^0 \to K^0_S K^0_S K^0_S$ signal MC, where one B from $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ is forced to decay into $B^0 \to K^0_S K^0_S K^0_S$,
- $B^+ \to K^0_S K^0_S K^+$ signal MC, where one B from $e^+e^- \to \Upsilon(4S) \to B^+B^-$ is forced to decay into $B^+ \to K^0_S K^0_S K^+$ or its charge conjugate, and

- *generic* MC, which includes one of $q\overline{q}(q = u, d, s, c)$, B^0 $\overline{B}^0$, and B^+B^- events.

The generic MC sample is equivalent to $700\,\mathrm{fb}^{-1}$. We use the EvtGen package to simulate the decay of hadrons and the KKMC and Pythia for $q\overline{q}$ [32][33][34]. The detector response to the generated particles are simulated with the Geant4 [35]. The digitization in the detectors is simulated with BASF2.

The digitized detector signals in data and simulation samples are both analyzed with BASF2.

4.2 K_S^0 reconstruction

We reconstruct K_S^0 candidates from combinations of oppositely charged tracks assumed to be pions. Unlike ordinary tracks originating from the IP, K_S^0 daughters are often produced far away from the IP because of the long K_S^0 lifetime ($c\tau = 2.7\,\mathrm{cm}$) and are not affected by multiple scattering through the beam pipe and detector materials if K_S^0 decays outside of them. Therefore, if a K_S^0 decay vertex is located outside the inner radius of the beam pipe ($r > 1\,\mathrm{cm}$), we re-fit its daughter tracks removing the extra material effect. We then determine the K_S^0 decay vertex by a kinematic fit using the refitted tracks and calculate the $\pi^+\pi^-$ invariant mass at the decay vertex.

We multiply scale factors to the covariance matrix of helix parameters of pions to correct an underestimation of the helix parameter uncertainties. The scale factors are determined using cosmic tracks (see Appendix F).

We select K_S^0 candidates based on the invariant mass $M_{\pi^+\pi^-}$ and a BDT classifier $\mathcal{O}_{K_S^0}$. The candidates should fulfill

$$(457.6\,\mathrm{MeV}/c^2 < M_{\pi^+\pi^-} < 537.6\,\mathrm{MeV}/c^2) \wedge (\mathcal{O}_{K_S^0} > 0.75), \tag{4.1}$$

so as to maximize the figure of merit for B^0 reconstruction (see Sec. 4.11).

4.2.1 K_S^0 BDT selection

We train a BDT to discriminate true K_S^0 using the following 22 input variables listed in Tab. 4.1, which include those of kinematics, PID, and number of hits in the VXD associated to the pion tracks. Particularly discriminating are the variables related to characteristic decay topology of K_S^0 such as the consistency of K_S^0 momentum and decay vertex directions (`cosVertexMomentum` and `ImpactXY`) and the displacement of K_S^0 decay vertex (`flightDistance` and `significanceOfDistance`). `decayAngle_D1,D2` should be essentially equivalent to `cosHelicityAngleMomentum` but in reality slightly differ from it and from each other.

We sample 100,000 true K_S^0 from signal MC and 100,000 fake K_S^0 from generic MC within a wide mass range $450\,\mathrm{MeV}/c^2 < M_{\pi^+\pi^-} < 550\,\mathrm{MeV}/c^2$ to form a training sample

Table 4.1: Input variables of K_S^0 BDT

name	explanation
`cosVertexMomentum`	cosine of angle between K_S^0 decay vertex and momentum vectors seen from IP
`flightDistance`	signed K_S^0 flight distance along the momentum direction
`significanceOfDistance`	significance of K_S^0 flight distance
`ImpactXY`	impact parameter of K_S^0 on the x-y plane
`cosHelicityAngleMomentum`	cosine of angle between the momentum difference of K_S^0 daughters in the K_S^0 rest frame and K_S^0 momentum in the lab frame
`decayAngle_D1,D2`	angle between $\pi^{+,-}$ momentum and reverted beam momentum in the K_S^0 rest frame
`px,py,pz`	K_S^0 momentum in x, y, and z direction
`p_D1,D2`	momentum magnitude of $\pi^{+,-}$
`muonID_NaNm1_D1,D2`	muon PID variable for $\pi^{+,-}$. 0,1, and -1 correspond to not-muon-like, muon-like, and no information.
`pionID_NaNm1_D1,D2`	pion PID variable for $\pi^{+,-}$. 0,1, and -1 correspond to not-pion-like, pion-like, and no information.
`daughtersDeltaZ`	z_0 difference between π^+ and π^-
`daughterAngle2body`	angle between π^+ and π^- momenta
`nPXDHits_D1,D2`	number of PXD hits associated to $\pi^{+,-}$ track
`nSVDHits_D1,D2`	number of SVD hits associated to $\pi^{+,-}$ track. r-ϕ and z hits are counted separately.

for the BDT. We make a testing sample in the same manner using the same amount of independent K_S^0 candidates. Figures 4.1 and 4.2 show the distribution of the variables for true and fake K_S^0 candidates in the testing sample. As shown in Fig. 4.3, the distributions of the BDT classifier for the training and testing samples agree with each other, which indicates no overtaining.

The list of input variables are inherited from a previous work with some modification [17]. In particular, we remove K_S^0 invariant mass variables and vertex position in absolute coordinates. By avoiding to use the mass, the BDT classifier does not bias the mass distribution as shown in Fig. 4.4. If we train a BDT with the K_S^0 mass, it biases the K_S^0 mass distribution as shown in Fig. 4.5. Removing the decay vertex coordinates does not result in visible change in performance for MC samples but the change may make the classifier more robust against the shift of IP observed in the real data.

Figure 4.6 show the signal efficiency and background rejection rate for the testing sample, which we calculate by scanning the threshold for the BDT selection and applying three different mass window denoted by the legend. The efficiency is defined as the probability for a true K_S^0 from the signal decay to survive the selection and the rejection rate is that for a fake K_S^0 candidate in generic MC. Our selection (Eq. (4.1)) provides the

efficiency of 95 % and rejection rate of 0.59 %.

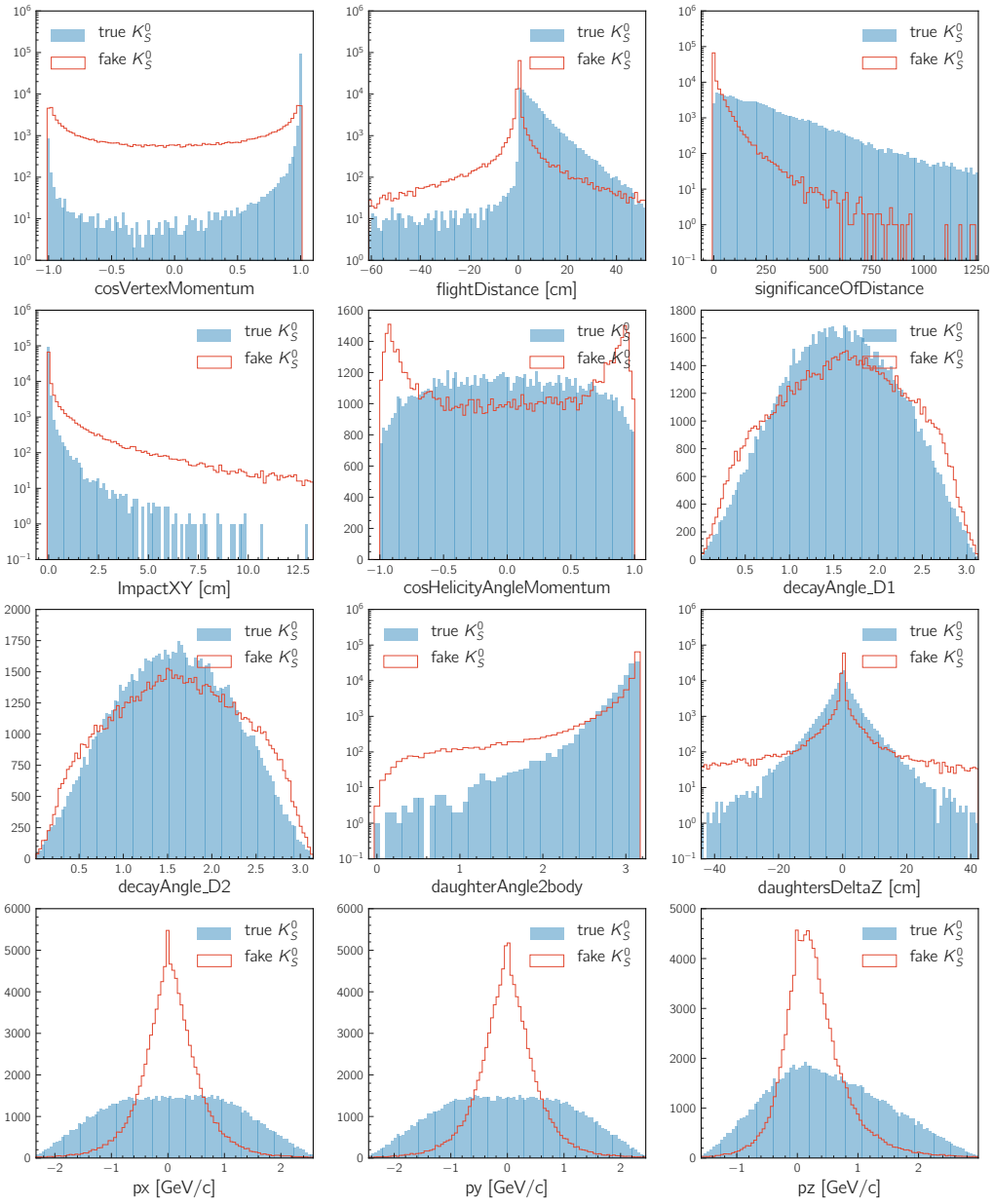


Figure 4.1: Distributions of K^0_S BDT input variables for true and fake K^0_S candidates in the testing sample (1)

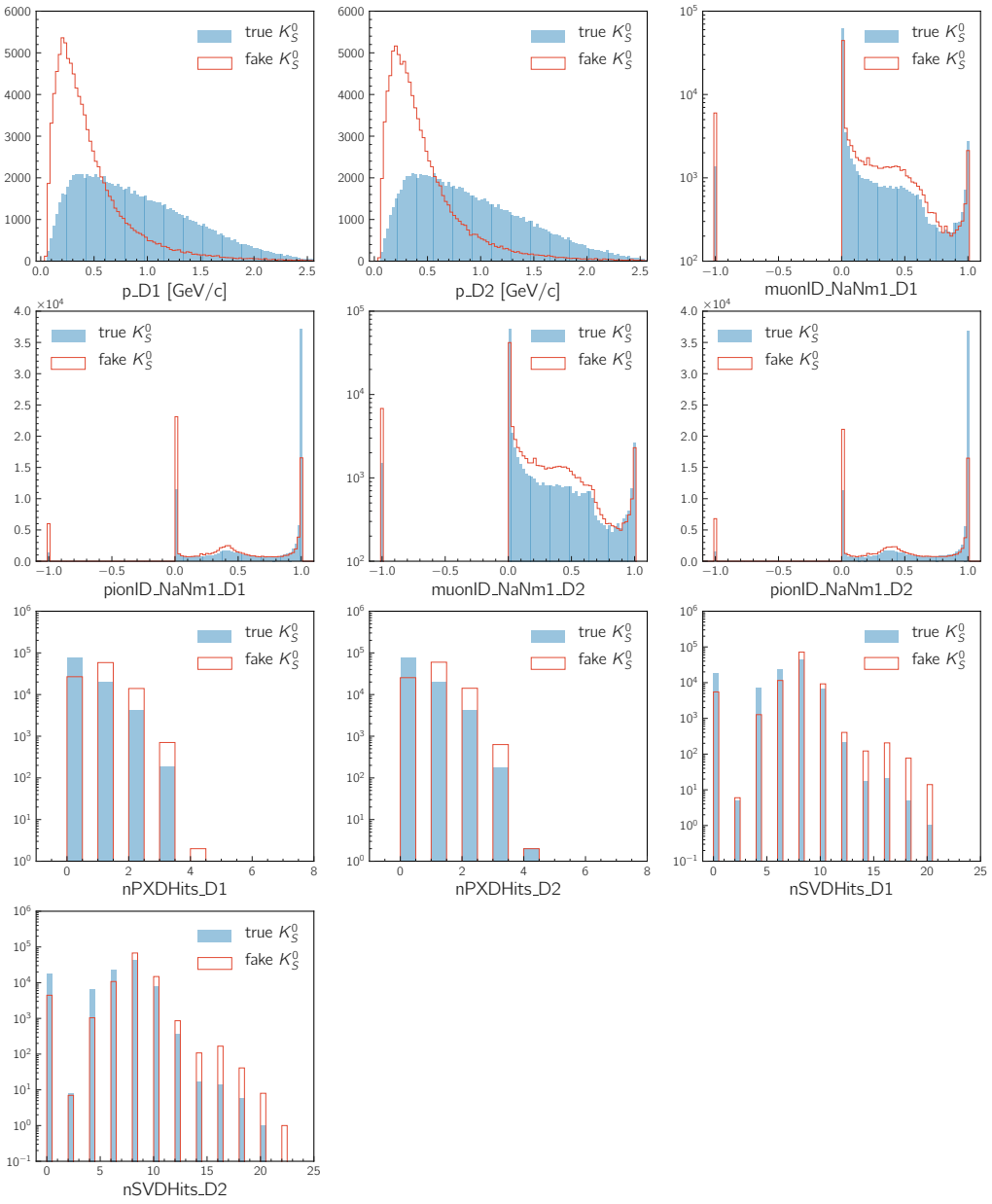


Figure 4.2: Distributions of K_S^0 BDT input variables for true and fake K_S^0 candidates in the testing sample (2)

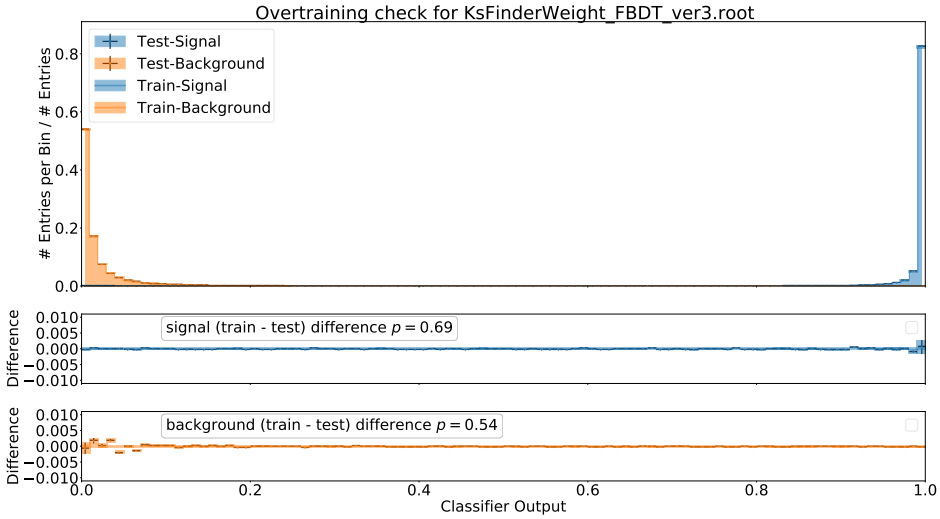


Figure 4.3: Distributions of K_S^0 BDT classifier $\mathcal{O}_{K_S^0}$. The distributions are shown separately for true ("signal") and fake ("background") K_S^0 candidates in the testing and training samples. The middle and bottom plots show the difference between the testing and training samples.

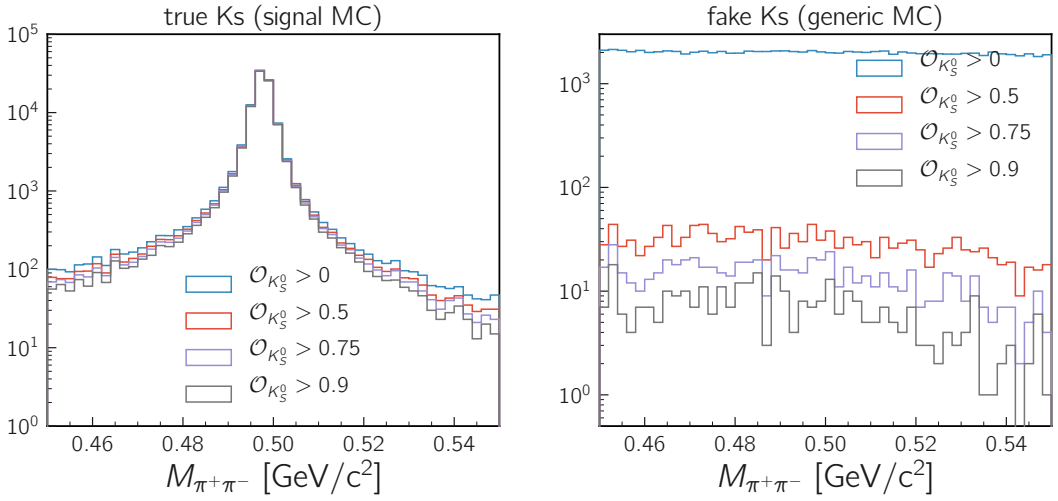


Figure 4.4: K_S^0 invariant mass distribution of true (left) and fake (right) K_S^0 candidates in the testing sample. Different selections about K_S^0 BDT classifier are applied.

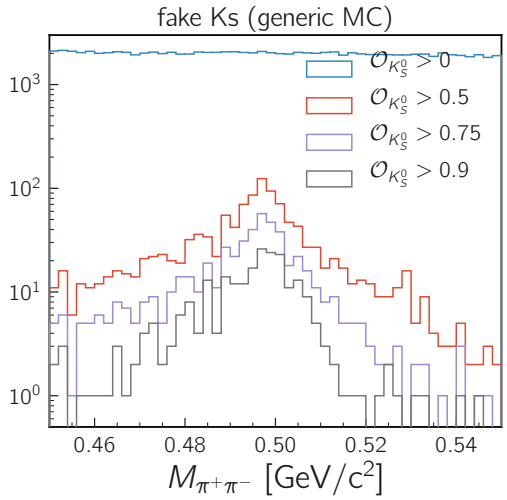


Figure 4.5: K_S^0 invariant mass distribution of fake K_S^0 candidates in the testing sample. Different selections about K_S^0 BDT classifier are applied. Here, the BDT is trained with the mass.

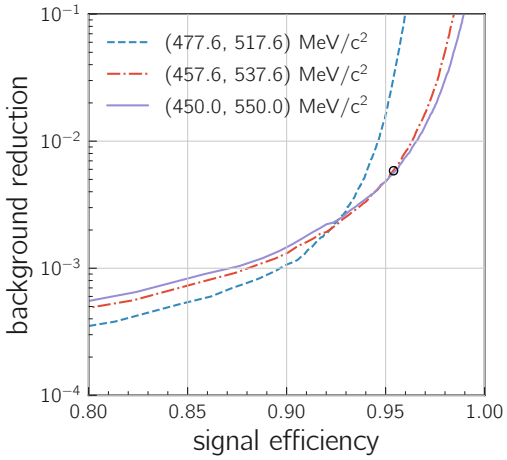


Figure 4.6: Signal efficiency and background reduction rate by K_S^0 selection using BDT and invariant mass. The curves are drawn by scanning the K_S^0 BDT threshold and applying the corresponding mass window in the legend. The black circle is our choice.

4.2.2 Off-diagonal K_S^0

We introduce a characteristics of K_S^0 candidates that indicate a probable contamination by a fake PXD hit and thus could lead to degradation of vertex resolution. We focus on which VXD layer has the innermost hit associated to a K_S^0 daughter track.

Figure 4.7 shows the fraction (in %) of true K_S^0 obtained from signal MC sample whose π^+ daughter have the innermost hit in the VXD layer in the row and π^- in the column. The true K_S^0 daughters should share the innermost layer, entering the diagonal elements of the matrix.

However there are considerable fraction of true K_S^0 in the off-diagonal elements, especially in (1, 3), (3, 1), (1, 4), and (4, 1). We call the K_S^0 candidates in these elements *off-diagonal* K_S^0. As shown in Fig. 4.8, this signature indicates a fake or missed Layer-1 hit. Because the fake Layer-1 hit lead to the deterioration of B^0 vertex resolution and incorrect estimation of the resolution, we classify the events including an off-diagonal K_S^0 as TI events and use them only for the time-integrated analysis.

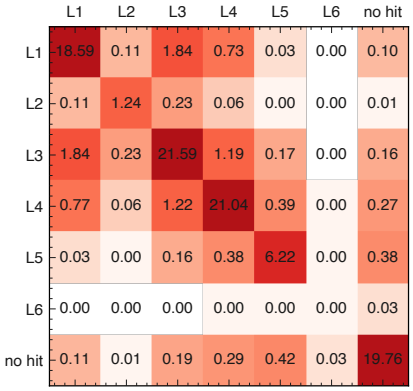


Figure 4.7: Matrix of innermost VXD layer of π^+ hit (column) and π^- (row) for true K_S^0 in signal MC. "L1" to "L6" denotes the layer numbers. The matrix element is the fraction of K_S^0 candidates whose π^+ daughter have the innermost hit in the VXD layer in the row and π^- in the column.

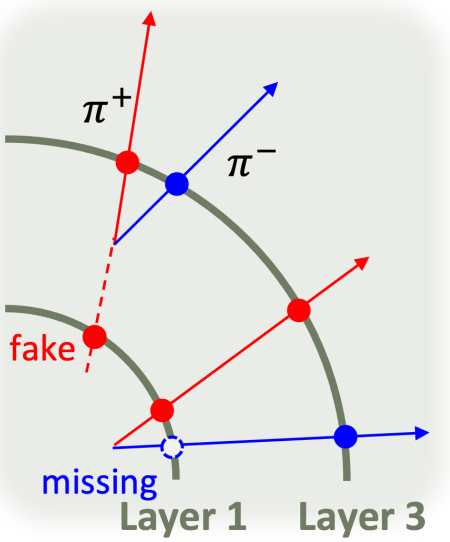


Figure 4.8: Illustration of off-diagonal K^0_S. The grey quarter circles represent the first and third VXD layers. The solid lines are the trajectories of pions produced from K^0_S. The filled (empty) circles are VXD hits assigned (not assigned) to the pions by the track reconstruction.

4.3 B^0 reconstruction

We reconstruct B_{CP} candidates from all possible combinations of three K^0_S candidates, and select them using the invariant mass M and the beam-energy-constrained mass $M_{\rm bc}$. $M_{\rm bc}$ is defined as

$$M_{\rm bc} = \sqrt{(E_{\rm beam}/2)^2 - \boldsymbol{p}_B^2}, \tag{4.2}$$

where $E_{\rm beam}$ and $\boldsymbol{p}_B$ are the beam energy and momentum of B_{CP} in the center-of-mass system. We select candidates within $(5.2\,{\rm GeV}/c^2 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2) \wedge (5.08\,{\rm GeV}/c^2 < M < 5.48\,{\rm GeV}/c^2)$. Figure 4.9 shows the distributions of $M_{\rm bc}$ and M for signal and background events in MC.

We reconstruct $B_{\rm tag}$ as the combination of the remaining tracks and clusters passing the following requirement to be discriminated from beam background. The tracks are required to be in the CDC acceptance $17^\circ < \theta < 150^\circ$, and to fulfill $|d_0| < 0.5\,{\rm cm}$ and $|z_0| < 3\,{\rm cm}$. We apply the correction to the helix parameter uncertainties of the tracks on the tag side as we do on the CP side. The clusters in forward ECL region are required to have an energy greater than 80 MeV; in barrel 30 MeV; and in backward 60 MeV.

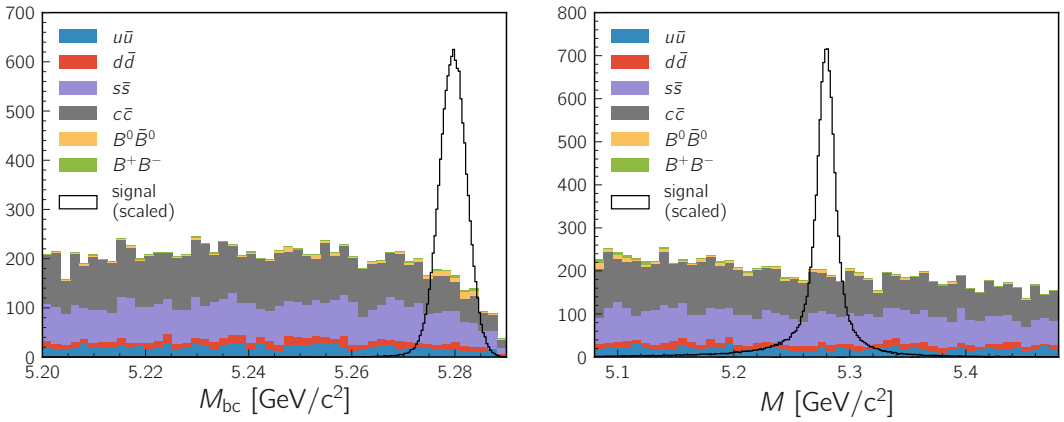


Figure 4.9: $M_{\rm bc}$ (left) and M (right) distributions of selected MC events. The stacked histograms show the background events from generic MC, which are separated according to its event type. The signal MC distributions are scaled arbitrarily.

4.4 Vertex reconstruction

We measure the decay vertex positions of B_{CP} and $B_{\rm tag}$ by performing kinematic fits with event-by-event constraints derived from the IP information. The profile of the IP is defined as three-dimensional Gaussian distribution and thus parametrized by a 3×3 covariance matrix and the center position. The IP profile is continuously measured using $e^+e^- \to \mu^+\mu^-$ events. The size and position of the IP are updated every around two hours and 30 minutes, respectively.

4.4.1 *CP*-side vertex reconstruction

For B_{CP}, we use the knowledge of the IP profile to define a virtual B_{CP} track to provide an additional constraint to determine the B_{CP} vertex position. The virtual track originates from the IP and points toward the reconstructed B_{CP} momentum as shown in Fig. 4.10.

We perform a least squares fitting to determine the vertex position. There are 53 parameters to be determined in the decay chain $\Upsilon(4S) \to B^0 \to K^0_S K^0_S K^0_S \to 3(\pi^+\pi^-)$:

- $\Upsilon(4S)$ decay vertex $\boldsymbol{x}_{\Upsilon(4S)}$ — (3),
- decay vertex position $\boldsymbol{x}$, energy E, momentum $\boldsymbol{p}$, and flight distance d of four intermediate states (B^0 and three K^0_S's) — (8×4), and
- momentum $\boldsymbol{p}$ of six pions in the final state — (3×6),

where the number of parameters are shown in the parentheses. On the other hand we have 61 constraints and measurements in total:

- IP position and covariance $\boldsymbol{x}_{IP}, \boldsymbol{V}_{IP}$ — (3),

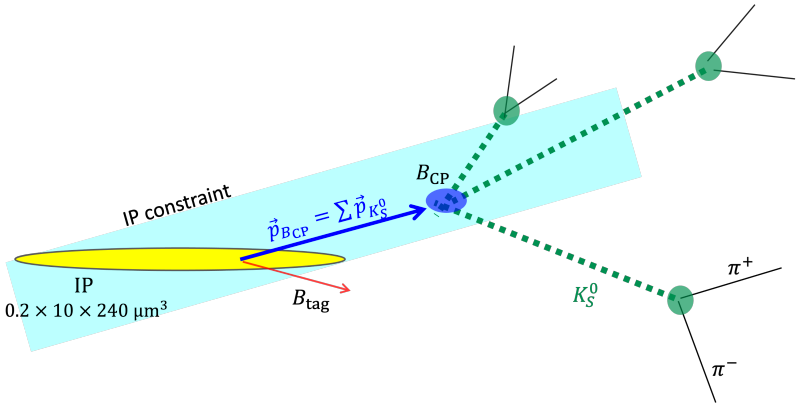


Figure 4.10: Concept of IP constraint

- geometric constraint for the flight of intermediate states f produced by i,

$$\boldsymbol{x}_i + d_f \frac{\boldsymbol{p}_f}{|\boldsymbol{p}_f|} = \boldsymbol{x}_f \tag{4.3}$$

  — (3×4),
- energy and momentum conservation in B^0 and K_S^0 decays — (4×4)
- measurement of helix parameters and their covariance matrix for six pions $\boldsymbol{h}, \boldsymbol{V}_h$ — (5×6),

where the number of constraints and measurements are shown in the parentheses. The number of degrees of freedom (ndf) in the fit is 8 $(= 61 - 53)$.

We define χ^2 as the compatibility of all kinematic variables on the *CP* side listed above. It is the sum of

- the measurement constraints for the helix parameters

$$\left(\boldsymbol{h}_\pi - \boldsymbol{h}(\boldsymbol{x}_{K_S^0}, \boldsymbol{p}_\pi)\right)^T \boldsymbol{V}_h^{-1} \left(\boldsymbol{h}_\pi - \boldsymbol{h}(\boldsymbol{x}_{K_S^0}, \boldsymbol{p}_\pi)\right), \tag{4.4}$$

  where $\boldsymbol{h}(\boldsymbol{x}_{K_S^0}, \boldsymbol{p}_\pi)$ is the helix parameters calculated from the fit parameters $\boldsymbol{x}_{K_S^0}$ and $\boldsymbol{p}_\pi$,
- the measurement constraint for the IP

$$(\boldsymbol{x}_{\Upsilon(4S)} - \boldsymbol{x}_{IP})^T \boldsymbol{V}_{IP}^{-1} (\boldsymbol{x}_{\Upsilon(4S)} - \boldsymbol{x}_{IP}), \tag{4.5}$$

  and
- the exact constraints $g(\boldsymbol{y}) = 0$ about flight and energy conservation implemented by the method of Lagrange multiplier as $2\lambda g(\boldsymbol{y})$, where λ is the Lagrange multiplier and $\boldsymbol{y}$ represents the parameters involved in the constraint. For example, the flight condition of Eq. (4.3) is implemented as $g(\boldsymbol{y} = \{\boldsymbol{x}_i, d_f, \boldsymbol{p}_f, \boldsymbol{x}_f\}) = \boldsymbol{x}_i + d_f \frac{\boldsymbol{p}_f}{|\boldsymbol{p}_f|} - \boldsymbol{x}_f = 0$

We minimize χ^2 with a Kalman filter to simultaneously determine all parameters [36].

We determine the B-meson decay vertex position and its uncertainty by minimizing χ^2, where the procedure is called a "vertex fit" For the time-dependent CP violation analysis, we are interested in the decay position projected onto the boost direction ℓ_{CP} and its uncertainty σ_ℓ^{CP}. We apply quite a loose selection on the uncertainty $\sigma_\ell^{CP} < 500\,\mu$m. Calculating the p-value from χ^2 and ndf, we find the distribution is not flat and events with bad fit quality concentrate at zero as shown in Fig 4.11. We require $p > 0.001 (\chi^2/\mathrm{ndf} < 3.27)$ for TD events.

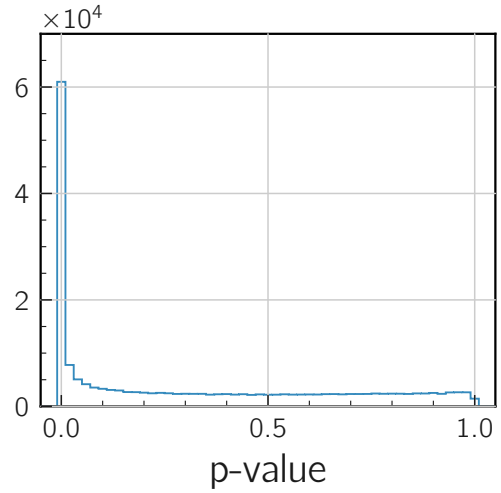
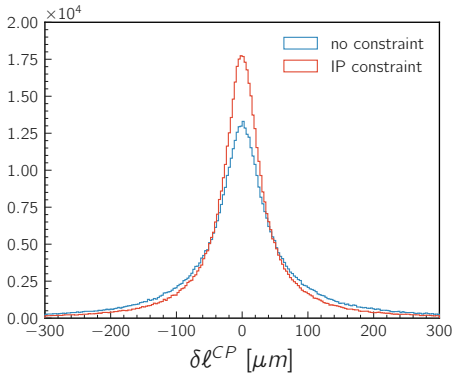


Figure 4.11: Distribution of p-value of B_{CP} vertex fit for signal MC events

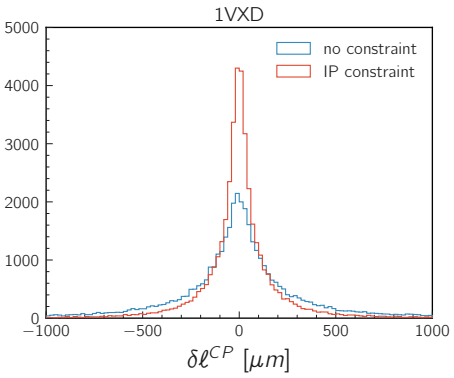
Figure 4.12 shows the residual of the fitted vertex position from the true value in the signal MC, separating the candidates according to the number of K_S^0 having VXD hits. The IP constraint improves the vertex position resolution especially when only one K_S^0 have VXD hits. The fraction of B^0 candidates with zero, one, two, and three K_S^0's having VXD hits are 0.4%, 8.0%, 37.7%, and 54.0% respectively.

4.4.2 Tag-side vertex reconstruction

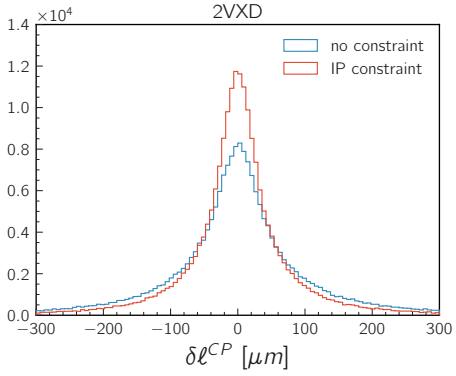
Unlike the CP side we do not know the decay chain or proper kinematic constraints among the tag-side particles. We perform a kinematic fit assuming all tracks are pions and originate from a single B_{tag} vertex. Here, we exclude tracks without a PXD hit and pairs of tracks whose invariant mass is within 10 MeV from the K_S^0 mass from the fit. The K_S^0 daughters should be removed because they do not originate from the B_{tag} vertex so they clearly violate the above assumption and bias the measurement. A similar inevitable bias arises from the daughters of charmed intermediate states, whose lifetimes are comparable to that of B^0 ($c\tau = 310\,\mu$m for D^+, $c\tau = 123\,\mu$m for D^0). We take into account the bias by a dedicated response function (see Sec. 5.3).



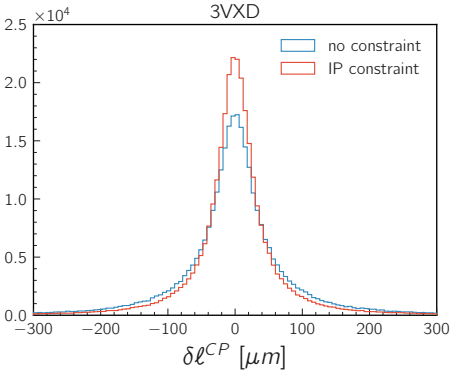
(a) signal events



(b) signal events with one K_S^0 having VXD hits, shown in a wider range than the others



(c) signal events with two K_S^0 having VXD hits



(d) signal events with three K_S^0 having VXD hits

Figure 4.12: Distributions of B_{CP} vertex position residual $\delta\ell^{CP}$ for signal MC events (a). The vertex fit is performed with or without the IP constraint for comparison. The distributions are separated according to the number of K_S^0 having associated VXD hits (b–d).

Similarly to the *CP* side we define a constraint as a virtual B_{tag} track called *Btube* using the IP profile. As illustrated in Fig. 4.13, we extrapolate back the B_{CP} track toward the IP to find the $\Upsilon(4S)$ decay vertex and expand it in the B_{tag} flight direction. The resulted long ellipsoid is regarded as the virtual track. The B_{tag} direction is calculated from the momenta of B_{CP} and beams.

Figure 4.14 shows the distributions of vertex position residual and the number of tracks used in the vertex fit for the signal MC events where the vertex fit is successful. The fit is performed with and without the Btube constraint. The Btube constraint does not only improve the resolution but also helps to save events where only one track is available for the vertex fit, which amounts to around 5 % of the reconstructed events. The asymmetry in the resolution arises from the charmed intermediate states.

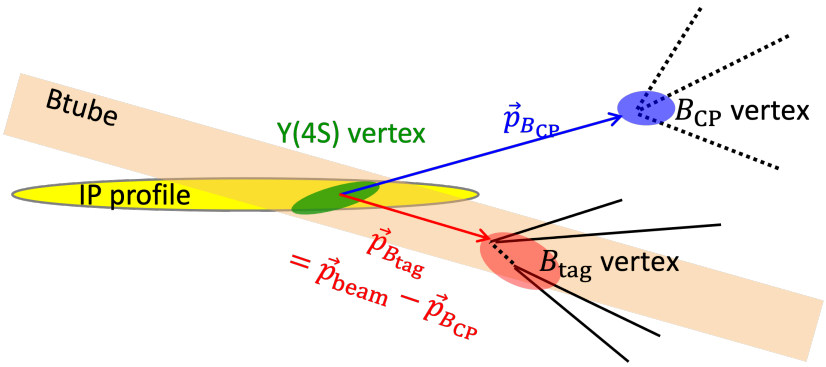


Figure 4.13: Concept of Btube constraint. The orange rectangle is the trimmed ellipsoid representing Btube.

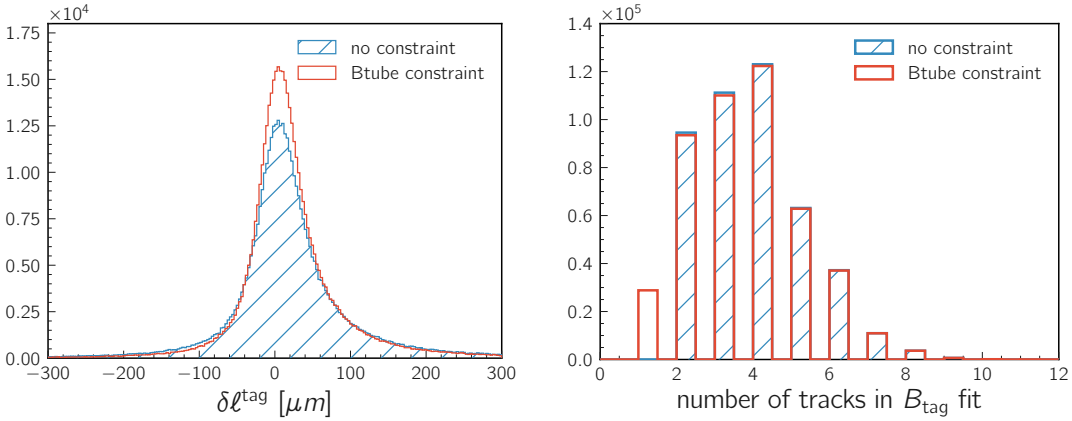


Figure 4.14: Distribution of B_{tag} vertex position residual (left) and the number of tracks used in vertex fit (right) for the signal MC events. The vertex fit is repeated with and without the Btube constraint. Only the candidates with successful vertex fit are shown.

We select candidates based on the tag-side vertex fit quality, requiring:

- success of the fit,
- $\sigma_\ell^{\text{tag}} < 500\,\mu\text{m}$, and
- $\left(\frac{\chi^2}{ndf}\right)^{\text{tag}} < 100$,

where σ_ℓ^{tag} and $\left(\frac{\chi^2}{ndf}\right)^{\text{tag}}$ are the vertex position uncertainty and reduced χ^2 estimated by the fit. The success of tag-side fit also requires that of the CP-side fit which is necessary for the construction of Btube.

Figure 4.15 shows the distribution of σ_ℓ^{CP} and σ_ℓ^{tag}. The first peak in the σ_ℓ^{CP} distribution is formed by events where at least one K_S^0 has PXD hits in its daughter tracks. In such a case, the vertex resolution is comparable between B_{CP} and B_{tag}. Otherwise, the B_{CP} resolution is worse in most of the cases.

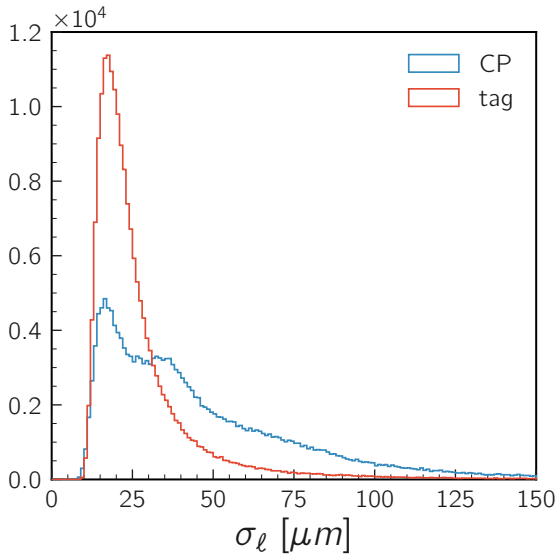


Figure 4.15: Distribution of σ_ℓ^{CP} and $\sigma_\ell^{\rm tag}$ for signal MC events

4.5 Flavor tag

As briefly mentioned in Sec. 3.2, we identify the $B_{\rm tag}$ flavor from its decay products using the flavor tagger. The flavor tagger adopts 13 categories of $B_{\rm tag}$ flavor signatures as shown in Fig. 4.16. Each category aims to catch the signature of a particle or correlation between particles which characterize the CKM-favored $b \to c \to s$ cascade decay or leptons from semileptonic decays of mesons. The flavor tagger software is implemented as BDTs in two steps as shown in Fig. 4.17. The first-step BDTs correspond to the 13 tagging categories. They use the particle identification information and kinematic variables of the tag-side tracks reconstructed on the hypothesis of the corresponding particle type to extract the $B_{\rm tag}$ flavor. Some categories also use the output from other categories. Then the second-step BDT combines the outputs from the first-step BDTs to get the final estimate of $B_{\rm tag}$ flavor q and its credibility r.

Figure 4.18 shows the distribution of the flavor tagger output qr in the signal MC. We confirm the tagging performance in the signal MC agrees with that of the independent measurement, from which we quote the values of w and Δw. The flavor tagging fails only if there is no track from $B_{\rm tag}$. Since our $B_{\rm tag}$ candidates are required to have at least one track for the success of vertex fit, the flavor tagger always estimate some results.

Categories	Targets for $\bar{B}^0$
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
Kinetic lepton	ℓ^-
Intermediate Kinetic Lepton	ℓ^+
Kaon	K^-
Kaon-Pion	K^-, π^+
Slow Pion	π^+
Maximum	ℓ^-, π^-
Fast-Slow-Correlated	ℓ^-, π^+
Fast Hadron	π^-, K^-
Lambda	Λ

$$\bar{B}^0 \to D^{*+}\, \ell^-\, \bar{\nu}_\ell$$
$$\hookrightarrow D^0\, \pi^+$$
$$\hookrightarrow X\, K^-$$

$$\bar{B}^0 \to D^+\, \pi^-(K^-)$$
$$\hookrightarrow \bar{K}^0\, \ell^+\, \nu_\ell$$

$$\bar{B}^0 \to \Lambda_c^+\, X^-$$
$$\hookrightarrow \Lambda\, \pi^+$$
$$\hookrightarrow p\, \pi^-$$

Figure 4.16: Flavor tagging categories and corresponding decay modes [1]

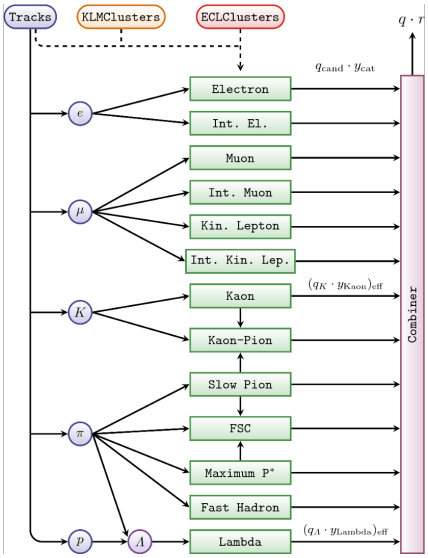


Figure 4.17: Concept of category-based flavor tagger [1]

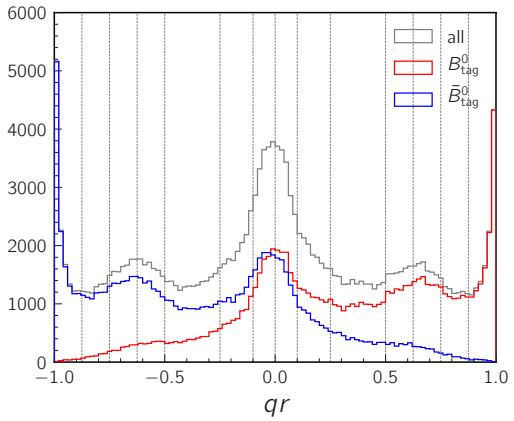


Figure 4.18: Distributions of flavor tagger output qr in selected signal MC events. The events are classified according to the true $B_{\rm tag}$ flavor. The dashed lines indicate the edges of r bin.

4.6 Continuum suppression

We use a BDT classifier $\mathcal{O}_{CS}$ to suppress the dominant $q\overline{q}$ background events using the different event topology between $q\overline{q}$ and $B\overline{B}$ events. The input discriminating variables are the following.

- 14 modified Fox-Wolfram moments called KSFW [27]
  The Fox-Wolfram moments are the moments of spherical harmonics for the momentum distribution in an event. For B physics, we use modified Fox-Wolfram moments that are computed using subsets of particles in the event, which are categorized according to their side (B_{CP} or $B_{\rm tag}$) and types (charged, neutral, or missing). Here, missing momentum is defined as an additional virtual particle. The moments are defined as

$$H^{so}_{xl} = \sum_i \sum_{jx} |\boldsymbol{p}_{jx}| P_l(\cos\theta_{i,jx}) \ (x = 0, 1, 2; l = 0, 2, 4), \tag{4.6}$$

$$H^{oo}_{l} = \sum_j \sum_k Q_j Q_k |\boldsymbol{p}_j||\boldsymbol{p}_k| P_l(\cos\theta_{j,k}) \ (l = 1, 3), \text{and} \tag{4.7}$$

$$H^{oo}_{l} = \sum_j \sum_k |\boldsymbol{p}_j||\boldsymbol{p}_k| P_l(\cos\theta_{j,k}) \ (l = 0, 2, 4), \tag{4.8}$$

  where i runs over the B_{CP} daughters, $j(x)$ and k runs over the $B_{\rm tag}$ (x=0:charged, 1:neutral, 2:missing) daughters, $P_l(\cos\theta_{i,j})$ is the lth order Legendre polynomial of the cosine of the angle between ith and jth particles, and $Q_i = \pm 1$ is the charge of ith particle.

- transverse energy and missing mass squared

$$E_T = \sum_i (p_T)_i, \tag{4.9}$$

$$M^2_{\rm miss} = \left(E_{\rm beam} - \sum_i E_i \right)^2 - \sum_i \boldsymbol{p}_i^2, \tag{4.10}$$

  where i runs over all particles in the events.
- magnitude of thrust of $B_{\rm tag}$ (`thrustOm`)

$$T = \frac{\sum_i |\boldsymbol{t} \cdot \boldsymbol{p}_i|}{\sum_i |\boldsymbol{p}_i|}, \tag{4.11}$$

  where $\boldsymbol{t}$ is a unit vector called thrust axis that maximizes T and i runs over the $B_{\rm tag}$ daughters.
- cosine of the angle between the thrust axes of B_{CP} and $B_{\rm tag}$ (`cosTBTO`)

We sample 100,000 events each from the signal and $q\overline{q}$ MC to form training and testing samples. We do not apply the K^0_S selection to the $q\overline{q}$ sample to keep a large number of events enough for training. Although this may slightly degrade the sensitivity of the BDT if the event topology changes through the K^0_S selection, it should not introduce any bias on the measurement. Figure 4.19 shows the distribution of the discriminating variables for the signal and $q\overline{q}$ events. We find the jet-like feature of the $q\overline{q}$ events: large thrust, aligned thrust axes of B candidates, and large higher-order moments.

The distribution of the output BDT classifier $\mathcal{O}_{CS}$ is shown in Fig. 4.20 for the training and testing samples. The consistency of the distributions indicates no significant overtraining. We apply a loose selection, $\mathcal{O}_{CS} > 0.1$, which rejects half of $q\overline{q}$ and retain 98% of signal. We then define a modified classifier,

$$\mathcal{O}'_{CS} = \log \frac{\mathcal{O}_{CS} - 0.1}{1 - \mathcal{O}_{CS}}, \tag{4.12}$$

and use it for signal extraction fit after selecting events within $-10 < \mathcal{O}'_{CS} < 10$ (see Sec. 5.2). Figure 4.21 shows the distribution of $\mathcal{O}'_{CS}$ for signal and background MC events with different event types.

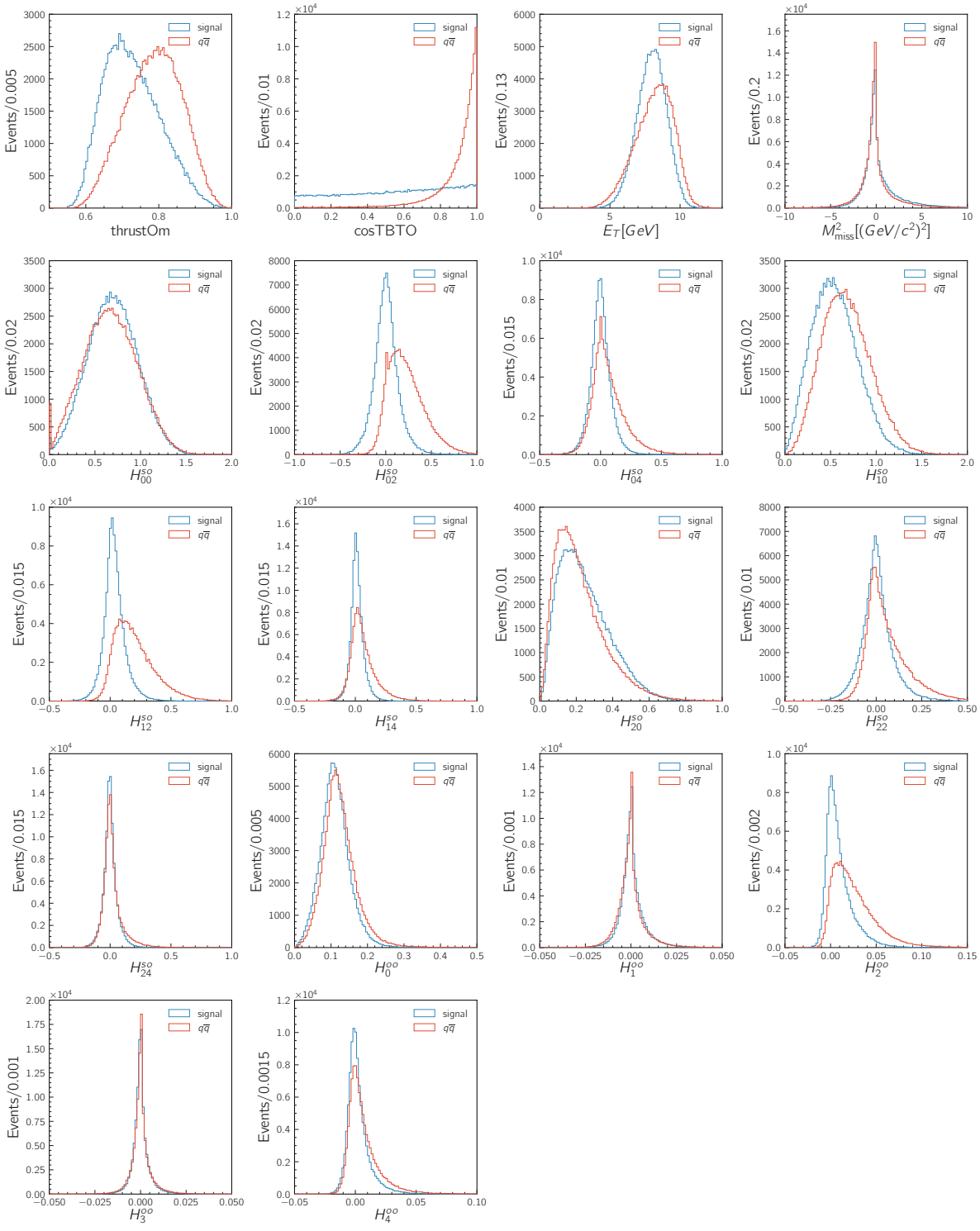


Figure 4.19: Distributions of input variables for continuum suppression BDT. The distributions are shown separately for signal and $q\bar{q}$ events in the testing sample.

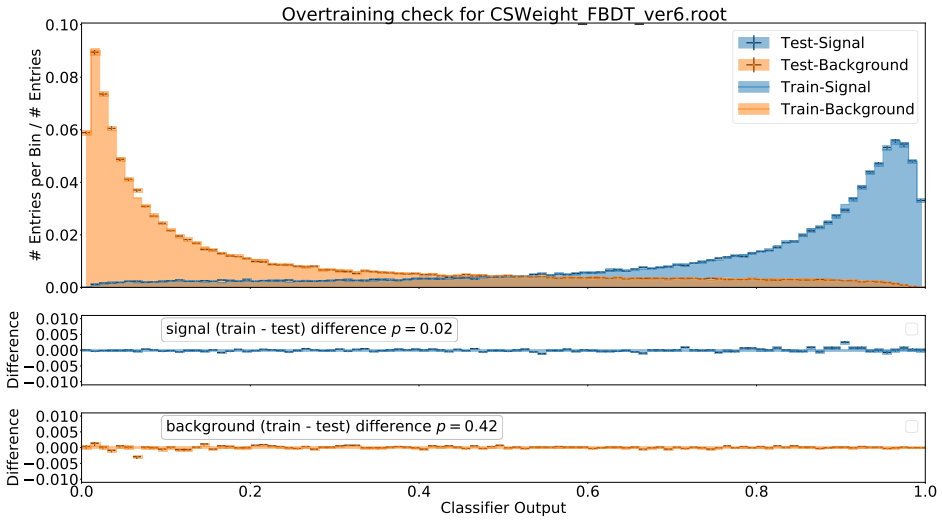


Figure 4.20: Distributions of continuum suppression BDT classifier $\mathcal{O}_{CS}$. The distributions are shown separately for signal and $q\bar{q}$ ("background") events in the testing and training samples. The middle and bottom plots show the difference between the testing and training samples.

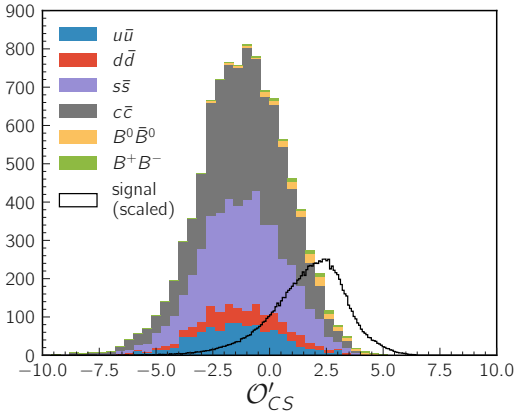


Figure 4.21: $\mathcal{O}'_{CS}$ distribution of selected MC events. The stacked histograms show the background events from generic MC, which are separated according to its event type. The signal MC distributions are scaled arbitrarily.

4.7 Best candidate selection

Figure 4.22 shows the distribution of candidate multiplicity for signal MC after we apply all selections described so far. The average candidate multiplicity is as small as 1.030. For events with multiple candidates we choose the candidate with the smallest B_{CP} vertex fit χ^2. After the best candidate selection, 98.3% of the reconstructed candidates in the signal MC are correctly reconstructed.

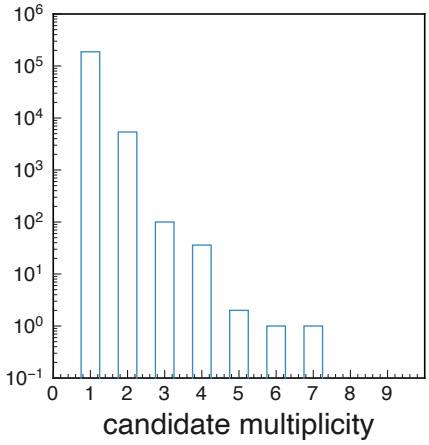


Figure 4.22: Distribution of candidate multiplicity for signal MC events that contain selected candidates

4.8 χ_{c0} veto

Apart from the non-resonant $B^0 \to K^0_S K^0_S K^0_S$ decay, there are contributions from quasi-two-body decays $B^0 \to X(\to K^0_S K^0_S) K^0_S$. Because our goal is to explore the NP effect in CP violation of $b \to s$ transition, we regard the f resonances such as $X = f_0(980), f_0(1710), f_0(2010)$ to be signal considering they originate from $b \to s$ and are CP even. On the other hand, $b \to c$ decays are problematic. The expected branching fraction or its upper limit for these decays are listed in Tab. 4.2. We expect the largest contribution from $\chi_{c0} K^0_S$, whose contribution is measured to be around 5% of $B^0 \to K^0_S K^0_S K^0_S$. Being CP even, it does not affect $\sin 2\phi_1$ measurement within the SM but it would dilute the possible NP effect in $b \to s$ transition. The dilution effect would be complicated due to the strong phase difference with respect to $B^0 \to K^0_S K^0_S K^0_S$.

We veto the $\chi_{c0} K^0_S$ decays by rejecting B^0 candidates if the invariant mass of any combination of two K^0_S fulfills

$$3378.8\,\mathrm{MeV}/c^2 < M_{K^0_S K^0_S} < 3447.1\,\mathrm{MeV}/c^2, \tag{4.13}$$

Table 4.2: Branching ratio or its upper limit of resonant background decays [7]. The numbers in the parentheses are the uncertainties of the last digit(s) or confidence level of the upper limits. The rightmost column shows the ratio of the branching fraction of $B^0 \to X(K^0_S K^0_S)K^0_S$ to that of $B^0 \to K^0_S K^0_S K^0_S$.

X	$\mathcal{B}(B^0 \to XK^0)$	$\mathcal{B}(X \to K^0_S K^0_S)$	$\mathcal{B}(B^0 \to XK^0)/2$ $\times\mathcal{B}(X \to K^0_S K^0_S)$	ratio to signal
D^0	$5.2(7) \times 10^{-5}$	$1.41(5) \times 10^{-4}$	3.7×10^{-9}	6.2×10^{-4}
χ_{c0}	$1.9(4) \times 10^{-4}$	$3.16(17) \times 10^{-3}$	3.0×10^{-7}	5.0×10^{-2}
χ_{c1}	$3.95(27) \times 10^{-4}$	$< 6 \times 10^{-5}$ (90%)	$< 1.2 \times 10^{-8}$	$< 2.0 \times 10^{-3}$
χ_{c2}	$< 1.5 \times 10^{-5}$ (90%)	$5.2(4) \times 10^{-4}$	$< 3.9 \times 10^{-9}$	$< 6.5 \times 10^{-4}$
η_c	$8.0(1.1) \times 10^{-4}$	$< 3.1 \times 10^{-4}$ (90%)	$< 1.2 \times 10^{-7}$	$< 2.0 \times 10^{-2}$
J/ψ	$8.91(21) \times 10^{-4}$	$< 1.4 \times 10^{-8}$ (95%)	$< 6.2 \times 10^{-12}$	1.0×10^{-6}
$\psi(2S)$	$5.8(5) \times 10^{-4}$	$< 4.6 \times 10^{-6}$ (95%)	$< 1.3 \times 10^{-9}$	2.2×10^{-4}

which covers 90% of a relativistic Breit-Wigner function,

$$f(M_{K^0_S K^0_S}) = \frac{k}{(M^2_{K^0_S K^0_S} - m^2_{\chi_{c0}})^2 + m^2_{\chi_{c0}}\Gamma^2}, \tag{4.14}$$

where

$$k = \frac{2\sqrt{2} m_{\chi_{c0}} \Gamma \gamma}{\pi \sqrt{m^2_{\chi_{c0}} + \gamma}}, \tag{4.15}$$

$$\gamma = \sqrt{m^2_{\chi_{c0}}(m^2_{\chi_{c0}} + \Gamma^2)}, \tag{4.16}$$

$$m_{\chi_{c0}} = 3414.71\,\mathrm{MeV}/c^2, \text{and} \tag{4.17}$$

$$\Gamma = 10.8\,\mathrm{MeV}/c^2 \text{ [7]}. \tag{4.18}$$

The χ_{c0} veto rejects around 7.5 % of the reconstructed signal MC events.

We neglect the decays via the other resonances because the rates are lower than the signal decay by three orders of magnitude or the decays are prohibited by P, CP conservation laws.

4.9 Background sources

Let us consider the sources of background events. As shown in Fig. 4.9, $q\overline{q}$ dominates the background and distributes gently in $M_{\rm bc}$ and M. The rest, $B\overline{B}$ background events, are classified into $B\overline{B}$ *combinatorial* and $B\overline{B}$ *peaking*. The $B\overline{B}$ combinatorial are the events where the B_{CP} candidate is reconstructed from the combination of decay products of both B mesons. In the $B\overline{B}$ peaking events the B_{CP} candidate is reconstructed solely from the decay products of either B meson. The $M_{\rm bc}$-M scatter plot of $B\overline{B}$ background events is shown in the left of Fig. 4.23. The $B\overline{B}$ combinatorial distribution is flat and

$q\overline{q}$-like while the $B\overline{B}$ peaking forms a peak in the low-M region. The $B\overline{B}$ peaking comes mostly from $B \to K^0_S K^0_S K^*$ decays. We will exclude the region contaminated by $B\overline{B}$ peaking, $(5.265\,\mathrm{GeV}/c^2 < M_{\mathrm{bc}} < 5.29\,\mathrm{GeV}/c^2) \wedge (5.08\,\mathrm{GeV}/c^2 < M < 5.1814\,\mathrm{GeV}/c^2)$, when estimating the signal yield to avoid biasing the result. The region is shown as red $\times$ in the right plot of Fig. 4.23.

As the remaining $B\overline{B}$ combinatorial are similar to the dominant $q\overline{q}$ component in terms of M_{bc} and M, we treat all background together. Though the $\mathcal{O}_{CS}$ distribution is different between $q\overline{q}$ and $B\overline{B}$, it does not affect the signal yield extraction. We determine the background $\mathcal{O}_{CS}$ PDF shape using the control sample data $B^+ \to K^0_S K^0_S K^+$, where we expect the same fraction of $B\overline{B}$ combinatorial to $q\overline{q}$ as $B^0 \to K^0_S K^0_S K^0_S$ based on the generic MC.

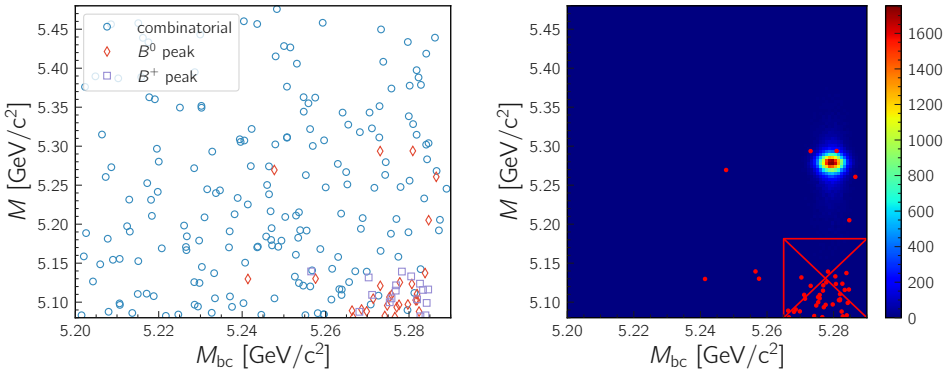


Figure 4.23: (left) M_{bc}-M scatter plot of $B\overline{B}$ background events in the $700\,\mathrm{fb}^{-1}$ generic MC that pass our selection. The events are classified to combinatorial background, peaking background from $B^0\overline{B}^0$, and peaking background from B^+B^-. (right) peaking $B\overline{B}$ background events in the generic MC (red dots) are overlaid to the signal MC events (histogram). The signal MC is arbitrarily scaled. The red $\times$ indicates the region excluded when we estimate the signal yield.

4.10 Summary of selection criteria

As explained in Chapter 3, we classify the reconstructed events into TD and TI after applying common selection criteria. Table 4.3 summarizes the baseline selection criteria required for both TD and TI events and the TD classification criteria additionally required for TD events.

Table 4.4 shows the cumulative and relative efficiencies of the baseline selection for signal MC, background MC, and data sample. The cumulative efficiency is normalized by the number of generated events for signal MC and by the number of events after selections on K^0_S candidate, M_{bc}, and M for background MC and data. The difference in the expected and observed events should come from the imperfection of $q\overline{q}$ simulation.

Table 4.3: Summary of baseline selection and TD classification criteria

purpose	criteria
baseline selection	$457.6 < M_{\pi^+\pi^-} < 537.6\,\mathrm{MeV}/c^2$
	$\mathcal{O}_{K^0_S} > 0.75$
	$5.2 < M_{\mathrm{bc}} < 5.29\,\mathrm{GeV}/c^2$
	$5.08 < M < 5.48\,\mathrm{GeV}/c^2$
	tag-side vertex fit succeeds
	$\sigma_\ell^{\mathrm{tag}} < 500\,\mu\mathrm{m}$
	$\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}} < 100$
	$\mathcal{O}_{CS} > 0.1$
	$-10 < \mathcal{O}'_{CS} < 10$
	highest CP-side vertex fit χ^2
	$M_{K^0_SK^0_S} < 3378.8\,\mathrm{MeV}/c^2$ or $3447.1\,\mathrm{MeV}/c^2 < M_{K^0_SK^0_S}$
TD-event classification	none of K^0_S from B_{CP} is off-diagonal
	at least one of K^0_S from B_{CP} has VXD hits
	$\sigma_\ell^{CP} < 500\,\mu\mathrm{m}$
	$\left(\frac{\chi^2}{ndf}\right)^{CP} < 3.27$
	$-30 < \Delta t < 30\,\mathrm{ps}$

Table 4.4: Relative (cumulative) efficiency (%) of baseline selections for $B^0 \to K^0_S K^0_S K^0_S$ decays. The efficiency is shown in separate columns for signal MC, background MC, and data sample. The cumulative efficiency is normalized to the number of all generated events for signal MC and to the number of events after the first selection for background MC and data. The last row shows the expected and observed number of events.

selection	signal MC	background MC	data
acceptance, K^0_S selection, M_{bc}, M	34.10 (34.10)	- (100.00)	- (100.00)
$\mathcal{O}_{CS} > 0.1$	98.14 (33.47)	50.52 (50.52)	52.61 (52.61)
$-10 < \mathcal{O}'_{CS} < 10$	100.00 (33.47)	99.97 (50.50)	100.00 (52.61)
tag vertex fit success	98.94 (33.11)	97.67 (49.32)	97.39 (51.24)
$\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}} < 100$	92.05 (30.48)	89.16 (43.98)	90.90 (46.58)
$\sigma_\ell^{\mathrm{tag}} < 500\,\mu\mathrm{m}$	99.90 (30.45)	99.74 (43.86)	99.71 (46.44)
best candidate selection	98.54 (30.00)	96.42 (42.29)	98.11 (45.56)
χ_{c0} veto	92.62 (27.79)	95.63 (40.44)	94.11 (42.88)
expected/observed number of events	110.5	2523 $\begin{pmatrix} q\overline{q} : 2458 \\ B\overline{B}\ \mathrm{comb} : 49.1 \\ B\overline{B}\ \mathrm{peak} : 16.3 \end{pmatrix}$	1903

Table 4.5: Relative (cumulative) efficiency (%) of TD classification criteria for $B^0 \to K_S^0 K_S^0 K_S^0$ decays. The efficiency is shown in separate columns for signal MC, background MC, and data sample. The cumulative efficiency is normalized to the number of events after the baseline selection. The last two rows show the expected and observed number of TD and TI events.

selection	signal MC	background MC	data
reject off-diagonal K_S^0	86.65 (86.65)	81.47 (81.47)	72.99 (72.99)
at least one K_S^0 with VXD hits	99.57 (86.27)	99.80 (81.31)	99.78 (72.83)
$\left(\frac{\chi^2}{ndf}\right)^{CP} < 3.27$	76.53 (66.02)	67.02 (54.49)	60.03 (43.72)
$\sigma_\ell^{CP} < 500\,\mu\text{m}$	99.85 (65.92)	99.84 (54.40)	99.64 (43.56)
$-30\,\text{ps} < \Delta t < 30\,\text{ps}$	99.95 (65.89)	99.90 (54.35)	100.00 (43.56)
TD fraction	65.89	54.35	43.56
expected/observed	72.8	1372	829
number of TD events		$\begin{pmatrix} q\overline{q} : 1344 \\ B\overline{B}\text{ comb} : 21.4 \\ B\overline{B}\text{ peak} : 6.5 \end{pmatrix}$	
expected/observed	37.7	1152	1074
number of TI events		$\begin{pmatrix} q\overline{q} : 1114 \\ B\overline{B}\text{ comb} : 27.7 \\ B\overline{B}\text{ peak} : 9.8 \end{pmatrix}$	

The cumulative and relative efficiencies of the TD classification criteria are summarized in Tab. 4.5. The cumulative efficiency is normalized by the number of events after the baseline selection. While the most of the above selections are loose, the ones about off-diagonal K_S^0 and $\left(\frac{\chi^2}{ndf}\right)^{CP}$ downgrade 35% of events to TI in the signal MC (See Appendix E for details).

The overall signal reconstruction efficiency including detector acceptance is 27.79%, for TD events 18.3% and for TI events 9.5% in the signal MC. With 200×10^6 $B\overline{B}$, we expect 110.5 (72.8 TD + 37.7 TI) events assuming the same fraction of $B^0\overline{B}^0$ and B^+B^-, $\mathcal{B}(B^0 \to K_S^0 K_S^0 K_S^0) = 6 \times 10^{-6}$, and $\mathcal{B}(K_S^0 \to \pi^+\pi^-) = 0.692$. As for background events, we expect 2523 (1372 TD + 1152 TI) events with $190\,\text{fb}^{-1}$.

In the data the fraction of TD is lower than the MC expectation, because of lower efficiencies in $\left(\frac{\chi^2}{ndf}\right)^{CP}$ and off-diagonal K_S^0 rejection. The data-MC difference of off-diagonal K_S^0 fraction is attributed to the difference of PXD efficiency and fake hit rate.

4.11 Figure of merit

The K_S^0 selection thresholds, $(457.6\,\text{MeV}/c^2 < M_{\pi^+\pi^-} < 537.6\,\text{MeV}/c^2) \wedge (\mathcal{O}_{K_S^0} > 0.75)$, are determined to maximize the figure of merit $FOM = S/\sqrt{S+B}$. Here, S and B are the expected numbers of signal and background events scaled to $200\,\text{fb}^{-1}$ in a tentative

Table 4.6: Breakdown of background events using the number of true K_S^0. Background B_{CP} candidates are classified according to the number of true K_S^0 in B_{CP} (0, 1, 2, 3) and the fraction of each class is shown in %. The fraction is shown before and after K_S^0 selection.

number of true K_S^0	before K_S^0 selection	after K_S^0 selection
0	65.0	0.2
1	29.4	3.8
2	5.2	26.2
3	0.3	69.7

signal region ($5.27\,\mathrm{GeV}/c^2 < M_{\mathrm{bc}} < 5.29\,\mathrm{GeV}/c^2$) $\wedge$ ($5.18\,\mathrm{GeV}/c^2 < M < 5.38\,\mathrm{GeV}/c^2$) $\wedge$ ($\mathcal{O}_{CS} > 0.5$) with all the baseline selections applied except for the ones regarding K_S^0, best candidate selection, and the χ_{c0} veto.

We calculate FOM while changing the $\mathcal{O}_{K_S^0}$ threshold and the width of $M_{\pi^+\pi^-}$ window in steps of 0.01 and 5 MeV as shown in Fig. 4.24. The $\mathcal{O}_{K_S^0}$ threshold of 0.75–0.9 and $M_{\pi^+\pi^-}$ window of 20–50 MeV/c^2 are favored within the statistical fluctuation.

To see the performance of the determined K_S^0 selection criteria, we count the number of true K_S^0 out of three K_S^0 candidates in each selected background events. As shown in Tab. 4.6, 78 % of background B_{CP} candidates surviving K_S^0 selection are reconstructed from random combinations of three true K_S^0 and thus irreducible.

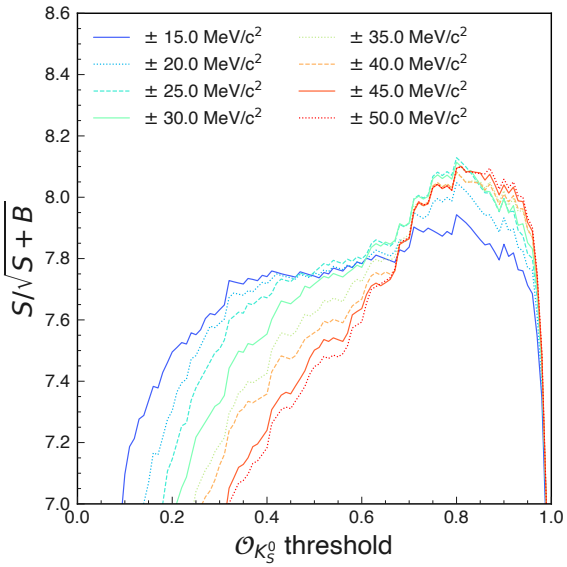
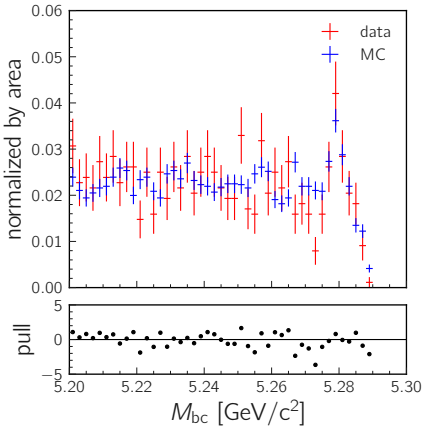


Figure 4.24: Result of FOM scan. Curves show the FOM as a function of $\mathcal{O}_{K_S^0}$ threshold with different $M_{\pi^+\pi^-}$ window widths denoted in the legend.

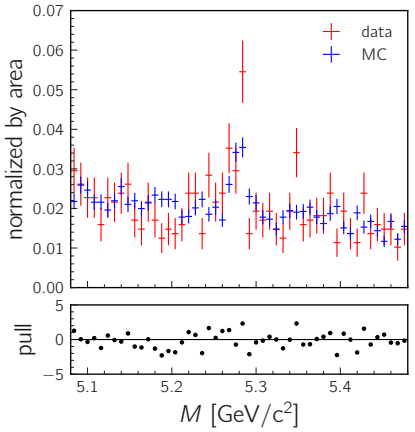
4.12 Data-MC comparison of variables

We compare the distribution of variables between data and MC. For the comparison we mix the background and signal MC samples equivalent to $700\,\mathrm{fb}^{-1}$ with the expected rate and normalize the distributions by area. Figure 4.25 and 4.26 provides the comparison for TD events, and Fig. 4.27 for TI events.

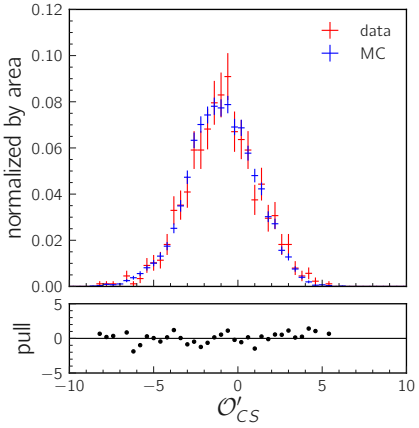
In general the distributions agree well within the statistical uncertainty while we find discrepancies in qr and $\sigma_\ell^{\mathrm{tag}}$ distributions. These discrepancies should arise from the background component, which dominates the distributions. Considering that, we use data sideband rather than MC sample to estimate the PDF of r distribution of background events in the following analysis (see Sec. 5.2.1) We attribute the discrepancy in $\sigma_\ell^{\mathrm{tag}}$ distribution to the difference in PXD efficiency and fake hit rate, which is also indicated by the data-MC difference in off-diagonal K_S^0 fraction. The imperfectness in $q\overline{q}$ simulation can be another reason of the discrepancy. In the following analysis we use $\sigma_\ell^{\mathrm{tag}}$ to estimate tag-side vertex resolution on an event-by-event basis but do not have to know the PDF of its distribution (see Sec. 5.3). Therefore, the discrepancy in the $\sigma_\ell^{\mathrm{tag}}$ distribution does not affect the analysis as long as $\sigma_\ell^{\mathrm{tag}}$ is estimated correctly. The same tendencies are seen in the control channel, which provide a larger data set and thus more decisive comparison (see Appendix. A.3).



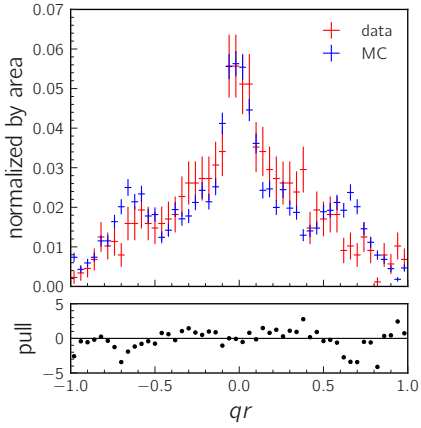
(a) $M_{\rm bc}$



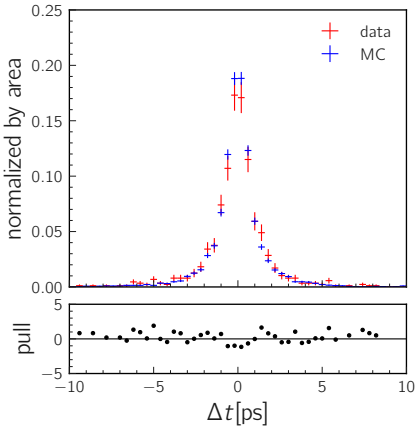
(b) M



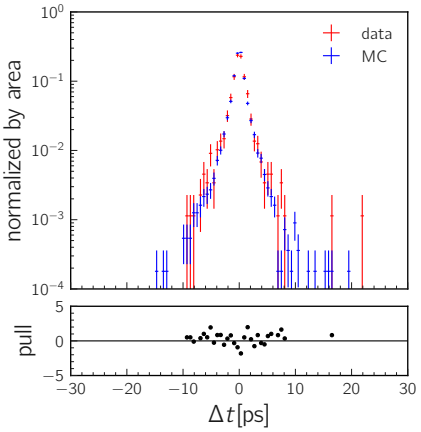
(c) $\mathcal{O}'_{CS}$



(d) qr

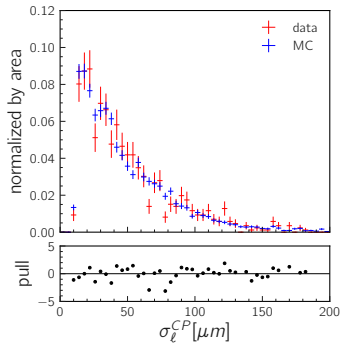


(e) Δt in linear scale

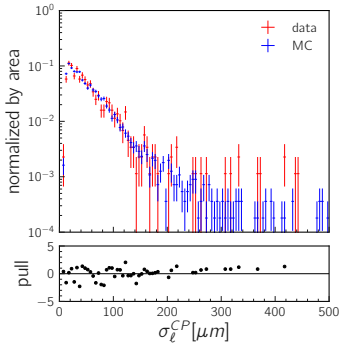


(f) Δt in log scale

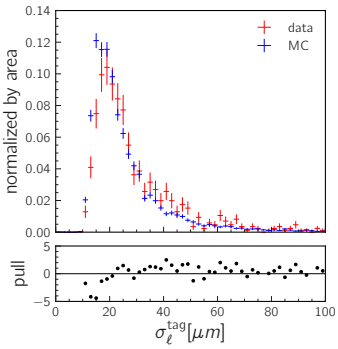
Figure 4.25: Data-MC comparison of distributions of $B^0 \to K^0_S K^0_S K^0_S$ TD events (1). The distributions are normalized by area. Pull is defined as the difference of data and MC distributions normalized by the statistical uncertainty.



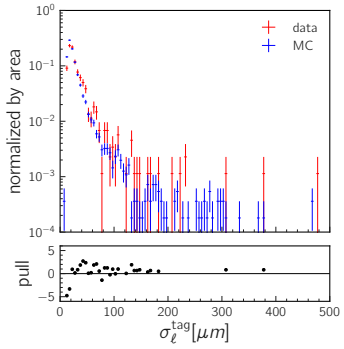
(a) σ_ℓ^{CP} in linear scale



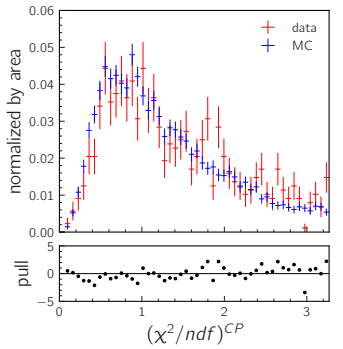
(b) σ_ℓ^{CP} in log scale



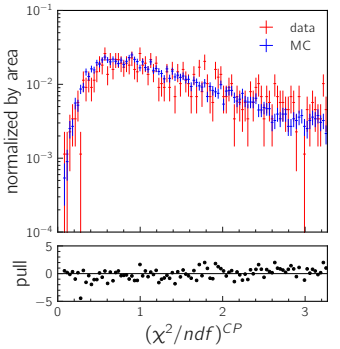
(c) $\sigma_\ell^{\mathrm{tag}}$ in linear scale



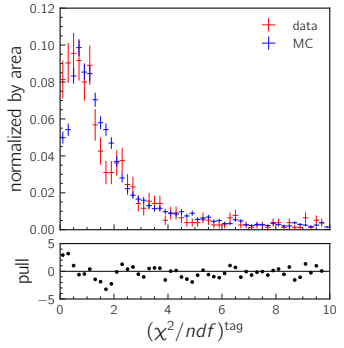
(d) $\sigma_\ell^{\mathrm{tag}}$ in log scale



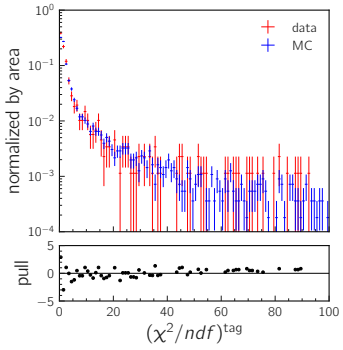
(e) $\left(\frac{\chi^2}{ndf}\right)^{CP}$ in linear scale



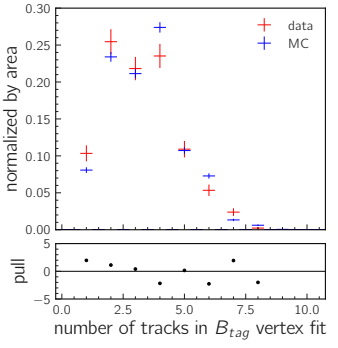
(f) $\left(\frac{\chi^2}{ndf}\right)^{CP}$ in log scale



(g) $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$ in linear scale

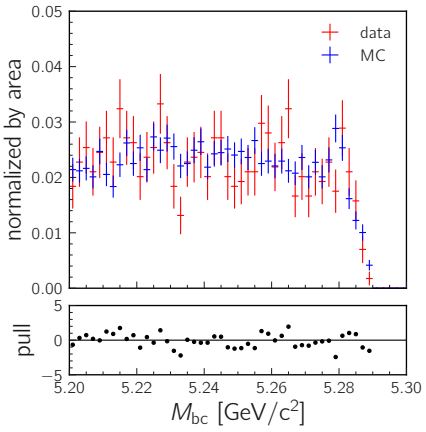


(h) $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$ in log scale

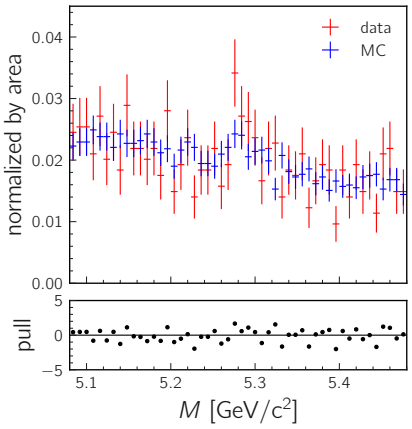


(i) number of tracks in tag-side vertex fit

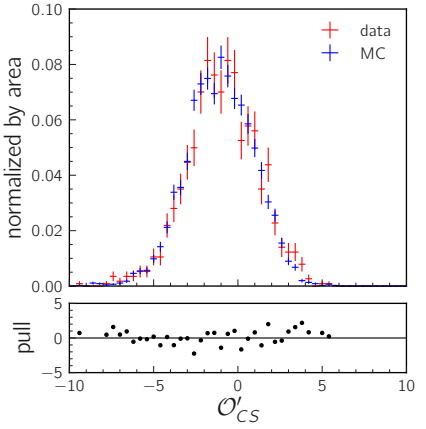
Figure 4.26: Data-MC comparison of distributions of $B^0 \to K_S^0 K_S^0 K_S^0$ TD events (2). The distributions are normalized by area. Pull is defined as the difference of data and MC distributions normalized by the statistical uncertainty.



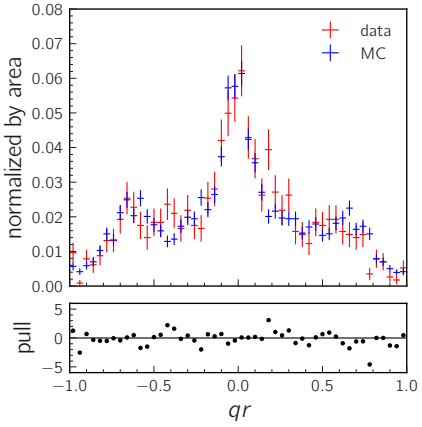
(a) $M_{\rm bc}$



(b) M



(c) $\mathcal{O}'_{CS}$



(d) qr

Figure 4.27: Data-MC comparison of distributions of $B^0 \to K^0_S K^0_S K^0_S$ TI events. The distributions are normalized by area. Pull is defined as the difference of data and MC distributions normalized by the statistical uncertainty.

Chapter 5

Estimation of CP asymmetries

5.1 Maximum likelihood fit

We estimate the CP violation parameters by an unbinned maximum likelihood fit ("CP fit") using events within the signal region defined as

$$\begin{aligned}(5.2708\,\mathrm{GeV}/c^2 < M_{\mathrm{bc}} < 5.2882\,\mathrm{GeV}/c^2)\wedge\\ (5.1814\,\mathrm{GeV}/c^2 < M < 5.3661\,\mathrm{GeV}/c^2)\wedge\\ (-3.945 < \mathcal{O}'_{CS} < 5.807).\end{aligned} \tag{5.1}$$

The likelihood function to be maximized is

$$L = \prod_i P^{\mathrm{TD}}(\Delta t_i, q_i) \prod_j P^{\mathrm{TI}}(q_j), \tag{5.2}$$

where i runs over TD events and j over TI events. The likelihood for a TD event is given by

$$P^{\mathrm{TD}}(\Delta t, q) = f_{\mathrm{sig}} \int d\Delta t' P^{\mathrm{TD}}_{\mathrm{sig}}(\Delta t', q) R(\Delta t - \Delta t') + (1 - f_{\mathrm{sig}}) P_{\mathrm{bkg}}(\Delta t), \tag{5.3}$$

$$P^{\mathrm{TD}}_{\mathrm{sig}}(\Delta t, q) = \frac{1}{4\tau_{B^0}} e^{-\frac{|\Delta t|}{\tau_{B^0}}} \left[1 - q\Delta w + q(1-2w)(S\sin(\Delta m_d \Delta t) + A\cos(\Delta m_d \Delta t))\right], \tag{5.4}$$

and for a TI event,

$$P^{\mathrm{TI}}(q) = f_{\mathrm{sig}} P^{\mathrm{TI}}_{\mathrm{sig}}(q) + \frac{1 - f_{\mathrm{sig}}}{2}, \tag{5.5}$$

$$P^{\mathrm{TI}}_{\mathrm{sig}}(q) = \frac{1}{2}\left[1 - q\Delta w + q(1-2w)A\frac{1}{1+x_d^2}\right]. \tag{5.6}$$

Here,

- f_{sig} is the signal probability calculated on an event-by-event basis called *signal fraction*,

Table 5.1: Parameters in the *CP* fit and the methods to determine them

parameters		method
$f_{\rm sig}$	$F_{\rm sig}$	signal extraction fit to data
	signal PDF $(M_{\rm bc}, M, \mathcal{O}'_{CS}, r)$	fit to signal MC
	background PDF $(M_{\rm bc}, M, \mathcal{O}'_{CS})$	signal extraction fit to control channel data
	background PDF (r)	$M_{\rm bc}$-sideband data of control channel
	signal/background PDF $(\cos\theta^*_B)$	theoretical assumption
$R(\delta\Delta t)$		fit to signal MC
$P_{\rm bkg}(\Delta t)$		fit to $M_{\rm bc}$-sideband data
$w, \Delta w$		calibration using $B^0 \to D^{(*)-}h^+$ decays [1]
$\tau_{B^0}, \Delta m_d$		world average from PDG [7]

- $R(\delta\Delta t)$ is the response function of Δt measurement called *resolution function*, and
- $P_{\rm bkg}(\Delta t)$ is the PDF of measured Δt for background events.

In the *CP* fit, S and A are the only free parameters. We fix the lifetime and mixing parameters to the world average and consider their uncertainties as a source of systematic uncertainty [7]. For testing purpose, we sum the PDF over $q = \pm 1$ and perform a fit to determine the lifetime τ_{B^0} ("lifetime fit"). Table 5.1 summarizes how the parameters other than S and A in the *CP* fit are fixed. We describe the details in the following sections.

One can include the flavor tagging asymmetry μ as well as Δw in the PDF to correct the bias. We rather ignore the effect in the fit model and consider it as a source of systematic uncertainty in this analysis because the uncertainty of the μ parameters are still large and μ is not significantly observed for most of r bin.

We use the MIGRAD and MINOS functions from the MINUIT package for the minimization of negative log likelihood and estimation of fit uncertainty [37].

5.2 Signal fraction

The signal fraction $f_{\rm sig}$ is calculated using the fraction of signal events in a given data set $F_{\rm sig}$ and five variables that discriminate signal and background:

$$\vec{x} = (M_{\rm bc}, M, \mathcal{O}'_{CS}, \cos\theta^*_B, rbin), \tag{5.7}$$

$$f_{\rm sig}(\vec{x}) = \frac{F_{\rm sig}P_{\rm sig}(\vec{x})}{F_{\rm sig}P_{\rm sig}(\vec{x}) + (1 - F_{\rm sig})P_{\rm bkg}(\vec{x})}, \tag{5.8}$$

where $\cos\theta^*_B$ is the cosine of the angle between B_{CP} momentum and the boost direction at the center-of-mass frame and $P_{\rm sig(bkg)}(\vec{x})$ is the PDF of $\vec{x}$ for signal (background) component. We describe the one-dimensional PDFs $P^x_{\rm sig(bkg)}(x)$ in Sec. 5.2.1. We define the five-dimensional PDF as the product of the one-dimensional PDFs assuming no

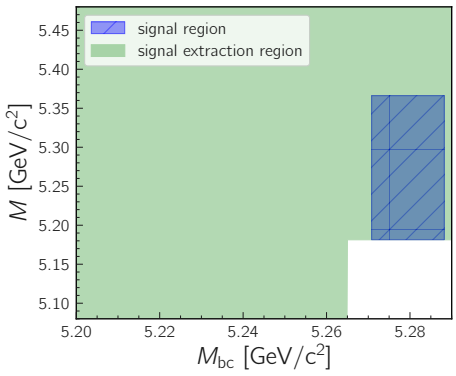


Figure 5.1: Signal extraction region on $M_{\rm bc}$-M plane
[Region used for signal extraction fit shown on $M_{\rm bc}$-M plane (green). The signal region is shown in blue.]

correlation between them:

$$P_{\rm sig}(\vec{x}) = P_{\rm sig}^{M_{\rm bc}}(M_{\rm bc})P_{\rm sig}^{M}(M)P_{\rm sig}^{\mathcal{O}'_{CS}}(\mathcal{O}'_{CS})P_{\rm sig}^{\cos\theta^*_B}(\cos\theta^*_B)P_{\rm sig}^{rbin}(rbin), \tag{5.9}$$

$$P_{\rm bkg}(\vec{x}) = P_{\rm bkg}^{M_{\rm bc}}(M_{\rm bc};E_{\rm beam})P_{\rm bkg}^{M}(M)P_{\rm bkg}^{\mathcal{O}'_{CS}}(\mathcal{O}'_{CS})P_{\rm bkg}^{\cos\theta^*_B}(\cos\theta^*_B)P_{\rm bkg}^{rbin}(rbin). \tag{5.10}$$

We determine the overall signal fraction $F_{\rm sig}$ by an unbinned maximum likelihood fit (signal extraction fit) to the three-dimensional distribution of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ using the following PDF:

$$\begin{aligned} P(M_{\rm bc},M,\mathcal{O}'_{CS};E_{\rm beam}) = {} & F_{\rm sig}P_{\rm sig}^{M_{\rm bc}}(M_{\rm bc})P_{\rm sig}^{M}(M)P_{\rm sig}^{\mathcal{O}'_{CS}}(\mathcal{O}'_{CS})+ \\ & (1-F_{\rm sig})P_{\rm bkg}^{M_{\rm bc}}(M_{\rm bc};E_{\rm beam})P_{\rm bkg}^{M}(M)P_{\rm bkg}^{\mathcal{O}'_{CS}}(\mathcal{O}'_{CS}), \end{aligned} \tag{5.11}$$

In the signal extraction fit we use the events within

$$(5.2\,{\rm GeV}/c^2 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2)\wedge \tag{5.12}$$

$$(5.08\,{\rm GeV}/c^2 < M < 5.48\,{\rm GeV}/c^2)\wedge \tag{5.13}$$

$$(-10 < \mathcal{O}'_{CS} < 10), \tag{5.14}$$

while excluding the region contaminated by the $B\overline{B}$ peaking background,

$$(5.265\,{\rm GeV}/c^2 < M_{\rm bc} < 5.29\,{\rm GeV}/c^2) \wedge (5.08\,{\rm GeV}/c^2 < M < 5.1814\,{\rm GeV}/c^2). \tag{5.15}$$

Figure 5.1 shows the fit region on the $M_{\rm bc}$-M plane. We normalize the PDFs of $M_{\rm bc}$ and M appropriately within the region. $F_{\rm sig}$ is defined as the fraction of signal events within the same region.

We estimate $F_{\rm sig}$ separately for TD and TI events because MC samples predict significantly different $F_{\rm sig}$ and signal M distribution between them.

Table 5.2: PDF models of $M_{\rm bc}$, M, $\mathcal{O}'_{CS}$, $\cos\theta^*_B$, and r bin distributions. The number of free parameters are shown in parentheses.

variable	signal	background
$M_{\rm bc}$	Gaussian (3)	ARGUS (1)
M	double Gaussian (5)	1D Chebyshev polynomial (1)
$\mathcal{O}'_{CS}$	bifurcated Gaussian (3)	bifurcated Gaussian + Gaussian (6)
$\cos\theta^*_B$	$1-\cos^2\theta^*_B$	uniform
r bin	histogram	histogram

5.2.1 1D PDFs

We model the PDFs for $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ as

$$P_{\rm sig}^{M_{\rm bc}}(M_{\rm bc}) = G(M_{\rm bc};\mu_{\rm sig}^{M_{\rm bc}},\sigma_{\rm sig}^{M_{\rm bc}}), \tag{5.16}$$

$$P_{\rm bkg}^{M_{\rm bc}}(M_{\rm bc};E_{\rm beam}) = {\rm ARGUS}(M_{\rm bc};a_{\rm bkg}^{M_{\rm bc}},E_{\rm beam}/2), \tag{5.17}$$

$$P_{\rm sig}^{M}(M) = f_{\rm 1,sig}^{M}G(M;\mu_{\rm 1,sig}^{M},\sigma_{\rm 1,sig}^{M}) + (1-f_{\rm 1,sig}^{M})G(M;\mu_{\rm 2,sig}^{M},\sigma_{\rm 2,sig}^{M}), \tag{5.18}$$

$$P_{\rm bkg}^{M}(M) = 1 + c_{\rm bkg}^{M}\frac{2M-M_{\rm max}-M_{\rm min}}{M_{\rm max}-M_{\rm min}}, \tag{5.19}$$

$$P_{\rm sig}^{\mathcal{O}'_{CS}}(\mathcal{O}'_{CS}) = {\rm bifG}(\mathcal{O}'_{CS};\mu_{\rm sig}^{\mathcal{O}'_{CS}},\sigma_{\rm L,sig}^{\mathcal{O}'_{CS}},\sigma_{\rm R,sig}^{\mathcal{O}'_{CS}}), \tag{5.20}$$

$$\begin{aligned}P_{\rm bkg}^{\mathcal{O}'_{CS}}(\mathcal{O}'_{CS}) = {}& f_{\rm 1,bkg}^{\mathcal{O}'_{CS}}G(\mathcal{O}'_{CS};\mu_{\rm 1,bkg}^{\mathcal{O}'_{CS}},\sigma_{\rm 1,bkg}^{\mathcal{O}'_{CS}}) \\ & + (1-f_{\rm 1,bkg}^{\mathcal{O}'_{CS}}){\rm bifG}(\mathcal{O}'_{CS};\mu_{\rm 2,bkg}^{\mathcal{O}'_{CS}},\sigma_{\rm 2L,bkg}^{\mathcal{O}'_{CS}},\sigma_{\rm 2R,bkg}^{\mathcal{O}'_{CS}}),\end{aligned} \tag{5.21}$$

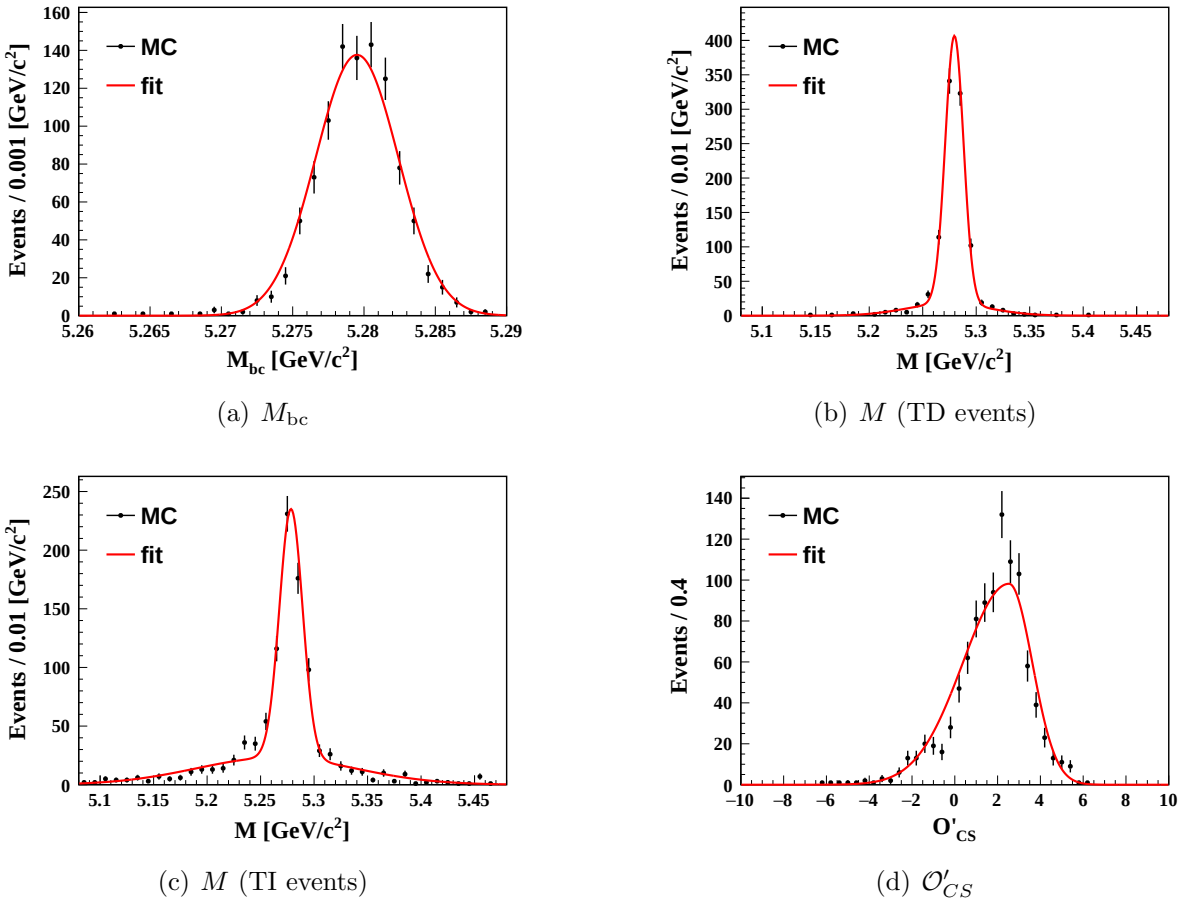
where $M_{\rm max} = 5.48\,{\rm GeV}/c^2$ and $M_{\rm min} = 5.08\,{\rm GeV}/c^2$ are the upper and lower limits of M, the other parameters are free parameters, and functions G, ARGUS, and bifG are defined as

$$G(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{5.22}$$

$${\rm ARGUS}(x;a,c) = x\sqrt{1-(x/c)^2}e^{a(1-(x/c)^2)}\ [38],\ \text{and} \tag{5.23}$$

$${\rm bifG}(x;\mu,\sigma_L,\sigma_R) = \begin{cases}\frac{2\sigma_L}{\sigma_L+\sigma_R}G(x;\mu,\sigma_L) & (x<\mu)\\ \frac{2\sigma_R}{\sigma_L+\sigma_R}G(x;\mu,\sigma_R) & (x\geq\mu)\end{cases} \tag{5.24}$$

We determine the parameters of signal PDFs of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ by an unbinned maximum likelihood fit to 1000 events sampled from the signal MC sample. We keep the number of events only one order of magnitude higher than the expected number in data so that we can reasonably simplify the PDF models. Because M distribution for TI events have a larger and broader tail than TD, we define a separate parameter set for M. Figure 5.2 shows the distributions and fitted functions. The signal region in Eq. (5.1) retains 99.73% of the signal PDF in each dimension. We determine the background PDF parameters by a signal extraction fit to the control sample $B^+ \to K^0_S K^0_S K^+$ simultaneously with its signal fraction (see Appendix A). The parameters of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ PDFs are summarized in Tab. 5.3.



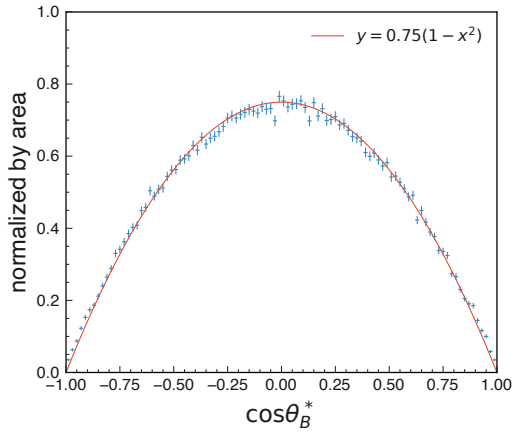
(a) $M_{\rm bc}$ (b) M (TD events)

(c) M (TI events) (d) $\mathcal{O}'_{CS}$

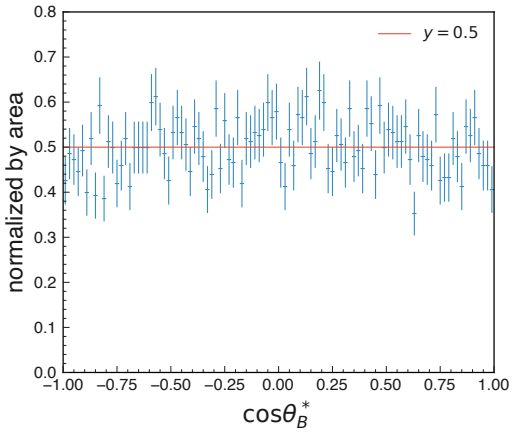
Figure 5.2: $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ distributions for signal MC events. M distributions are shown separately for TD and TI events. The red curves show the fitted PDFs.

Table 5.3: Parameters of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ PDFs

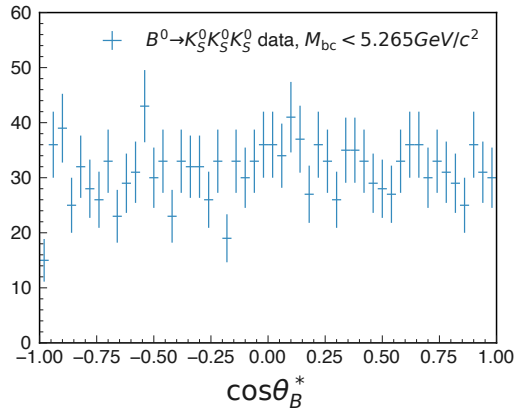
parameter	value
$\mu_{\rm sig}^{M_{\rm bc}}$	5.27951 ± 0.00009 GeV/c^2
$\sigma_{\rm sig}^{M_{\rm bc}}$	0.002899 ± 0.000065 GeV/c^2
$f_{\rm 1,sig}^{M}$(TD)	0.833 ± 0.021
$\mu_{\rm 1,sig}^{M}$(TD)	5.27962 ± 0.00034 GeV/c^2
$\sigma_{\rm 1,sig}^{M}$(TD)	0.00853 ± 0.00032 GeV/c^2
$\mu_{\rm 2,sig}^{M}$(TD)	5.2737 ± 0.0031 GeV/c^2
$\sigma_{\rm 2,sig}^{M}$(TD)	0.0384 ± 0.0027 GeV/c^2
$f_{\rm 1,sig}^{M}$(TI)	0.574 ± 0.024
$\mu_{\rm 1,sig}^{M}$(TI)	5.27845 ± 0.00057 GeV/c^2
$\sigma_{\rm 1,sig}^{M}$(TI)	0.01081 ± 0.00063 GeV/c^2
$\mu_{\rm 2,sig}^{M}$(TI)	5.2641 ± 0.0038 GeV/c^2
$\sigma_{\rm 2,sig}^{M}$(TI)	0.0740 ± 0.0035 GeV/c^2
$\mu_{\rm sig}^{\mathcal{O}'_{CS}}$	2.53 ± 0.10
$\sigma_{\rm L,sig}^{\mathcal{O}'_{CS}}$	2.160 ± 0.075
$\sigma_{\rm R,sig}^{\mathcal{O}'_{CS}}$	1.091 ± 0.062
$a_{\rm bkg}^{M_{\rm bc}}$	-23.3 ± 1.3
$c_{\rm bkg}^{M}$	-0.208 ± 0.019
$f_{\rm 1,bkg}^{\mathcal{O}'_{CS}}$	0.869 ± 0.033
$\mu_{\rm 1,bkg}^{\mathcal{O}'_{CS}}$	-1.118 ± 0.041
$\sigma_{\rm 1,bkg}^{\mathcal{O}'_{CS}}$	1.780 ± 0.038
$\mu_{\rm 2,bkg}^{\mathcal{O}'_{CS}}$	2.25 ± 0.30
$\sigma_{\rm 2L,bkg}^{\mathcal{O}'_{CS}}$	4.61 ± 0.38
$\sigma_{\rm 2R,bkg}^{\mathcal{O}'_{CS}}$	0.74 ± 0.21



(a) signal MC



(b) background MC



(c) $M_{\rm bc}$-sideband data

Figure 5.3: Distributions of $\cos\theta_B^*$ in signal MC (a), background MC (b), and $M_{\rm bc}$-sideband data (c). The red curve shows the signal (a) and background (b,c) PDFs.

We define the PDFs of $\cos\theta_B^*$ as

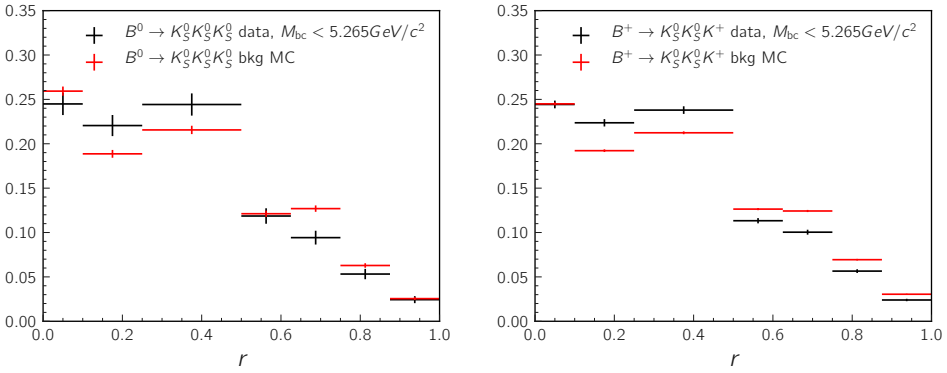
$$P_{\rm sig}^{\cos\theta_B^*}(\cos\theta_B^*) = \frac{3}{4}(1 - \cos^2\theta_B^*) \text{ and} \tag{5.25}$$

$$P_{\rm bkg}^{\cos\theta_B^*}(\cos\theta_B^*) = 0.5 \ . \tag{5.26}$$

The signal PDF is the theoretical distribution where two pseudoscalar mesons are produced from a transversely polarized vector meson. Figure 5.3 shows the PDFs together with $\cos\theta_B^*$ distributions for signal and background MC samples and $M_{\rm bc}$-sideband data defined as $M_{\rm bc} < 5.265\,{\rm GeV}/c^2$.

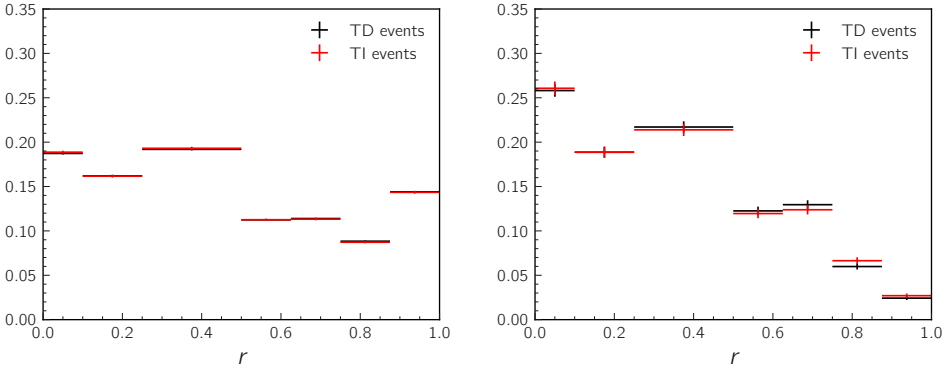
We use histogram PDFs for r bin. The signal PDF is sampled from MC. Because we find a disagreement in the rbin distributions between data and MC as shown in Fig. 5.4, we use the r bin distribution in the $M_{\rm bc}$ sideband ($M_{\rm bc} < 5.265\,{\rm GeV}/c^2$) of control sample data $B^+ \to K_S^0 K_S^0 K^+$ as the background PDF. We also confirm that TD and TI events

have similar r bin distribution as shown in Fig. 5.5 since we keep the same tag-side vertex cut for both categories.



(a) $B^0 \to K^0_S K^0_S K^0_S$ (b) $B^+ \to K^0_S K^0_S K^+$

Figure 5.4: r bin distributions in $M_{\rm bc}$-sideband data and background MC. The distributions are shown each for $B^0 \to K^0_S K^0_S K^0_S$ (a) and $B^+ \to K^0_S K^0_S K^+$ (b).



(a) signal MC (b) background MC

Figure 5.5: r bin distributions of TD and TI events from $B^0 \to K^0_S K^0_S K^0_S$ signal MC (a) and background MC (b)

5.2.2 Results of signal extraction fit

Figure 5.6 and 5.7 shows the fit result of three-dimensional fits to TD and TI events. Table 5.4 summarizes the obtained values of $F_{\rm sig}$ and purity in the signal region.

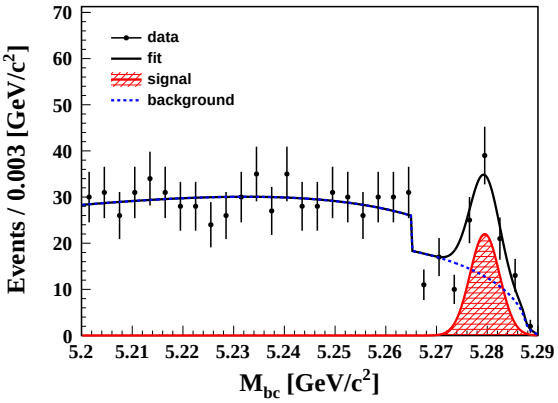
The total number of observed events, 103^{+12}_{-11}, is consistent with our expectation, 106 ± 9, which is based on:

- $N_{B\bar{B}} = (198 \pm 3.0) \times 10^6$,

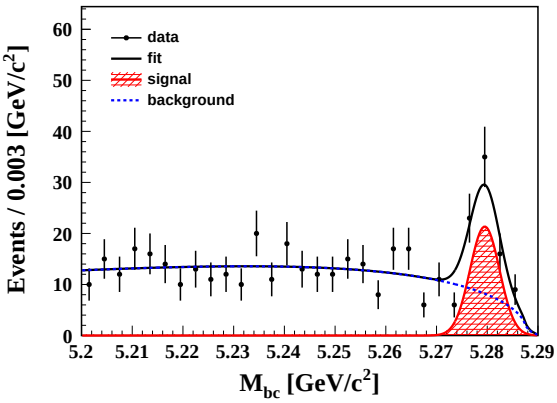
Table 5.4: Results of signal extraction fit to $B^0 \to K^0_S K^0_S K^0_S$ data. $N_{\rm sig}$ is the number of signal events.

	TD events	TI events	total
Number of events in the fit	788	1005	1793
$F_{\rm sig}$	$0.067603^{+0.010114}_{-0.009328}$	$0.049967^{+0.008149}_{-0.007487}$	-
$N_{\rm sig}$	$53.27^{+8.19}_{-7.59}$	$50.22^{+8.34}_{-7.69}$	$103.49^{+11.69}_{-10.80}$
$N_{\rm sig}$ in signal region	52.9	48.1	101.0
purity in signal region	53.7 %	44.7 %	49.0 %

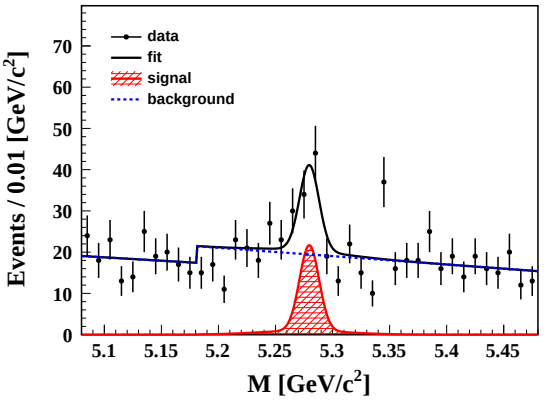
- $\mathcal{B}(\Upsilon(4S) \to B^0\overline{B}^0) = 0.486 \pm 0.006$,
- $\mathcal{B}(B^0 \to K^0_S K^0_S K^0_S) = (6.0 \pm 0.5) \times 10^{-6}$,
- $\mathcal{B}(K^0_S \to \pi^+\pi^-) = 0.692$, and
- reconstruction efficiency of 27.8 % in MC,



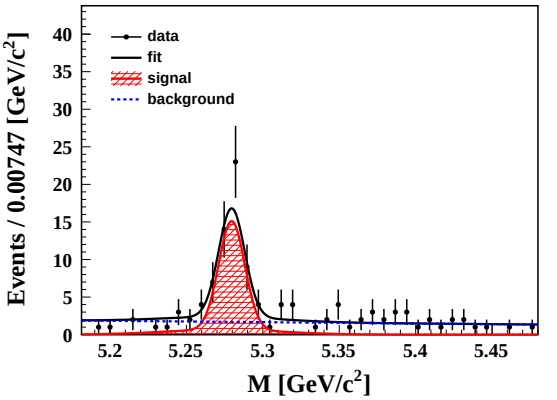
(a) $M_{\rm bc}$



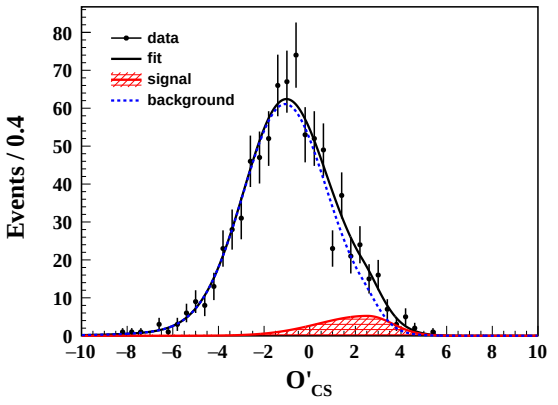
(b) $M_{\rm bc}$ in signal region of M



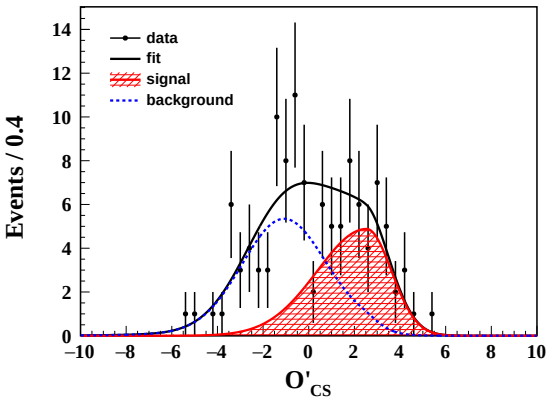
(c) M



(d) M in signal region of $M_{\rm bc}$

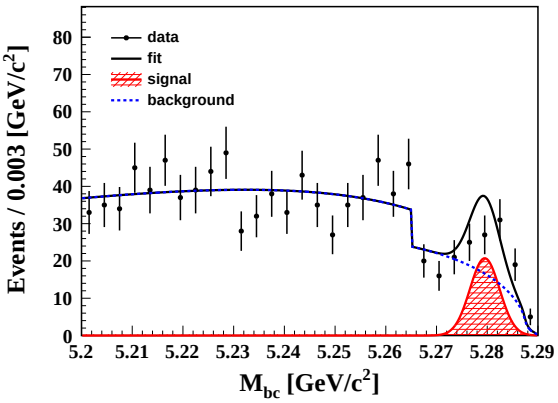


(e) $\mathcal{O}'_{CS}$

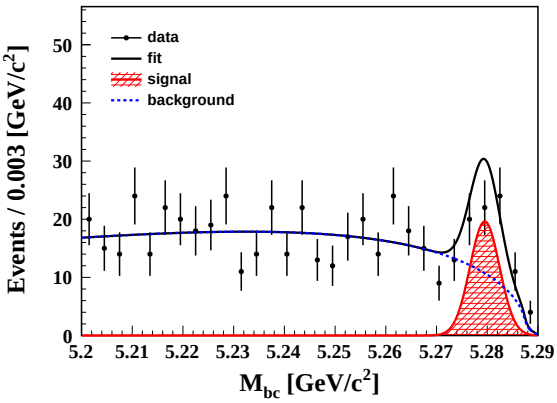


(f) $\mathcal{O}'_{CS}$ in signal region of $M_{\rm bc}$

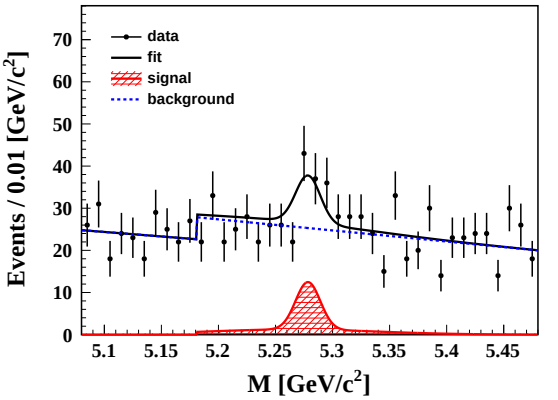
Figure 5.6: Results of signal extraction fit to TD events in $B^0 \to K^0_S K^0_S K^0_S$ data. The black dots with error bars show the distribution. The black line shows the fitted PDF. The red and blue lines show the signal and background components of the PDF. In the right column, projections to the signal region of M or $M_{\rm bc}$ are shown. The step structure in (a,c) originates from the non-rectangular fit region shown in Fig. 5.1.



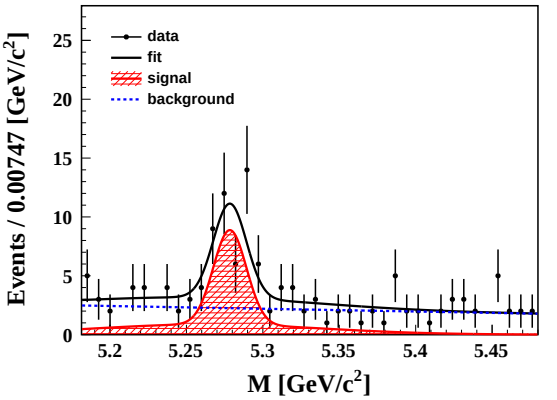
(a) $M_{\rm bc}$



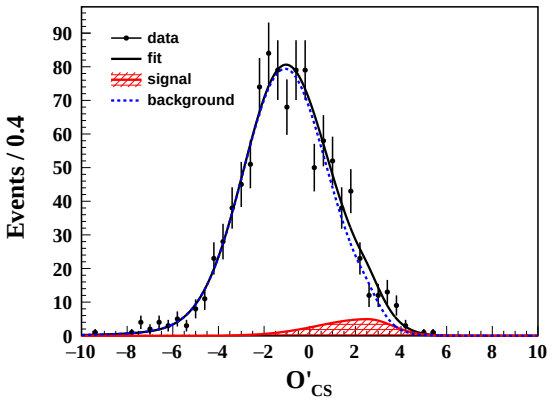
(b) $M_{\rm bc}$ in M signal region



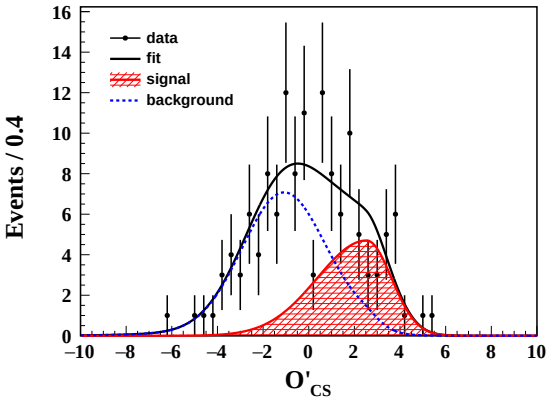
(c) M



(d) M in $M_{\rm bc}$ signal region



(e) $\mathcal{O}'_{CS}$



(f) $\mathcal{O}'_{CS}$ in $M_{\rm bc}$ signal region

Figure 5.7: Results of signal extraction fit to TI events in $B^0 \to K^0_S K^0_S K^0_S$ data. The black dots with error bars show the distribution. The black line shows the fitted PDF. The red and blue lines show the signal and background components of the PDF. In the right column, projections to the signal region of M or $M_{\rm bc}$ are shown. The step structure in (a,c) originates from the non-rectangular fit region shown in Fig. 5.1.

5.3 Resolution function

We define the resolution function as the convolution of four functions [39],

$$R = R_{\rm k} \otimes R_{\rm rec} \otimes R_{\rm asc} \otimes R_{\rm np}, \tag{5.27}$$

where

- $R_{\rm k}$ corrects an approximation of boost factor,
- $R_{\rm rec}$ stands for the detector resolution of B_{CP} vertex position,
- $R_{\rm asc}$ stands for the detector resolution of $B_{\rm tag}$ vertex position, and
- $R_{\rm np}$ stands for the bias on $B_{\rm tag}$ vertex position due to secondary tracks from long-lived intermediate states including a charm quark.

The correction of boost factor requires estimation of $\cos\theta^*_B$ and $E_{\rm beam}$. The detector resolution and the secondary-track bias strongly depend on χ^2/ndf and σ_ℓ of the vertex fit, so we use them as conditional variables of $R_{\rm rec}$, $R_{\rm asc}$, and $R_{\rm np}$ to describe the resolution well and to eventually improve the sensitivity to CP violation.

In this section we use TD events from the signal MC but with a loose cut, $\left(\frac{\chi^2}{ndf}\right)^{CP} < 100$, to determine the parameters of resolution function. As the parameter determination fully relies on the MC, we validate the parameters using $B^+ \to K^0_S K^0_S K^+$ data, for which we set most of the resolution function parameters common to $B^0 \to K^0_S K^0_S K^0_S$. Although the resolution function works fine for $B^0 \to K^0_S K^0_S K^0_S$ up to $\left(\frac{\chi^2}{ndf}\right)^{CP} < 100$, we find that the common parameters cannot describe the resolution for $B^+ \to K^0_S K^0_S K^+$ when $\left(\frac{\chi^2}{ndf}\right)^{CP}$ is large (see Appendix A). Therefore we conservatively apply the tight selection of $\left(\frac{\chi^2}{ndf}\right)^{CP} < 3.27$ to TD events in the CP fit. Within the $\left(\frac{\chi^2}{ndf}\right)^{CP}$ range, the common resolution function parameters work for both $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$.

5.3.1 Kinematic approximation

In Eq. (3.7) we approximate the boost factors of B mesons by that of $\Upsilon(4S)$ as the B's are almost at rest at the center-of-mass frame. They are correctly expressed as:

$$(\beta\gamma)_{CP} = \beta\gamma\frac{E^*_B}{m_B} + \gamma\frac{p^*_B\cos\theta^*_B}{m_B} \equiv (a_k + c_k)\beta\gamma \tag{5.28}$$

$$(\beta\gamma)_{\rm tag} = \beta\gamma\frac{E^*_B}{m_B} - \gamma\frac{p^*_B\cos\theta^*_B}{m_B} \equiv (a_k - c_k)\beta\gamma, \tag{5.29}$$

where E^*_B and p^*_B are the energy and momentum magnitude of B meson at the center-of-mass frame calculated from $E_{\rm beam}$:

$$E^*_B = E_{\rm beam}/2, \tag{5.30}$$

$$p^*_B = \sqrt{(E_{\rm beam}/2)^2 - m^2_B}. \tag{5.31}$$

Without vertex resolution, measured Δt would be

$$\Delta t = \frac{\ell_{CP} - \ell_{\rm tag}}{\beta\gamma} \tag{5.32}$$

$$= \frac{(\beta\gamma)_{CP}}{\beta\gamma} t_{CP} - \frac{(\beta\gamma)_{\rm tag}}{\beta\gamma} t_{\rm tag} \tag{5.33}$$

$$= a_k \Delta t^{\rm true} + c_k (t_{CP} + t_{\rm tag}), \tag{5.34}$$

where t_{CP} and $t_{\rm tag}$ are the true proper decay time of B_{CP} and $B_{\rm tag}$. $a_k \sim 1.002$ accounts for a relativistic effect and $c_k \sim 0.22 \cos\theta^*_B$ for correction due to nonzero B momentum at the center-of-mass system. We calculate a_k from the beam energy calibration and measure c_k on an event-by-event basis.

The PDF for the measured Δt is given by

$$P(\Delta t) = \int d\Delta t^{\rm true} E_{\rm f}(\Delta t^{\rm true}) R_k(\Delta t - \Delta t^{\rm true}) \tag{5.35}$$

$$= \frac{1}{2a_k\tau_B} \exp\left(-\frac{|\Delta t|}{(a_k \pm c_k)\tau_B}\right) \ (+\text{for } \Delta t \geq 0, - \text{ for } \Delta t < 0), \tag{5.36}$$

where $E_{\rm f} \equiv \frac{1}{2\tau_B} \exp\left(-\frac{|\Delta t^{true}|}{\tau_B}\right)$ is the true PDF for Δt. The expression of $R_{\rm k}$ is in Ref. [39].

5.3.2 *CP*-side resolution function

We define the *CP*-side resolution function $R_{\rm rec}$ as

$$R_{\rm rec}\left(\delta\ell^{CP}; \left(\frac{\chi^2}{ndf}\right)^{CP}, \sigma_\ell^{CP}\right) = (1 - f_{\rm tail}^{CP}) G\left(\delta\ell^{CP}; \mu = 0, \sigma = (s_{\rm main}^{CP,0} + s_{\rm main}^{CP,1}\left(\frac{\chi^2}{ndf}\right)^{CP})\sigma_\ell^{CP}\right)$$
$$+ f_{\rm tail}^{CP} G\left(\delta\ell^{CP}; \mu = 0, \sigma = (s_{\rm tail}^{CP,0} + s_{\rm tail}^{CP,1}\left(\frac{\chi^2}{ndf}\right)^{CP})\sigma_\ell^{CP}\right), \tag{5.37}$$

where $\delta\ell^{CP}$ is the residual of measured vertex position from the true position and $f_{\rm tail}^{CP}$ and $s_{\rm main(tail)}^{CP,0(1)}$ are free parameters. We determine the parameters by fitting $R_{\rm rec}$ to the distribution of $\delta\ell^{CP}$, $\left(\frac{\chi^2}{ndf}\right)^{CP}$, and σ_ℓ^{CP} in the signal MC. The parameters are shown in Tab. 5.5. Figure 5.8 shows the $\delta\ell^{CP}$ distribution and the fitted function. Dividing the distribution in bins of σ_ℓ^{CP} and $\left(\frac{\chi^2}{ndf}\right)^{CP}$ as in Figs. 5.9 and 5.10, we find that the resolution strongly depends on these variables and $R_{\rm rec}$ describe the dependence well.

$R_{\rm rec}$ can also cover well the difference in number of PXD hits on the *CP* side (See Appendix D).

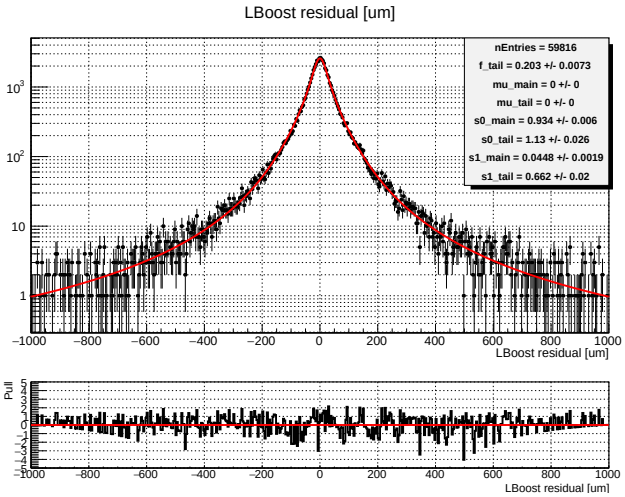
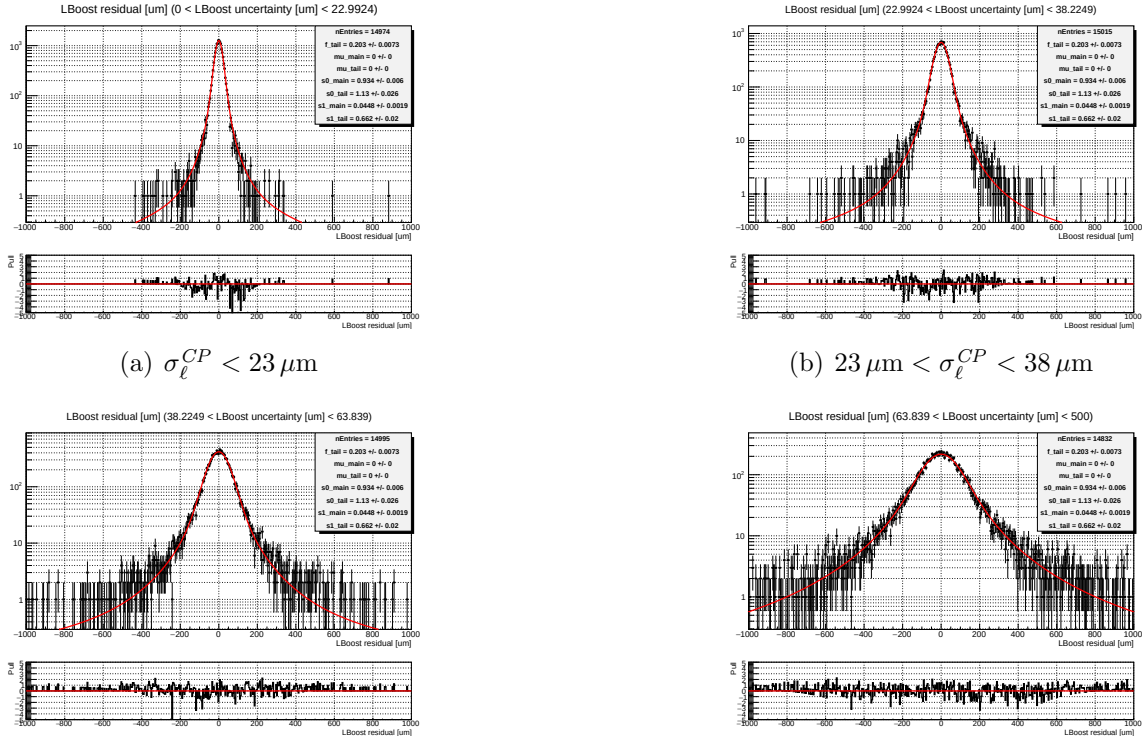


Figure 5.8: $\delta\ell^{CP}$ distribution of signal MC events and R_{rec} fitted to the distribution. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.



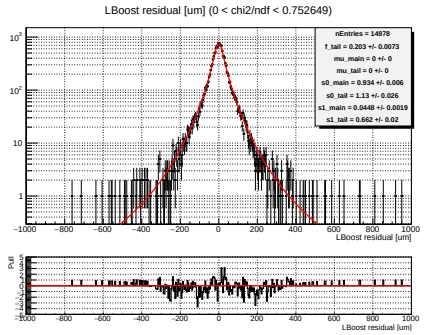
(a) $\sigma_\ell^{CP} < 23\,\mu\mathrm{m}$

(b) $23\,\mu\mathrm{m} < \sigma_\ell^{CP} < 38\,\mu\mathrm{m}$

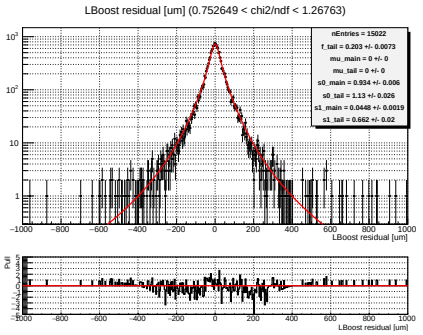
(c) $38\,\mu\mathrm{m} < \sigma_\ell^{CP} < 64\,\mu\mathrm{m}$

(d) $64\,\mu\mathrm{m} < \sigma_\ell^{CP}$

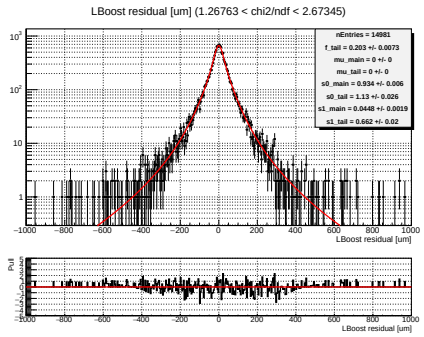
Figure 5.9: $\delta\ell^{CP}$ distribution of signal MC events shown in bins of σ_ℓ^{CP}. R_{rec} is overlaid to the distribution, whose parameters are fixed by the fit shown in Fig. 5.8. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.



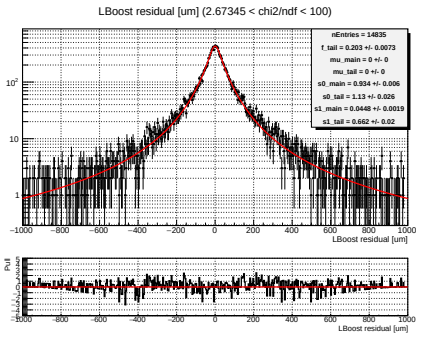
(a) $\left(\frac{\chi^2}{ndf}\right)^{CP} < 0.75$



(b) $0.75 < \left(\frac{\chi^2}{ndf}\right)^{CP} < 1.27$



(c) $1.27 < \left(\frac{\chi^2}{ndf}\right)^{CP} < 2.67$



(d) $2.67 < \left(\frac{\chi^2}{ndf}\right)^{CP}$

Figure 5.10: $\delta\ell^{CP}$ distribution of signal MC events shown in bins of $\left(\frac{\chi^2}{ndf}\right)^{CP}$. $R_{\rm rec}$ is overlaid to the distribution, whose parameters are fixed by the fit shown in Fig. 5.8. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.

Table 5.5: *CP*-side resolution function parameters fixed by the fit to signal MC shown in Fig. 5.8

parameter	value
$s^{CP,0}_{\rm main}$	0.9344 ± 0.0060
$s^{CP,1}_{\rm main}$	0.0448 ± 0.0019
$s^{CP,0}_{\rm main}$	1.126 ± 0.026
$s^{CP,1}_{\rm main}$	0.6619 ± 0.020
$f^{CP}_{\rm tail}$	0.2030 ± 0.0073

5.3.3 tag-side resolution functions

We describe the tag-side resolution by the convolution of two functions, $R_{\rm asc}$ and $R_{\rm np}$. $R_{\rm asc}$ takes the same form as $R_{\rm rec}$ but with independent parameters:

$$R_{\rm asc}\left(\delta\ell^{\rm tag};(\frac{\chi^2}{ndf})^{\rm tag},\sigma_\ell^{\rm tag}\right)=(1-f_{\rm tail}^{\rm tag})G\left(\delta\ell^{\rm tag};\mu=0,\sigma=(s_{\rm main}^{\rm tag,0}+s_{\rm main}^{\rm tag,1}(\frac{\chi^2}{ndf})^{\rm tag})\sigma_\ell^{\rm tag}\right)$$
$$+f_{\rm tail}^{\rm tag}G\left(\delta\ell^{\rm tag};\mu=0,\sigma=(s_{\rm tail}^{\rm tag,0}+s_{\rm tail}^{\rm tag,1}(\frac{\chi^2}{ndf})^{\rm tag})\sigma_\ell^{\rm tag}\right). \tag{5.38}$$

We define $R_{\rm np}$ as the sum of a delta function and asymmetric exponential tails:

$$R_{\rm np}\left(\delta\ell^{\rm tag};(\frac{\chi^2}{ndf})^{\rm tag},\sigma_\ell^{\rm tag}\right)=f_\delta\delta(\delta\ell^{\rm tag})+(1-f_\delta)f_{\rm p}E_{\rm p}(\delta\ell^{\rm tag};\tau=\tau'\sigma_\ell^{\rm tag})$$
$$+(1-f_\delta)(1-f_{\rm p})E_{\rm n}(\delta\ell^{\rm tag};\tau=\tau'\sigma_\ell^{\rm tag}), \tag{5.39}$$

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

$$E_{\rm p}(x;\tau)=\begin{cases}\frac{1}{\tau}\exp(-\frac{x}{\tau}) & (x>0)\\ 0 & (x\le 0)\end{cases} \tag{5.40}$$

$$E_{\rm n}(x;\tau)=\begin{cases}0 & (x>0)\\ \frac{1}{\tau}\exp(+\frac{x}{\tau}) & (x\le 0)\end{cases}, \tag{5.41}$$

and $f_\delta, f_{\rm p}$, and τ' are expressed as

$$f_\delta=\begin{cases}0 & (f_\delta^0+f_\delta^1(\frac{\chi^2}{ndf})^{\rm tag}<0)\\ f_\delta^0+f_\delta^1(\frac{\chi^2}{ndf})^{\rm tag} & (0\le f_\delta^0+f_\delta^1(\frac{\chi^2}{ndf})^{\rm tag}\le 1)\\ 1 & (f_\delta^0+f_\delta^1(\frac{\chi^2}{ndf})^{\rm tag}>1)\end{cases}, \tag{5.42}$$

$$f_{\rm p}=\begin{cases}0 & (f_{\rm p}^0+f_{\rm p}^1\sigma_\ell^{\rm tag}<0)\\ f_{\rm p}^0+f_{\rm p}^1\sigma_\ell^{\rm tag} & (0\le f_{\rm p}^0+f_{\rm p}^1\sigma_\ell^{\rm tag}\le 1)\\ 1 & (f_{\rm p}^0+f_{\rm p}^1\sigma_\ell^{\rm tag}>1)\end{cases}, \tag{5.43}$$

$$\tau'=\begin{cases}\tau^0+\tau^1(\frac{\chi^2}{ndf})^{\rm tag} & (\tau^0+\tau^1(\frac{\chi^2}{ndf})^{\rm tag}\le\tau^{\rm max})\\ \tau^{\rm max} & (\tau^0+\tau^1(\frac{\chi^2}{ndf})^{\rm tag}>\tau^{\rm max})\end{cases}, \tag{5.44}$$

using seven free parameters $f_\delta^{0,1}, f_{\rm p}^{0,1},$ and $\tau^{0,1,\rm max}$.

As well as $R_{\rm rec}$, we determine the 12 parameters by fitting $R_{\rm asc}\otimes R_{\rm np}$ to the tag-side distribution. The parameters are shown in Tab. 5.6. Figures 5.11, 5.12 and 5.13 shows the $\delta\ell^{\rm tag}$ distribution and the fitted function over entire range and in bins of $\sigma_\ell^{\rm tag}$ and $(\frac{\chi^2}{ndf})^{\rm tag}$.

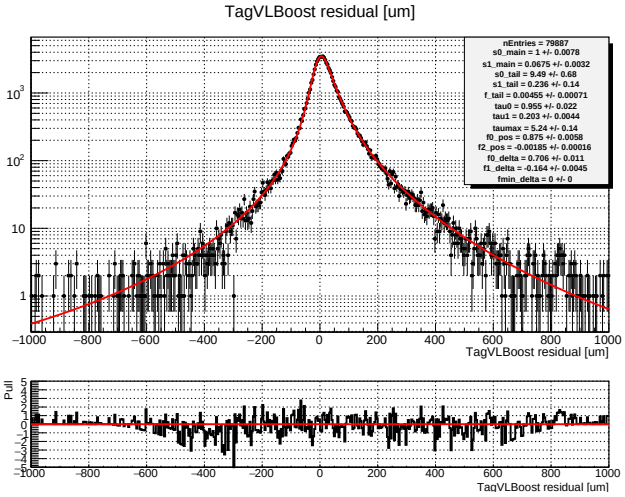
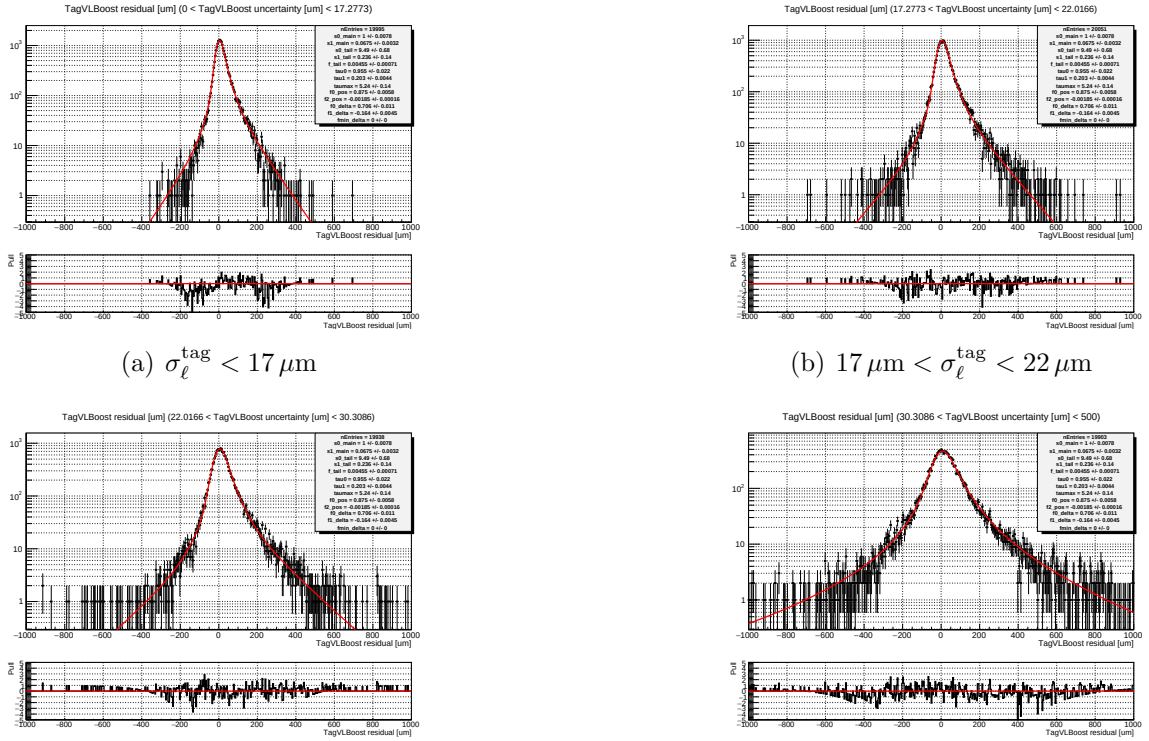


Figure 5.11: $\delta\ell^{\mathrm{tag}}$ distribution of signal MC events and $R_{\mathrm{asc}} \otimes R_{\mathrm{np}}$ fitted to the distribution. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.



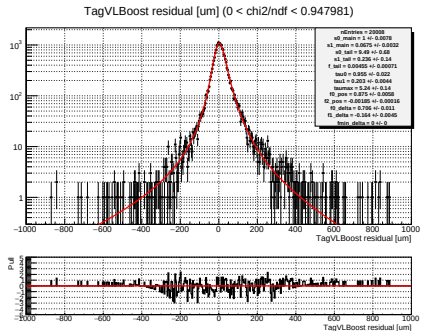
(a) $\sigma_\ell^{\mathrm{tag}} < 17\,\mu\mathrm{m}$

(b) $17\,\mu\mathrm{m} < \sigma_\ell^{\mathrm{tag}} < 22\,\mu\mathrm{m}$

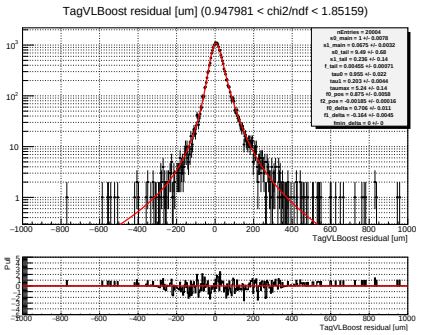
(c) $22\,\mu\mathrm{m} < \sigma_\ell^{\mathrm{tag}} < 30\,\mu\mathrm{m}$

(d) $30\,\mu\mathrm{m} < \sigma_\ell^{\mathrm{tag}}$

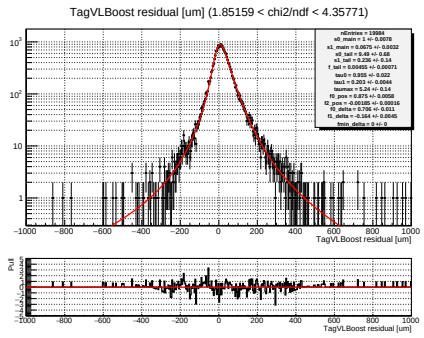
Figure 5.12: $\delta\ell^{\mathrm{tag}}$ distribution of signal MC events shown in bins of $\sigma_\ell^{\mathrm{tag}}$. $R_{\mathrm{asc}} \otimes R_{\mathrm{np}}$ is overlaid to the distribution, whose parameters are fixed by the fit shown in Fig. 5.11. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.



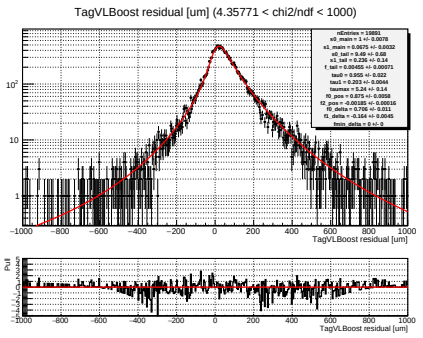
(a) $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}} < 0.95$



(b) $0.95 < \left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}} < 1.85$



(c) $1.85 < \left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}} < 4.36$



(d) $4.36 < \left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$

Figure 5.13: $\delta\ell^{\mathrm{tag}}$ distribution of signal MC events shown in bins of $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$. $R_{\mathrm{asc}} \otimes R_{\mathrm{np}}$ is overlaid to the distribution, whose parameters are fixed by the fit shown in Fig. 5.11. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.

Table 5.6: Tag-side resolution function parameters fixed by the fit to signal MC shown in Fig. 5.11

parameter	value
$s_{\mathrm{main}}^{\mathrm{tag},0}$	1.0000 ± 0.0078
$s_{\mathrm{main}}^{\mathrm{tag},1}$	0.0675 ± 0.0032
$s_{\mathrm{main}}^{\mathrm{tag},0}$	9.49 ± 0.68
$s_{\mathrm{main}}^{\mathrm{tag},1}$	0.24 ± 0.14
$f_{\mathrm{tail}}^{\mathrm{tag}}$	0.00455 ± 0.00071
τ^0	0.955 ± 0.022
τ^1	0.2025 ± 0.0044
τ^{max}	5.24 ± 0.14
f_{p}^0	0.8748 ± 0.0058
f_{p}^1	$-0.00185 \pm 0.00016\,\mu\mathrm{m}^{-1}$
f_{δ}^0	0.706 ± 0.011
f_{δ}^1	-0.164 ± 0.045

5.3.4 Lifetime fit

To confirm that the resolution function correctly describes the Δt resolution in MC we extract B^0 lifetime using about 60,000 TD events from the signal MC sample, where the $\left(\frac{\chi^2}{ndf}\right)^{CP}$ selection is loose: $\left(\frac{\chi^2}{ndf}\right)^{CP} < 100$. We calculate Δt-like variables using MC information to incorporate the resolution effects step by step: $\Delta t^{\rm true}$, $\frac{\ell_{CP}^{\rm true}-\ell_{\rm tag}^{\rm true}}{\beta\gamma c}$, $\frac{\ell_{CP}-\ell_{\rm tag}^{\rm true}}{\beta\gamma c}$, $\frac{\ell_{CP}^{\rm true}-\ell_{\rm tag}}{\beta\gamma c}$, and Δt. For each variable, B^0 lifetime is determined by maximizing the appropriate PDF:

$$P\left(\Delta t^{\rm true}\right) = E_{\rm f}, \tag{5.45}$$

$$P\left(\frac{\ell_{CP}^{\rm true}-\ell_{\rm tag}^{\rm true}}{\beta\gamma c}\right) = E_{\rm f}\otimes R_{\rm k}, \tag{5.46}$$

$$P\left(\frac{\ell_{CP}-\ell_{\rm tag}^{\rm true}}{\beta\gamma c}\right) = E_{\rm f}\otimes R_{\rm k}\otimes R_{\rm rec}, \tag{5.47}$$

$$P\left(\frac{\ell_{CP}^{\rm true}-\ell_{\rm tag}}{\beta\gamma c}\right) = E_{\rm f}\otimes R_{\rm k}\otimes R_{\rm asc}\otimes R_{\rm np}, \text{ and} \tag{5.48}$$

$$P\left(\Delta t\right) = E_{\rm f}\otimes R_{\rm k}\otimes R_{\rm rec}\otimes R_{\rm asc}\otimes R_{\rm np}. \tag{5.49}$$

Table 5.7 summarizes the results. The lifetimes agree with each other, which indicates the resolution function works appropriately to take into account each effect.

However, the overall results seem slightly shorter than the MC input 1.519 ps. We consider this to be a bias that occurs owing to K_S^0 BDT selection in the following mechanism. The BDT relies on the variables about the flight length and direction of K_S^0 which assumes that K_S^0 is produced at the IP. When B^0 flies for long distance, the assumption is violated and the K_S^0 selection efficiency slightly gets worse. We leave the bias as it is so small and does not affect the *CP* violation measurement.

Table 5.7: Results of lifetime fits to signal MC events using Δt variables with partial resolution effects

variable	τ_{B^0} [ps]
$\Delta t^{\rm true}$	1.509 ± 0.006
$\frac{\ell_{CP}^{\rm true}-\ell_{\rm tag}^{\rm true}}{\beta\gamma c}$	1.510 ± 0.006
$\frac{\ell_{CP}-\ell_{\rm tag}^{\rm true}}{\beta\gamma c}$	1.507 ± 0.007
$\frac{\ell_{CP}^{\rm true}-\ell_{\rm tag}}{\beta\gamma c}$	1.517 ± 0.007
Δt	1.513 ± 0.008

5.4 Background Δt distribution

Considering the background events are dominated by $q\overline{q}$ events, we regard the true Δt distribution for them is similar to a delta function so define the Δt PDF in a similar form as $R_{\rm rec}$ and $R_{\rm asc}$:

$$P_{\rm bkg}(\Delta t; X^2, \sigma_{\Delta t}) = (1 - f_{\rm tail}^{\rm bkg}) G\left(\Delta t; \mu = \mu_{\rm main}^{\rm bkg}, \sigma = (s_{\rm main}^{\rm bkg,0} + s_{\rm main}^{\rm bkg,1} X^2)\sigma_{\Delta t}\right) + f_{\rm tail}^{\rm bkg} G\left(\Delta t; \mu = \mu_{\rm tail}^{\rm bkg}, \sigma = (s_{\rm tail}^{\rm bkg,0} + s_{\rm tail}^{\rm bkg,1} X^2)\sigma_{\Delta t}\right), \quad (5.50)$$

where $X^2 \equiv \frac{1}{2}\left(\frac{\chi^2}{ndf}\right)^{CP} + \frac{1}{2}\left(\frac{\chi^2}{ndf}\right)^{\rm tag}$, $\sigma_{\Delta t} \equiv \sqrt{(\sigma_\ell^{CP})^2 + (\sigma_\ell^{\rm tag})^2}/(\beta\gamma)$ is the Δt uncertainty, and the function has seven free parameters, $\left\{ f_{\rm tail}^{\rm bkg}, \mu_{\rm main(tail)}^{\rm bkg}, s_{\rm main(tail)}^{\rm bkg,0(1)} \right\}$.

We determine the background Δt PDF parameters by a fit to sideband data $M_{\rm bc} < 5.265\,{\rm GeV}/c^2$. The fit results are shown in Fig. 5.14 and the parameters are in Tab. 5.8. We compare the background Δt distributions within different $M_{\rm bc}$ ranges in the MC as shown in Fig. 5.15. No visible correlation between Δt and $M_{\rm bc}$ supports the extrapolation of the Δt distribution in the $M_{\rm bc}$ sideband to the signal region.

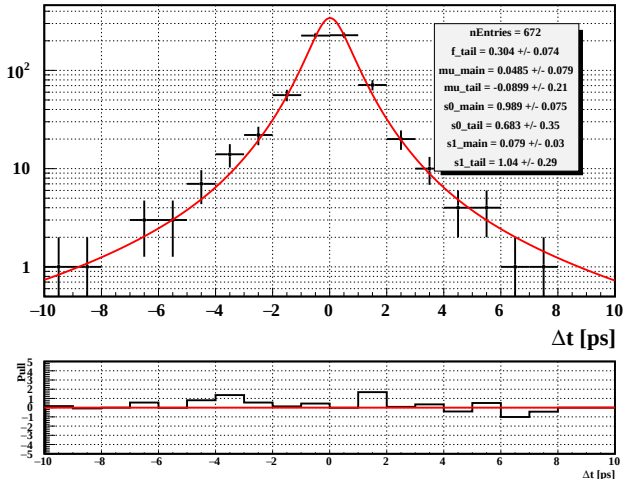


Figure 5.14: Δt distribution of sideband data and fitted $P_{\rm bkg}(\Delta t)$. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.

Table 5.8: Background Δt PDF parameters

parameter	value
$f_{\rm tail}^{\rm bkg}$	0.304 ± 0.074
$\mu_{\rm main}^{\rm bkg}$	0.049 ± 0.079
$\mu_{\rm tail}^{\rm bkg}$	-0.09 ± 0.21
$s_{\rm main}^{\rm bkg,0}$	0.989 ± 0.075
$s_{\rm tail}^{\rm bkg,0}$	0.68 ± 0.35
$s_{\rm main}^{\rm bkg,1}$	0.079 ± 0.030
$s_{\rm tail}^{\rm bkg,1}$	1.0 ± 0.29

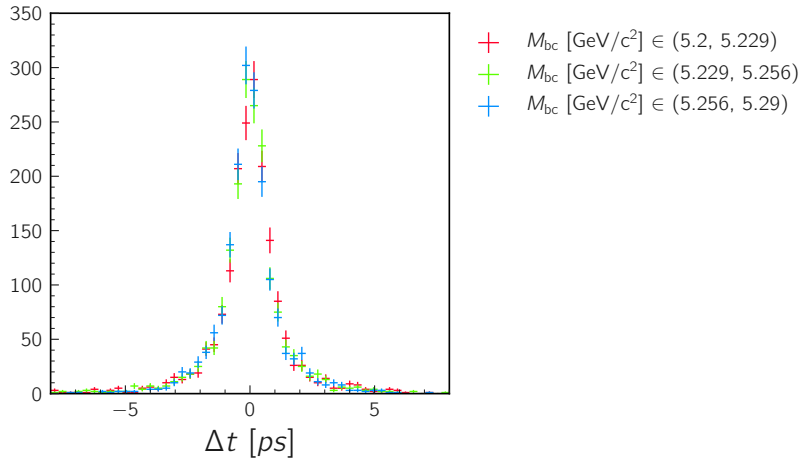


Figure 5.15: Δt distributions of background MC events shown in different bins of $M_{\rm bc}$

5.5 Validation

We perform various tests of the analysis procedure as follows:

- ensemble test
  to examine the fit program,
- linearity test of *CP* fit using high-statistics signal MC
  to examine the resolution function and flavor tagger output,
- lifetime fit and *CP* fit using 700 fb^{-1} generic MC
  to examine the PDF models for signal fraction and to spot overlooked effects such as unknown background components, correlations between variables (Punzi effect [40]), etc.,
- lifetime fit and *CP* fit using control sample data $B^+ \to K^0_S K^0_S K^+$
  to spot a bias due to data-MC difference of resolution function, flavor tagger, PDF shapes etc.,
- lifetime fit using $B^0 \to K^0_S K^0_S K^0_S$ data
  to spot a bias due to data-MC difference of resolution function, PDF shapes etc.

5.5.1 Ensemble test

We generate "toy MC" samples to perform an ensemble test of the fitter in the following procedure:

1. choose input *CP* asymmetries $(S^{\rm input}, A^{\rm input})$
2. fix the equivalent statistics and calculate the expected number of events for each event type (signal/background and TD/TI) by scaling the yields obtained in MC

3. fix randomly the observed number of signal and background events based on the Poisson statistics and the expected values

4. generate randomly a set of variables ($M_{\rm bc}$, M, $\mathcal{O}'_{CS}$, $\cos\theta^*_B$, r bin, q, Δt, $\left(\frac{\chi^2}{ndf}\right)^{CP}$, σ_ℓ^{CP}, $\left(\frac{\chi^2}{ndf}\right)^{\rm tag}$, $\sigma_\ell^{\rm tag}$) for each event based on the PDF of corresponding event type. Here, χ^2/ndf and σ_ℓ are sampled from the MC distributions: we sort the values in the MC sample, randomly choose two consecutive values, and take a random value within the interval between the two.

The toy MC samples are generated in the following configurations:

- expected number of signal and background events are set to $200\,{\rm fb}^{-1}$ equivalent or $2\,{\rm ab}^{-1}$, and
- *CP* asymmetries are varied:
  $(S_{\rm input}, A_{\rm input}) = (-1,0), (-0.8,0), ..., (+1,0),$ and $(0,-1), (0,-0.8), ...(0,+1)$.

For each configuration we generate 10,000 samples for $200\,{\rm fb}^{-1}$ equivalent and 500 samples for $2\,{\rm ab}^{-1}$. We assume flavor symmetry of the flavor tagging efficiency by setting $\mu = 0$ in all r bins in the event generation.

We estimate S and A for each experiment and fit a Gaussian function to the pull distribution of them for each configuration. Here, the pull is defined as the difference of the estimated value from the true value divided by the estimation uncertainty: $S_{\rm pull} \equiv (S_{\rm fit} - S_{\rm input})/\delta S$, $A_{\rm pull} \equiv (A_{\rm fit} - A_{\rm input})/\delta A$. Figure 5.16 show the mean and standard deviation of the fitted Gaussian functions as a function of input *CP* asymmetries. With the larger statistics, the pull distributions are consistent with the standard normal distribution.

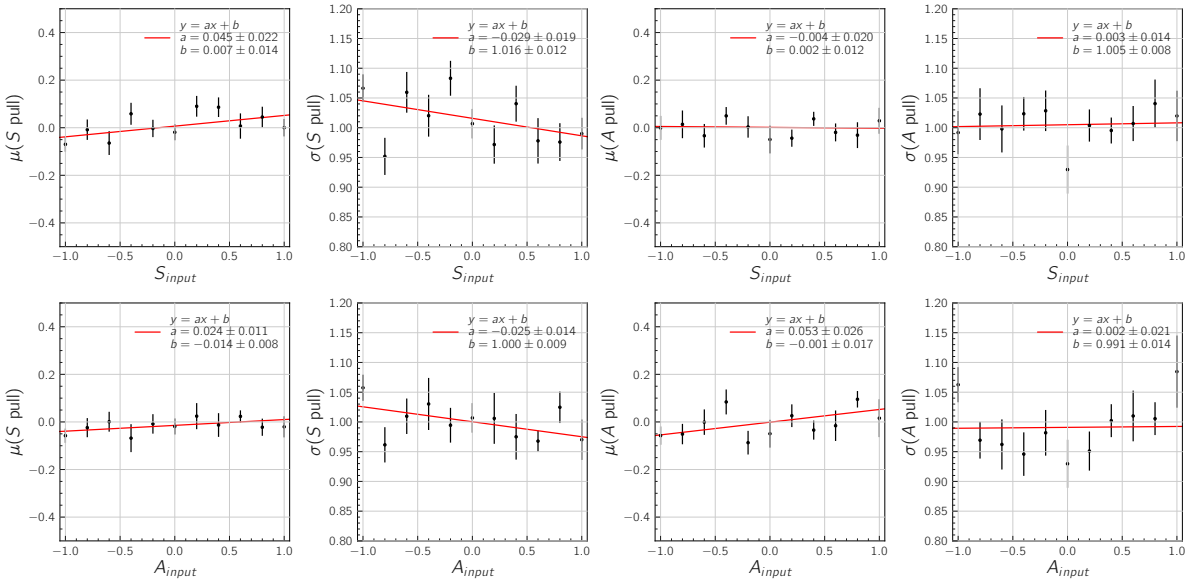
With the lower statistics, we observe a bias in the pull mean especially for configurations with large *CP* violation. We also find the underestimation of fit uncertainty as the pull width larger than unity. We consider this bias to be a specific issue with low statistics as it is not significant in the $2\,{\rm ab}^{-1}$ sample. We use a frequentist approach [41] [42] to estimate the confidence interval for the *CP* asymmetries, taking into account the bias (See Appendix G).

5.5.2 Linearity test of *CP* fit using signal MC

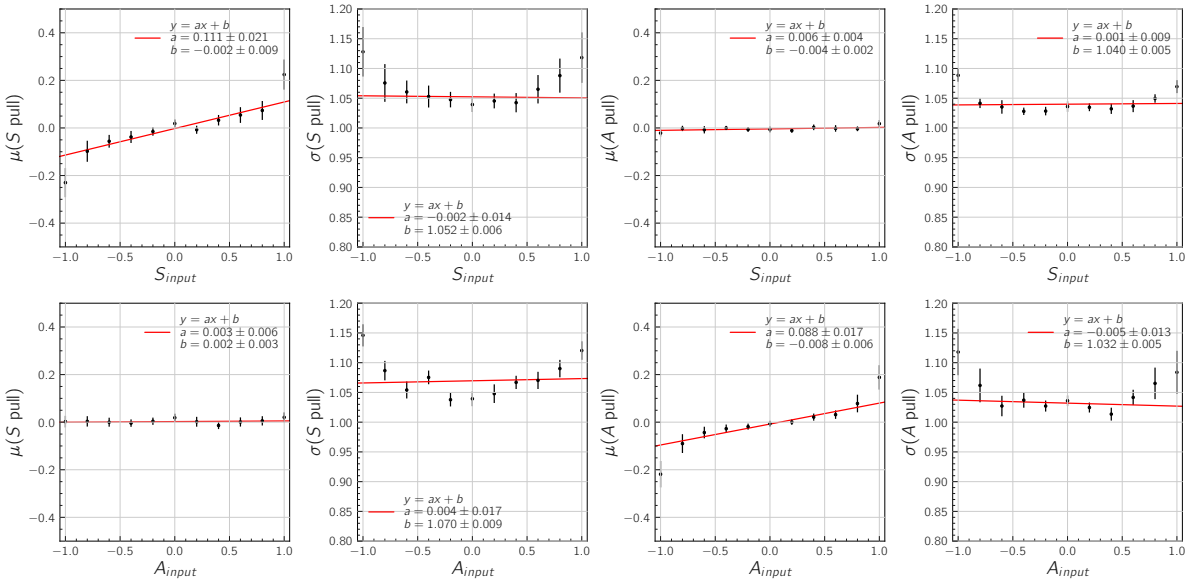
We generate signal MC samples with nonzero *CP* asymmetries to test the linearity of the *CP* fit results. For each of the following parameter sets 20000 events are simulated, which is roughly equivalent to $10\,{\rm ab}^{-1}$:

$$(S_{\rm input}, A_{\rm input}) = (-1,0), (-0.8,0), ..., (+1,0), \text{and } (0,-1), (0,-0.8), ...(0,+1). \quad (5.51)$$

We determine *CP* asymmetries for each sample, using only the signal events. Figure 5.17 shows the fit results. The good linearity indicates no issue in the resolution function or flavor tagger that would dilutes or exaggerates *CP* violation.



(a) $2\,\mathrm{ab}^{-1}$



(b) $200\,\mathrm{fb}^{-1}$

Figure 5.16: Results of ensemble test for *CP* fit. For $2\,\mathrm{ab}^{-1}$-equivalent (a) and $200\,\mathrm{fb}^{-1}$-equivalent (b) toy MC samples, the mean (μ) and standard deviation (σ) of pull distribution of the fitted *CP* asymmetries are shown as functions of input *CP* asymmetries. The error bars indicate the uncertainty of Gaussian fits to the pull distributions. A linear function is fitted to each plot by the method of least squares.

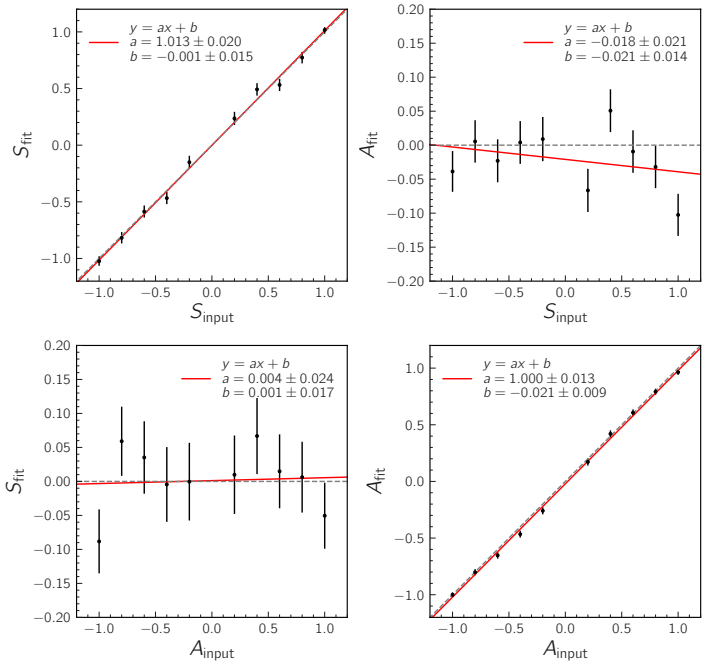


Figure 5.17: Results of linearity test of *CP* fit. Each data point shows the result of *CP* fit using signal MC generated with the input *CP* asymmetry in the horizontal axis. The error bar indicates the uncertainty of each *CP* fit. A linear function is fitted to each plot by the method of least squares. In the upper (lower) plots $A_{\rm input}$ ($S_{\rm input}$) is fixed to be zero while $S_{\rm input}$ ($A_{\rm input}$) is varied.

However, a small offset is seen in the $A_{\rm fit}$ vs $A_{\rm input}$ plot. We attribute the offset to the small asymmetry of the flavor tagging efficiency μ. Such an offset is not seen in the toy MC study with $\mu = 0$ described in the previous subsection. The bias due to μ is considered as a systematic uncertainty.

5.5.3 Lifetime fit and *CP* fit using 700 fb^{-1} generic MC

We apply the whole analysis procedure using the 700 fb^{-1} generic MC sample as follows:

1. perform signal extraction fit on $B^+ \to K^0_S K^0_S K^+$ sample to determine $F_{\rm sig}$ and background PDF shapes for $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$,
2. perform signal extraction fit to $B^0 \to K^0_S K^0_S K^0_S$ sample using the background PDFs,
3. obtain background r bin PDF from $B^+ \to K^0_S K^0_S K^+$ sideband,
4. perform background Δt fit to $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$ sideband, and
5. perform lifetime fit and *CP* fit to $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$.

Relevant distributions are shown in Appendix I. The resulted lifetimes and *CP* asymmetries are consistent with the MC input as shown in Tab. 5.10. $F_{\rm sig}$ values shown in Tab. 5.9 are also consistent with the true value calculated from MC information.

Note that the branching fraction of $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$ are almost doubled by a mistake in the generic MC sample, so $F_{\rm sig}$ is larger than what we expect in the real data. We also eliminate $B \to X(\to K^0_S K^0_S) K (X = \eta_c, \eta_c(2S), J/\psi)$ decays, which are prohibited but exist in the MC, from the sample beforehand.

Table 5.9: Results of signal extraction fits to generic MC. $N^{\rm true}_{\rm sig}$ and $N^{\rm true}_{\rm bkg}$ is the number of signal (true $B^+ \to K^0_S K^0_S K^+$ events for $B^+ \to K^0_S K^0_S K^+$ sample) and background events in the fit region. $F^{\rm true}_{\rm sig}$ is defined as $N^{\rm true}_{\rm sig}/(N^{\rm true}_{\rm sig} + N^{\rm true}_{\rm bkg})$.

mode	parameter	fitted value	$F^{\rm true}_{\rm sig}$	$N^{\rm true}_{\rm sig}$	$N^{\rm true}_{\rm bkg}$
$B^+ \to K^0_S K^0_S K^+$	$a^{M_{\rm bc}}_{\rm bkg}$	-28.07 ± 0.52			
	$c^{M}_{\rm bkg}$	$-0.2473^{+0.0077}_{-0.0076}$			
	$f^{\mathcal{O}'_{CS}}_{\rm 1,bkg}$	$0.567^{+0.081}_{-0.125}$			
	$\mu^{\mathcal{O}'_{CS}}_{\rm 1,bkg}$	$-1.470^{+0.084}_{-0.137}$			
	$\sigma^{\mathcal{O}'_{CS}}_{\rm 1,bkg}$	$1.558^{+0.075}_{-0.127}$			
	$\mu^{\mathcal{O}'_{CS}}_{\rm 2,bkg}$	$0.78^{+0.22}_{-0.32}$			
	$\sigma^{\mathcal{O}'_{CS}}_{\rm 2L,bkg}$	$3.15^{+0.17}_{-0.22}$			
	$\sigma^{\mathcal{O}'_{CS}}_{\rm 2R,bkg}$	$1.314^{+0.089}_{-0.073}$			
	$F_{\rm sig}$	0.03107 ± 0.00085	0.0300	1579	50996
$B^0 \to K^0_S K^0_S K^0_S$ TD event	$F_{\rm sig}$	$0.1094^{+0.0045}_{-0.0044}$	0.1065	585	4908
$B^0 \to K^0_S K^0_S K^0_S$ TI event	$F_{\rm sig}$	$0.0594^{+0.0042}_{-0.0041}$	0.0529	232	4150

Table 5.10: Results of lifetime and *CP* fits to generic MC. MC input values are shown in parentheses.

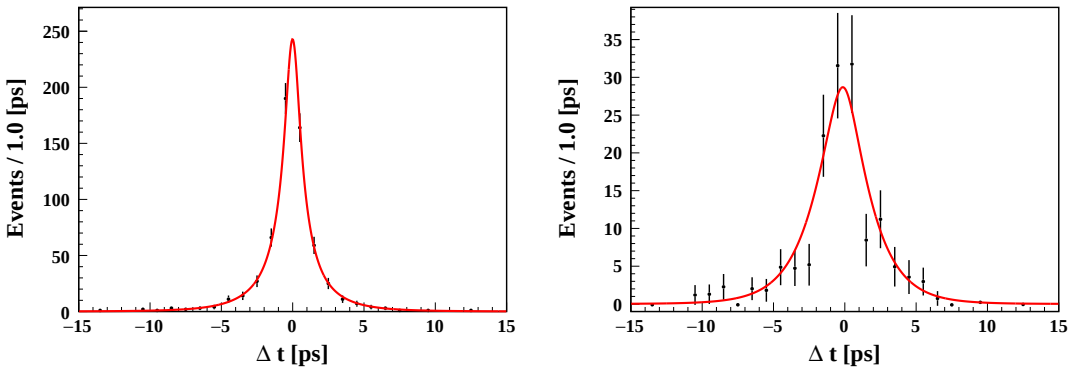
mode	τ_B [ps]	S	A
$B^+ \to K^0_S K^0_S K^+$	$1.586^{+0.052}_{-0.050}$	-0.014 ± 0.088	fixed at 0
	(1.638)	(0)	(0)
$B^0 \to K^0_S K^0_S K^0_S$	$1.435^{+0.075}_{-0.073}$	0.05 ± 0.16	0.001 ± 0.087
	(1.519)	(0)	(0)

5.5.4 Lifetime fit and *CP* fit using control sample

We extract B^+ lifetime and ostensible mixing-induced *CP* asymmetry from the $B^+ \to K^0_S K^0_S K^+$ data by separate fits. The fit results are as expected and exhibit no issue in the analysis procedure.

The extracted lifetime,

$$\tau_{B^+} = 1.53^{+0.19}_{-0.17}\,{\rm ps}, \tag{5.52}$$



(a) Δt distribution and fitted PDF

(b) ${}_s\mathcal{P}lot$ for signal and signal PDF

Figure 5.18: Results of lifetime fit to $B^+ \to K^0_S K^0_S K^+$ data. The left plot shows the Δt distribution and the fitted PDF. The right plot shows the signal component of Δt distribution using the ${}_s\mathcal{P}lot$ technique and the signal PDF.

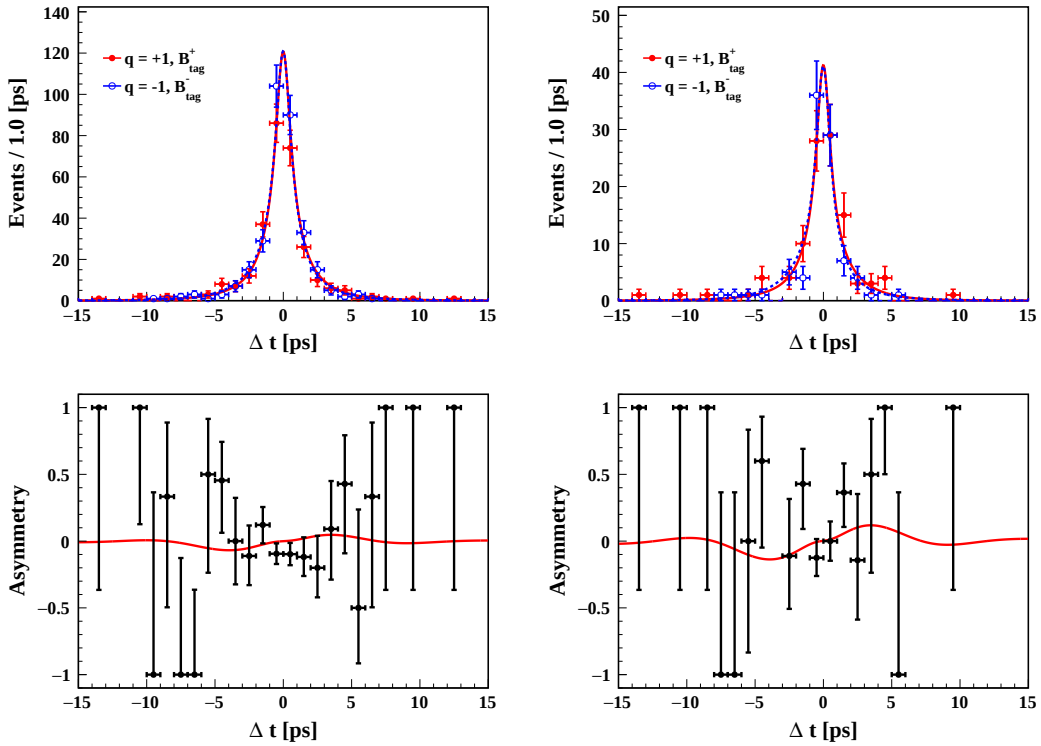
is consistent with the world average 1.638±0.004 ps[7]. Figure 5.18 shows the Δt distribution and the distribution for the signal component extracted with the ${}_s\mathcal{P}lot$ technique [43]. The ${}_s\mathcal{P}lot$ is a technique to reweight the data sample to decompose its distribution of a control variable using the information of variables that discriminate the components. We calculate the weight (sWeight) for the ${}_s\mathcal{P}lot$ using $M_{\rm bc}$, M, $\mathcal{O}'_{CS}$, r bin, and $\cos\theta^*_B$ as the discriminating variables. The ${}_s\mathcal{P}lot$ is shown only to visualize the signal component and is not used for any analysis.

We then perform a *CP* fit to determine S while fixing A to be zero. Because charged B mesons do not mix with each other, the ostensible mixing-induced asymmetry is expected to be zero. We fix the lifetime and mixing parameters to be $\tau_{B^+} = 1.638\,\text{ps}$ and $\Delta m_d = 0.507\,\text{ps}^{-1}$ in the fit. We fix w and Δw to the values determined by Ref. [1].

The obtained value,

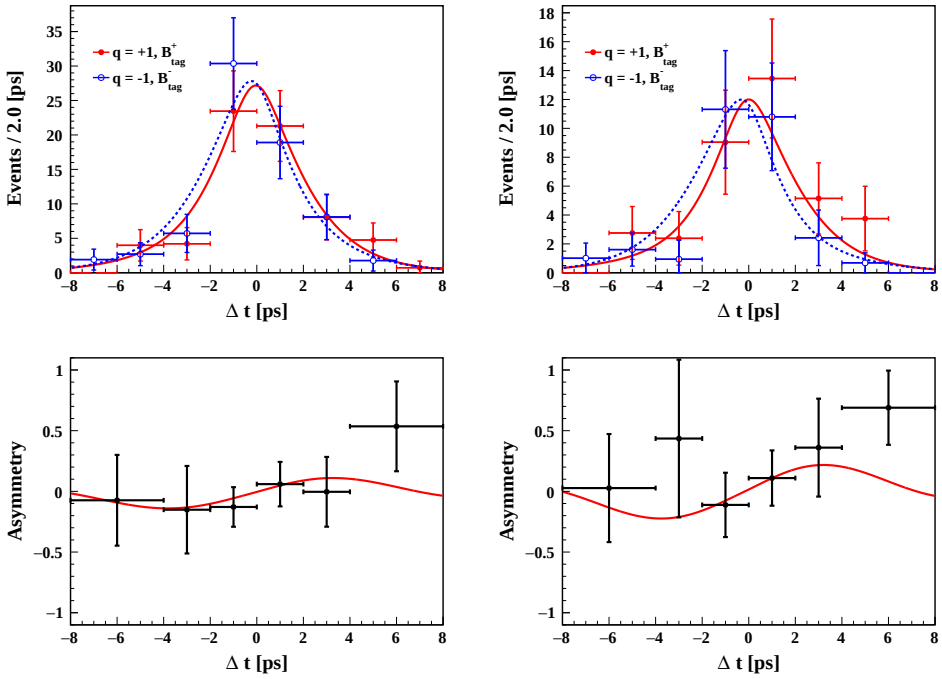
$$S = 0.37^{+0.31}_{-0.33}, \tag{5.53}$$

is consistent with null asymmetry in 1.1 σ. Figure 5.19 shows the Δt distributions of $q = +1$ and $q = -1$ events and their asymmetry as well as the fitted PDF for each flavor and the asymmetry of the PDFs. The error bars of the asymmetry plot show the Clopper-Pearson interval of 68% confidence level. Figure 5.20 shows the ${}_s\mathcal{P}lot$ for the signal component and signal PDFs. We calculate the error bars of the asymmetry plot by propagating the uncertainty of sWeights.



(a) no selection on r (b) $r > 0.5$

Figure 5.19: Result of *CP* fit to $B^+ \to K^0_S K^0_S K^+$ data (1). The upper plots show the Δt distribution and fitted PDF separately for $q = +1$ and $q = -1$ events. The lower plots show the asymmetry of them between the different flavors. The right plots show only the events fulfilling $r > 0.5$.



(a) no selection on r (b) $r > 0.5$

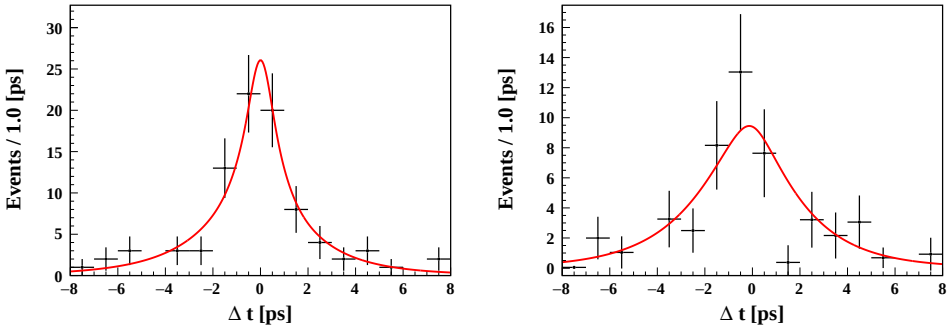
Figure 5.20: Result of *CP* fit to $B^+ \to K^0_S K^0_S K^+$ data (2). The upper plots show the signal component of Δt distribution extracted with the ${}_s\mathcal{P}lot$ technique and signal PDF separately for $q = +1$ and $q = -1$ events. The lower plots show the asymmetry of them between the different flavors. The right plots show only the events fulfilling $r > 0.5$.

5.5.5 Lifetime fit using $B^0 \to K^0_S K^0_S K^0_S$ data

We determine the B^0 lifetime using the $B^0 \to K^0_S K^0_S K^0_S$ data to be

$$\tau_{B^0} = 1.90^{+0.35}_{-0.29}\,\text{ps}, \tag{5.54}$$

which is consistent with the world average $1.519 \pm 0.004\,\text{ps}$ at 1.3σ level. The lifetime fit results are shown in Fig. 5.21.



(a) Δt distribution and fitted PDF (b) ${}_s\mathcal{P}lot$ for signal and signal PDF

Figure 5.21: Results of lifetime fit to $B^0 \to K^0_S K^0_S K^0_S$ data. The left plot shows the Δt distribution and the fitted PDF. The right plot shows the signal component of Δt distribution using the ${}_s\mathcal{P}lot$ technique and the signal PDF.

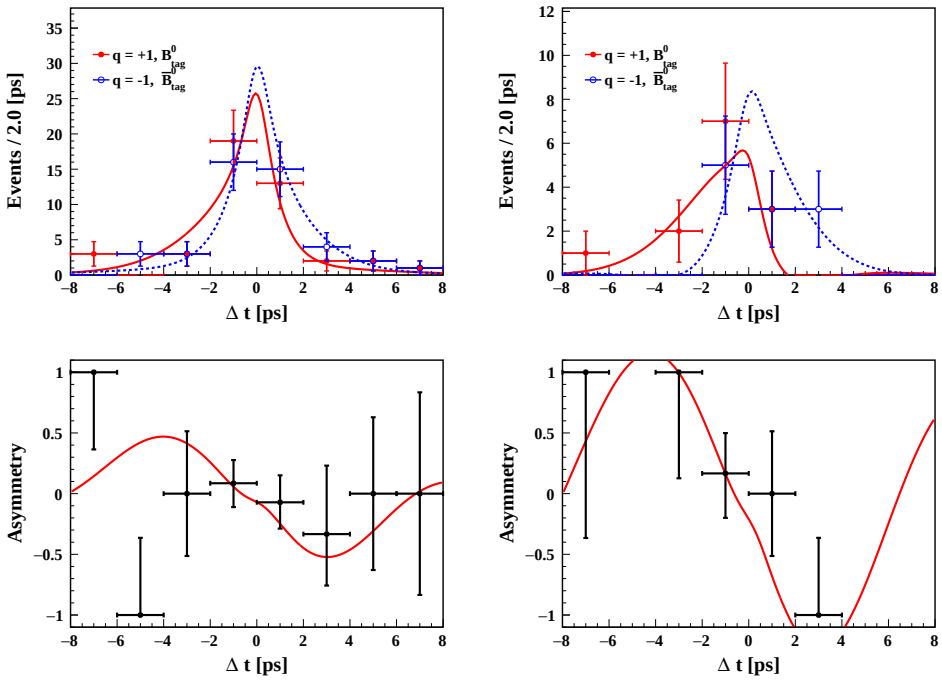
5.6 Fit results

We determine the CP asymmetries in $B^0 \to K^0_S K^0_S K^0_S$ data as

$$S = -1.86^{+0.60}_{-0.52}(\text{MINOS}), \text{ and} \tag{5.55}$$
$$A = -0.22^{+0.22}_{-0.21}(\text{MINOS}), \tag{5.56}$$

where the uncertainties are estimated by MINOS. Figures 5.22 and 5.23 show the fit results in the same manner as in Sec. 5.5.4. Only TD events are shown in the plots. Figure 5.24 shows the negative log likelihood function normalized by its minimum value $-2\log(L/L_{\text{min}})$ as functions of S and A.

Although the central value is far beyond the physical boundary of Eq. (1.41), the results are not extremely unlikely. Performing pseudo experiments on the assumption of $(S_{\text{input}}, A_{\text{input}}) = (-0.7, 0)$, which is expected in the SM, we find that the probability to obtain a result less frequent than our result, $P(S_{\text{fit}}, A_{\text{fit}}) < P(S_{\text{fit}} = -1.86, A_{\text{fit}} = -0.22)$, is 15%. Here, $P(S_{\text{fit}}, A_{\text{fit}})$ is the PDF of S and A obtained by the pseudo experiments and defined as $P(S_{\text{fit}}, A_{\text{fit}}) = P(S_{\text{fit}}, A_{\text{fit}}|S_{\text{input}} = -0.7, A_{\text{input}} = 0)$ using $P(S_{\text{fit}}, A_{\text{fit}}|S_{\text{input}}, A_{\text{input}})$ given in Appendix G. The reason of such a large value is that there are no high-r events



(a) no selection on r (b) $r > 0.5$

Figure 5.22: Result of CP fit to $B^0 \to K^0_S K^0_S K^0_S$ data (1). The plots contain only TD events. The upper plots show the Δt distribution and fitted PDF separately for $q = +1$ and $q = -1$ events. The lower plots show the asymmetry between the different flavors. The right plots show only the events fulfilling $r > 0.5$.

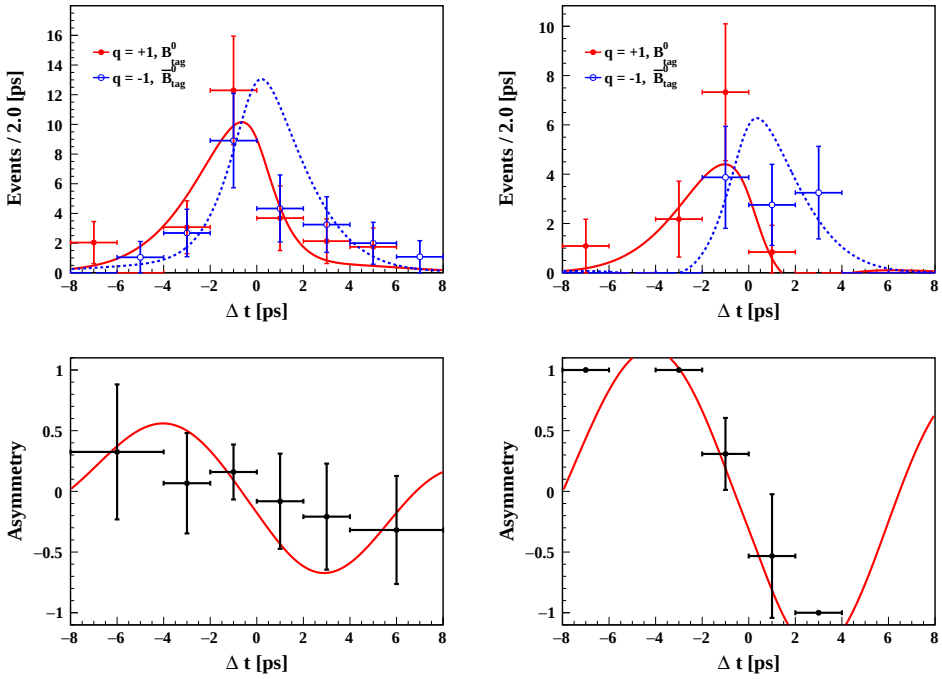
with unfavored flavor in the region where the asymmetry curve oscillates the largest ($|\Delta t| \sim 4\,\mathrm{ps}$). We determine the most probable values within the physical region to be $S = -0.98$ and $A = -0.18$ by maximizing a likelihood function of S and A obtained from toy MC samples (see Appendix G).

As we find the underestimation of the fit uncertainty from the ensemble test, we estimate statistical uncertainties in a parametric bootstrap method. We generate pseudo experiments using the most probable values as input. Figure 5.25 shows the distributions of S and A obtained in the pseudo experiments. We define the distance between the input value and 18 (84) percentile of the distribution as the upper (lower) uncertainty. With this method, we update the result to be

$$S = -1.86^{+0.91}_{-0.46}, \text{ and} \tag{5.57}$$

$$A = -0.22^{+0.30}_{-0.27}. \tag{5.58}$$

If we use only TD events in the fit, we obtain $S = -1.84^{+0.60}_{-0.51}$ and $A = -0.07^{+0.30}_{-0.28}$, where the uncertainties are estimated by MINOS. Comparing the MINOS uncertainties to Eqs. (5.55) and (5.56), we find the inclusion of TI events improves the uncertainty for A by a factor 0.75.



(a) no selection on r (b) $r > 0.5$

Figure 5.23: Result of *CP* fit to $B^0 \to K^0_S K^0_S K^0_S$ data (2). The upper plots show the signal component of Δt distribution extracted with the ${}_s\mathcal{P}lot$ technique and signal PDF separately for $q = +1$ and $q = -1$ events. The plots contain only TD events. The lower plots show the asymmetry of them between the different flavors. The right plots show only the events fulfilling $r > 0.5$.

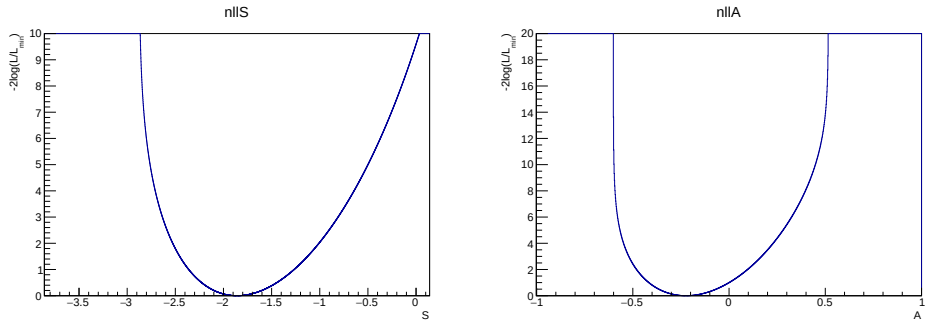


Figure 5.24: Negative log likelihood function normalized by its minimum value as functions of S (left) and A (right). A (S) is fixed to the optimal value in Eq. (5.56)((5.55)) in the left (right) plot.

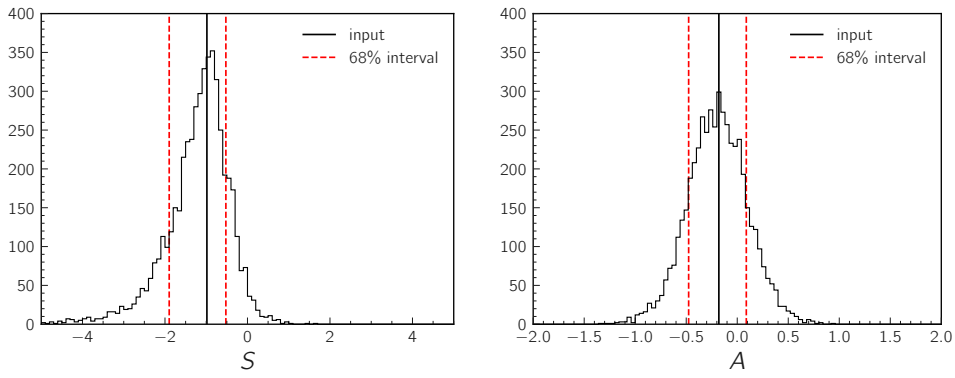


Figure 5.25: Distribution of S (left) and A (right) in the pseudo experiments using the most probable values of S and A as input. The black solid line is the input value. The red dashed lines are 18 and 84 percentiles of the distribution.

5.7 Systematic uncertainties

Tab. 5.11 summarizes the systematic uncertainties in our measurement. We estimate systematic uncertainties in different methods depending on their sources as follows.

- *parameter uncertainty*
  In the *CP* fit, many parameters are fixed by independent measurements on real data or MC. To take into account the uncertainty of the parameters, we repeat *CP* fits adding a random Gaussian deviation to these parameters. The standard deviation of the Gaussian is determined by the uncertainty of the corresponding parameter. For a parameter fixed by MC, we conservatively double the Gaussian deviation. We fit a Gaussian function to the distribution of S and A obtained by the repeated *CP* fits and regard the standard deviation of the fitted function as a systematic uncertainty. Another approach is to perform a *CP* fit using an alternative parameter sets obtained by yet another measurement. We take the deviation of S and A from those obtained with the nominal parameters as a systematic uncertainty.
- *reconstruction uncertainty*
  Some sources of uncertainty are in the reconstruction process such as event selection and Δt measurement. We reconstruct the data in different configurations considering the uncertainty and take the deviation of *CP* fit results from the original result.
- *model uncertainty*
  Our measurement relies on the resolution function model that we arbitrarily choose. We perform *CP* fits using alternative models to evaluate the deviation of fit results from the original model.
- *bias due to effects not considered in fit model*
  Some known effects are not considered in our fit model, which may results in bias in the measurement. To evaluate the possible bias, we generate toy MC samples with and without simulating such effects and examine the difference of fit results between the ensembles.

Vertex reconstruction We consider reconstruction uncertainties due to beam parameter calibration, helix uncertainty correction, and vertex quality selection.

Beam parameters We vary the following beam parameters according to the calibration uncertainty: the size, position, and angle of ellipsoid defining the IP profile, and the magnitude and angle of boost vector.

Helix uncertainty correction We use a data sample reconstructed without the correction. *CP* fit is done with the same resolution function parameters as default.

Table 5.11: Systematic uncertainties

source	δS	δA
Vertex reconstruction	0.0255	0.0221
Detector misalignment	0.0022	0.0036
Flavor tagging	0.0789	0.0308
Fit bias	0.0031	0.0017
Physics parameters	0.0080	0.0002
Signal fraction	0.0112	0.0067
Background Δt	0.0106	0.0012
Resolution function parameter	0.0124	0.0057
Resolution function model	0.0011	0.0026
Tag-side interference	0.0139	0.0150
total	0.0868	0.0420

Vertex quality selection We vary the thresholds for the selection about vertex quality as follows:

- $\chi^2/ndf < 100 \to \chi^2/ndf < 80, 120$ for tag side,
- $\chi^2/ndf < 3.27 \to \chi^2/ndf < 3.1, 3.4$ for *CP* side,
- $\sigma_\ell < 500\,\mu\text{m} \to \sigma_\ell < 400, 600\,\mu\text{m}$, and
- $|\Delta t| < 30\,\text{ps} \to |\Delta t| < 20, 50\,\text{ps}$.

The motivation for this item is to evaluate the effect due to possible data-MC difference in the selection variables. Because some events are added or removed to the data sample by the changes, the deviation of the fit results involves also a statistical fluctuation, which is critical for our analysis because of low statistics. The cut on the *CP*-side χ^2 especially gives large statistical fluctuation because the cut threshold lies in the bulk of the distribution, not the tail. We use $B^+ \to K^0_S K^0_S K^+$ data sample instead of $B^0 \to K^0_S K^0_S K^0_S$ to mitigate the statistical fluctuation and quote the results for $B^0 \to K^0_S K^0_S K^0_S$. The statistical fluctuation will be less significant with more statistics in future.

Detector misalignment We consider the reconstruction uncertainty due to detector misalignment. We generate signal decays and simulate the detector response to them using four different configurations of misaligned detector geometry. The detector misalignment can significantly change the Δt uncertainty and thus the event weight for each event, so the deviation in the fit results also involves statistical fluctuation. Therefore we use high-statistics signal MC for the evaluation of this uncertainty to suppress the fluctuation.

Flavor tagging We consider parameter uncertainty of w and Δw and the bias due to μ.

Parameter uncertainty of w and Δw Not only adding Gaussian deviations to the parameters, we also consider a possible time evolution of the flavor tag performance because the flavor tag performance was calibrated using only one third ($63.8\,\mathrm{fb}^{-1}$) of our data set. We use alternative set of parameters determined by a recent measurement of τ_{B^0} and Δm_d to evaluate the time evolution [44]. This is the largest systematic uncertainty in our measurement because of large statistical uncertainty on the measured w and Δw. It will be reduced by a calibration using a larger data set.

Bias due to μ We generate toy MC with random μ within its measured uncertainty and evaluate the bias.

Fit bias We consider the bias from two known correlations between variables, M and $\left(\frac{\chi^2}{ndf}\right)^{CP}$, and $\mathcal{O}'_{CS}$ and r bin. The correlation might result in under- or overestimation of f_{sig} and thus dilution or exaggeration of CP asymmetries. We simulate these effects in toy MC and found the deviation of fit results as a function of input CP asymmetries as shown in Fig. 5.26. We observe a significant slope only for A with correlation between $\mathcal{O}'_{CS}$ and r bin. When we do not observe a significant slope, we take the fit uncertainty as the potential size of systematic uncertainty.

Physics parameters We consider the parameter uncertainties of $\tau_{B^0} = 1.519 \pm 0.004\,\mathrm{ps}$ and $\Delta m_d = 0.5065 \pm 0.0019\,\mathrm{ps}^{-1}$ fixed at the world average [7]. Taking into account the small bias on τ_{B^0} due to our K^0_S BDT selection, we assume a large uncertainty of $0.016\,\mathrm{ps}$ on τ_{B^0}.

Signal fraction We consider the parameter uncertainties of F_{sig} and PDF parameters of M_{bc}, M, $\mathcal{O}'_{CS}$, and r bin.

Background Δt We consider the parameter uncertainty of the background Δt PDF.

Resolution function parameter We consider the parameter uncertainties of resolution function parameters.

Resolution model We define alternative resolution function models for which we turn off the χ^2/ndf dependence by setting $s^{CP,1}_{\mathrm{main}} = s^{CP,1}_{\mathrm{tail}} = s^{\mathrm{tag},1}_{\mathrm{main}} = s^{\mathrm{tag},1}_{\mathrm{tail}} = f^1_\delta = \tau^1 = 0$ and introduce mean parameters $\mu^{\mathrm{tag}}_{\mathrm{main}}$ and $\mu^{\mathrm{tag}}_{\mathrm{tail}}$ to the Gaussian functions of R_{asc}. Therefore

the alternative functions are expressed as:

$$R'_{\rm rec}(\delta\ell^{CP};\sigma_\ell^{CP}) = (1-f_{\rm tail}^{CP})G(\delta\ell^{CP};\mu=0,\sigma=s_{\rm main}^{CP,0}\sigma_\ell^{CP}) + f_{\rm tail}^{CP}G(\delta\ell^{CP};\mu=0,\sigma=s_{\rm tail}^{CP,0}\sigma_\ell^{CP}), \quad (5.59)$$

$$R'_{\rm asc}(\delta\ell^{\rm tag};\sigma_\ell^{\rm tag}) = (1-f_{\rm tail}^{\rm tag})G(\delta\ell^{\rm tag};\mu=\mu_{\rm main}^{\rm tag}\sigma_\ell^{\rm tag},\sigma=s_{\rm main}^{\rm tag,0}\sigma_\ell^{\rm tag}) + f_{\rm tail}^{\rm tag}G(\delta\ell^{\rm tag};\mu=\mu_{\rm tail}^{\rm tag}\sigma_\ell^{\rm tag},\sigma=s_{\rm tail}^{\rm tag,0}\sigma_\ell^{\rm tag}), \text{ and} \quad (5.60)$$

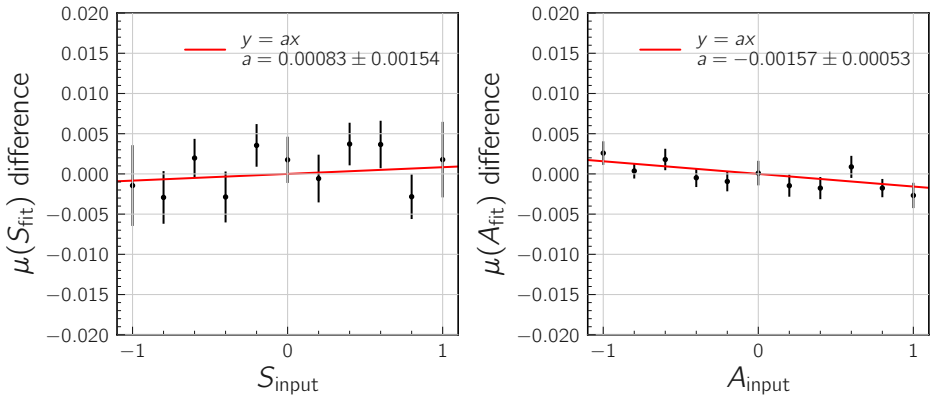
$$R'_{\rm np}(\delta\ell^{\rm tag};\sigma_\ell^{\rm tag}) = f_\delta^0\delta(\delta\ell^{\rm tag})+(1-f_\delta^0)f_{\rm p}E_{\rm p}(\delta\ell^{\rm tag};\tau=\tau^0\sigma_\ell^{\rm tag}) + (1-f_\delta^0)(1-f_{\rm p})E_{\rm n}(\delta\ell^{\rm tag};\tau=\tau^0\sigma_\ell^{\rm tag}), \quad (5.61)$$

where

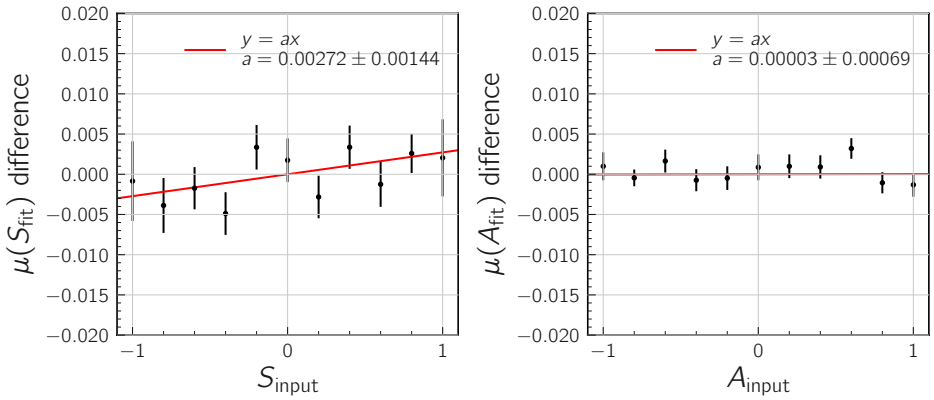
$$f_{\rm p} = \begin{cases} 0 & (f_{\rm p}^0 + f_{\rm p}^1\sigma_\ell^{\rm tag} < 0) \\ f_{\rm p}^0 + f_{\rm p}^1\sigma_\ell^{\rm tag} & (0 \le f_{\rm p}^0 + f_{\rm p}^1\sigma_\ell^{\rm tag} \le 1) \\ 1 & (f_{\rm p}^0 + f_{\rm p}^1\sigma_\ell^{\rm tag} > 1) \end{cases}, \quad (5.62)$$

and the functions have 12 free parameters, $f_{\rm tail}^{CP}$, $s_{\rm main}^{CP,0}$, $s_{\rm tail}^{CP,0}$, $f_{\rm tail}^{\rm tag}$, $\mu_{\rm main}^{\rm tag}$, $s_{\rm main}^{\rm tag,0}$, $\mu_{\rm tail}^{\rm tag}$, $s_{\rm tail}^{\rm tag,0}$, f_δ^0,$f_{\rm p}^0$, $f_{\rm p}^1$, and τ^0. We determined the parameters by fitting $R'_{\rm rec}$ and $R'_{\rm asc} \otimes R'_{\rm np}$ to $\delta\ell^{CP}$ and $\delta\ell^{\rm tag}$ distributions from signal MC as shown in Fig. 5.27.

We simulate 50,000 signal events with CP violation $S = -0.7, A = 0$ to evaluate the model uncertainty.

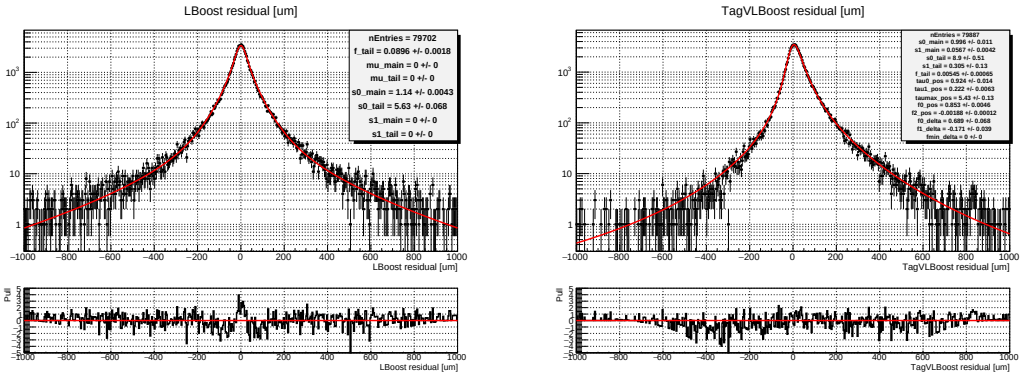


(a) correlation between $\mathcal{O}'_{CS}$ and r bin



(b) correlation between M and χ^2

Figure 5.26: Possible bias due to correlation between $\mathcal{O}'_{CS}$ and r bin (a), and between M and $\left(\frac{\chi^2}{ndf}\right)^{CP}$ (b). For every data point, we generate 8000 toy MC samples using the input CP parameter on the horizontal axis and obtain the mean of fitted CP parameters by fitting Gaussian functions. The toy MC samples are generated with and without the correlation and the difference of the mean between the two ensembles is plotted. In the left (right) plot $A_{\rm input}(S_{\rm input})$ is fixed to be zero while $S_{\rm input}(A_{\rm input})$ is varied. A linear function without an offset is fitted to each plot.



(a) R'_{rec} fitted to MC $\delta\ell^{CP}$ distribution (b) $R'_{\mathrm{asc}} \otimes R'_{\mathrm{np}}$ fitted to MC $\delta\ell^{\mathrm{tag}}$ distribution

Figure 5.27: *CP*-side (a) and tag-side (b) vertex position residual distribution of signal MC events and alternative resolution functions R'_{rec} and $R'_{\mathrm{asc}} \otimes R'_{\mathrm{np}}$. The resolution functions are fitted to the residual distributions. The bottom plot shows the difference of the distribution and the fit curve normalized by the statistical uncertainty.

Tag-side interference We consider the possible bias due to tag-side interference described in Sec. 1.2.2.

We simulate the tag-side interference (TSI) effect in toy MC modifying the Δt PDF of Eq. (3.1) as [45, 8]

$$P(\Delta t, q = \pm 1) = \frac{1}{4\tau_{B^0}} e^{-\frac{|\Delta t|}{\tau_{B^0}}} \frac{1 + |\lambda_f|^2}{2} \big[(1 - S' \cdot C_{\mp}) + (qS - A \cdot S_{\mp}) \sin(\Delta m_d \Delta t) \\ + (qA + S \cdot S_{\mp}) \cos(\Delta m_d \Delta t)) \big], \tag{5.63}$$

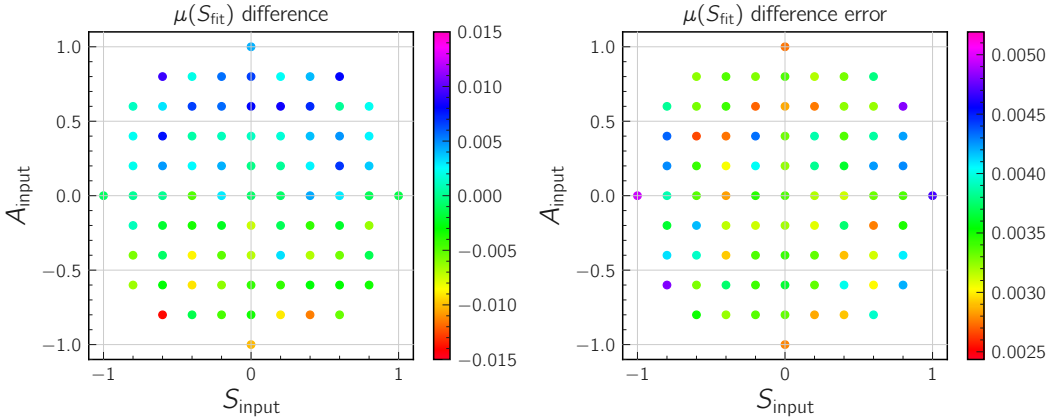
where $S' \equiv 2\mathrm{Re}(\lambda_f)/(1+|\lambda_f|^2)$, $S_{\pm} \equiv 2r' \sin(2\phi_1 + \phi_3 \pm \delta')$, $C_{\pm} \equiv 2r' \cos(2\phi_1 + \phi_3 \pm \delta')$, and r' (δ') is the effective value of amplitude ratio (strong phase difference) of doubly CKM-suppressed transition $b \to u\bar{c}d$ relative to CKM-favored transition $\bar{b} \to \bar{c}u\bar{d}$.

We use the following input values for the simulation:

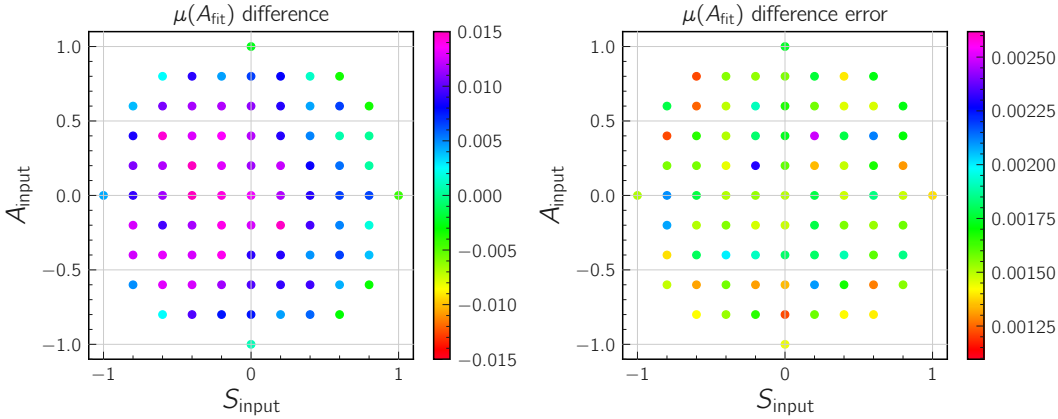
- $\phi_1 = 22.14^{+0.69}_{-0.67}$ degrees [5],
- $\phi_3 = 72.1^{+5.4}_{-5.7}$ degrees [5],
- $S_+ = 0.0096 \pm 0.0073$ [46], and
- $S_- = -0.0067 \pm 0.0073$ [46].

Here we use the measurement of $S_{\pm}$ by the Belle neglecting the difference of flavor tagging efficiency between Belle II and Belle. In each toy MC sample we pick a set of $(\phi_1, \phi_3, S_+, S_-)$ with random Gaussian deviations according to the size of the uncertainties.

Figure 5.28 shows the deviation of fit results as a function of input CP asymmetries and its uncertainty due to limited statistics of the toy MC samples. We conservatively take the largest deviation over the whole physical region as the systematic uncertainty because the statistical uncertainties of S and A are large in the current analysis and the TSI effect should significantly change within the allowed region. In future with more data we will be able to precisely estimate the input parameters and include the TSI effect in the fit model.



(a) Bias of S



(b) Bias of A

Figure 5.28: Bias of *CP* asymmetries due to TSI (left) and its uncertainty (right) shown on the plane of input *CP* asymmetries. For every data point, we generate 8000 toy MC samples without TSI and 5000 with TSI using the corresponding input *CP* parameters, and obtain the mean of fitted *CP* parameter (S in the upper plots, A in the lower plots) by fitting a Gaussian function. The difference of the mean between the two ensembles and its uncertainty are indicated by the color in left and right plots, respectively.

Chapter 6

Discussion

6.1 Significance of the measurement

We determine the *CP* asymmetries in $B^0 \to K^0_S K^0_S K^0_S$ to be

$$S = -1.86^{+0.91}_{-0.46}(\text{stat}) \pm 0.09(\text{syst}), \text{ and} \tag{6.1}$$

$$A = -0.22^{+0.30}_{-0.27}(\text{stat}) \pm 0.04(\text{syst}). \tag{6.2}$$

Figure 6.1 shows the confidence region calculated based on the fit results using a frequentist approach based on the Feldman-Cousins likelihood-ratio ordering [41] [42] (see Appendix G). Here, the systematic uncertainties are also taken into account. Our result is consistent with the SM expectation $(S, A) = (-\sin 2\phi_1, 0) = (-0.7, 0)$ at around 1σ confidence level, and confirms *CP* violation at 96.6%. It is also consistent with the previous measurements by Belle, BaBar, and Belle II [15][16][17]:

$$\left.\begin{aligned} S &= -0.71 \pm 0.23(\text{stat}) \pm 0.05(\text{syst}) \ (1.2\sigma) \\ A &= 0.12 \pm 0.16(\text{stat}) \pm 0.05(\text{syst}) \ (1.0\sigma) \end{aligned}\right\} \text{Belle} \tag{6.3}$$

$$\left.\begin{aligned} S &= -0.94^{+0.24}_{-0.21}(\text{stat}) \pm 0.06(\text{syst}) \ (1.0\sigma) \\ A &= 0.17 \pm 0.18(\text{stat}) \pm 0.04(\text{syst}) \ (1.1\sigma) \end{aligned}\right\} \text{BaBar} \tag{6.4}$$

$$\left.\begin{aligned} S &= -0.82 \pm 0.85(\text{stat}) \pm 0.07(\text{syst}) \ (0.8\sigma) \\ A &= -0.21 \pm 0.28(\text{stat}) \pm 0.06(\text{syst}) \ (0.0\sigma) \end{aligned}\right\} \text{Belle II} \tag{6.5}$$

Here, the values in the parentheses are the difference from our result in the unit of uncertainties summed in quadrature. For the calculation of the summed uncertainties, we ignore the correlation with the previous measurement by Belle II, which is based on a part of our data set.

Let us now interpret the result to derive a model-independent constraint on possible NP contribution. We consider a NP decay amplitude contributing to $B^0 \to K^0_S K^0_S K^0_S$ decay so that Eq. (1.31) becomes

$$\lambda_f = e^{-2i\phi_1} \frac{r_{\text{SM}} e^{-i\phi_{\text{SM}}} e^{i\delta_{\text{SM}}} + r_{\text{NP}} e^{-i\phi_{\text{NP}}} e^{i\delta_{\text{NP}}}}{r_{\text{SM}} e^{i\phi_{\text{SM}}} e^{i\delta_{\text{SM}}} + r_{\text{NP}} e^{i\phi_{\text{NP}}} e^{i\delta_{\text{NP}}}}, \tag{6.6}$$

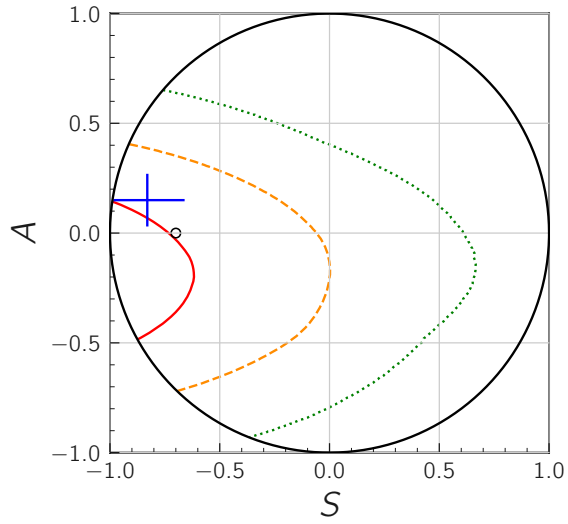


Figure 6.1: Confidence region of S and A from this measurement. The solid, dashed, and dotted lines enclose the region where the true values of S and A in $B^0 \to K^0_S K^0_S K^0_S$ decays should exist with the probability of 68.27%, 95.45%, and 99.73% given the constraint of physical region $S^2 + A^2 \leq 1$. The cross indicates the world average of (S, A) in $B^0 \to K^0_S K^0_S K^0_S$ measured by Belle and BaBar and their uncertainties. The open circle indicates the SM expectation $(S, A) = (-\sin 2\phi_1, 0)$ based on measurements in $B^0 \to (c\bar{c})K^0$[10].

where $r_{SM(NP)}, \phi_{SM(NP)}$ and $\delta_{SM(NP)}$ are the magnitude, weak phase, and strong phase of SM (NP) decay amplitude. Defining $\delta \equiv \delta_{\rm NP} - \delta_{\rm SM}$ and using $\phi_{\rm SM} \simeq 0$, we obtain the constraint on $\{r_{\rm NP}/r_{\rm SM}, \phi_{\rm NP}, \delta\}$ shown in Fig. 6.2. As long as $\sin\phi_{\rm NP}$ is large, the result put a constraint of $r_{\rm NP}/r_{\rm SM} \lesssim 0.15$–$0.25$ with 90% confidence level at best depending on the strong phase difference.

We combine our result with the world average in the Bayesian method, where we define the prior PDF of S and A as a 2D Gaussian function representing the world average, $S = -0.83 \pm 0.17$ and $A = 0.15 \pm 0.12$ with 7% correlation [10], and update it with the likelihood function from our measurement given in Appendix G. Figure 6.3 shows contours of the resulting posterior PDF. The PDF is normalized in the physical region. The central value moves to $S = -0.91$ and $A = 0.09$.

6.2 Comparison with Belle

We compare the performance of the measurement with Belle to evaluate and understand the source of improvement.

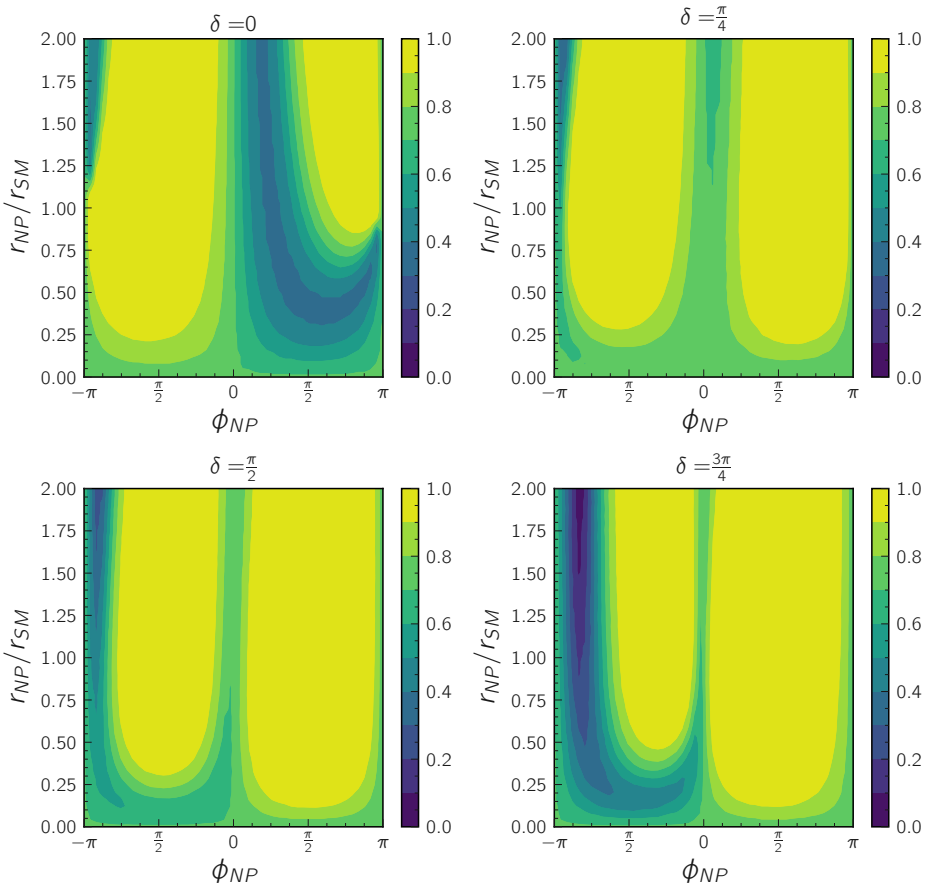


Figure 6.2: Confidence level to exclude parameters of new physics decay amplitude $(r_{\rm NP}/r_{\rm SM}, \phi_{\rm NP}, \delta)$ derived by the measured CP asymmetries in $B^0 \to K^0_S K^0_S K^0_S$ decays. Each plot shows the confidence level on $(r_{\rm NP}/r_{\rm SM})$-$\phi_{\rm NP}$ plane for fixed δ.

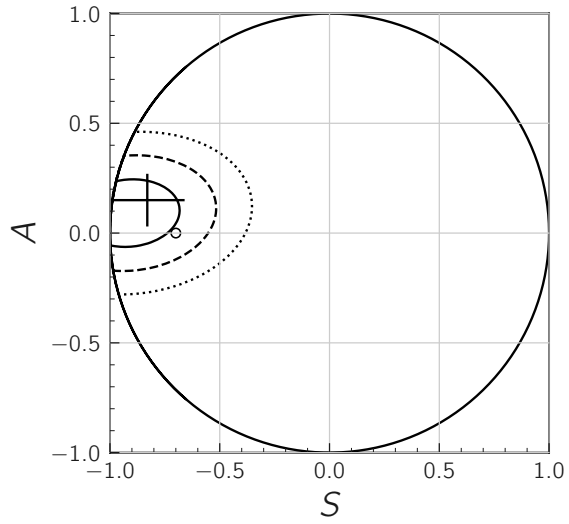


Figure 6.3: Posterior PDF of S and A from the combination of the world average and this measurement. The solid, dashed and dotted lines are contours of equal probability density which enclose 68.27%, 95.45%, and 99.73% of the PDF. The cross indicates the world average of (S, A) in $B^0 \to K^0_S K^0_S K^0_S$ measured by Belle and BaBar and their uncertainties. The open circle indicates the SM expectation $(S, A) = (-\sin 2\phi_1, 0)$ based on measurements in $B^0 \to (c\bar{c})K^0$[10].

Table 6.1: Fraction of signal MC events in Belle [47] and Belle II divided according to $N_{\mathrm{VXDK^0_S}}$

$N_{\mathrm{VXDK^0_S}}$	Belle	Belle II
0	12%	0.4%
1	35%	8%
2	39%	38%
3	14%	54%

6.2.1 VXD acceptance

Δt resolution in $B^0 \to K^0_S K^0_S K^0_S$ decays largely depends on the number of K^0_S's that have associated hits in the vertex detector ($N_{\mathrm{VXDK^0_S}}$). Table 6.1 compares the fraction of signal MC events with $N_{\mathrm{VXDK^0_S}} = 0, 1, 2, 3$ between Belle and Belle II. At Belle II, over 99% of signal events have at least one K^0_S in the VXD and thus potentially provides sufficient Δt resolution for time-dependent *CP* violation analysis. On average, $N_{\mathrm{VXDK^0_S}}$ is 2.46 in Belle II and 1.55 in Belle. We confirm 1.6 times higher K^0_S reconstruction efficiency in the Belle II vertex detector thanks to the enlarged detector volume.

Table 6.2: Comparison of signal reconstruction performance with Belle [15]. $N_{\rm sig}^{\rm TD(TI)}$ is the signal yield for TD(TI) events. Signal yield and purity are calculated within the signal region.

	Belle	Belle II
$N_{B\overline{B}}$ $[10^6]$	772	197
$N_{\rm sig}/N_{B\overline{B}}$ $[10^{-6}]$	0.46	0.51
$N_{\rm sig}^{\rm TD}/N_{B\overline{B}}[10^{-6}]$	0.33	0.27
$N_{\rm sig}^{\rm TI}/N_{B\overline{B}}[10^{-6}]$	-	0.24
purity for TD events	74%	54%
purity for TI events	-	45%

6.2.2 Signal yield

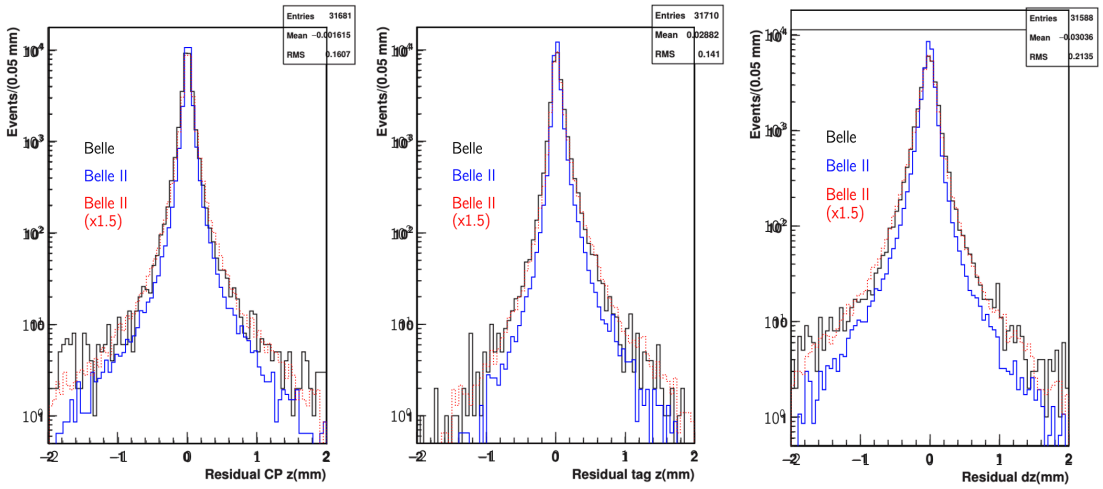
The signal yield and purity in the signal region at Belle and Belle II are shown in Tab. 6.2. The improvement in the signal yield per number of $B\overline{B}$ pair and the degradation in the purity is attributed to the looser continuum suppression selection in our analysis. Our reconstruction efficiency for TD events is slightly worse than Belle because of the tight selection on χ^2 and off-diagonal K_S^0 to assure good validation of Δt measurement performance. At Belle, events with good Δt quality were used for time-dependent fit and the other events were discarded without being used for time-integrated fit. Most of the discarded events were intrinsically unusable for time-dependent analysis because none of K_S^0 in the event had SVD hits. On the other hand, in our case the TI events have potentially good Δt resolution. We expect most of them will be retrieved for time-dependent use in near future by improvements in the analysis procedure (see Appendix. E).

6.2.3 Flavor tag and Δt resolution

As for the flavor tagging performance, the effective tagging efficiencies are measured to be the same within uncertainty between Belle ($30.1 \pm 0.4\%$) and Belle II ($30.0 \pm 1.3\%$) [1]. However, although it is not yet reflected in our analysis, we expect around 1.1 times higher effective efficiency in future analysis thanks to recent updates in the flavor tagger.

Concerning Δt resolution, Belle II should benefit from the larger and finer vertex detector to compensate the degradation due to 1.5 times smaller boost factor. At first we compare the vertex resolution in signal MC selecting good-resolution events that fulfill $N_{\rm VXDK_S^0} \geq 2$. Then, considering the smaller boost factor, we multiply a scale factor of 1.5 to the vertex resolution of Belle II to effectively compare the Δt resolution.

Fig. 6.4 shows the distributions of *CP*-side and tag-side vertex position residual $\delta\ell^{CP}, \delta\ell^{\rm tag}$, and the residual of vertex position difference $\delta\ell^{CP} - \delta\ell^{\rm tag}$ in signal MC of Belle and Belle II. On both *CP* and tag sides, the vertex resolution in Belle II has been clearly improved from Belle. Then comparison of $\ell^{CP} - \ell^{\rm tag}$ distribution between Belle and scaled Belle II indicates that the Δt distribution is almost the same for the selected



(a) *CP*-side vertex resolution (b) tag-side vertex resolution (c) vertex-position-difference resolution

Figure 6.4: Comparison of vertex position resolution between Belle and Belle II [47]. The distributions of *CP*-side (left) and tag-side (center) vertex position residual $\delta\ell^{CP}, \delta\ell^{\text{tag}}$, and the residual of vertex position difference $\delta\ell^{CP} - \delta\ell^{\text{tag}}$ (right) are shown. The red histogram shows the Belle II distribution scaled by 1.5 to take into account the smaller boost factor. Signal MC events fulfilling $N_{\text{VXDK}^0_S} \geq 2$ are shown. In Belle, the boost direction is equivalent to z direction. All distributions are normalized to have the same area.

events. Recalling that the selection $N_{\text{VXDK}^0_S} \geq 2$ picks up only 53% of Belle sample and 92% of Belle II as can be seen from Tab. 6.1, we conclude the Δt resolution has been improved on the whole sample.

6.2.4 Expected statistical uncertainty

Figure 6.5 shows the statistical uncertainties for S and A from Belle results and expected values in our analysis. Room for improvement in Belle II is shown as "all TD", for which we assume that all reconstructed events are usable for time-dependent fit. The expected uncertainties are derived by generating toy MC samples and fitting a Gaussian function to the distribution of the fit results. The width of the Gaussian is taken as the expected uncertainty. It should be noted that the expected uncertainty of S shown in the figure is smaller than that obtained by the parametric bootstrap method. This is because the latter assumes input *CP* asymmetries at the physics boundary, which results in a broad tail in the distribution of fitted S.

Extrapolating the current Belle II analysis, we expect similar and better sensitivities per statistics for S and A, respectively, as the consequence of the similar TD-event yield and the inclusion of TI events for extra time-integrated use. If we succeed in promoting all TI events to TD events in future, 1.3 times smaller statistical uncertainties are expected

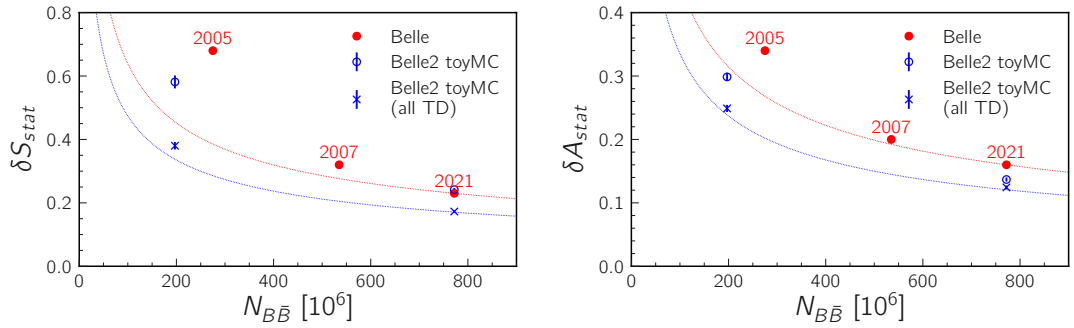


Figure 6.5: Comparison of expected statistical uncertainty in Belle II with the Belle results as a function of $N_{B\bar{B}}$ [48][49][15]. The solid lines are drawn to guide eyes and are proportional to $1/\sqrt{N_{B\bar{B}}}$. The red numbers indicates the year when the corresponding Belle analysis was published.

for both S and A at the same statistics as Belle.

6.3 Prospect

Based on the knowledge learnt from our analysis, we discuss on possible improvements in the analysis to be done in the near term and expected sensitivity with $50\,\mathrm{ab}^{-1}$, the target integrated luminosity at Belle II.

6.3.1 Retrieval of TI events

Large room for improvement lies in the treatment of TI events. Currently roughly half of the signal events are not used for time-dependent fit but only for time-integrated fit as TI events because of the presence of off-diagonal K_S^0 or large χ^2 of the CP-side vertex fit.

As we discuss in Appendix E, we expect both problems can be overcome in the near future by improving the reconstruction algorithm. Off-diagonal K_S^0 can be recovered by modifying the K_S^0 reconstruction software so that it identifies obvious fake hits associated to K_S^0 daughter tracks and repeats track reconstruction after removing them. Regarding large-χ^2 events, we have revealed that the large χ^2 originates from poorly reconstructed K_S^0. By removing such K_S^0 from the vertex fit, we can properly estimate the vertex position uncertainty and χ^2 for these events without degrading the resolution. Such modification to the analysis tools is possible in short term and will retrieve most of the TI events as TD events.

Table 6.3: Projection of uncertainties

	190 fb^{-1}		$50\,\text{ab}^{-1}$	
Source	δS	δA	δS	δA
Vertex reconstruction	**0.025**	**0.022**	**0.018**	**0.005**
Detector misalignment	0.002	0.004	0.002	0.004
Flavor tagging	**0.079**	**0.031**	**0.026**	**0.010**
Fit bias	0.003	0.002	0.003	0.002
Physics parameters	0.008	0.000	0.008	0.000
Signal fraction	**0.011**	**0.007**	**0.001**	**0.000**
Background Δt	**0.011**	**0.001**	**0.001**	**0.000**
Resolution function parameter	0.012	0.006	0.012	0.006
Resolution function model	0.001	0.003	0.001	0.003
Tag-side interference	**0.014**	**0.015**	**0.000**	**0.004**
Systematic uncertainty	0.087	0.042	0.035	0.014
Statistical uncertainty (TD+TI)	$^{+0.91}_{-0.46}$	$^{+0.30}_{-0.27}$	0.027	0.016
Statistical uncertainty (all TD)	-	-	0.022	0.014
Total uncertainty (TD+TI)	$^{+0.91}_{-0.47}$	$^{+0.30}_{-0.27}$	0.045	0.021
Total uncertainty (all TD)	-	-	0.042	0.020

6.3.2 Projection of uncertainties

Table 6.3 shows the projection of uncertainties. We argue the reason of the improvements below.

Statistical uncertainty The expected statistical uncertainties at $50\,\text{ab}^{-1}$ are estimated in the same method as Sec. 6.2.4. We consider two scenarios: assuming no improvement in the treatment of TI event ("TD+TI"), and assuming all TI events are retrieved as TD events ("all TD").

Vertex reconstruction The systematic uncertainty due to vertex reconstruction is currently dominated by helix uncertainty correction and vertex quality selection. While the former is irreducible, we expect the latter to be reduced because it involves statistical fluctuation as mentioned in Sec. 5.7. Therefore, we assume the helix uncertainty correction dominates this systematic uncertainty source.

Flavor tagging The systematic uncertainties due to flavor tagging currently come from large statistical uncertainties in the measurement of w and Δw. We assume that the systematic uncertainties of w and Δw are irreducible and the uncertainties of w and Δw will be reduced to one third of the current values. Consequently, the corresponding systematic uncertainties will be reduced by the same rate.

Signal fraction and background Δt The parameters describing the shape of $M_{\rm bc}$, M, $\mathcal{O}'_{CS}$ and background Δt PDFs are currently determined by a fit to the data sample or MC sample of similar statistics to the data. We expect the uncertainties will scale by $1/\sqrt{\mathcal{L}}$ with the integrated luminosity $\mathcal{L}$.

Tag-side interference In the current analysis we take the largest deviation of *CP* asymmetries within the physical region as the systematic uncertainty. However, with more precise input parameters we will be able to incorporate the TSI into our fit model and correct the effect. Then, instead of the deviation due to TSI itself, we consider the uncertainty of the correction, which arises from the precision of input parameters $\{\phi_1, \phi_3, S_+, S_-\}$. The dominant uncertainty sources, S_+ and S_-, are measured using semileptonic decays of B^0, where the TSI does not exist. Assuming that the measurement uncertainties scale by $1/\sqrt{\mathcal{L}}$, the uncertainties on the correction will be negligible.

Resolution function parameter As we already use high-statistics signal MC sample to determine the resolution function parameters in the current analysis, we do not expect the uncertainties to be reduced by higher statistics. However, in the future the larger data set will enable data-driven validation of the resolution function such as the lifetime fits and *CP* fit in the control channel and self validation of $2K^0_S$ resolution discussed in Appendix B. These validations will make us more confident about the resolution function or allow us to introduce data-driven correction. It may also be possible to determine the resolution function parameters based on yet another control channel.

Remarks There should also be improvements in the items that we do not mention above. For example, more precise determination of τ_{B^0}, Δm_d, and ϕ_3 will further reduce the uncertainties about physics parameters and tag-side interference. Yet these will not significantly change the total uncertainties in Tab. 6.3.

Table 6.3 tells us that the systematic uncertainties will be comparable or larger than the statistical uncertainties with $50\,{\rm ab}^{-1}$. We briefly review the prospect for other $b \to sq\overline{q}$ decays, $B^0 \to \phi K^0_S$ and $\eta' K^0_S$, which have larger branching ratios than $B^0 \to K^0_S K^0_S K^0_S$. The size of systematic uncertainties for these decays should be similar to $B^0 \to K^0_S K^0_S K^0_S$. For $B^0 \to \phi K^0_S$ decays, we expect the statistical uncertainties of $\delta S = 0.025$ and $\delta A = 0.017$, which are similar to $B^0 \to K^0_S K^0_S K^0_S$, at $50\,{\rm ab}^{-1}$ by extrapolating the $5\,{\rm ab}^{-1}$ expectation in Ref. [18] with $1/\sqrt{\mathcal{L}}$. For $B^0 \to \eta' K^0_S$ decays, the statistical uncertainties are expected to be as small as $\delta S = 0.04$ and $\delta A = 0.03$ already at $5\,{\rm ab}^{-1}$ so the sensitivity will be limited by the systematic uncertainties much earlier than $B^0 \to K^0_S K^0_S K^0_S$ [18]. The saturation of the sensitivity will increase the relative importance of $B^0 \to K^0_S K^0_S K^0_S$ in the future. Here we consider only the main decay modes $\phi K^0_S \to (K^+K^-)(\pi^+\pi^-)$ and $\eta' K^0_S \to (\rho\gamma)(\pi^+\pi^-)$ for simplicity.

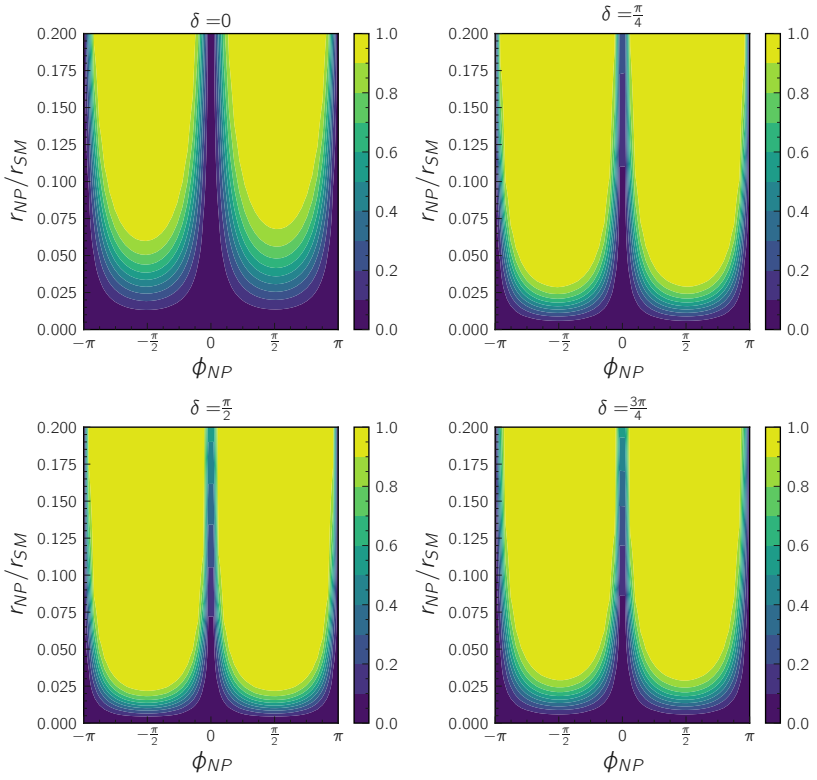


Figure 6.6: Expected sensitivity to new physics parameters with $50\,\mathrm{ab}^{-1}$. Confidence level to exclude parameters of new physics decay amplitude $(r_{\mathrm{NP}}/r_{\mathrm{SM}}, \phi_{\mathrm{NP}}, \delta)$ is shown. Each plot shows the confidence level on $(r_{\mathrm{NP}}/r_{\mathrm{SM}})$-$\phi_{\mathrm{NP}}$ plane for fixed δ.

6.3.3 Constraints on new physics

Based on the discussions in the previous subsections, we expect the total uncertainties of $\delta S = 0.042$ and $\delta A = 0.020$ at $50\,\mathrm{ab}^{-1}$. Let us update the NP constraints in Sec. 6.1 using the uncertainties. Figure 6.6 shows the constraints we will obtain in the case we observe no deviation from the SM prediction. As long as the NP has a sizeable $\sin\phi_{\mathrm{NP}}$, the measurement will be sensitive to NP contributions whose amplitude is 2–7% of the SM amplitude at 90% confidence level.

It would be interesting to reinterpret the model-independent constraint to constrain parameters in specific models such as the mass insertion parameters in the SUSY model introduced in Sec. 1.3.2. However, we do not find a theoretical study about the SUSY contribution to $B^0 \to K^0_S K^0_S K^0_S$ decays while there are some for $B^0 \to \phi K^0_S$ and $B^0 \to \eta' K^0_S$. Here, we refer to a calculation for $B^0 \to \phi K^0_S$ assuming a similar contribution to $B^0 \to K^0_S K^0_S K^0_S$. If we assume the same uncertainty for $B^0 \to \phi K^0_S$ as expected for $B^0 \to K^0_S K^0_S K^0_S$ and follow the calculation in Ref. [50], the model-independent constraint is translated to $|(\delta^d_{RL(LR)})_{23}| < O(10^{-3})$ when the masses of squark and gluino are 3 TeV. We hope for progress in the calculation of NP contributions in $B^0 \to K^0_S K^0_S K^0_S$ decays in the future.

Chapter 7

Conclusion

$B^0 \to K^0_S K^0_S K^0_S$ decay is mediated by $b \to sq\overline{q}$ penguin transition within the Standard Model. Comparison of *CP*-violating parameters between the penguin-dominated decays and $b \to c\bar{c}s$ decays such as $B^0 \to J/\psi K^0_S$ provides a probe for new physics which contributes to the $b \to s$ transition. We report a measurement of time-dependent *CP* violation in $B^0 \to K^0_S K^0_S K^0_S$ decays, using a data set containing 198×10^6 $B\overline{B}$ pairs collected at the Belle II experiment from 2019 to 2021. We obtain

$$S = -1.86^{+0.91}_{-0.46}(\text{stat}) \pm 0.09(\text{syst}), \text{ and}$$
$$A = -0.22^{+0.30}_{-0.27}(\text{stat}) \pm 0.04(\text{syst}).$$

The result is consistent with the SM expectation $(S, A) = (-0.70 \pm 0.02, 0)$ based on the measurements in $B^0 \to (c\bar{c})K^0$ and previous measurements at Belle, BaBar, and Belle II.

We have established the analysis procedure anticipating the high-precision measurement using $50\,\text{ab}^{-1}$ data to be taken at the Belle II experiment. Below we list major improvements from the previous work. We have revealed that uncertainties of fundamental track parameters are underestimated and developed a method to correct the uncertainties as described in Appendix F. For the validation of the resolution function parameters, we have developed a technique to emulate the vertex resolution of $B^0 \to K^0_S K^0_S K^0_S$ decays using $B^+ \to K^0_S K^0_S K^+$ as a control channel. We have improved the reconstruction efficiency and thus the sensitivity to *CP* violation by including the TI events characterized by poor vertex reconstruction performance and the events contaminated by $q\overline{q}$ background. These events were discarded in the previous analysis. Accordingly, the signal fraction is extended by adding the continuum suppression BDT classifier as a conditional variable to take into account the $q\overline{q}$ contamination. We have also resolved possible small biases in the measurement. The improvement includes the use of r and $\cos\theta^*_B$ in signal fraction estimation to avoid the bias due to assuming the same distributions for these variables between signal and background components, and the removal of the peaking $B\overline{B}$ background component that was neglected in the previous analyses.

Appendix A

Control channel $B^+ \to K^0_S K^0_S K^+$

We use a control channel of $B^+ \to K^0_S K^0_S K^+$ decays. The main purposes of the control channel study are to determine the background PDFs for signal extraction fit and to validate the resolution function by lifetime fit and *CP* fit. We intend to make as many parameters common between the two decay modes as possible, which are

- $M_{\rm bc}$, M, $\mathcal{O}'_{CS}$, and r bin PDF parameters for background events,
- $R_{\rm rec}$ parameters,
- $R_{\rm asc}$ parameters, and
- a part of $R_{\rm np}$ parameters.

In the *CP* fit to the control sample, we determine only S and fix A to be zero. We intentionally leave the direct *CP* violation unmeasured, which is not our target. The results of lifetime fit and *CP* fit using the control channel is described in Sec. 5.5. We describe the reconstruction algorithm, signal extraction, resolution function and background Δt distribution of $B^+ \to K^0_S K^0_S K^+$ in the following sections.

A.1 Event selection and vertex reconstruction

We keep the reconstruction flow and selection criteria for $B^+ \to K^0_S K^0_S K^+$ the same as $B^0 \to K^0_S K^0_S K^0_S$ except that

- we select K^+ based on its PID variable requiring $kaonID > 0.5$,
- we do not veto $B^+ \to X(K^0_S K^0_S) K^+$ decays,
- we discard TI events and only use TD events,
- we scale up d_0 and z_0 uncertainties of K^+ by a factor of 1000, and
- we convert $\left(\frac{\chi^2}{ndf}\right)^{CP}$ to correct the difference of ndf between $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$.

We determine the PID selection criteria by maximizing the figure of merit. We do not need to care the contamination by the resonance because the *CP*-violating parameter S

in $B^+ \to K^0_S K^0_S K^+$ is not physically meaningful. As we fix A to be zero, the TI events would not contribute to the measurement. The latter two prescriptions are to emulate the vertex fit performance of $B^0 \to K^0_S K^0_S K^0_S$. $B^+ \to K^0_S K^0_S K^+$ vertex fit without K^+ track would emulate that of $B^0 \to K^0_S K^0_S K^0_S$ where one or more K^0_S do not have VXD hits and thus do not contribute to the fit. However, in this way we cannot use the IP constraint as it requires the B_{CP} to be fully reconstructed. As a workaround we spoil the K^+ track by setting extremely large position uncertainties. The spoiled track does not directly contribute to the vertex fit but only provides the momentum information needed to define the IP constraint. The effective *ndf* of such vertex fit is then smaller by 2 than the ostensible value, $7 - 2 = 5$. For comparison with $B^0 \to K^0_S K^0_S K^0_S$ vertex fit where *ndf* is 8, we convert the χ^2 as

$$\chi^2 \to F^{-1}(F(\chi^2; 5); 8), \tag{1.1}$$

where $F(x; n)$ is a cumulative χ^2 distribution with ndf of n.

Table A.1 shows the efficiency of each selection and expected number of events for signal and background events. The cumulative efficiency is normalized by the number of generated events for signal and by the number of events after selections on K^0_S candidate, $M_{\rm bc}$, and M for background. The expected signal yield does not include $B^+ \to X K^0_S$ and assumes $200 \times 10^6 B\overline{B}$, the equal fraction of $B^0\overline{B}^0$ and B^+B^-, $\mathcal{B}(B^+ \to K^0_S K^0_S K^+) = 1.08 \times 10^{-6}$, and $\mathcal{B}(K^0_S \to \pi^+\pi^-) = 0.691$. The expected background yield is scaled to $200\,\mathrm{fb}^{-1}$.

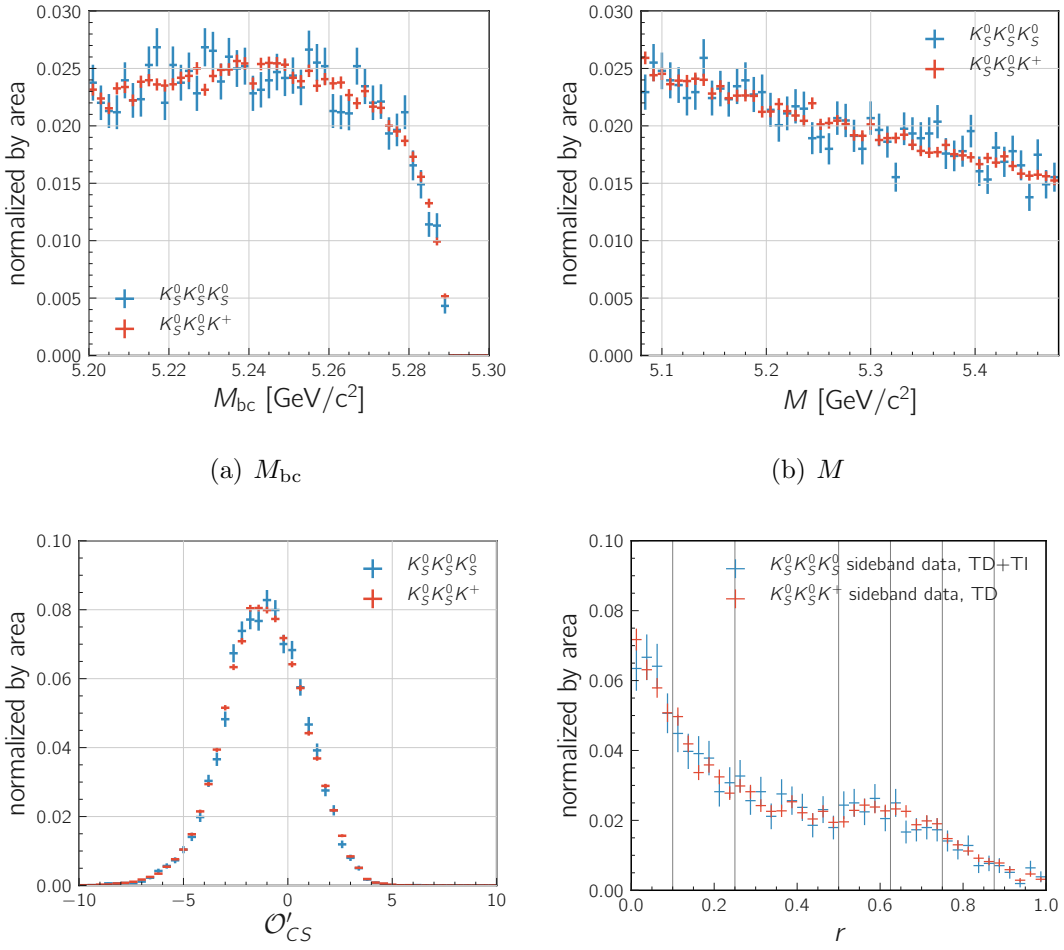
We observe a significant data-MC discrepancy in the kaonID efficiency. Since we do not need to know the reconstruction efficiency we leave the discrepancy as it is. We do not expect either that the difference in the kaonID performance can significantly affect Δt resolution as we only use the momentum information of K^+ for the vertex fit.

A.2 Similarity to $B^0 \to K^0_S K^0_S K^0_S$

We use the $B^+ \to K^0_S K^0_S K^+$ decays to determine background PDFs parameters of $M_{\rm bc}$, M, $\mathcal{O}'_{CS}$, and r bin. The agreement of the distributions between $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$ shown in Fig. A.1 verifies the strategy. The figure shows the $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ distributions from MC sample and r bin distribution from data sideband, $M_{\rm bc} > 5.265\,\mathrm{GeV}/c^2$.

We confirm that the r distribution is not correlated with $M_{\rm bc}$ in the background MC for both of $B^+ \to K^0_S K^0_S K^+$ and $B^0 \to K^0_S K^0_S K^0_S$ as shown in Fig. A.2.

The vertex fit performance is also reproduced well as shown in Figs A.3 and A.4, where we compare σ_ℓ^{CP} distribution and the core part of $\left(\frac{\chi^2}{ndf}\right)^{CP}$ distribution. The samples are divided according to the number of K^0_S's with VXD hits. However, we observe a discrepancy in the tail part of $\left(\frac{\chi^2}{ndf}\right)^{CP}$ as shown in Fig. A.5, which we do not use for the time-dependent analysis.



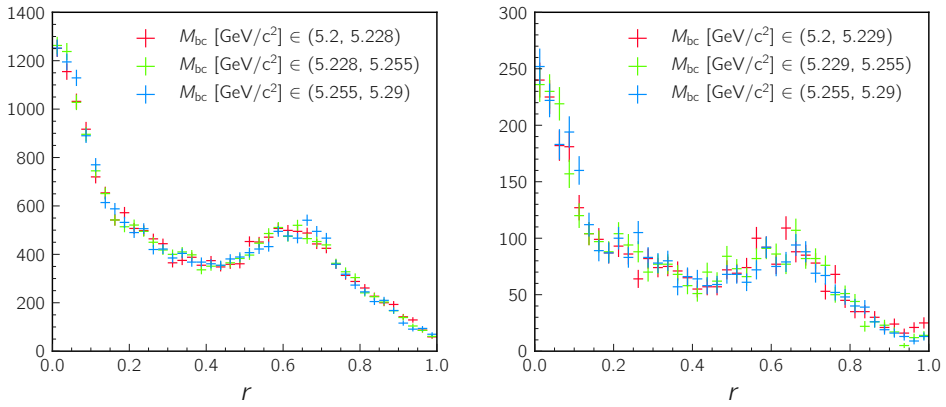
(a) M_{bc} (b) M

(c) $\mathcal{O}'_{CS}$ (d) r bin

Figure A.1: Comparison of background M_{bc}, M, and $\mathcal{O}'_{CS}$ distributions between $B^0 \to K^0_S K^0_S K^0_S$ and $B^+ \to K^0_S K^0_S K^+$. Background MC samples are shown for M_{bc}, M, and $\mathcal{O}'_{CS}$, and sideband data for r bin.

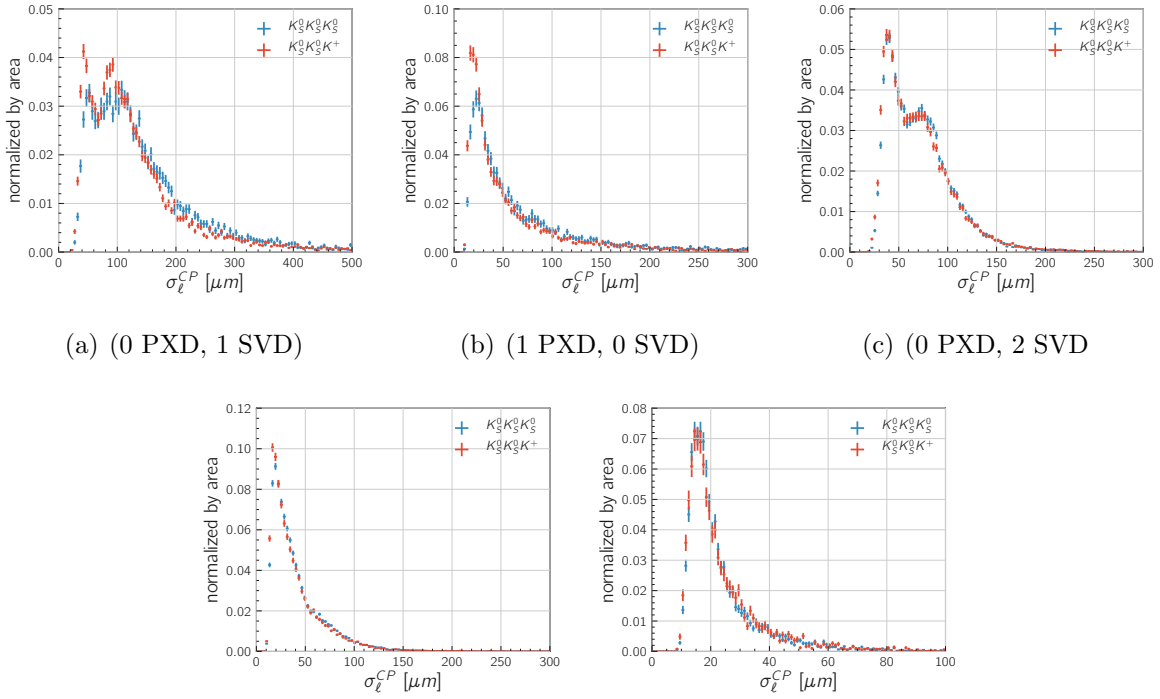
Table A.1: Relative (cumulative) efficiency [%] for $B^+ \to K^0_S K^0_S K^+$

selection	signal MC	background MC	data
K^0_S selection, $M_{\rm bc}$, M	– (100.00)	– (100.00)	– (100.00)
$kaonID > 0.5$	90.39 (90.39)	73.43 (73.43)	65.82 (65.82)
$\mathcal{O}_{CS} > 0.1$	97.93 (88.52)	49.19 (36.12)	51.77 (34.08)
$-10 < \mathcal{O}'_{CS} < 10$	100.00 (88.52)	99.98 (36.11)	99.98 (34.07)
tag vertex fit success	99.02 (87.65)	98.70 (35.64)	97.96 (33.38)
$\left(\frac{\chi^2}{ndf}\right)^{\rm tag} < 100$	93.35 (81.82)	90.71 (32.33)	91.54 (30.55)
$\sigma_\ell^{\rm tag}[\,\mu{\rm m}] < 500$	99.91 (81.75)	99.85 (32.28)	99.81 (30.49)
flavor tag success	100.00 (81.75)	100.00 (32.28)	100.00 (30.49)
best candidate selection	98.97 (80.91)	95.25 (30.74)	96.90 (29.55)
reject off-diagonal K^0_S	91.17 (73.77)	87.05 (26.76)	82.38 (24.34)
at least one K^0_S with VXD hits	96.29 (71.03)	96.87 (25.92)	96.39 (23.46)
B_{CP} vertex fit success	100.00 (71.03)	100.00 (25.92)	100.00 (23.46)
$\left(\frac{\chi^2}{ndf}\right)^{CP} < 3.27$	83.52 (59.32)	77.31 (20.04)	71.42 (16.76)
$\sigma_\ell^{CP}[\,\mu{\rm m}] < 500$	99.44 (58.99)	99.27 (19.89)	99.33 (16.65)
$-30 < \Delta t[\,{\rm ps}] < 30$	99.97 (58.97)	99.93 (19.88)	99.92 (16.63)
TD fraction	72.89	64.67	56.29
Expected/Observed yield with $190\,{\rm fb}^{-1}$	208	14800	9214



(a) $B^+ \to K^0_S K^0_S K^+$ (b) $B^0 \to K^0_S K^0_S K^0_S$

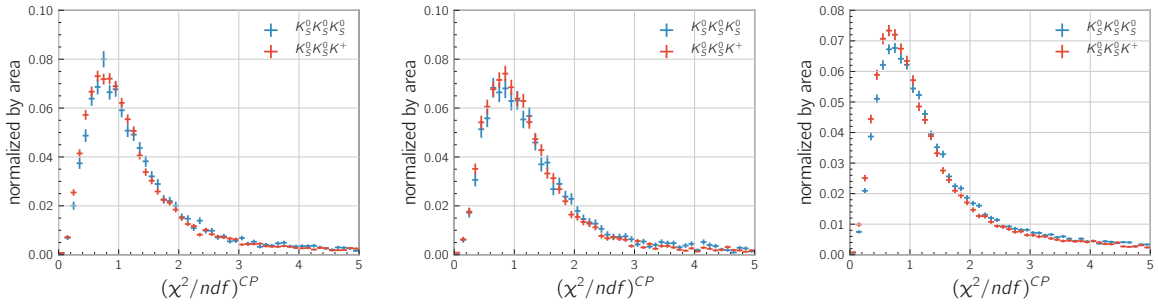
Figure A.2: r distribution in MC for different $M_{\rm bc}$ bins



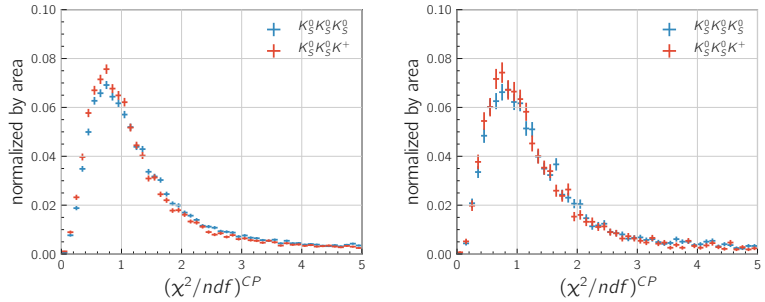
(a) (0 PXD, 1 SVD) (b) (1 PXD, 0 SVD) (c) (0 PXD, 2 SVD

(d) (1 PXD, 1 SVD) (e) (2 PXD, 0 SVD)

Figure A.3: Comparison of σ_ℓ^{CP} between $B^0 \to K_S^0K_S^0K_S^0$ and $B^+ \to K_S^0K_S^0K^+$, divided according to (number of K_S^0 with PXD hits, number of K_S^0 with SVD hits)

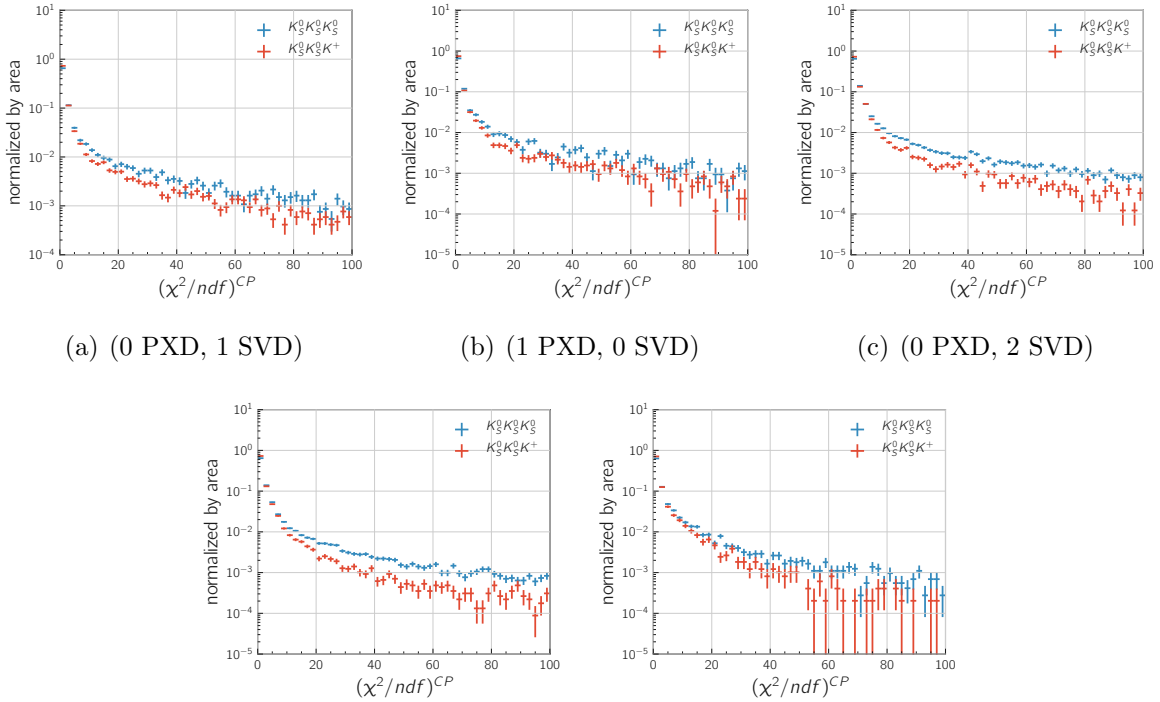


(a) (0 PXD, 1 SVD) (b) (1 PXD, 0 SVD) (c) (0 PXD, 2 SVD)



(d) (1 PXD, 1 SVD) (e) (2 PXD, 0 SVD)

Figure A.4: Comparison of $\left(\frac{\chi^2}{ndf}\right)^{CP}$ between $B^0 \to K^0_S K^0_S K^0_S$ and $B^+ \to K^0_S K^0_S K^+$, divided according to (number of K^0_S with PXD hits, number of K^0_S with SVD hits)



(a) (0 PXD, 1 SVD) (b) (1 PXD, 0 SVD) (c) (0 PXD, 2 SVD)

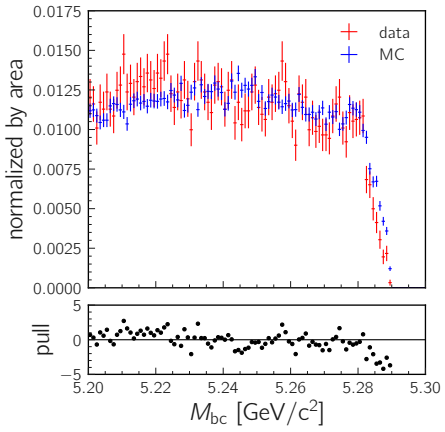
(d) (1 PXD, 1 SVD) (e) (2 PXD, 0 SVD)

Figure A.5: Comparison of $\left(\frac{\chi^2}{ndf}\right)^{CP}$ between $B^0 \to K^0_S K^0_S K^0_S$ and $B^+ \to K^0_S K^0_S K^+$, divided according to (number of K^0_S with PXD hits, number of K^0_S with SVD hits)

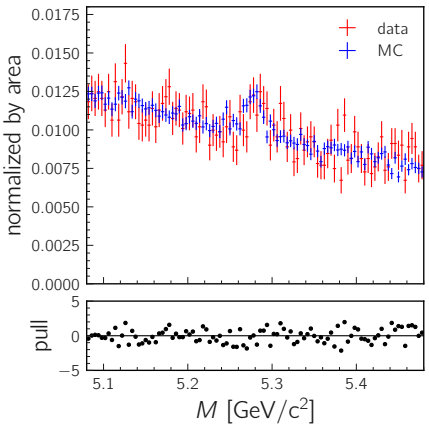
A.3 Data-MC comparison

Thanks to its larger data set than the signal channel, the control channel provides more detailed comparison of distributions between data and MC. Figures A.6 and A.7 compares important distributions between data and MC samples for $B^+ \to K^0_S K^0_S K^+$. For the comparison we mix the background and signal MC samples equivalent to $700\,\mathrm{fb}^{-1}$ with the expected rate and normalize the distributions by area.

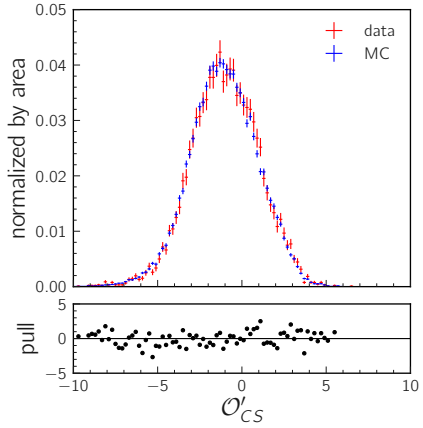
As we observe a significant discrepancy in the r bin distribution, we take the background r bin distribution from the sideband data for signal fraction estimation. We also find the discrepancy in M_{bc} endpoint, which is due to slightly lower beam energy in data. The number of tracks used in the tag-side vertex fit differs between data and MC, which could one of the causes of small discrepancies in $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$ and $\sigma_\ell^{\mathrm{tag}}$. Because we require a PXD hit for tracks in the tag-side vertex fit, lower Layer-1 efficiency in data may change the distribution. However, it cannot fully explain the behavior, for example, the equal fraction of single-track events so we suspect the imperfect $q\overline{q}$ simulation as another cause.



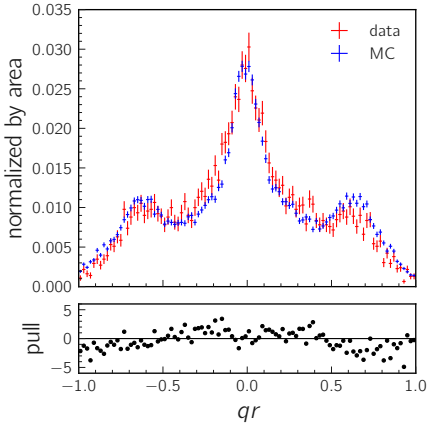
(a) $M_{\rm bc}$



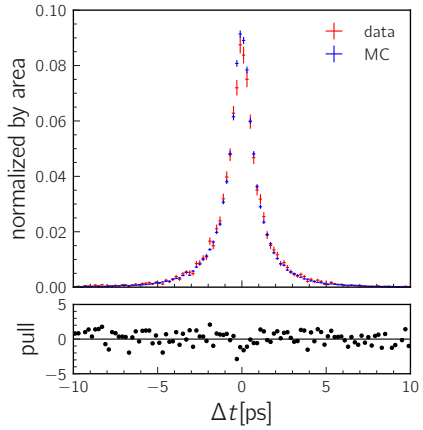
(b) M



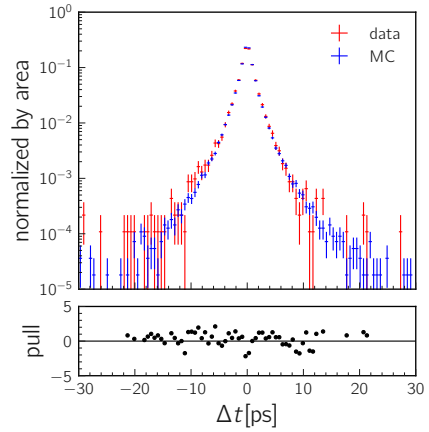
(c) $\mathcal{O}'_{CS}$



(d) qr

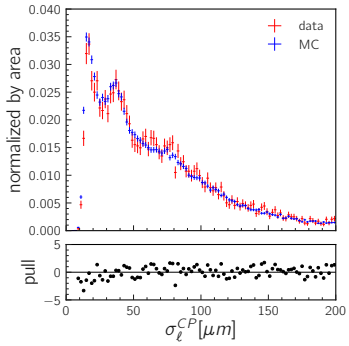


(e) Δt in linear scale

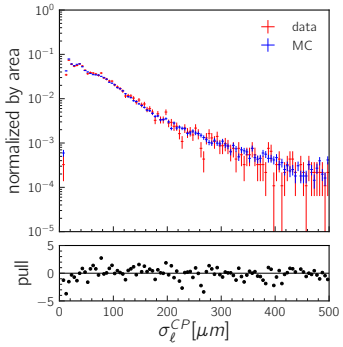


(f) Δt in log scale

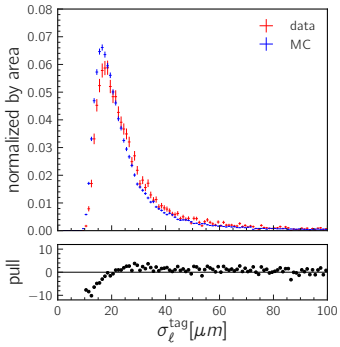
Figure A.6: Data-MC comparison of $B^+ \to K^0_S K^0_S K^+$ distributions (1)



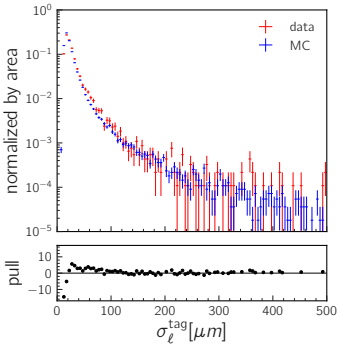
(a) σ_ℓ^{CP} in linear scale



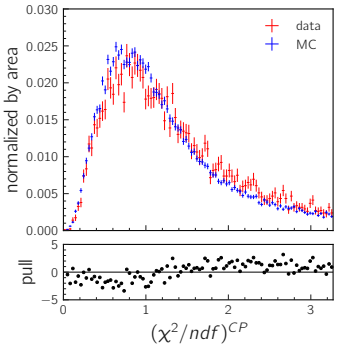
(b) σ_ℓ^{CP} in log scale



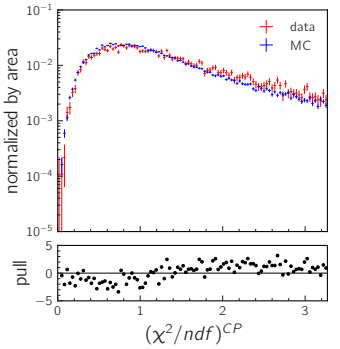
(c) $\sigma_\ell^{\mathrm{tag}}$ in linear scale



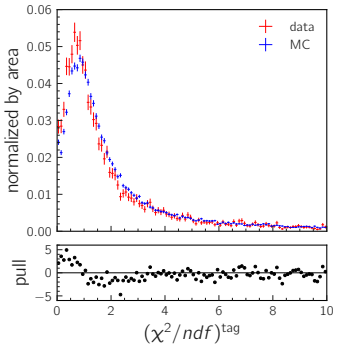
(d) $\sigma_\ell^{\mathrm{tag}}$ in log scale



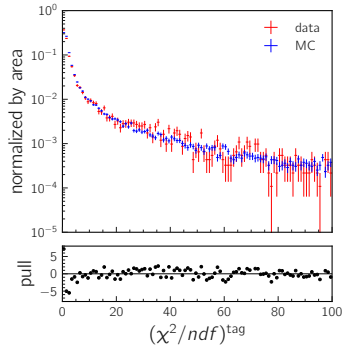
(e) $\left(\frac{\chi^2}{ndf}\right)^{CP}$ in linear scale



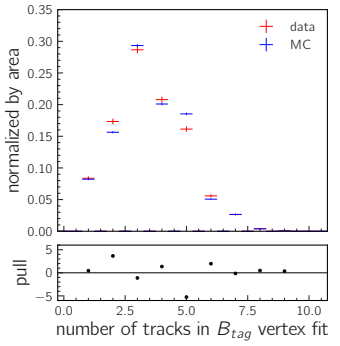
(f) $\left(\frac{\chi^2}{ndf}\right)^{CP}$ in log scale



(g) $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$ in linear scale



(h) $\left(\frac{\chi^2}{ndf}\right)^{\mathrm{tag}}$ in log scale



(i) number of tracks in tag-side vertex fit

Figure A.7: Data-MC comparison of $B^+ \to K_S^0 K_S^0 K^+$ distributions (2)

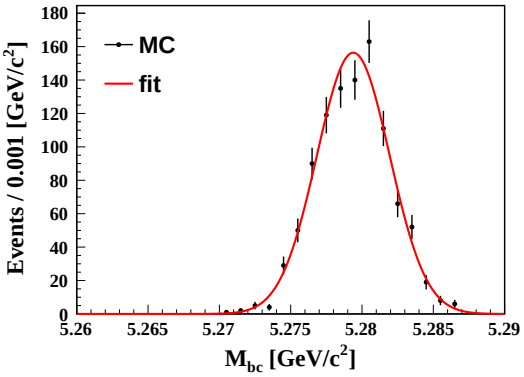
A.4 Signal extraction

We use the same PDF models of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ as those for $B^0 \to K^0_S K^0_S K^0_S$ described in Sec. 5.2. We fix the shape parameters for signal events by fits to 1D MC distributions while we determine the background shape parameters simultaneously with $F_{\rm sig}$ in the 3D signal extraction fit.

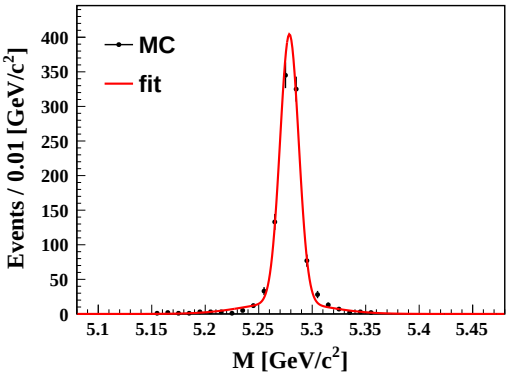
Figure A.8 show the 1D PDFs fitted to signal and background distributions. The result of the signal extraction fit to the $B^+ \to K^0_S K^0_S K^+$ data are shown in Fig. A.9. We obtain the background parameters listed in Tab. 5.3 and $F_{\rm sig} = 0.017243 \pm 0.001659$, which corresponds to 149^{+15}_{-14} signal events. We expect 216 signal events with around 5 % uncertainty assuming:

- $N_{B\bar{B}} = (198 \pm 3.0) \times 10^6$,
- $BR(\Upsilon(4S) \to B^+B^-) = 0.514 \pm 0.006$,
- $BR(B^+ \to K^0_S K^0_S K^+) = (1.05 \pm 0.04) \times 10^{-5}$,
- $BR(B^+ \to \chi_{c0} K^0_S) \cdot BR(\chi_{c0} \to K^0_S K^0_S) = (1.51^{+0.15}_{-0.13}) \times 10^{-4} \cdot (3.16 \pm 0.17) \times 10^{-3}$,
- $BR(K^0_S \to \pi^+\pi^-) = 0.692$, and
- reconstruction efficiency of 21.2 % in MC.

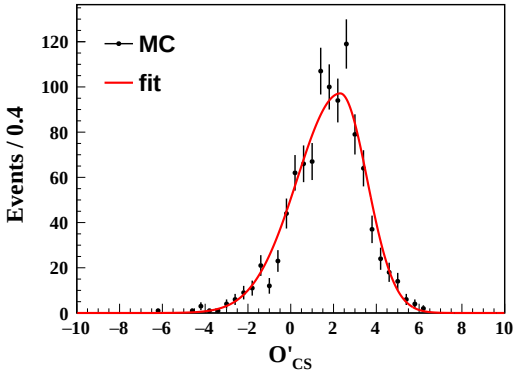
Considering that the selection efficiency in data is around 0.8 times shorter than MC as shown in Tab. A.1, the observed number of events is around 1.4 σ smaller than the expectation.



(a) signal $M_{\rm bc}$

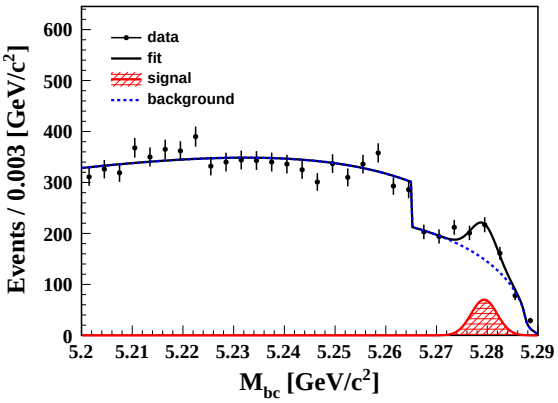


(b) signal M

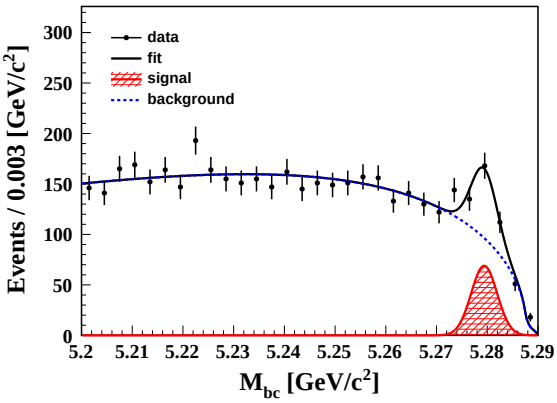


(c) signal $\mathcal{O}'_{CS}$

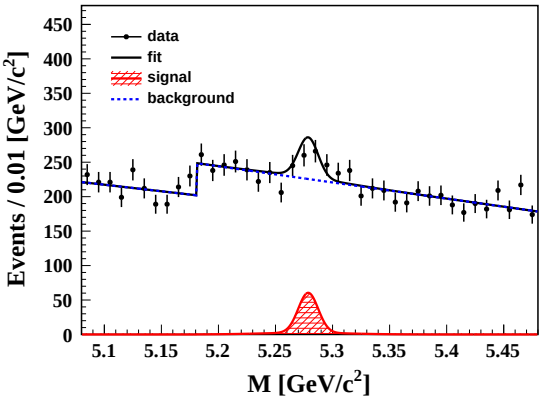
Figure A.8: MC distributions of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ for signal events



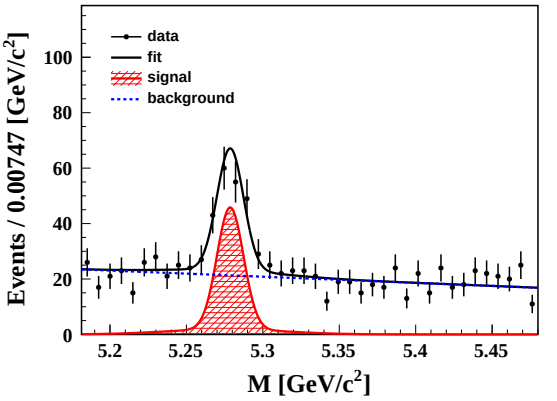
(a) $M_{\rm bc}$



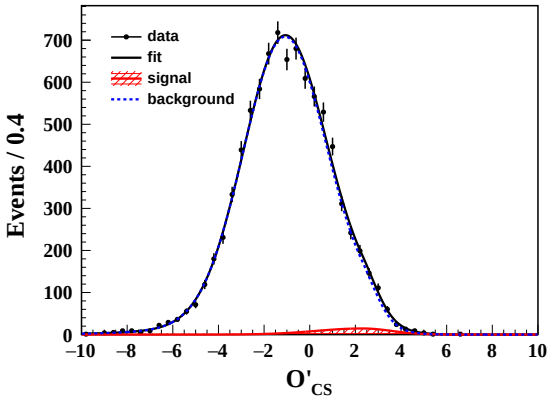
(b) $M_{\rm bc}$ in M signal region



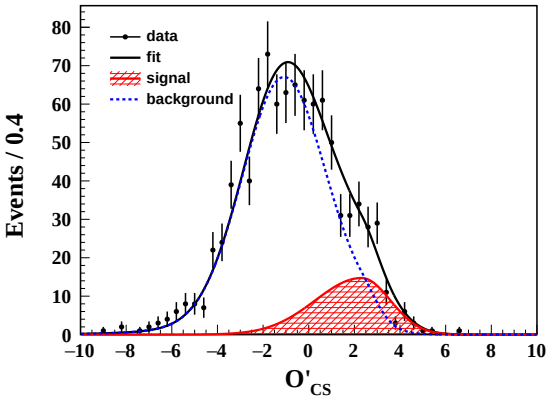
(c) M



(d) M in $M_{\rm bc}$ signal region



(e) $\mathcal{O}'_{CS}$



(f) $\mathcal{O}'_{CS}$ in $M_{\rm bc}$ signal region

Figure A.9: Results of signal extraction fit for $B^+ \to K^0_S K^0_S K^+$ data

We check the feasibility of the simultaneous fit using toy MC. We generate 1000 pseudo-experiments each with 155 signal and 9000 background events expected and perform the signal extraction fits to the samples. Figure A.10 shows the pull distribution of F_{sig} and background shape parameters, where we find no issues.

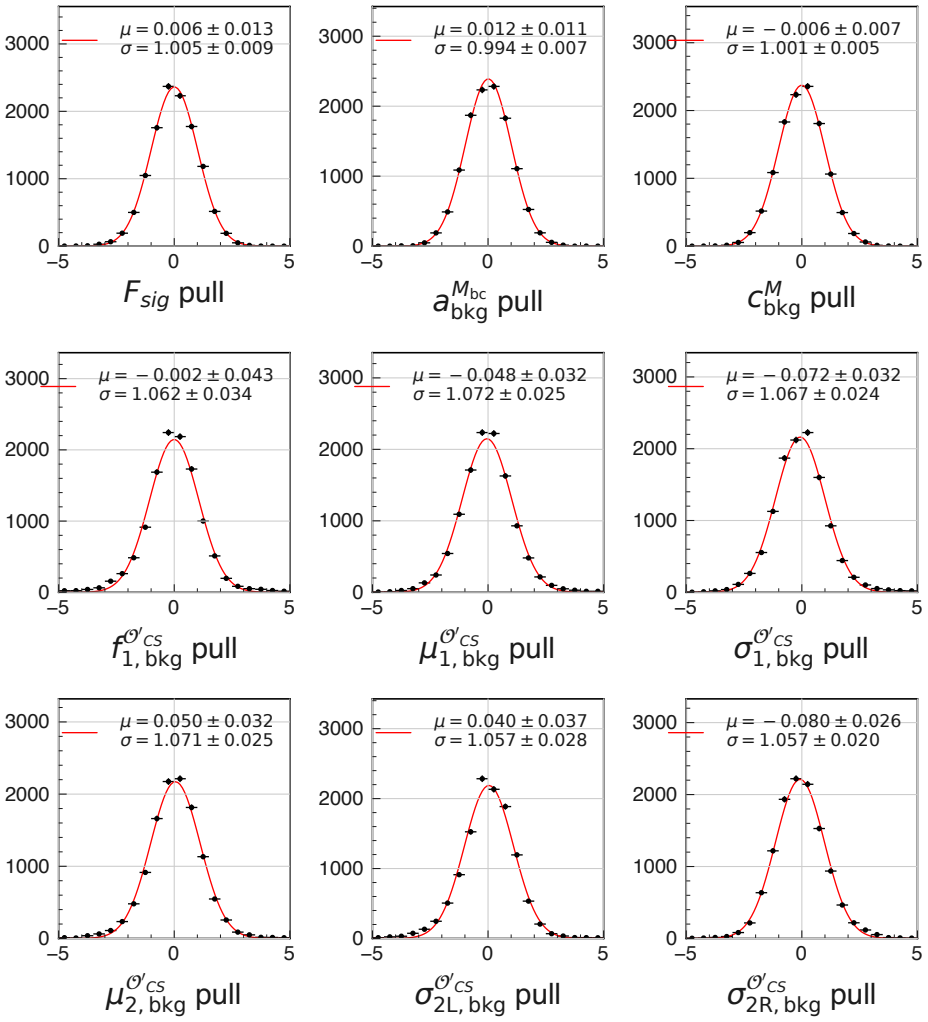


Figure A.10: Pull distributions of $F_{\rm sig}$ and background shape parameters from toy MC study

A.5 Δt resolution function

We study the Δt resolution in the control channel to understand the condition where the resolution can be described by the same resolution function parameters as $B^0 \to K^0_S K^0_S K^0_S$.

The $R_{\rm rec}$ parameters determined on $B^0 \to K^0_S K^0_S K^0_S$ signal MC also describe $B^+ \to K^0_S K^0_S K^+$ vertex resolution well as shown in Fig. A.11 where $R_{\rm rec}$ is overlaid to the MC $\delta\ell^{CP}$ distribution. It is also true in each bin of σ_ℓ^{CP} as in Fig. A.12. However, when we compare them in bins of $\left(\frac{\chi^2}{ndf}\right)^{CP}$ as in Fig. A.13, we see a clear discrepancy in the three largest $\left(\frac{\chi^2}{ndf}\right)^{CP}$ bins. This is the reason we use the tight selection $\chi^2/ndf < 3.27$ (corresponding to p-value > 0.001) for TD events so that the CP-side resolutions of $B^0 \to K^0_S K^0_S K^0_S$ and $B^+ \to K^0_S K^0_S K^+$ can be described by the common resolution function.

We suspect that the discrepancy arises due to the different number of tracks in the fit. Sometimes tracks are poorly reconstructed to make the χ^2 huge but do not really deteriorate the vertex resolution. This insight is verified by a study described in Appendix E.2. Having more tracks, $B^0 \to K^0_S K^0_S K^0_S$ picks up such tracks more frequently, which results in a better resolution in the large χ^2 region.

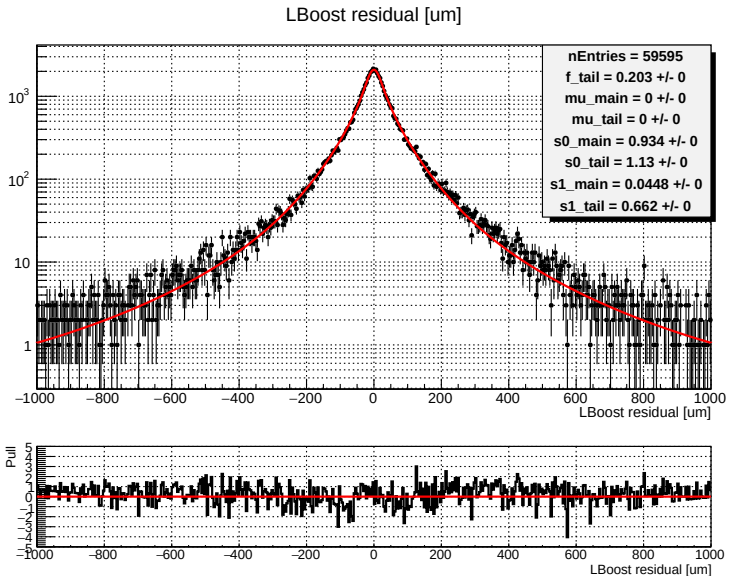
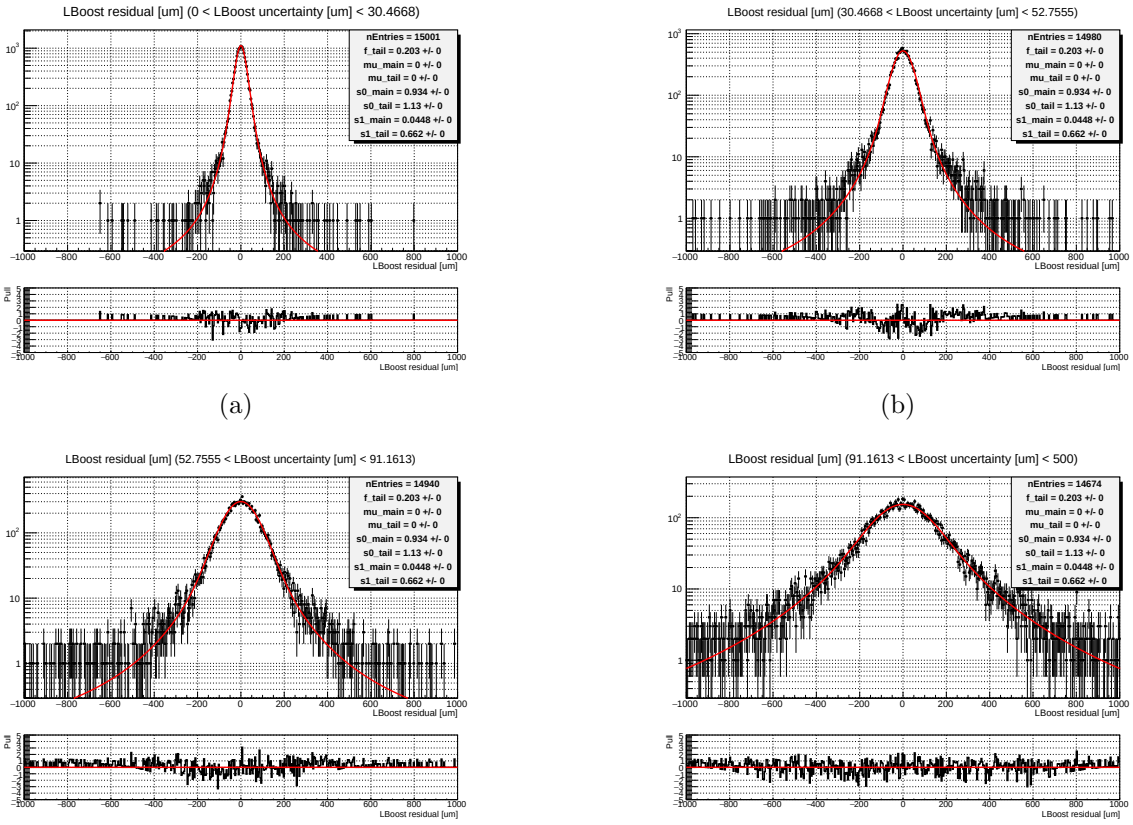


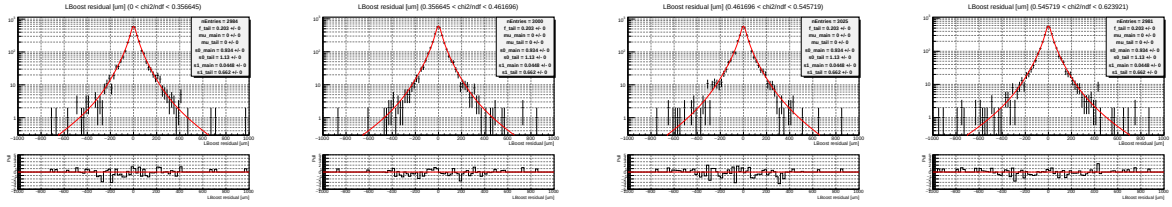
Figure A.11: R_{rec} overlaid on MC $\delta\ell^{CP}$ distribution



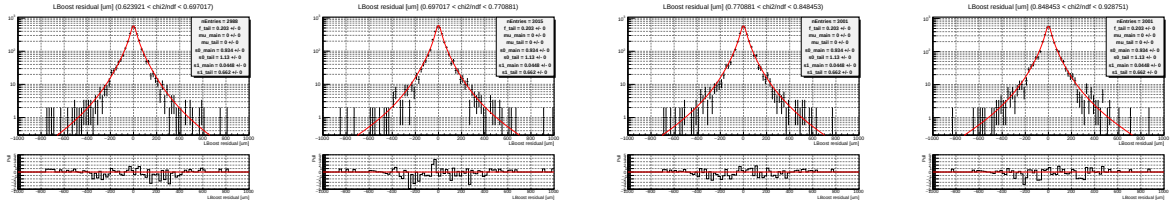
(a) (b)

(c) (d)

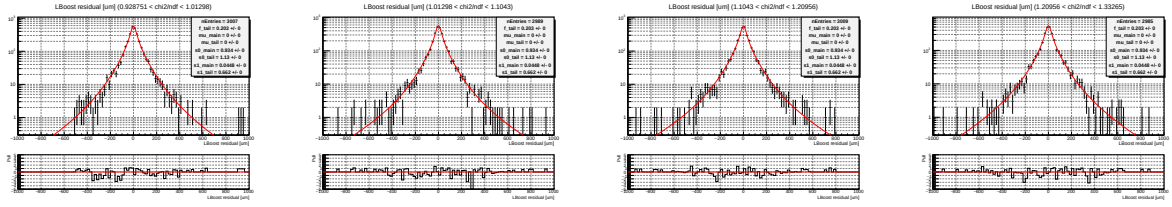
Figure A.12: R_{rec} overlaid on MC $\delta\ell^{CP}$ distribution in bins of σ_ℓ^{CP}



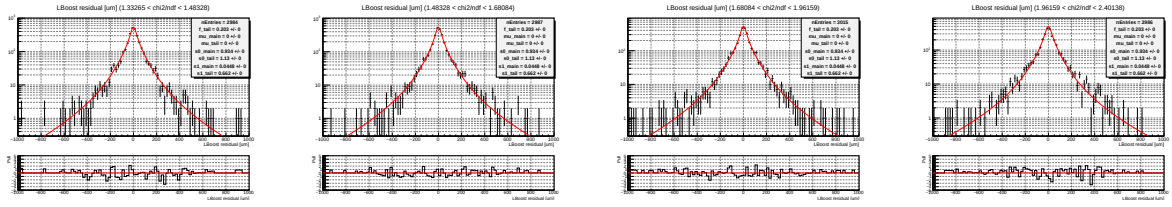
(a) (b) (c) (d)



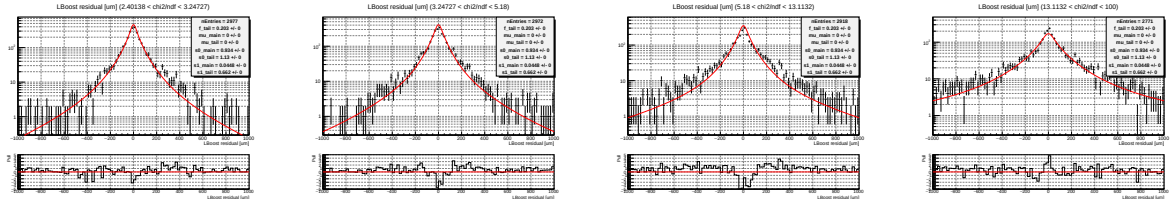
(e) (f) (g) (h)



(i) (j) (k) (l)



(m) (n) (o) (p)



(q) (r) (s) (t)

Figure A.13: R_{rec} overlaid on MC $\delta\ell^{CP}$ distribution in bins of $\left(\frac{\chi^2}{ndf}\right)^{CP}$, each plot containing 5% of reconstructed events

On the tag side, $B^+ \to K^0_S K^0_S K^+$ suffers less bias due to non-primary tracks than $B^0 \to K^0_S K^0_S K^0_S$ because $B^\pm$ produces fewer $D^\pm$ and more $D^0/\bar{D}^0$. We determine an independent set of the lifetime parameters (τ^0, τ^1, and $\tau^{\rm max}$) in $R_{\rm np}$ by fitting $R_{\rm asc} \otimes R_{\rm np}$ to the $\delta\ell^{\rm tag}$ distribution of $B^+ \to K^0_S K^0_S K^+$ MC. The other parameters in $R_{\rm np}$ and $R_{\rm asc}$ are kept the same as $B^0 \to K^0_S K^0_S K^0_S$. Figure A.14 shows the MC $\delta\ell^{\rm tag}$ distribution and the fitted function, and Figs. A.15 and A.16 show those in bins of $\sigma_\ell^{\rm tag}$ and $\left(\frac{\chi^2}{ndf}\right)^{\rm tag}$. The tag-side resolution is described well with most of the parameters common with $B^0 \to K^0_S K^0_S K^0_S$.

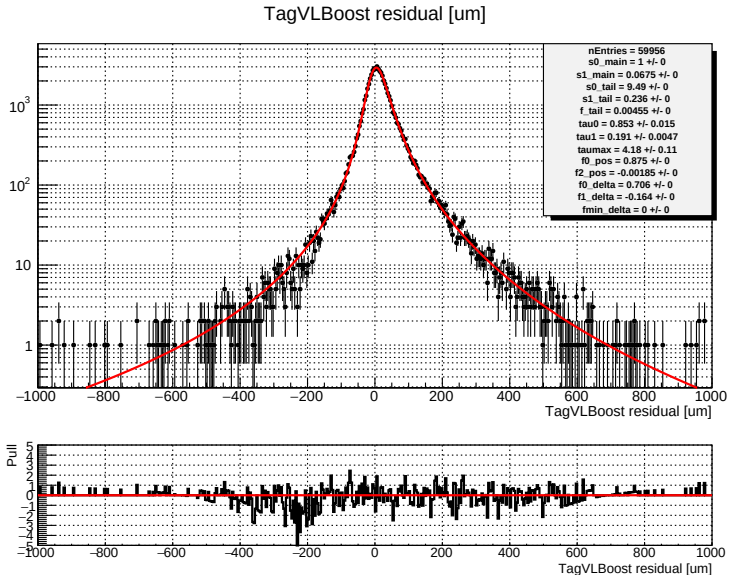
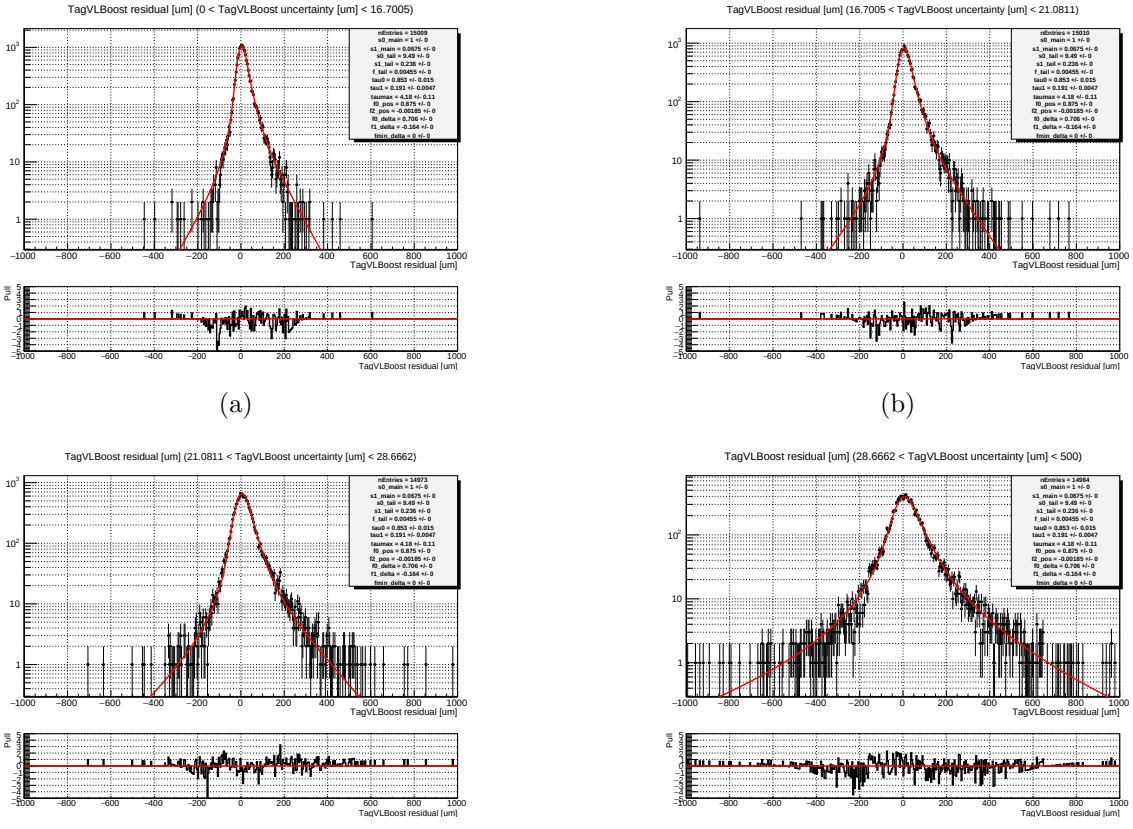


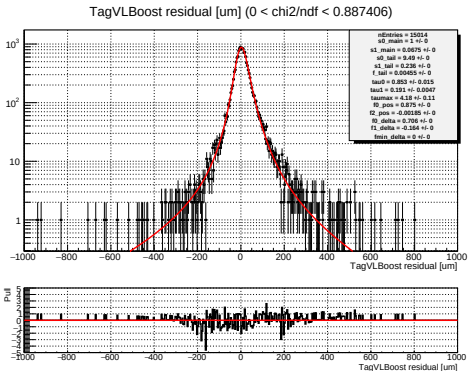
Figure A.14: $R_{\text{asc}} \otimes R_{\text{np}}$ fitted to MC $\delta\ell^{\text{tag}}$ distribution



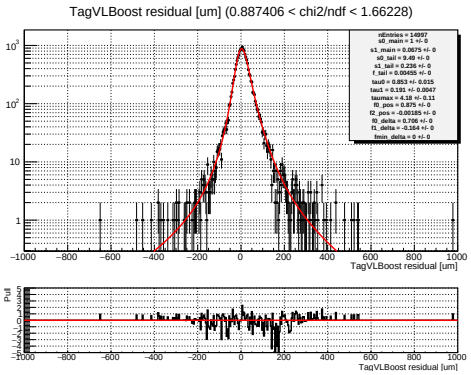
(a) (b)

(c) (d)

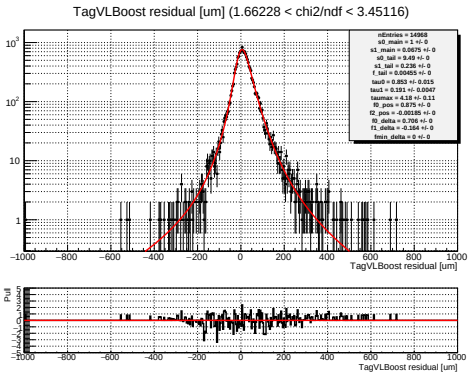
Figure A.15: $R_{\text{asc}} \otimes R_{\text{np}}$ fitted to MC $\delta\ell^{\text{tag}}$ distribution in bins of σ_ℓ^{tag}



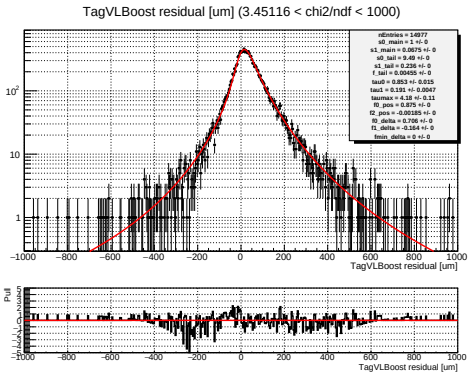
(a)



(b)



(c)



(d)

Figure A.16: $R_{\rm asc} \otimes R_{\rm np}$ fitted to MC $\delta\ell^{\rm tag}$ distribution in bins of $\left(\frac{\chi^2}{ndf}\right)^{\rm tag}$

As in Sec. 5.3.4 we perform lifetime fits to the $B^+ \to K^0_S K^0_S K^+$ MC sample. The fits are done in different ranges of $\left(\frac{\chi^2}{ndf}\right)^{CP}$; (0,100), (0,3.27), and (3.27,100). We summarize the results in Tab. A.2. The lifetime is clearly biased when the large $\left(\frac{\chi^2}{ndf}\right)^{CP}$ region is used in the fit. When the *CP*-side resolution in that region is excluded, the fitted lifetimes are around 1.62–1.63 ps and shorter than the MC input 1.638 ps by around 10 fs. This should be the bias due to the K^0_S selection which is also found in $B^0 \to K^0_S K^0_S K^0_S$ lifetime fit.

A.6 Background Δt distribution

We use the same background Δt model as $B^0 \to K^0_S K^0_S K^0_S$ with a different set of parameters. We determine the parameters by a fit to sideband of $B^+ \to K^0_S K^0_S K^+$ data, $M_{\rm bc} < 5.265\,{\rm GeV}/c^2$. Figure A.17 shows the fit result.

We confirm that the Δt is not correlated with $M_{\rm bc}$ in the background MC as shown in Fig. A.18 to verify the extrapolation of Δt distribution from the sideband to signal region.

Table A.2: Results of lifetime fits τ_{B^+} [ps] in different ranges of $\left(\frac{\chi^2}{ndf}\right)^{CP}$, with fit uncertainties of the last digit(s) in parentheses

variable	(0,100)	(0,3.27)	(3.27,100)
Δt^{true}	1.629(6)	1.625(7)	1.655(18)
$\frac{\ell^{true}_{CP}-\ell^{true}_{\rm tag}}{\beta\gamma c}$	1.629(6)	-	-
$\frac{\ell_{CP}-\ell^{true}_{\rm tag}}{\beta\gamma c}$	1.640(8)	1.621(8)	1.866(29)
$\frac{\ell^{true}_{CP}-\ell_{\rm tag}}{\beta\gamma c}$	1.628(7)	-	-
Δt	1.639(8)	1.619(8)	1.877(31)

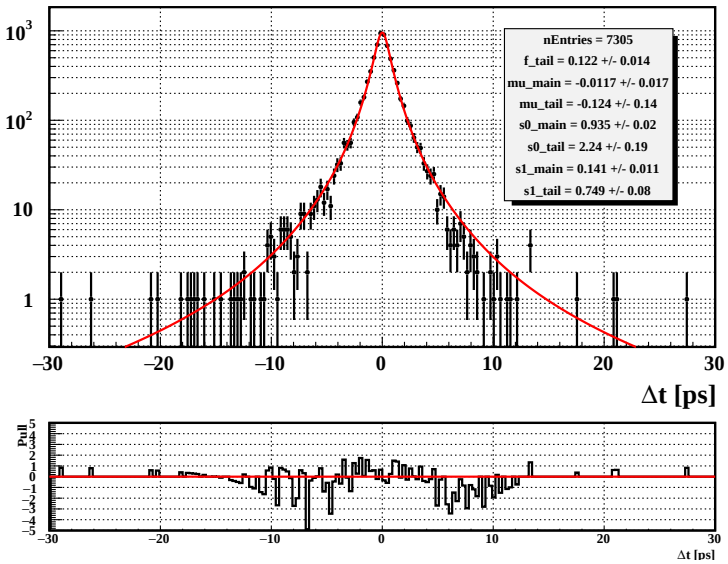


Figure A.17: Δt distribution in $B^+ \to K^0_S K^0_S K^+$ sideband data and fitted $P_{\rm bkg}(\Delta t)$

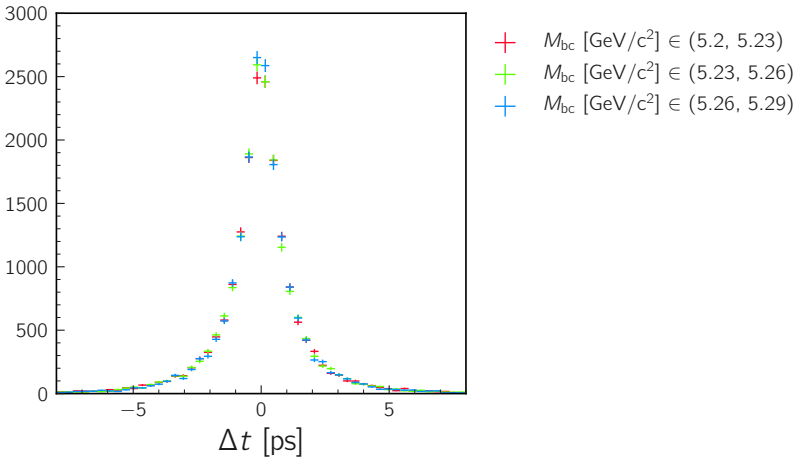


Figure A.18: Δt distribution in $B^+ \to K^0_S K^0_S K^+$ background MC for different $M_{\rm bc}$ bins

Appendix B

Data-MC comparison of CP-side vertex resolution using two K^0_S tracks

We attempt to compare the CP-side vertex resolution between data and MC.

To extract information on the vertex resolution, we perform vertex fit using only two K^0_S out of three. Using the i-th and j-th K^0_S in the vertex fit, we obtain a vertex position ℓ^{CP}_{ij}. The difference between vertex positions, $\ell^{CP}_{ij} - \ell^{CP}_{jk}$, should give some information on the vertex resolution using three K^0_S. We do not use the IP constraint for a technical difficulty and restrict the study to only the events where all of three K^0_S have VXD hits.

At first we use the signal MC to compare the two-K^0_S resolution $\ell^{CP}_{ij} - \ell^{CP}_{jk}$ with the three-K^0_S resolution (obtained as the vertex position residual from the MC truth, $\ell^{CP} - \ell^{CP}_{\rm true}$). Figure B.1 shows the comparison of the distributions, where the two-K^0_S resolution is stacked for $(i,j) = (0,1),(1,2),(2,0)$ and the three-K^0_S resolution is multiplied by an arbitrary factor $\sqrt{3}$[1]. We find that the two-K^0_S resolution reproduces the three-K^0_S resolution to some extent.

Then we compare the two-K^0_S resolution between data and MC as in Fig. B.2. The reasonable agreement indicates that the Δt resolution in the data is reproduced by the MC without a significant difference.

[1] $\sqrt{3}$ assumes the equal resolution between the three K^0_S and no correlation between ℓ^{CP}_{ij} and ℓ^{CP}_{jk}, which is clearly incorrect.

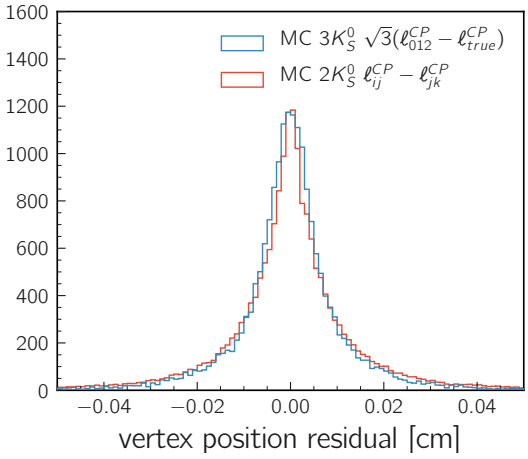


Figure B.1: Vertex resolution using $3K_s^0$ and difference of $2K_s^0$ vertex fits in the signal MC

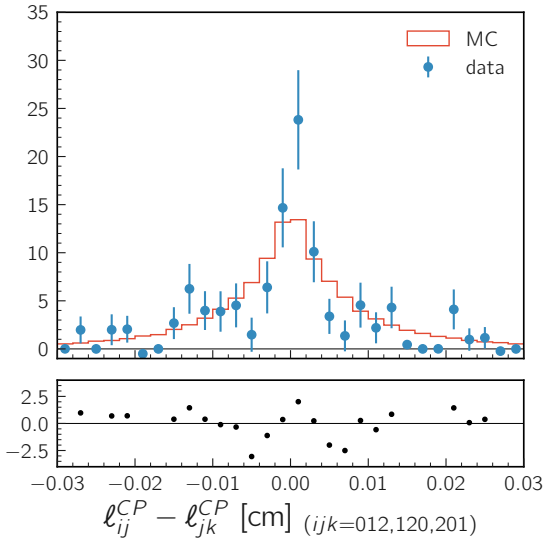


Figure B.2: Data-MC comparison of $\ell_{ij}^{CP} - \ell_{jk}^{CP}$. ${}_s\mathcal{P}lot$ is shown as the data distribution.

Appendix C

Correlation between variables used for signal extraction

B analyses often use a variable $\Delta E \equiv E_{\rm beam}/2 - E_B$ for signal yield extraction, where $E_{\rm beam}$ and E_B is the energy of beam and B in the center-of-mass frame. However we replace it with B invariant mass M since we observe a sizable correlation between ΔE and $M_{\rm bc}$ as shown in Fig. C.1. Within the plotted region these variables have the Pearson correlation coefficient of -0.12.

$M_{\rm bc}$ and ΔE are originally invented so that they are only weakly correlated with each other. The correlation is generally weak because the uncertainty of $M_{\rm bc}$ is dominated by the fluctuation of beam energy and that of ΔE by the detector resolution. However for decay modes with good ΔE resolution like $B^0 \to K^0_S K^0_S K^0_S$, the contribution from beam energy spread in ΔE uncertainty cannot be neglected so ΔE and $M_{\rm bc}$ show a visible correlation originated from beam energy spread.

We confirm that the correlation coefficients between $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ are small as shown in Tab. C.1. The correlations are weak for both signal and $q\overline{q}$ samples. We do not observe a correlation in the 2D histograms shown in Fig. C.2 either.

Table C.1: Correlation coefficients between $M_{\rm bc}$, M and $\mathcal{O}'_{CS}$ for signal and $q\overline{q}$ MC events. The numbers in the parentheses are the coefficients calculated within $(5.27 < M_{\rm bc}[\,{\rm GeV}/c^2] < 5.29) \wedge (5.18 < M[\,{\rm GeV}/c^2] < 5.38)$.

variables	signal	$q\overline{q}$
$M_{\rm bc}$, M	0.014 (-0.018)	0.001 (0.049)
$M_{\rm bc}$, $\mathcal{O}'_{CS}$	0.002 (-0.001)	-0.008 (-0.003)
M, $\mathcal{O}'_{CS}$	0.000 (0.000)	0.011 (0.026)

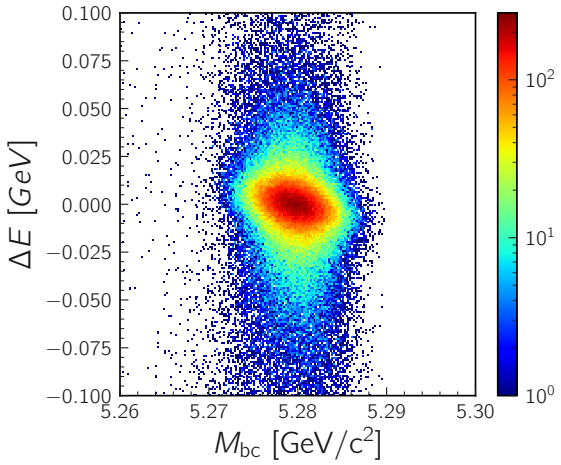
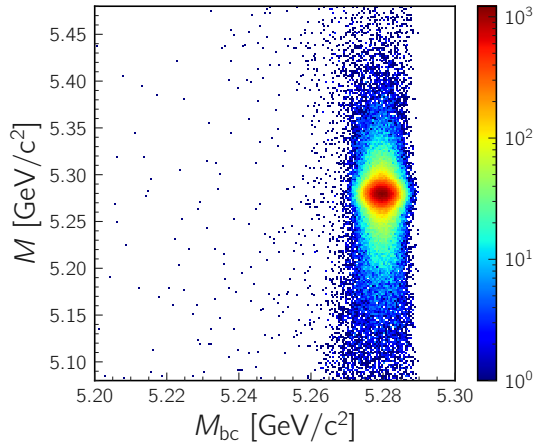
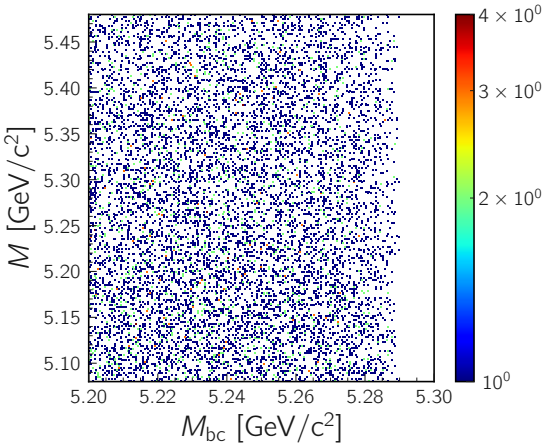


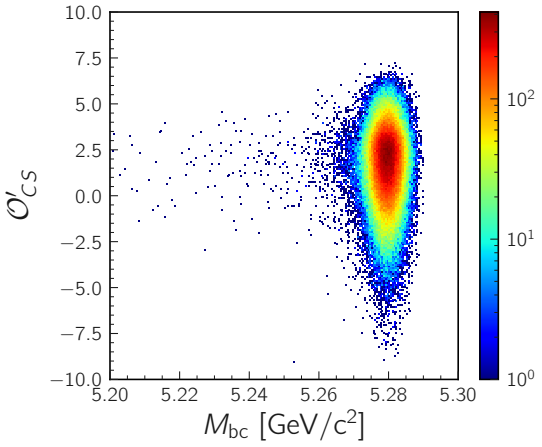
Figure C.1: 2D histogram of M_{bc} and ΔE for $B^0 \to K^0_S K^0_S K^0_S$ signal MC events



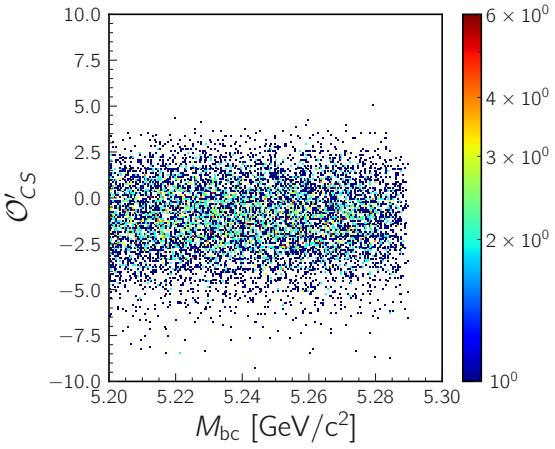
(a) signal $M_{\rm bc}$ vs M



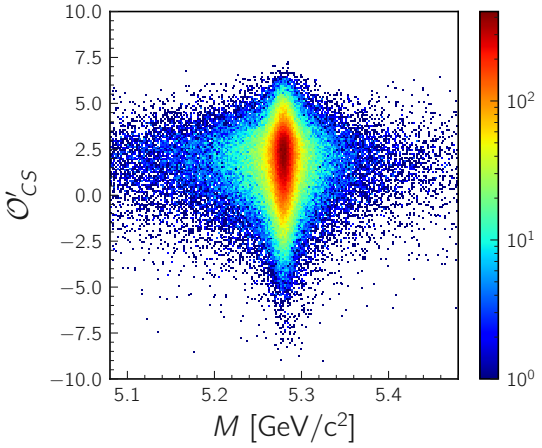
(b) $q\bar{q}$ $M_{\rm bc}$ vs M



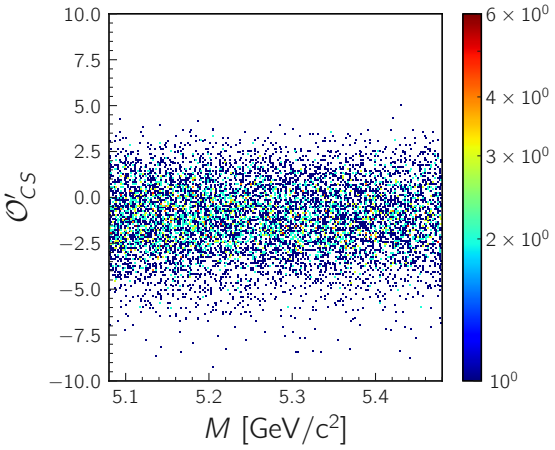
(c) signal $M_{\rm bc}$ vs $\mathcal{O}'_{CS}$



(d) $q\bar{q}$ $M_{\rm bc}$ vs $\mathcal{O}'_{CS}$



(e) signal M vs $\mathcal{O}'_{CS}$



(f) $q\bar{q}$ $\mathcal{O}'_{CS}$ vs M

Figure C.2: 2D histograms of $M_{\rm bc}$, M, and $\mathcal{O}'_{CS}$ for signal and $q\bar{q}$ MC

Appendix D

CP-side vertex resolution and number of K_S^0 with PXD hits

The vertex resolution of B_{CP} strongly depends on the number of K_s^0 having associated VXD hits, especially PXD hits. We divide TD events from the signal MC sample into three categories according to the number of K_s^0 that have PXD hits:

- events where two or three K_s^0 have PXD hits,
- events where one K_s^0 has PXD hits, and
- events where no K_s^0 has PXD hits

and show the distributions of σ_ℓ^{CP} and $\delta\ell^{CP}$ for each of them in Fig. D.1. Despite the strong dependence, the CP-side resolution function $R_{\rm rec}$ covers all categories well with a common parameter set as shown in Fig. D.2.

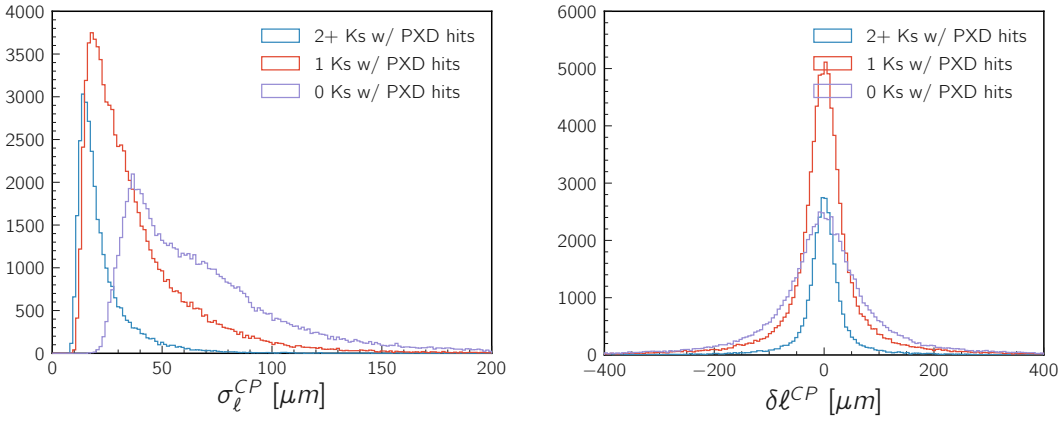
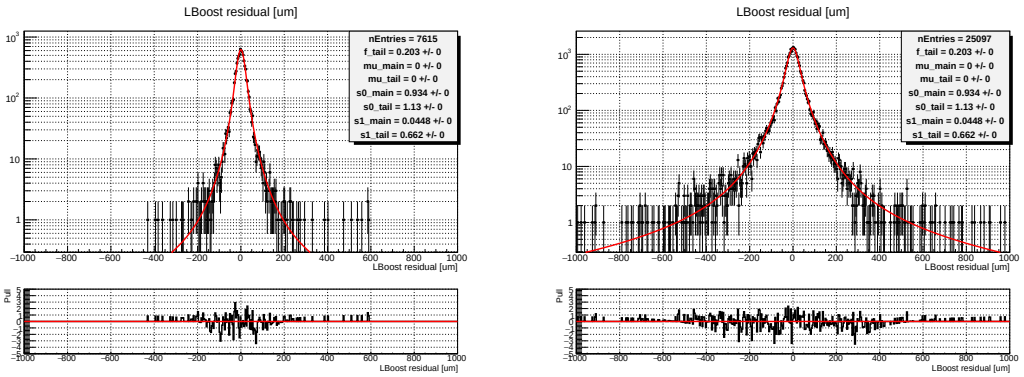
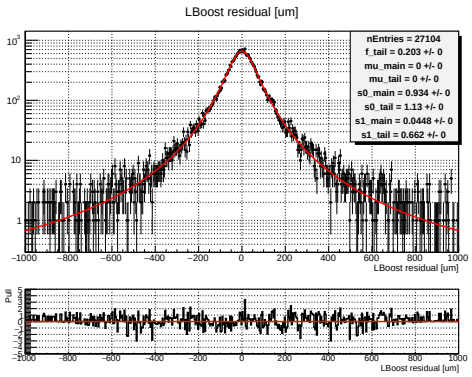


Figure D.1: σ_ℓ^{CP} and $\delta\ell^{CP}$ distributions of signal MC events classified according to the number of K_S^0 that has PXD hits



(a) two or three K_S^0 have PXD hits

(b) one K_S^0 has PXD hits



(c) no K_S^0 has PXD hits

Figure D.2: $R_{\rm rec}$ fitted to MC $\delta\ell^{CP}$ distribution, classified according to the number of K_S^0 that has PXD hits

Appendix E

Room for improvement

The current analysis adopts tight selection criteria about *CP*-side vertex reconstruction quality and discard Δt information from roughly a half of signal events, which are classified as TI events due to off-diagonal K^0_S and large χ^2. We discuss the improvement in the situation in near term to recover the vertex reconstruction quality for these events and exploit the Δt information as TD events.

E.1 off-diagonal K^0_S

As described in Sec. 4.2.2, off-diagonal K^0_S are the K^0_S candidates whose daughter tracks do not share the same innermost VXD layer to have a hit in. We particularly focus on the candidates where one daughter has its innermost hit in Layer 1 and the other in Layer 3 or 4. They arise in two ways:

- K^0_S decays inside Layer 1 and we miss a Layer-1 hit for one daughter, or
- K^0_S decays between Layer 3 (4) and the next inner layer and we assign an irrelevant (fake) Layer-1 hit to one daughter.

In the current analysis 13.3 % of reconstructed signal MC events are classified as TI events due to the presence of off-diagonal K^0_S. The fraction is even larger in the real data. We exclude these events from the time-dependent analysis because they are likely to be affected by fake Layer-1 hits, which leads to wrong estimation of the vertex position uncertainty. As shown in Fig. E.1, the pull distribution of B_{CP} vertex position becomes significantly broader if B_{CP} includes an off-diagonal K^0_S. If we do not exclude the events affected by off-diagonal K^0_S, $R_{\rm rec}$ can not describe the vertex resolution well anymore as in Fig. E.2 due to the wrongly estimated uncertainty. The systematic uncertainty arising from the poor resolution modelling is hard to quantify.

Though it is not yet reflected to the analysis, we have already modified the K^0_S reconstruction algorithm to solve the issue. In the new algorithm, after finding an initial K^0_S decay vertex, the hits assigned to the daughter tracks are examined if they are located on

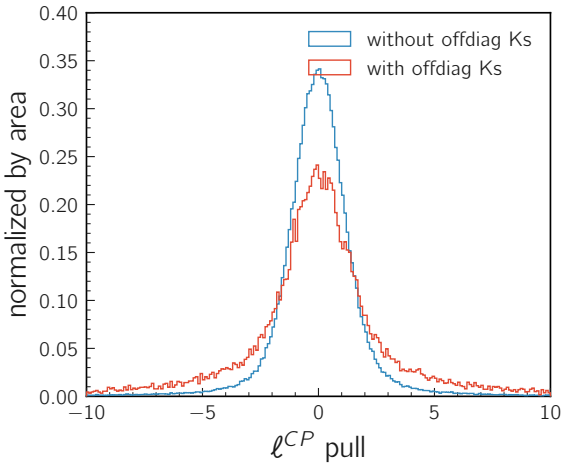


Figure E.1: Pull distribution of B_{CP} vertex position for signal MC sample, divided according to if the B_{CP} candidate has an off-diagonal K_S^0

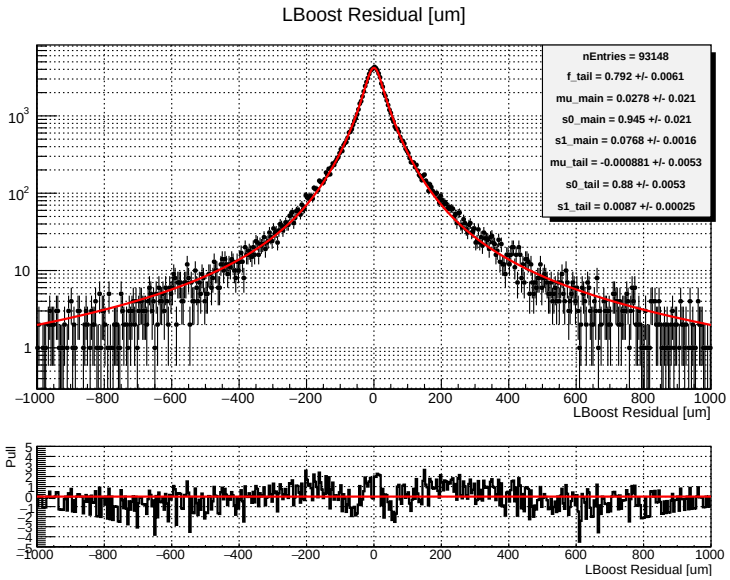


Figure E.2: CP-side resolution function fitted to signal MC sample including off-diagonal K_S^0 events (Event selections and vertex fit constraint are not up to date for this plot.)

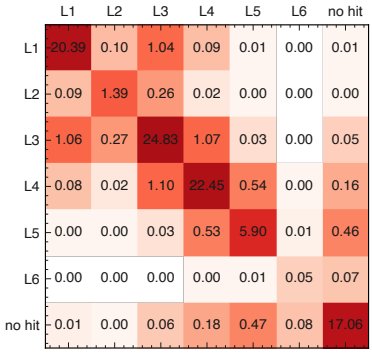
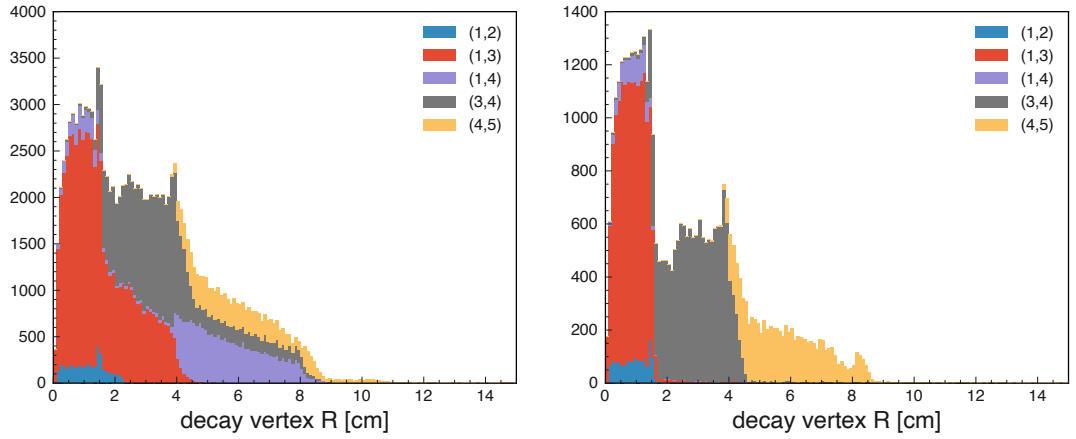


Figure E.3: Innermost VXD layers where K_S^0 daughter tracks are detected in the modified software release



(a) Current software

(b) Modified software

Figure E.4: Radial coordinate of decay vertices for off-diagonal K_S^0 reconstructed with the current and modified software releases. The legend denotes the innermost VXD layers of daughter tracks.

the inner or outer side with respect to the K_S^0 vertex. The hits located on the inner side, which are obviously fake, are then removed and the track is refitted. Figure E.3 shows the fraction of innermost VXD layers of the K_S^0 daughters reconstructed with the modifed software. Compared to Fig. 4.7, the off-diagonal K_S^0's are reduced to half. Figure E.4 shows the radial coordinate of off-diagonal K_S^0 decay vertices. We confirm that the off-diagonal K_S^0 with obvious fake hits are removed by the modification. The remaining ones with missing hits can be safely used without the wrong estimation of vertex position uncertainty. Therefore, from the next round of this analysis the rejection of off-diagonal K_S^0 will not be needed anymore.

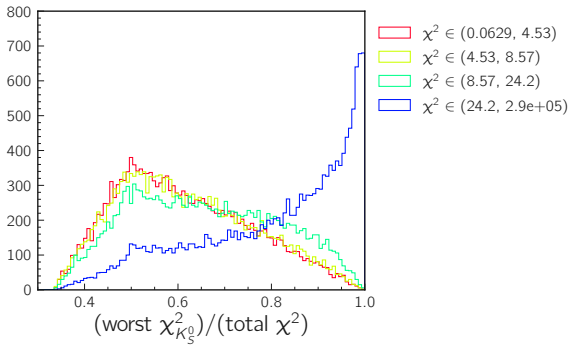
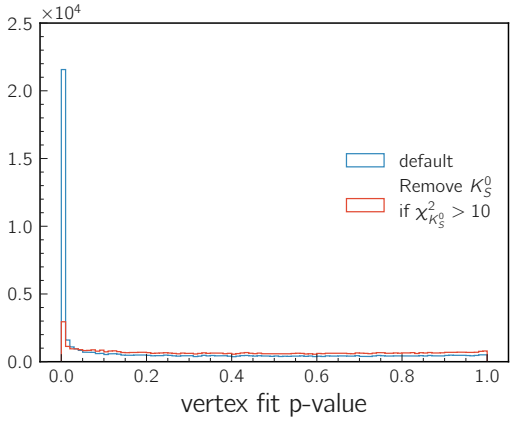


Figure E.5: Ratio of largest $\chi^2_{K^0_S}$ to total χ^2

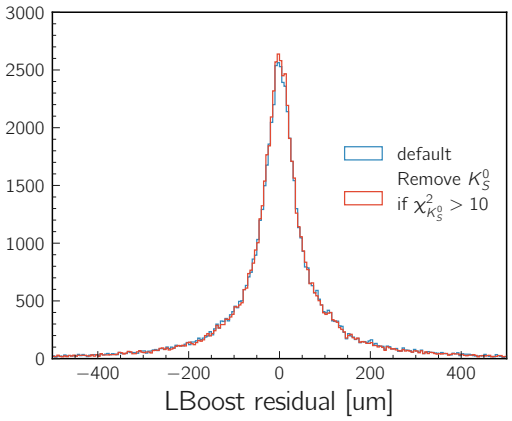
E.2 Removal of large-χ^2 K^0_S

As shown in Fig. 4.11, there are a considerable fraction of signal events where the *CP*-side vertex fit p-value is unnaturally large. We suspect that one of K^0_S's is poorly reconstructed in these events, for example, due to a wrong hit assigned to pion tracks. We study this issue without using the IP constraint in the *CP*-side vertex fit for a technical reason. We quantify the reconstruction quality of K^0_S as the partial χ^2 of the B_{CP} vertex fit, which corresponds to the χ^2 component of its daughter tracks. We refer to the partial χ^2 as $\chi^2_{K^0_S}$ and the K^0_S having the largest $\chi^2_{K^0_S}$ among the three K^0_S's on the *CP* side as the worst K^0_S. Figure E.5 shows the ratio of $\chi^2_{K^0_S}$ of the worst K^0_S to total χ^2 in bins of total χ^2. When the total χ^2 is large, it is often dominated by the single worst K^0_S. This supports our suspicion.

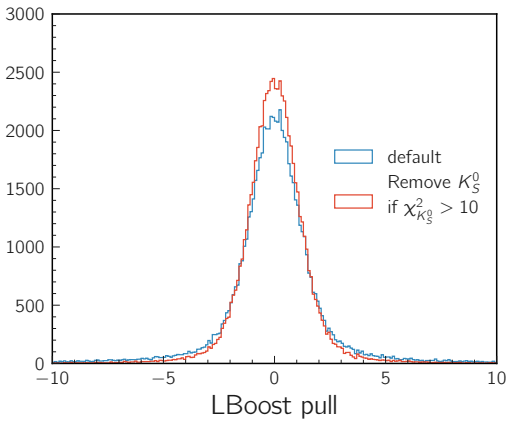
We try to improve the vertex fit quality by removing the worst K^0_S from the fit if $\chi^2_{K^0_S}$ exceeds a threshold of 10. Figure E.6 shows the distribution of p-value, B_{CP} vertex position residual, and vertex position pull for the cases where all three K^0_S's are used in the fit and the worst K^0_S is removed. We find that the p-value distribution is greatly improved and becomes almost flat by the prescription. It is notable that the resolution does not deteriorate even though a K^0_S is removed from the fit while the pull distribution gets sharper. This suggests that the poorly reconstructed K^0_S does not almost contribute to the resolution but leads to underestimation of vertex position uncertainty. Using this technique, we will be able to reduce the large-χ^2 events and prevent the wrong estimation of vertex fit uncertainty due to poorly reconstructed K^0_S. It will recover most of the TI events due to large χ^2 and make our resolution function more decay-mode universal, which is convenient for its validation.



(a) vertex fit p-value



(b) vertex position resolution



(c) vertex position pull

Figure E.6: Comparison of B_{CP} vertex fit performance with and without the prescription of removing the worst K_S^0

Appendix F

Helix parameter uncertainty correction

In this chapter, d_0 and z_0 are shown in the unit of cm unless specified.

F.1 Motivation

An MC study on the reconstruction of $B^0 \to J/\psi K^0_S$ vertex reveals that the vertex position uncertainty estimated from the vertex fit is underestimated. Figure F.1 shows the width of the pull distribution of vertex position as a function of its uncertainty.[1] It indicates severe underestimation of vertex uncertainty especially when the estimated uncertainty is small. Also the overall level of pull width is apart from unity. Among the helix parameters of the muon tracks shown in Fig. F.2, a similar correlation between pull and uncertainty is observed for d_0 and z_0. The correlation is supposed to arise from imperfect estimation of PXDCluster position uncertainty. The position of particle intersection at the PXD is reconstructed by merging the cluster of adjacent pixel hits, which is referred to as PXDCluster. The PXDCluster position and its uncertainty is estimated based on a center-of-gravity method using the amount of collected charge in each pixels belonging to the PXDCluster. The sum of collected charge in the PXDCluster is referred to as cluster charge. Figure F.3 shows the cluster charge dependence of PXDCluster position resolution and estimated uncertainty studied in MC. The current formula of uncertainty assumes that the position uncertainty decreases inversely proportionally to cluster charge, which is not correct and leads to underestimation of position uncertainty especially for clusters with large charge and thus small estimated uncertainty.

As a temporary measure, we apply a correction described in the following sections to helix parameter uncertainties so that the correlation is mitigated and the pull width becomes closer to one as well. We introduce different corrections for tracks that have a

[1] Pull is defined as the residual of estimated value from the true value normalized by its uncertainty. "Width" refers to the half range of [16, 84] % quantiles of a distribution hereafter.

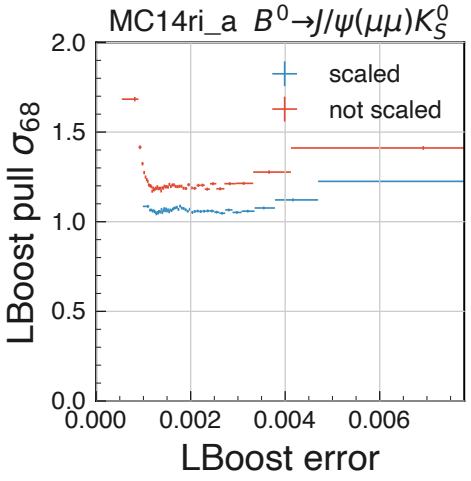


Figure F.1: Width of $B^0 \to J/\psi K^0_S$ vertex position pull distribution as a function of its uncertainty in cm. B^0 vertex is reconstructed from two muon tracks with ("scaled") and without helix uncertainty normalization ("not scaled").

PXD hit ("PXD tracks") and tracks that do not have a PXD hit but have a SVD hit ("SVD tracks"). PXD tracks are further classified into tracks produced inside the beam pipe and outside.

F.2 Correction to PXD tracks

For PXD tracks, we apply a constant scale factor c_h to each helix uncertainty ε_h,

$$\varepsilon_h \to \varepsilon_h^{\text{scaled}} = c_h \cdot \varepsilon_h \ (h = \phi_0, \omega, \tan\lambda). \tag{6.1}$$

In addition to the constant scaling, we limit d_0 and z_0 uncertainties so that they do not become smaller than the intrinsic resolution:

$$\varepsilon_h \to \varepsilon_h^{\text{scaled}} = \max\left(c_h \cdot \varepsilon_h, \sigma_h^{\text{best}}(a_h, b_h; \tilde{p}_h)\right) \ (h = d_0, z_0). \tag{6.2}$$

The "best" resolution σ_h^{best} is given by a pseudo-momentum dependent formula,

$$\sigma^{\text{best}}(a_h, b_h; \tilde{p}_h) = \sqrt{a_h^2 + (b_h/\tilde{p}_h)^2}, \tag{6.3}$$

where $\tilde{p}_h = p\beta \sin^{\frac{3}{2}(\frac{5}{2})}\theta$ for $h = d_0(z_0)$. The lower limit of the best resolution deals with the steep rising edge observed in Fig. F.2. The best resolution is fixed below 0.5 GeV/c,

$$\sigma^{\text{best}}(a_h, b_h; \tilde{p}_h < 0.5) = \sigma^{\text{best}}(a_h, b_h; \tilde{p}_h = 0.5\,\text{GeV}/c) \tag{6.4}$$

These low-momentum tracks do not suffer from the severe underestimation of d_0 and z_0 uncertainties as multiple scattering effect dominates the uncertainties. If we kept

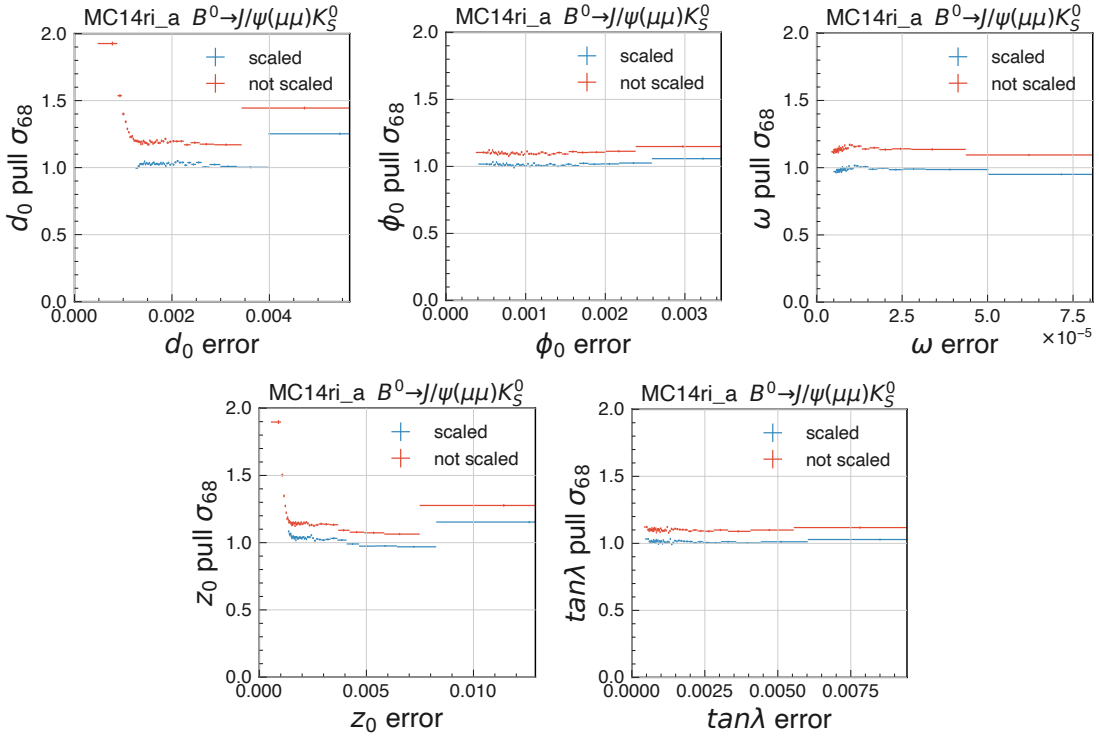


Figure F.2: Width of pull distribution as a function its uncertainty for each helix parameter with and without helix uncertainty normalization. Muons from $B^0 \to J/\psi(\mu\mu)K^0_S$ MC sample are used.

the momentum dependence $\sigma^{\rm best}$ would become too large for the low-momentum tracks, resulting in overscaling.

The correction requires nine parameters in total: five c's and two sets of a and b. The undesired correlation between pull and uncertainty is resolved by the prescription as shown in Figs. F.1 and F.2.

Long-lived particles such as K^0_S often decay outside the beam pipe or within its thickness. In that case they are affected by no or less multiple scattering through the pipe.Because our correction parameters, especially b, are also sensitive to the multiple scattering, we define a separate parameter set for displaced K^0_S daughter tracks. The displaced K^0_S is defined as K^0_S decaying at $r > 1\,\text{cm}$. For the K^0_S decaying inside the beam pipe, we apply the common scale factors to their daughter tracks as the prompt tracks produced at the IP.

F.2.1 Parameter determination for prompt tracks using MC

In order to determine the nine parameters an MC sample is used where a muon is generated in each event with the following settings:

- uniform momentum between [0.2, 3.0] GeV/c,

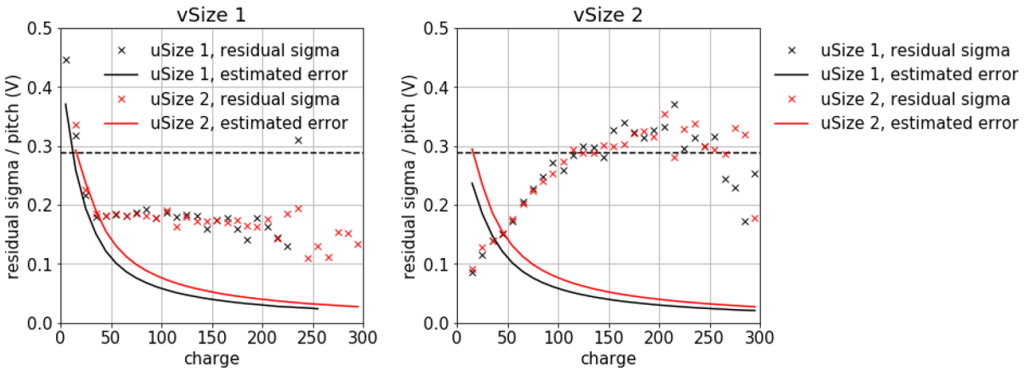


Figure F.3: PXDCluster position resolution in v (z) direction normalized by the pixel pitch vs cluster charge. Crosses are the standard deviations of cluster position residual and solid lines are the estimated position uncertainties. "uSize" and "vSize" indicate the cluster size in u (r-ϕ) and v (z) directions, respectively.

- uniform $\cos\theta$ between [17, 150] degrees.

The scale factors c_{ϕ_0}, c_ω and $c_{\tan\lambda}$ are determined by the widths of their pull distributions. Regarding the other two requiring the special prescription, $h = d_0, z_0$, the parameters a_h, b_h and c_h are determined in the following procedure.

The reconstructed muon tracks are split into ten bins of pseudo-momentum $\tilde{p}_h$ from 0.5 to 3.0 GeV/c, in each of which the width of h residual distribution is plotted as a function of h uncertainty as in Figs. F.4 and F.5. The residual tends to improve with the estimated uncertainty but saturates when the uncertainty is too small. The saturation level depends on the pseudo-momentum while the slope of the distribution is similar among different pseudo-momentum bins. The best resolution σ_i^{best} for i-th bin and the scaling factor c_h are obtained as the saturation level and the slope by fitting a kink function $y = \max(c \cdot x, \sigma_i^{\text{best}})$ to the residual vs uncertainty plots. The kink functions are fitted simultaneously with a common c_h. The resulted best resolution is shown in Fig. F.6. The parameters a_h and b_h are otained by fitting Eq. (6.3) to the data points.

The obtained parameters are listed in Tab. F.3 Applying the correction we can make the pull widths of helix parameters closer to unity and also less dependent on their uncertainties for d_0 and z_0 as shown in Fig. F.2.

F.2.2 Parameter determination for displaced K_S^0 using MC

For displaced K_S^0 daughter tracks we use basically the same correction scheme as the prompt tracks but with a different set of correction parameters. We determine the correction parameters in the same way as in Sec. F.2.1. We do not need the cutoff defined in Eq. 6.4 for z_0 uncertainty because the multiple scattering effect is small for displaced K_S^0 tracks.

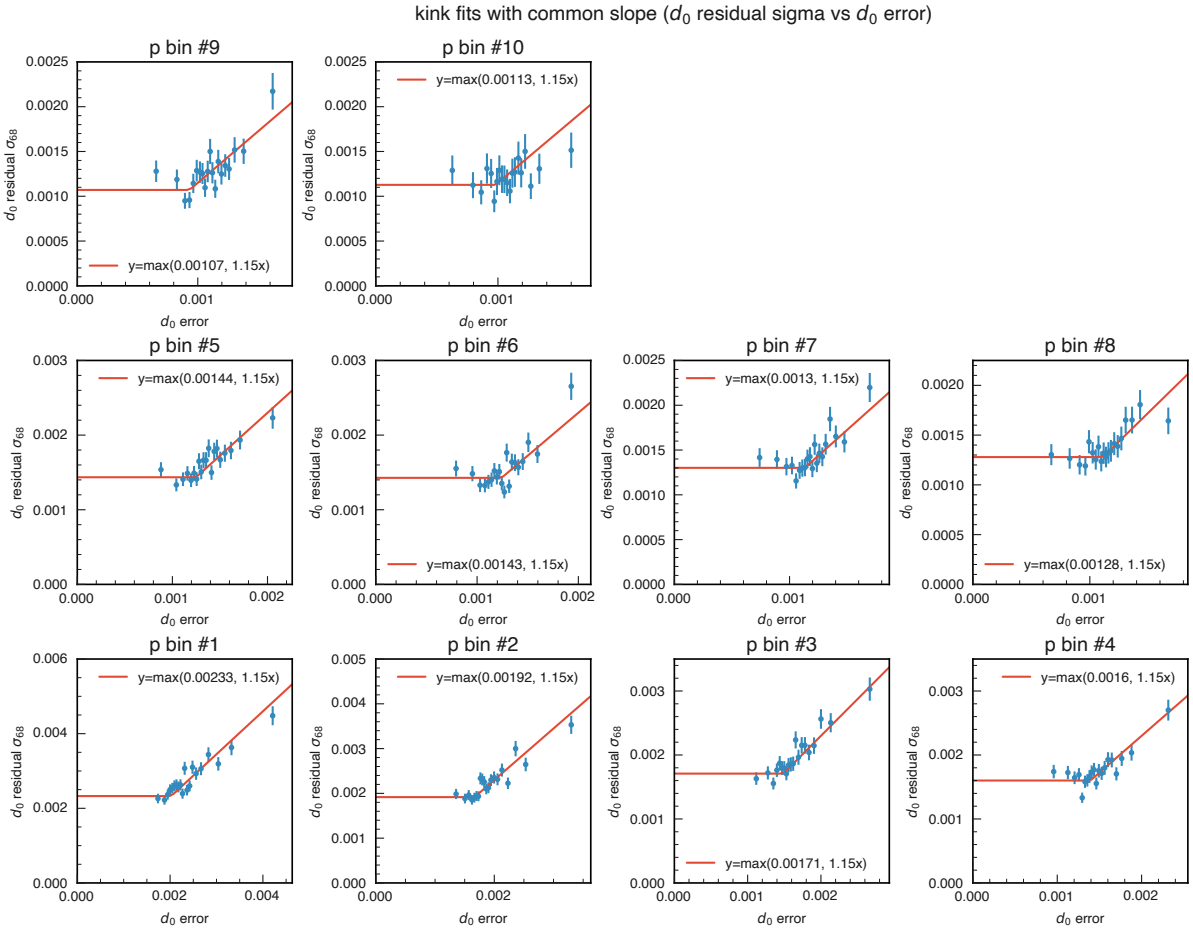


Figure F.4: d_0 resolution vs d_0 uncertainty for prompt tracks in ten pseudo-momentum bins

We generate an MC sample containing one K_S^0 in each event with the following settings:

- uniform momentum between [0.2, 4.0] GeV/c,
- uniform $\cos\theta$ between [17, 150] degrees.

and select displaced K_S^0 daughter tracks that have a Layer-1 hit. We also require that the K_S^0 daughters do not share the common PXD hit. The results of kink function fitting for d_0 and z_0 are shown in Figs. F.7 and F.8. Figure F.9 shows the best resolution. Compared to the prompt tracks, displaced K_S^0 daughter tracks have similar a parameters but significantly smaller b parameters because of the absence of multiple scattering through the beam pipe. The correction parameters for displaced K_S^0 daughters are listed in Tab. F.4

We demonstrate the correction performance by applying the correction to the K_S^0 daughter tracks in the MC sample. Figure F.10 shows the helix pull width as a function of its uncertainty for the K_S^0 daughter tracks produced inside the beam pipe, for which we

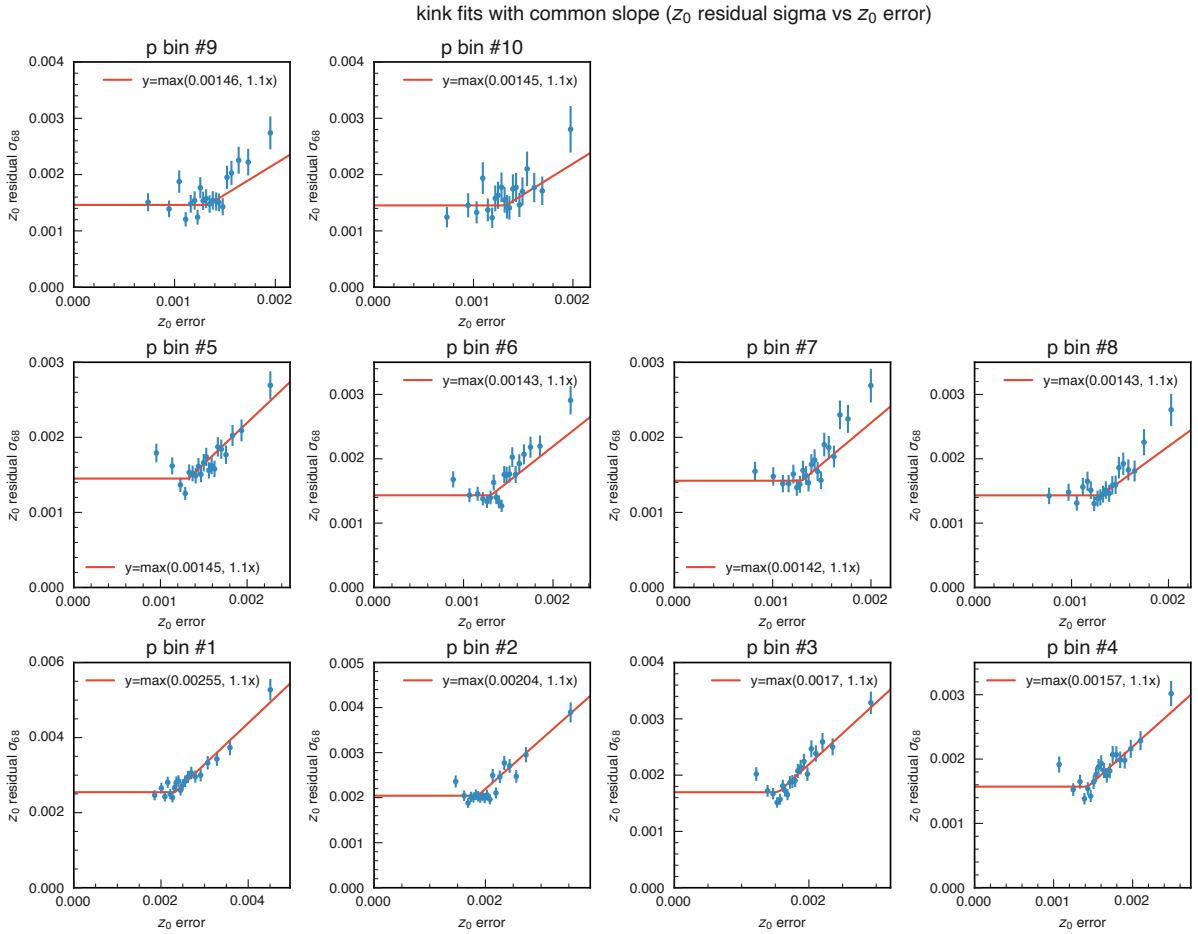


Figure F.5: z_0 resolution vs z_0 uncertainty for prompt tracks in ten pseudo-momentum bins

apply the correction factors for prompt tracks. Figure F.11 shows the result of correction for the displaced K^0_S daughter tracks.

F.2.3 Data-MC correction

For the application to real data, we consider the data-MC difference of the correction parameters using cosmic data and MC samples. The cosmic data were taken during beam collisions and processed with two calibration conditions, "prompt" corresponding to initial processing and "proc12" reprocessed with detector alignment calibration. In these samples, cosmic rays are separately reconstructed as upper and lower tracks and the tracks are required to have a hit in PXD and four in SVD. Impact parameter resolution and helix parameter pull are evaluated by comparison of the upper and lower tracks. Residual and pull of a helix parameter are defined as $(h^{\rm up} - h^{\rm down})/\sqrt{2}$ and $(h^{\rm up} - h^{\rm down})/\sqrt{\varepsilon^2_{h,\rm up} + \varepsilon^2_{h,\rm down}}$, respectively.

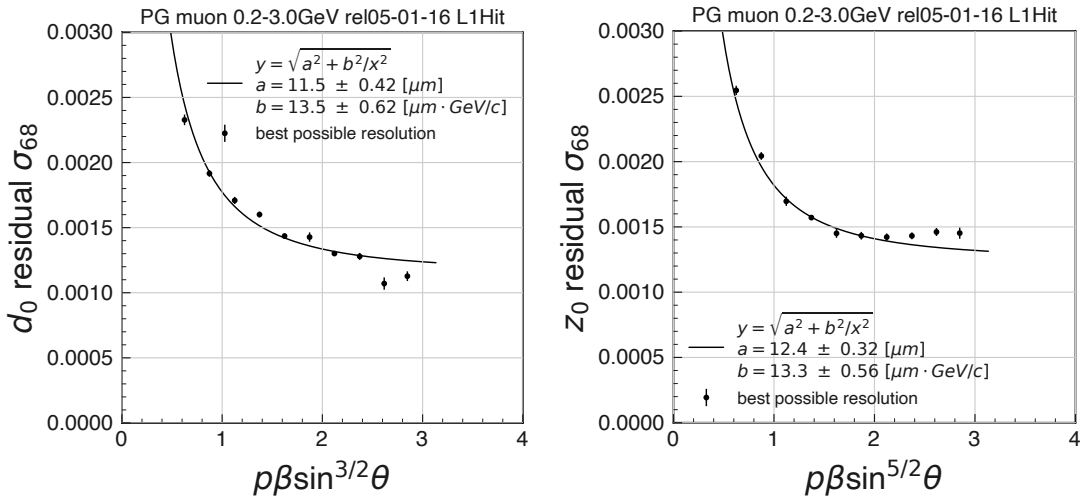


Figure F.6: d_0 (left) and z_0 (right) best resolution vs pseudo-momentum for prompt tracks

(1) Impact parameter resolution

In terms of impact parameter resolution, we are interested in cosmic tracks affected by multiple scattering through the beam pipe as ordinary B meson daughter tracks coming from IP and reside within the detector acceptance. Therefore, the following selections are applied to the cosmic events:

- $|z_0| > 0.3\,\mathrm{cm}$, rejecting background tracks coming from IP
- $-2 < z_0 < 4\,\mathrm{cm}$, rejecting tracks which penetrate titanium parts of the beam pipe and are significantly affected by multiple scattering
- $|d_0| < 1\,\mathrm{cm}$, requiring tracks to be scattered through the beam pipe.

The latter two selections significantly change the impact parameter resolution. Figure F.12 shows the impact parameter resolution in each cosmic sample fitted by Eq. (6.3). The difference in the fit parameters between data and MC is summarized in Tab. F.1. We use the results to obtain the correction parameters for data,

$$a^2_{\mathrm{data}} = a^2_{\mathrm{MC}} + (a^2_{\mathrm{cosmic,data}} - a^2_{\mathrm{cosmic,MC}})$$
$$b_{\mathrm{data}} = b_{\mathrm{MC}} \times \frac{b_{\mathrm{cosmic,data}}}{b_{\mathrm{cosmic,MC}}},$$

where a_{MC} and b_{MC} are the correction parameters obtained in Sec. F.2.1 and $(a^2_{\mathrm{cosmic,data}} - a^2_{\mathrm{cosmic,MC}})$ and $\frac{b_{\mathrm{cosmic,data}}}{b_{\mathrm{cosmic,MC}}}$ the data-MC difference in impact parameter resolution.

(2) Scaling factors

Figure F.13 shows the pull distributions of helix parameters obtained from the cosmic samples. The data-MC difference of scaling factors c is estimated as that of the pull

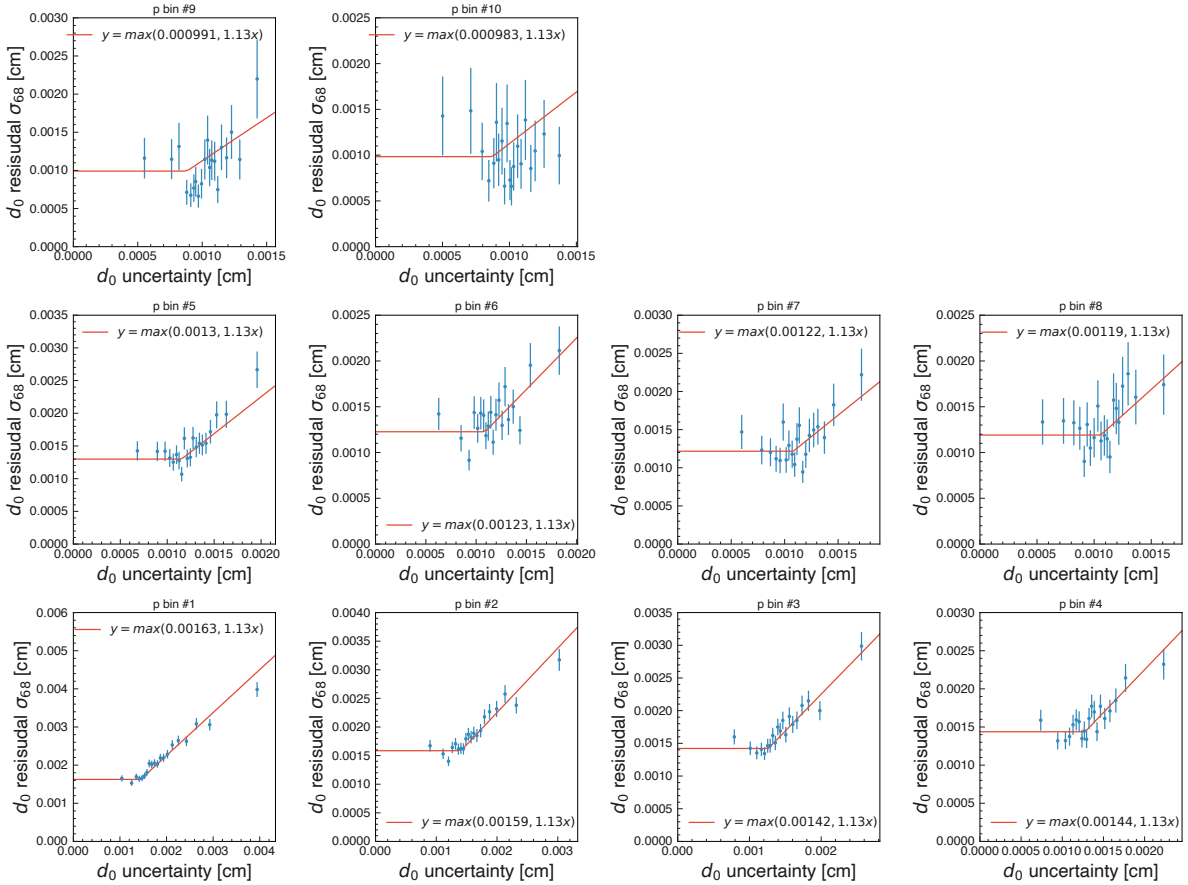


Figure F.7: d_0 resolution vs d_0 uncertainty for displaced K_S^0 daughter tracks in ten pseudo-momentum bins

widths. The pull widths of d_0 and z_0 are calculated in a low pseudo-momentum region $[0.5, 1.0]\,\mathrm{GeV}/c$, where the uncertainty is not too much underestimated. Using the results in Tab. F.2, the scaling factors for data are obtained as

$$c_{\mathrm{data}} = c_{\mathrm{MC}} \times \frac{c_{\mathrm{cosmic,data}}}{c_{\mathrm{cosmic,MC}}}, \tag{6.5}$$

where c_{MC} is the scaling factor in Sec. F.2.1 and $\frac{c_{\mathrm{cosmic,data}}}{c_{\mathrm{cosmic,MC}}}$ is the ratio of pull width.

F.2.4 Summary of correction parameters

Tables F.3 and F.4 summarizes the correction parameters obtained from MC and corrected by the data-MC difference.

	d_0	z_0
$\sqrt{a^2_{\mathrm{proc12}} - a^2_{\mathrm{MC}}}$ [μm]	4.0	5.3
$\sqrt{a^2_{\mathrm{prompt}} - a^2_{\mathrm{MC}}}$ [μm]	6.9	7.9
$b_{\mathrm{proc12}}/b_{\mathrm{MC}}$	1.034	1.040
$b_{\mathrm{prompt}}/b_{\mathrm{MC}}$	1.027	1.032

Table F.1: Difference in impact parameter resolution between cosmic data and MC

	d_0	ϕ_0	ω	z_0	$\tan\lambda$
proc12	1.014	0.899	1.138	1.025	0.894
prompt	1.043	0.908	1.138	1.034	0.898

Table F.2: Ratio of the width of helix pull distribution between cosmic data and MC

data set	c_{d_0}	c_{ϕ_0}	c_{ω}	c_{z_0}	$c_{\tan\lambda}$	a_{d_0} [μm]	b_{d_0} [$\mu\mathrm{m}\cdot$ GeV/c]	a_{z_0} [μm]	b_{z_0} [$\mu\mathrm{m}\cdot$ GeV/c]
MC	1.150	1.086	1.152	1.096	1.087	11.5	13.5	12.4	13.3
proc12	1.166	0.976	1.310	1.124	0.971	12.2	13.9	13.5	13.8
prompt	1.199	0.986	1.310	1.134	0.976	13.4	13.8	14.8	13.7

Table F.3: Summary of the correction parameters for prompt tracks

data set	c_{d_0}	c_{ϕ_0}	c_{ω}	c_{z_0}	$c_{\tan\lambda}$	a_{d_0} [μm]	b_{d_0} [$\mu\mathrm{m}\cdot$ GeV/c]	a_{z_0} [μm]	b_{z_0} [$\mu\mathrm{m}\cdot$ GeV/c]
MC	1.126	1.059	1.206	1.067	1.048	11.74	7.79	13.5	0.00058
proc12	1.142	0.952	1.372	1.093	0.936	12.41	8.057	14.5	6.062
prompt	1.174	0.962	1.372	1.104	0.941	13.6	7.997	15.66	6.015

Table F.4: Summary of the correction parameters for displaced K^0_S daughter tracks

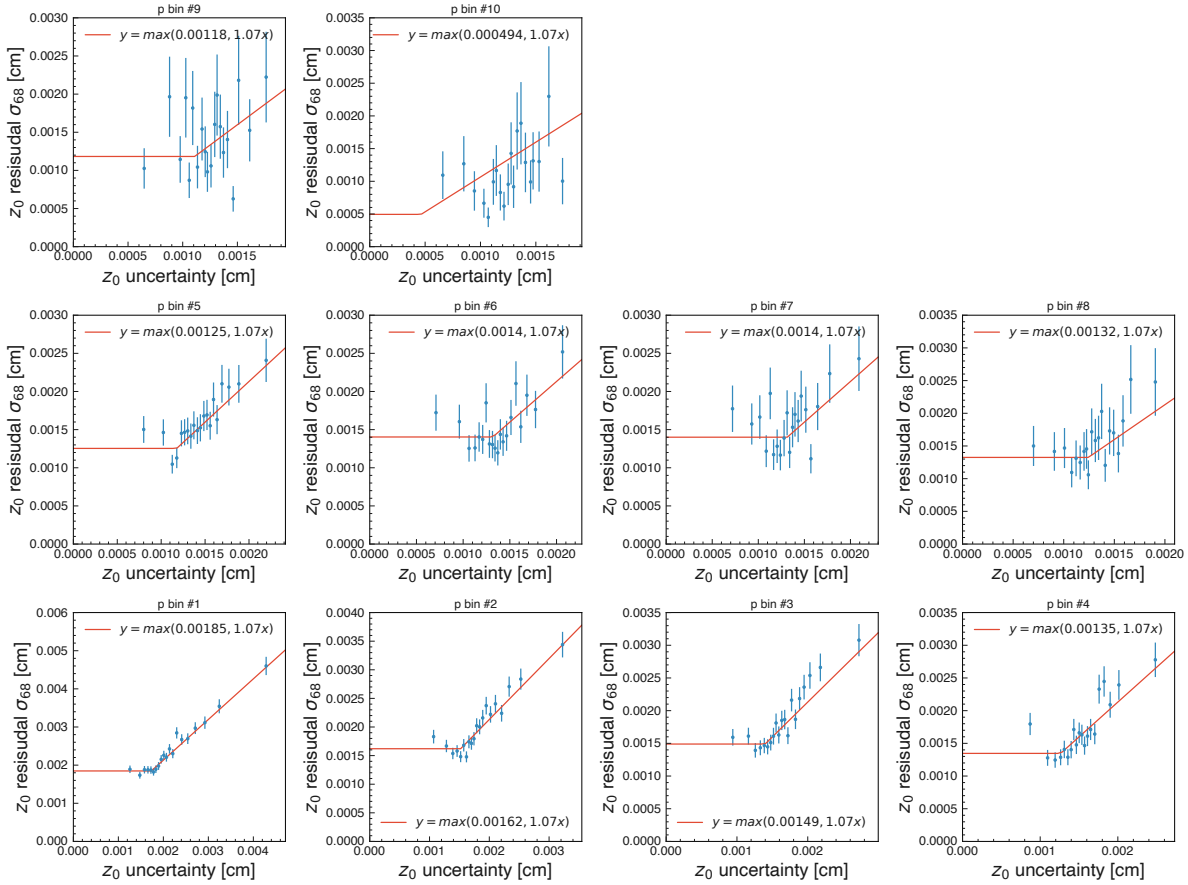


Figure F.8: z_0 resolution vs z_0 uncertainty for displaced K_S^0 daughter tracks in ten pseudo-momentum bins

F.3 Correction to SVD tracks

F.3.1 Parameter determination

Many of K_S^0 daughters do not have PXD hits but have SVD hits because of the long K_S^0 lifetime. Because the source of position uncertainty for the SVD tracks is different from the PXD tracks, they require different correction. We simply apply constant scaling factors to helix parameter uncertainties for SVD tracks, which do not depend on the track momentum. We collect SVD tracks from displaced K_S^0 in $B^0 \to K_S^0 K_S^0 K_S^0$ signal MC to have tracks of reasonably broad momentum range considering that the correction will be applied to K_S^0 produced in $B\overline{B}$ events. We determine the scaling factor for each helix parameter uncertainty as the width of helix pull distribution of the collected SVD tracks. The scaling factors are in Tab. F.6.

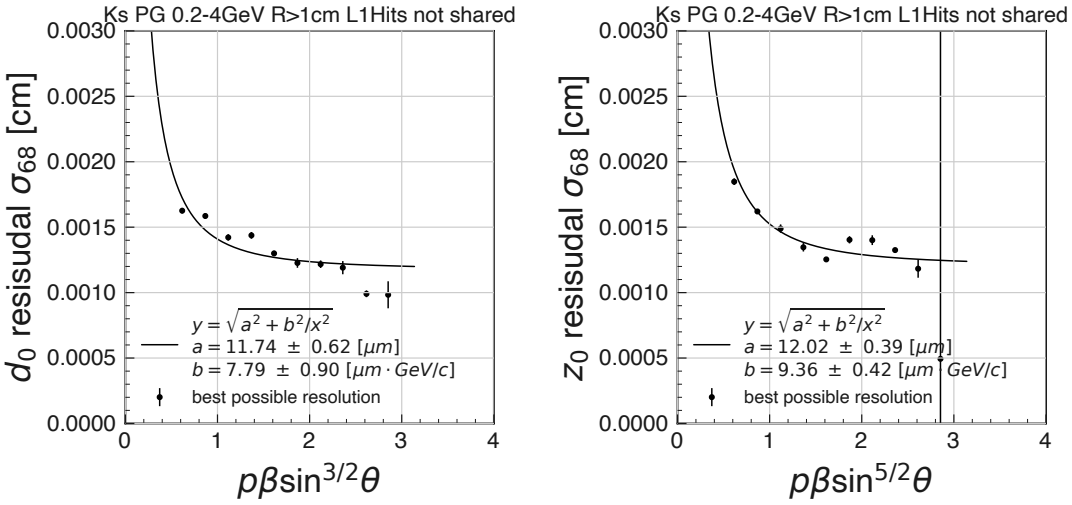


Figure F.9: d_0 (left) and z_0 (right) best resolution vs pseudo-momentum for displaced K_S^0 daughter tracks

d_0	ϕ_0	ω	z_0	$\tan\lambda$
0.922	0.946	1.189	1.000	0.945

Table F.5: Ratio of the width of helix pull distribution between cosmic data and MC

F.3.2 data-MC correction

For data-MC correction of the SVD-track parameters we use cosmic events as in Sec. F.2.3 but reconstruct the tracks without using PXD hits. We require the cosmic tracks

- to have six or more SVD hits (corresponding to three SVD layers),
- $|z_0| > 0.3\,\text{cm}$,
- $-8\,\text{cm} < z_0 < 12\,\text{cm}$,
- $|d_0| < 2.5\,\text{cm}$, and
- $p < 3\,\text{GeV}/c$.

The second selection removes background coming from beam collision, the third removes tracks scattered by PXD mount blocks, and the fourth removes tracks going through Layer 3 in r-ϕ direction that are somehow reconstructed with large helix pull. Figure F.14 shows the pull distribution of helix parameters for the cosmic SVD tracks. We determine the data-MC correction factor as the ratio of the pull widths between data and MC as in Eq. (6.5). The correction factors and the corrected parameters are listed in Tab. F.5 and F.6.

We only define a single set of parameters for data because the c parameters are not sensitive to the calibration as indicated by the data-MC comparison of PXD tracks in Tab. F.2.

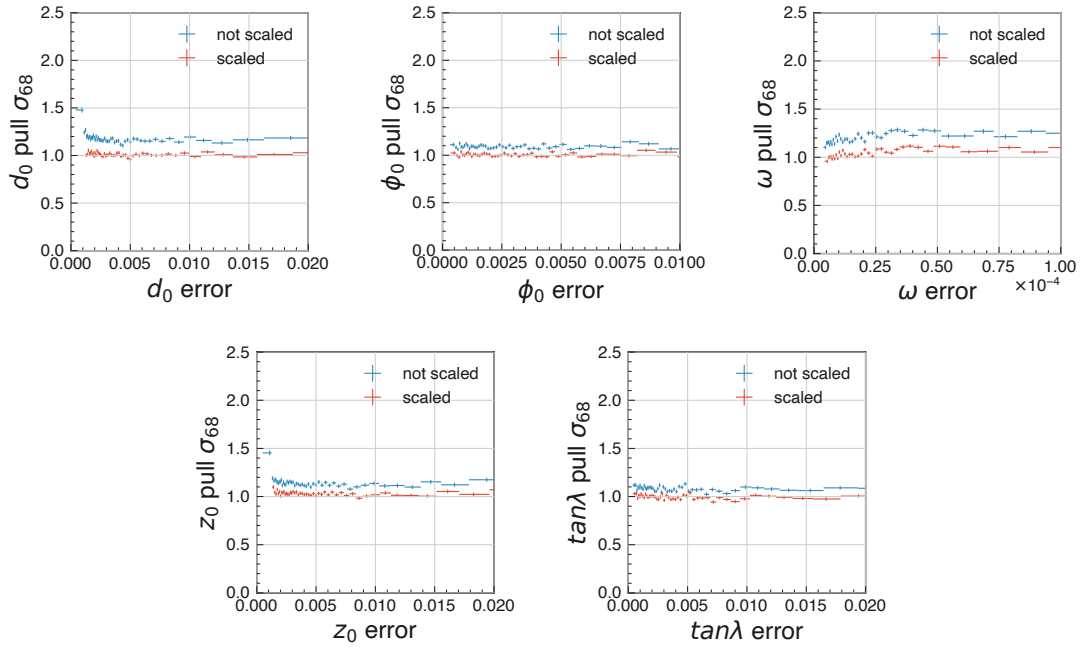


Figure F.10: Width of pull distribution as a function of the uncertainty for each helix parameter with and without helix uncertainty normalization. K^0_S daughter tracks produced inside the beam pipe are selected.

data set	c_{d_0}	c_{ϕ_0}	c_{ω}	c_{z_0}	$c_{\tan\lambda}$
MC	1.079	1.080	1.209	1.088	1.083
data	0.994	1.022	1.438	1.088	1.023

Table F.6: Summary of the correction parameters for SVD tracks

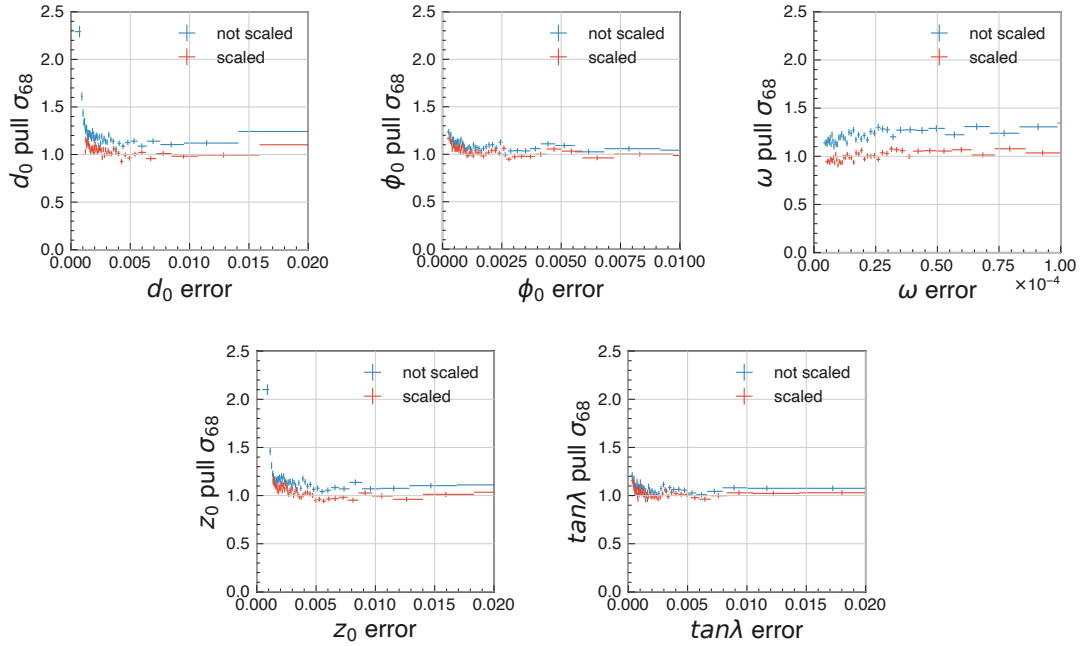


Figure F.11: Width of pull distribution as a function of the uncertainty for each helix parameter with and without helix uncertainty normalization. Displaced K_S^0 daughter tracks are selected.

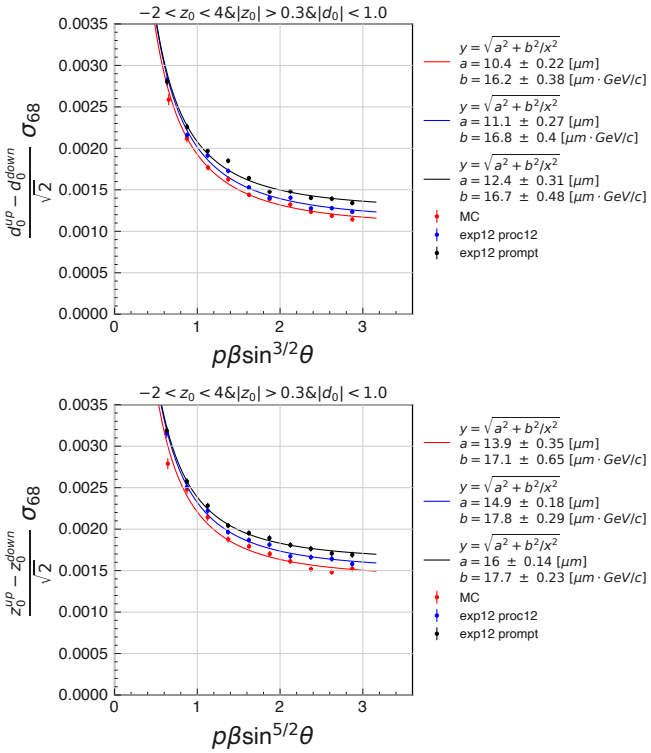


Figure F.12: Impact parameter resolution of cosmic tracks in data and MC samples

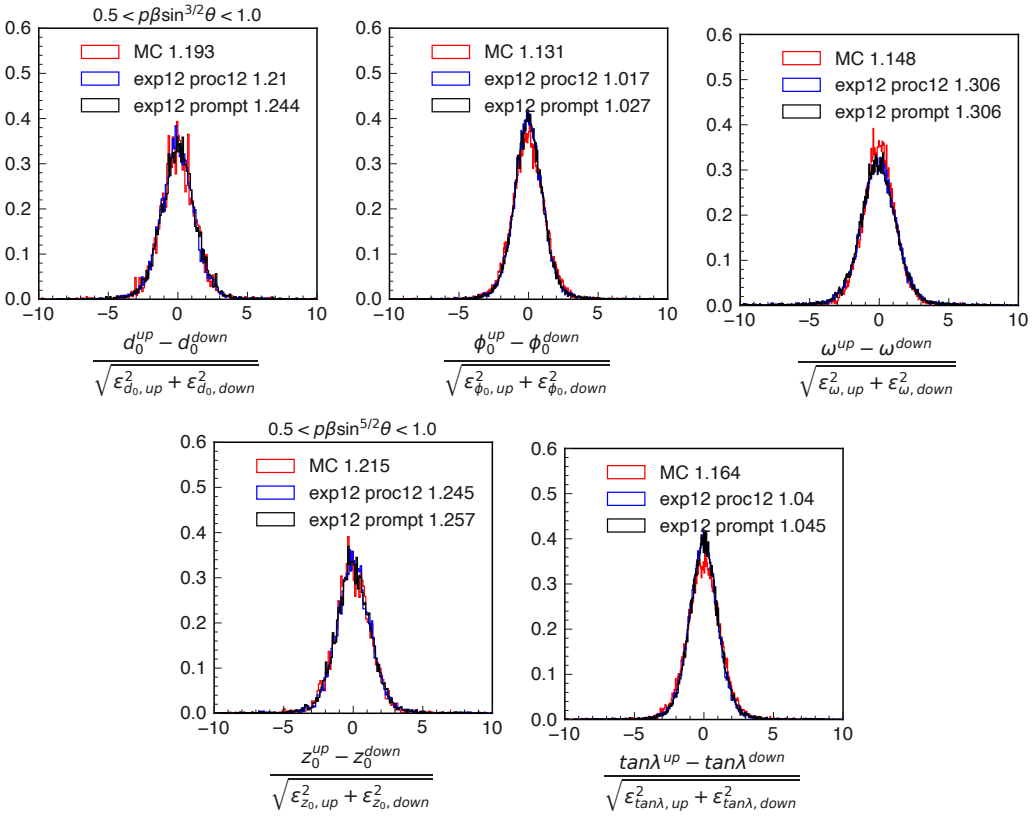


Figure F.13: Pull distributions of helix parameters of PXD tracks in cosmic samples

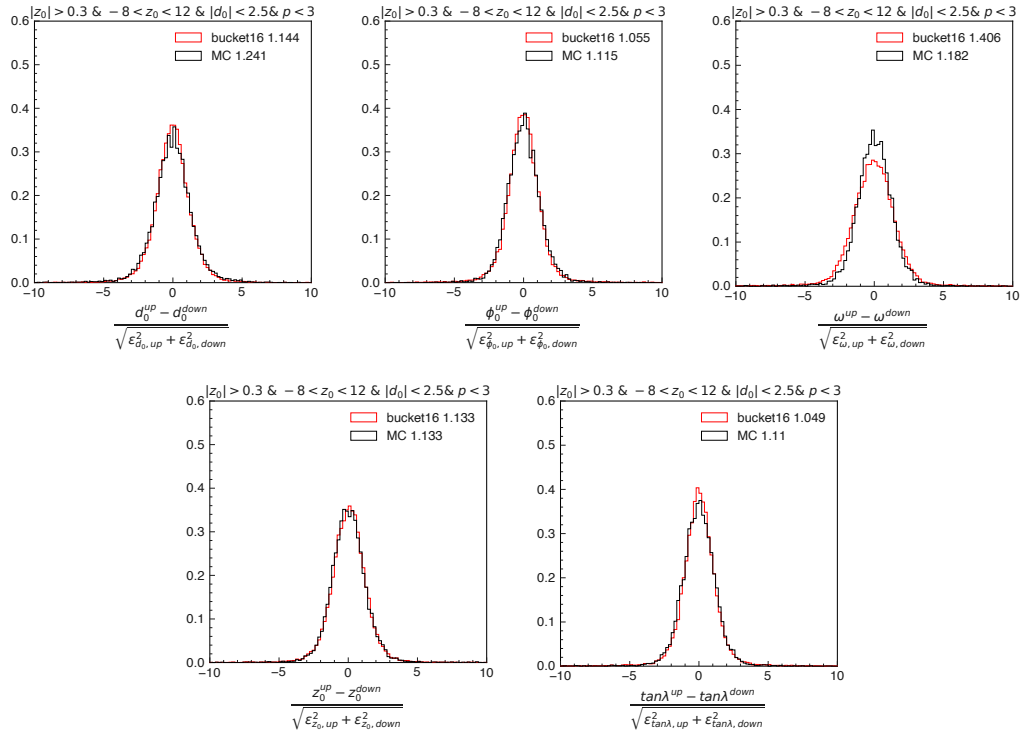


Figure F.14: Pull distributions of helix parameters of SVD tracks in cosmic samples

Appendix G

Feldman-Cousins confidence interval

From the ensemble test in Sec. 5.5.1 we find the *CP* fit results are biased due to low statistics and the uncertainties obtained by MINOS are not fully reliable. We estimate the confidence interval using a frequentist approach based on the Feldman-Cousins likelihood-ratio ordering [41][42]. Confidence interval is defined as the interval where the true value of the parameter of interest exists with a given probability (confidence level).

Given the true *CP* asymmetries $\vec{\theta} = (S_{\rm input}, A_{\rm input})$ as conditional variables, we construct a PDF for measured values $\vec{x} = (S_{\rm fit}, A_{\rm fit})$ as a sum of two two-dimensional Gaussian functions:

$$P(S_{\rm fit}, A_{\rm fit}|S_{\rm input}, A_{\rm input}) = f_m G(S_{\rm fit}; \mu_{mS}, \sigma_{mS}) G(A_{\rm fit}; \mu_{mA}, \sigma_{mA}) \tag{7.1}$$

$$+(1 - f_m) G(S_{\rm fit}; \mu_{tS}, \sigma_{tS}) G(A_{\rm fit}; \mu_{tA}, \sigma_{tA}), \tag{7.2}$$

where

$$f_m = f_{m0} + f_{m2S} S^2_{\rm input} + f_{m2A} A^2_{\rm input}, \tag{7.3}$$

$$\mu_{mS} = \mu_{mS0} + \mu_{mS1S} S_{\rm input} + \mu_{mS3S} S^3_{\rm input}, \tag{7.4}$$

$$\sigma_{mS} = \sigma_{mS0} + \sigma_{mS2S} S^2_{\rm input} + \sigma_{mS2A} A^2_{\rm input}, \tag{7.5}$$

$$\mu_{tS} = \mu_{tS0} + \mu_{tS1S} S_{\rm input}, \tag{7.6}$$

$$\sigma_{tS} = \sigma_{tS0}, \tag{7.7}$$

$$\mu_{mA} = \mu_{mA0} + \mu_{mA1A} A_{\rm input}, \tag{7.8}$$

$$\sigma_{mA} = \sigma_{mA0} + \sigma_{mA2A} A^2_{\rm input}, \tag{7.9}$$

$$\mu_{tA} = \mu_{tA0} + \mu_{tA1A} A_{\rm input}, \text{ and} \tag{7.10}$$

$$\sigma_{tA} = \sigma_{tA0}, \tag{7.11}$$

and the PDF contains 19 parameters. We determine the parameters by an unbinned maximum likelihood fit to the distribution of ($S_{\rm fit}$, $A_{\rm fit}$, $S_{\rm input}$, $A_{\rm input}$) obtained from toy MC samples, where we vary the $S_{\rm input}$ and $A_{\rm input}$ in 0.2 steps within the physical boundary $S^2_{\rm input} + A^2_{\rm input} \leq 1$ as shown in Fig. G.1 and repeat 8000 experiments at each ($S_{\rm input}$, $A_{\rm input}$). Table G.1 show the fit results. As shown in Figs. G.2–G.9 the fitted PDF describes well the distributions of toy MC results.

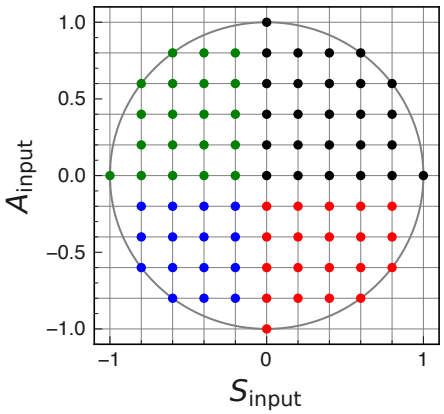


Figure G.1: Input *CP* asymmetries (S_{input}, A_{input}) for toy MC samples

Table G.1: PDF parameters

parameter	value
f_{m0}	$0.8920^{+0.0022}_{-0.0023}$
f_{m2S}	$-0.0788^{+0.0045}_{-0.0046}$
f_{m2A}	$-0.1408^{+0.0045}_{-0.0045}$
μ_{mS0}	$-0.00225^{+0.00083}_{-0.00083}$
μ_{tS0}	$0.0027^{+0.0043}_{-0.0043}$
μ_{mA0}	$0.00072^{+0.00042}_{-0.00042}$
μ_{tA0}	$-0.0063^{+0.0016}_{-0.0016}$
μ_{mS1S}	$0.9698^{+0.0038}_{-0.0038}$
μ_{mS3S}	$0.0705^{+0.0068}_{-0.0068}$
μ_{tS1S}	$2.047^{+0.012}_{-0.012}$
μ_{mA1A}	$0.99590^{+0.00087}_{-0.00088}$
μ_{tA1A}	$1.1967^{+0.0031}_{-0.0031}$
σ_{mS0}	$0.5228^{+0.0015}_{-0.0015}$
σ_{mS2S}	$0.0092^{+0.0033}_{-0.0033}$
σ_{mS2A}	$-0.0187^{+0.0016}_{-0.0016}$
σ_{tS0}	$1.1476^{+0.0040}_{-0.0039}$
σ_{mA0}	$0.27023^{+0.00045}_{-0.00045}$
σ_{mA2A}	$-0.0417^{+0.0012}_{-0.0012}$
σ_{tA0}	$0.3538^{+0.0012}_{-0.0011}$

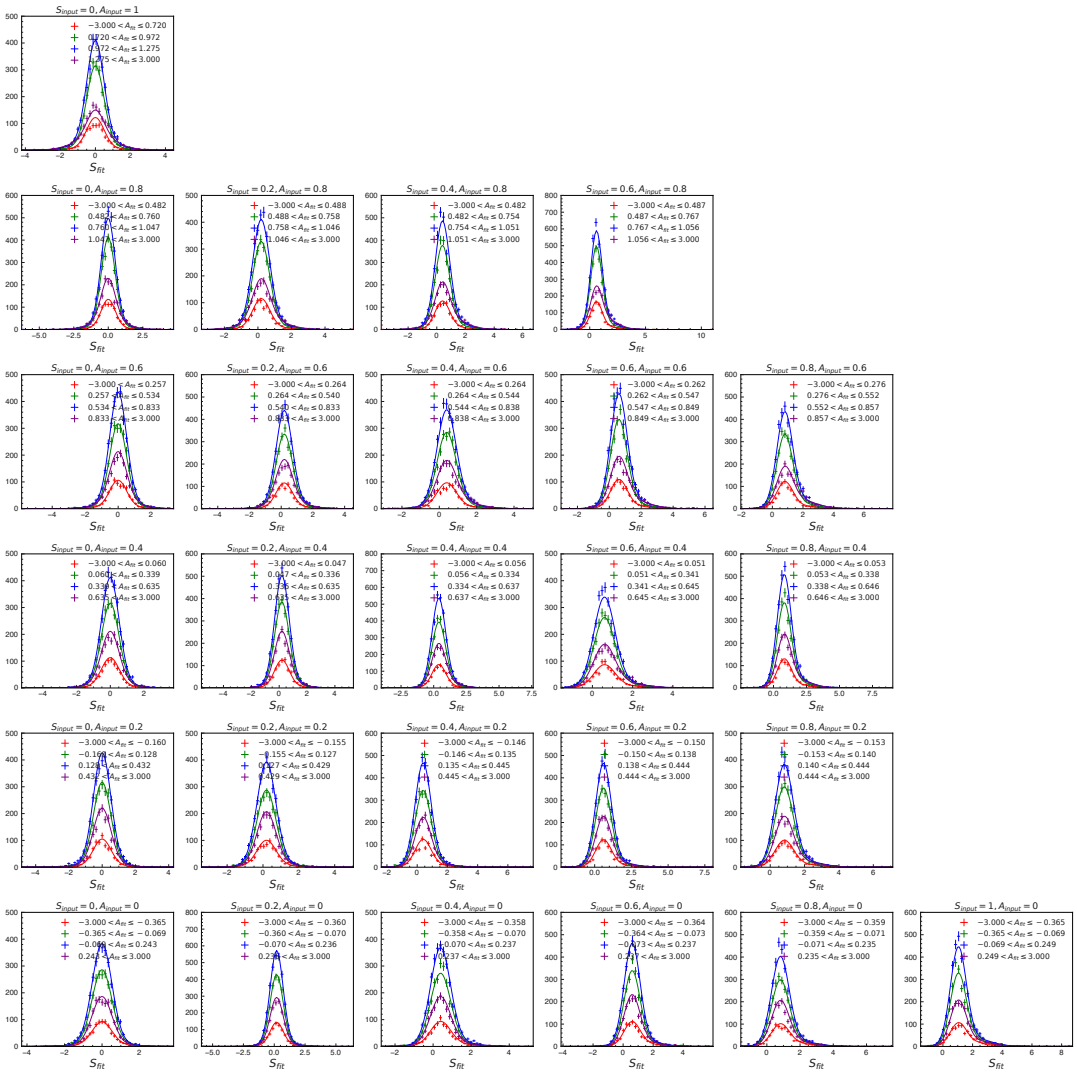


Figure G.2: Fitted PDF and toy MC distribution projected onto S_{fit} in arbitrary bins of A_{fit} with (S_{input}, A_{input}) at black dots in Fig. G.1

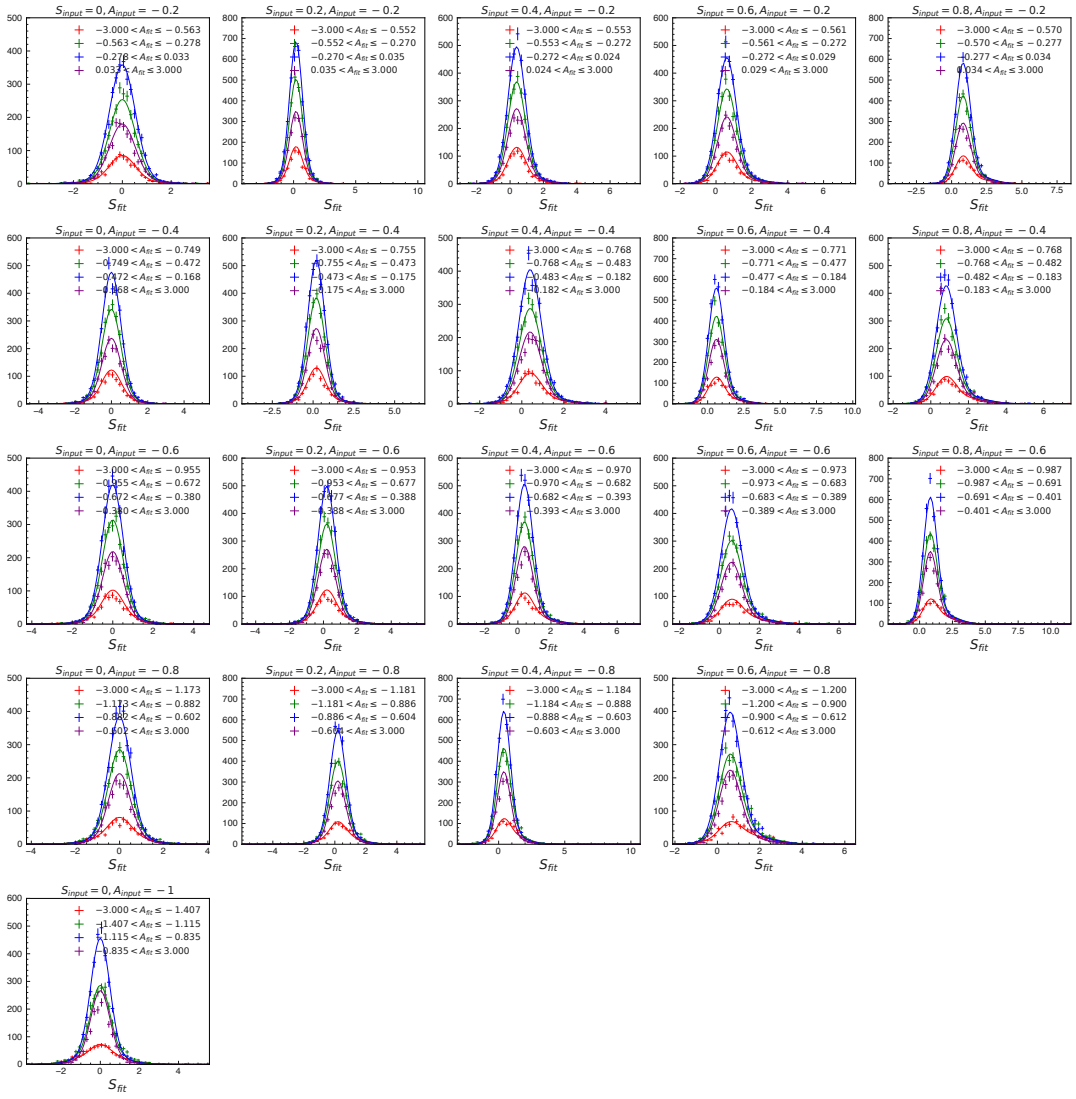


Figure G.3: Fitted PDF and toy MC distribution projected onto S_{fit} in arbitrary bins of A_{fit} with (S_{input}, A_{input}) at red dots in Fig. G.1

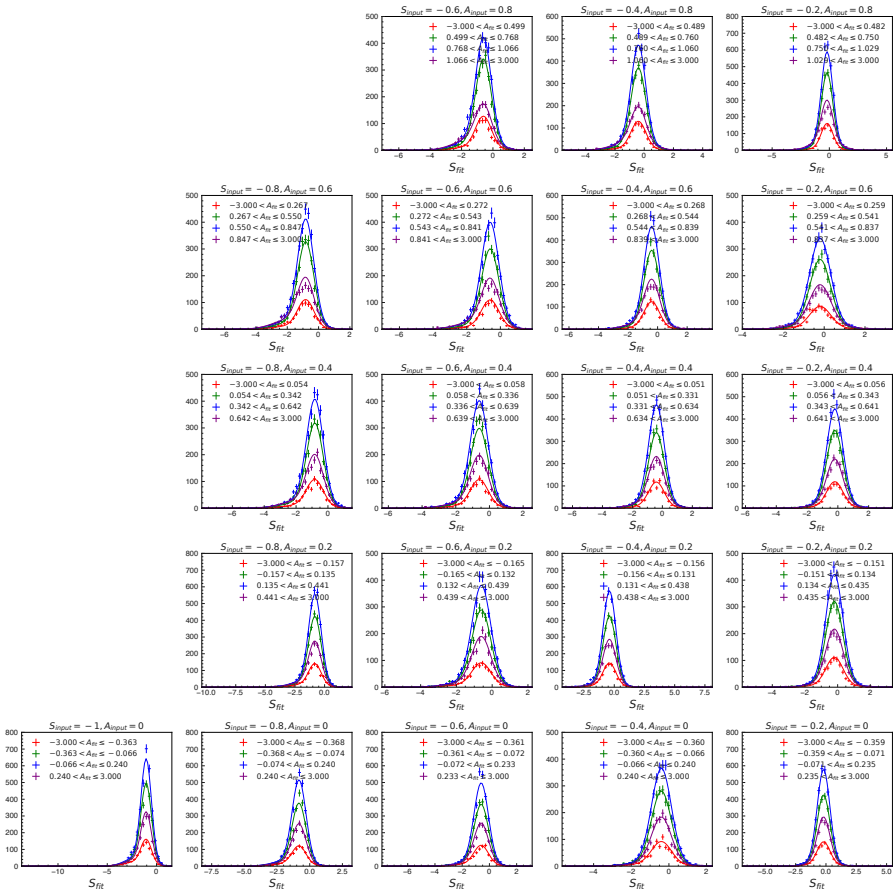


Figure G.4: Fitted PDF and toy MC distribution projected onto $S_{\rm fit}$ in arbitrary bins of $A_{\rm fit}$ with ($S_{\rm input}$, $A_{\rm input}$) at green dots in Fig. G.1

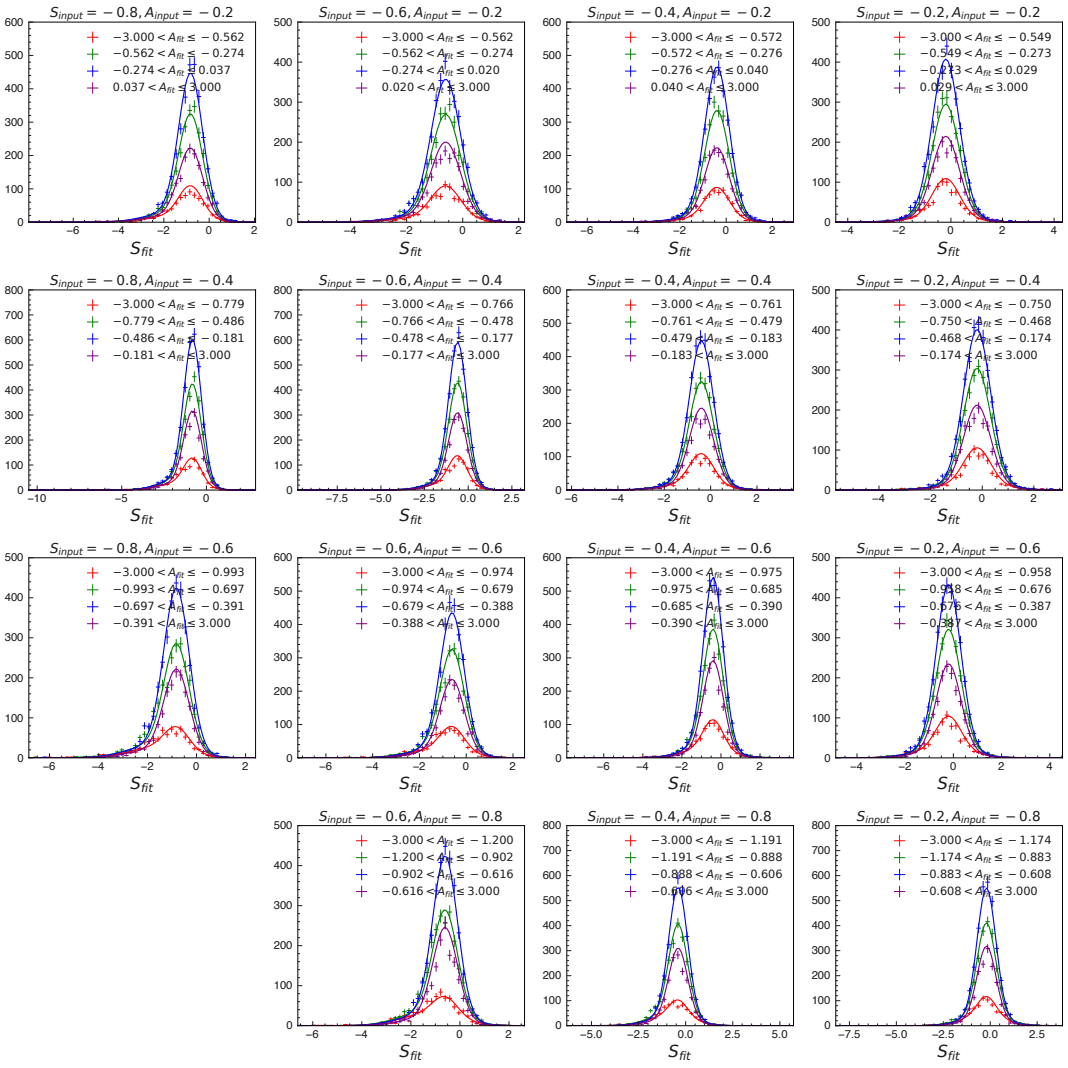


Figure G.5: Fitted PDF and toy MC distribution projected onto S_{fit} in arbitrary bins of A_{fit} with (S_{input}, A_{input}) at blue dots in Fig. G.1

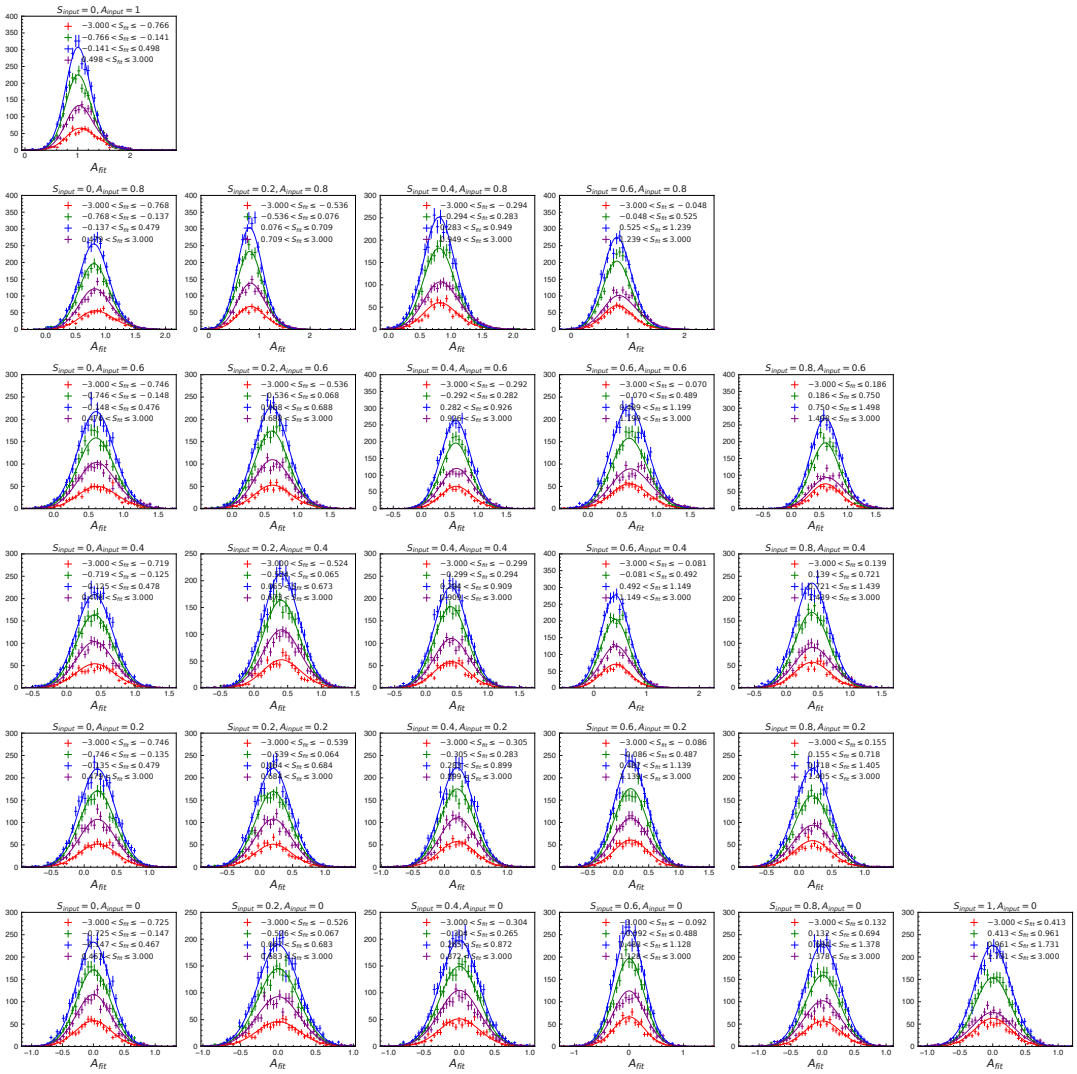


Figure G.6: Fitted PDF and toy MC distribution projected onto $A_{\rm fit}$ in arbitrary bins of $S_{\rm fit}$ with ($S_{\rm input}$, $A_{\rm input}$) at black dots in Fig. G.1

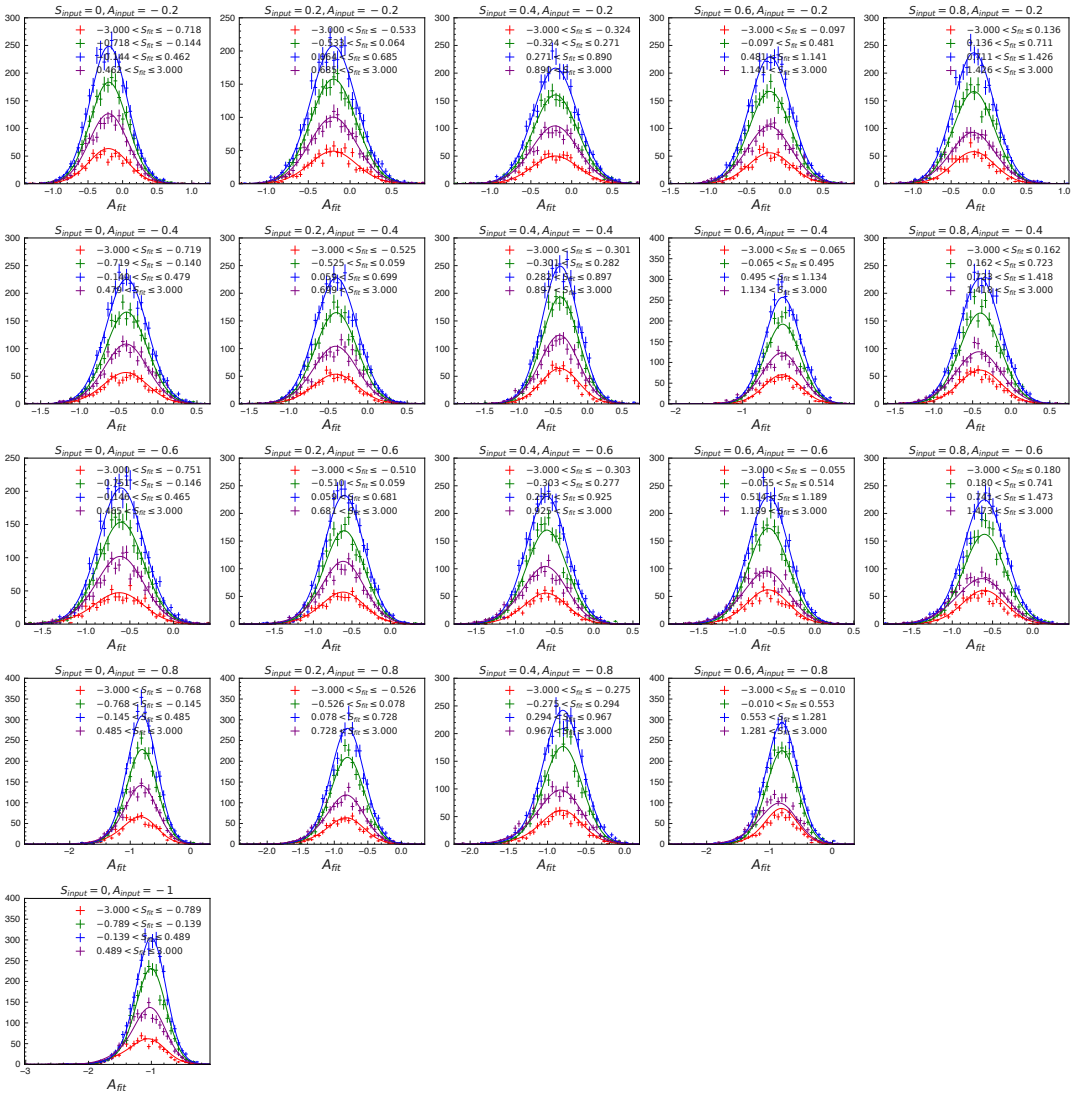


Figure G.7: Fitted PDF and toy MC distribution projected onto A_{fit} in arbitrary bins of S_{fit} with (S_{input}, A_{input}) at red dots in Fig. G.1

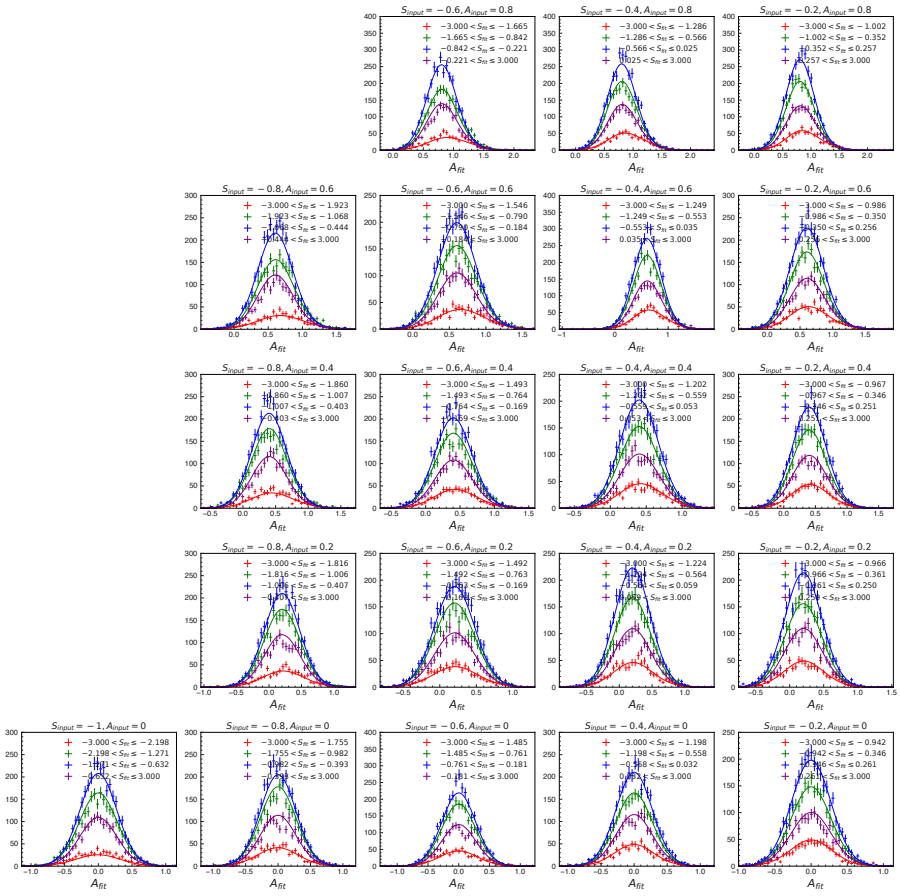


Figure G.8: Fitted PDF and toy MC distribution projected onto $A_{\rm fit}$ in arbitrary bins of $S_{\rm fit}$ with ($S_{\rm input}$, $A_{\rm input}$) at green dots in Fig. G.1

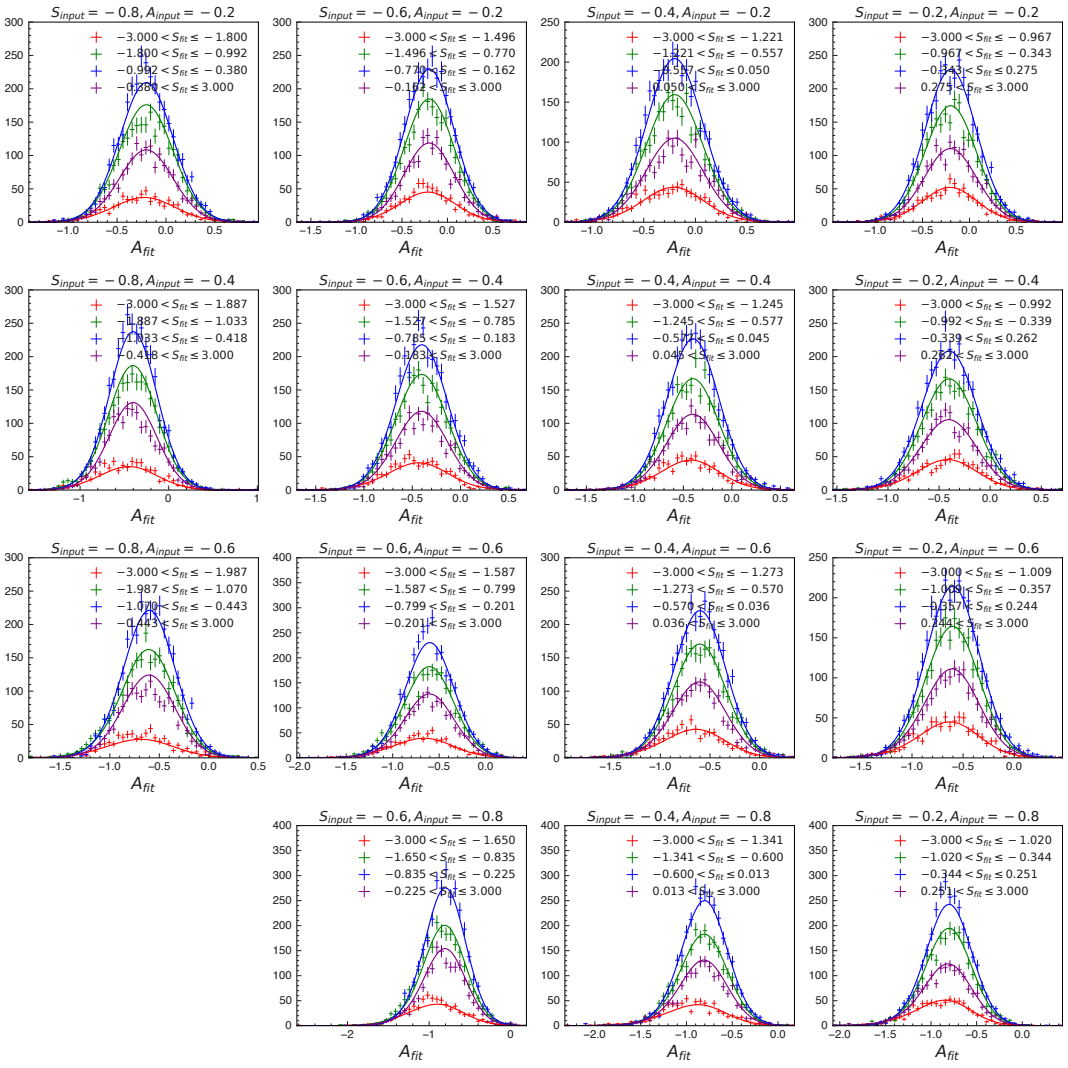


Figure G.9: Fitted PDF and toy MC distribution projected onto A_{fit} in arbitrary bins of S_{fit} with (S_{input}, A_{input}) at blue dots in Fig. G.1

Once we measure a set of CP asymmetries on a data sample $\vec{x}_0 = (S_0, A_0)$, the confidence level to reject an arbitrary set of CP asymmetries $\vec{\theta}$ is

$$\int_{LR(\vec{x_0}|\vec{\theta}) \geq LR(\vec{x}|\vec{\theta})} d\vec{x} P(\vec{x}|\vec{\theta}), \tag{7.12}$$

where LR is likelihood ratio defined as

$$LR(\vec{x}|\vec{\theta}) = \frac{P(\vec{x}|\vec{\theta})}{P(\vec{x}|\vec{\theta}_{\rm best})}, \tag{7.13}$$

using $\vec{\theta}_{\rm best}(\vec{x})$ which maximizes $P(\vec{x}|\vec{\theta})$ for given $\vec{x}$. Due to the complexity of the PDF, we numerically compute $\vec{\theta}_{\rm best}(\vec{x})$ using MIGRAD. Figure G.10 shows the distribution of $P(\vec{x}|\vec{\theta}_{\rm best})$. $P(\vec{x}|\vec{\theta}_{\rm best})$ corresponds to the outermost line of the PDF when $S_{\rm input}$ and $A_{\rm input}$ are varied as shown in Fig. G.11. In the $S_{\rm fit}$ direction, the PDF has a broader tail when $|S_{\rm input}| \sim 1$ so that the height of the main peak gets lower. On the other hand, in the $A_{\rm fit}$ direction, the PDF gets narrower when $|A_{\rm input}|$ is larger so the outline is downwardly convex at the center.

To draw confidence contours for the measurement $\vec{x}_0$, we compute the integral of Eq. (7.12) at arbitrary values of $\vec{\theta}$ with Monte Carlo method, where we pseudorandomly generate $\vec{x}$ based on the PDF and estimate the probability for $\vec{x}$ to fulfill $LR(\vec{x_0}|\vec{\theta}) \geq LR(\vec{x}|\vec{\theta})$. Figure G.12 shows an example of the contours assuming $\vec{x}_0 = (-1.1, 0.5)$.

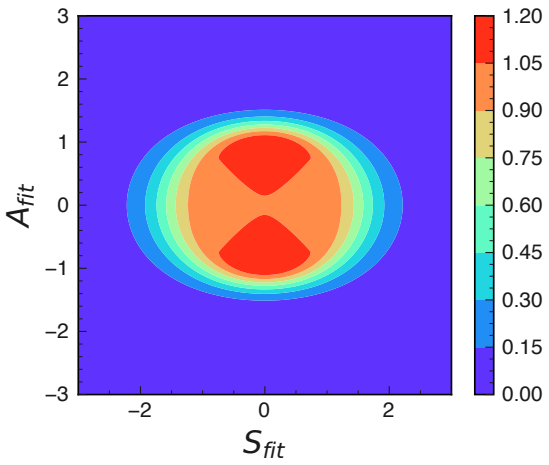


Figure G.10: $P(\vec{x}|\vec{\theta}_{\text{best}})$

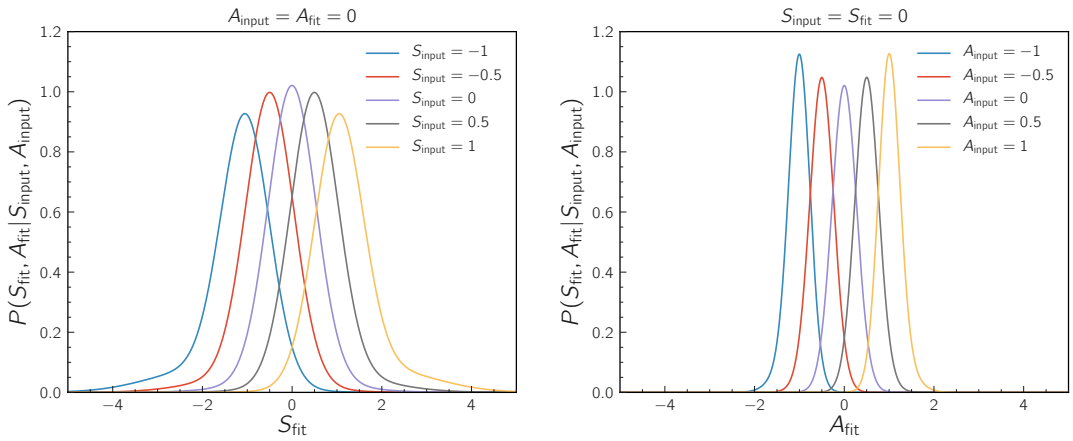


Figure G.11: $P(S_{\text{fit}}, A_{\text{fit}}|S_{\text{input}}, A_{\text{input}})$ projected to S_{fit} (left) and A_{fit} (right). In the left (right) plot, S_{input} (A_{input}) is varied while A_{fit} and A_{input} (S_{fit} and S_{input}) are fixed to zero.

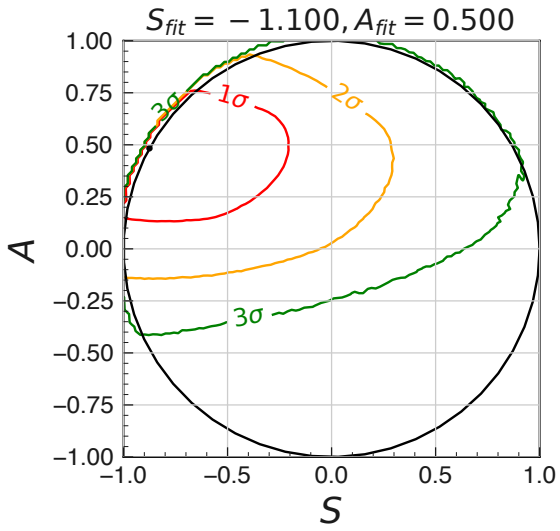


Figure G.12: Example of confidence contours assuming a fit results of $S_0 = -1.1$ and $A_0 = 0.5$, which excludes the phase space at 1σ, 2σ, and 3σ confidence levels

Appendix H

Data-MC comparison of K^0_S BDT selection efficiency

Comparing the expected and observed number of events in Tabs. 4.4 and A.1, we find around 30% deficit in data both for $B^0 \to K^0_S K^0_S K^0_S$ and $B^+ \to K^0_S K^0_S K^+$ decays. Motivated by the difference, we study the efficiency of K^0_S BDT selection, $\mathcal{O}_{K^0_S} > 0.75$, in data.

We fit the following function to K^0_S mass distribution:

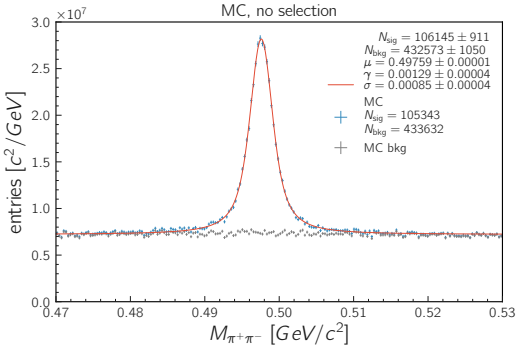
$$f(M_{\pi^+\pi^-}) = s \cdot V(M_{\pi^+\pi^-}; \mu, \sigma, \gamma) + b, \tag{8.1}$$

where $V(x; \mu, \sigma, \gamma) = \int_{-\infty}^{\infty} G(x - x'; \mu, \sigma) \frac{\gamma}{\pi(x'^2 + \gamma^2)} dx'$ is the Voigt function and s, b, μ, σ, and γ are free parameters. The number of true K^0_S, $N_{\rm sig}$, is calculated as the integral of the Voigt function within the fit range. We define the efficiency as the ratio of $N_{\rm sig}$ with and without the K^0_S BDT selection.

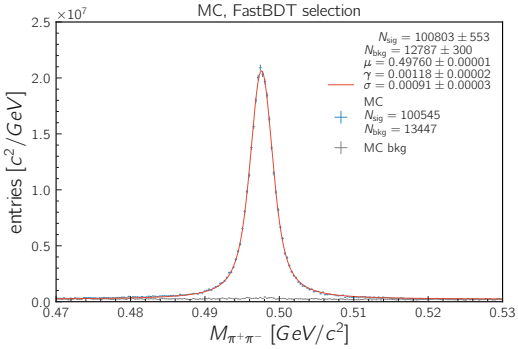
For the K^0_S efficiency study, we use $B^0 \to K^0_S K^0_S K^0_S$ sample with only the $M_{\rm bc}$ and M selection and $470 < M_{\pi^+\pi^-}[\,{\rm MeV}/c^2] < 530$. The K^0_S mass range justifies the flat background distribution in MC. Figure H.1 show the fit results to data and MC distributions with the K^0_S BDT selection applied or not. The estimated K^0_S selection efficiencies in data and MC match each other and that obtained by counting MC truth information as shown in Tab. H.1. As the efficiency is found to be similar, we attribute the data-MC difference to imperfect simulation of $q\bar{q}$ fragmentation in MC.

Table H.1: K^0_S BDT selection efficiency in data and MC

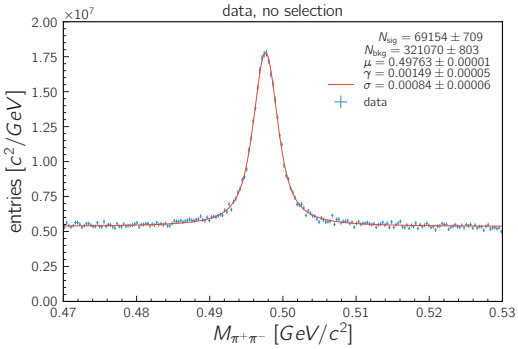
sample	$N_{\rm sig}$ without selection	$N_{\rm sig}$ with selection	efficiency
MC count	105343	100545	95.4 %
MC fit	106145 ± 911	100803 ± 553	95.0 ± 0.9 %
data fit	69154 ± 709	66942 ± 424	96.8 ± 1.1%



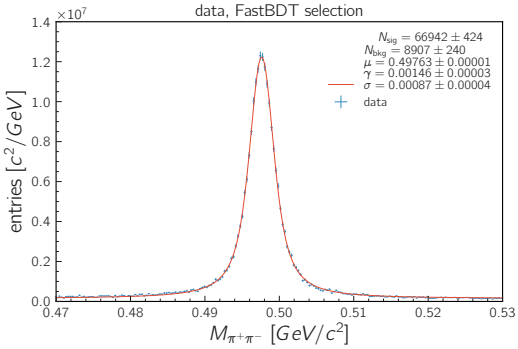
(a) MC without selection



(b) MC with BDT selection



(c) Data without selection



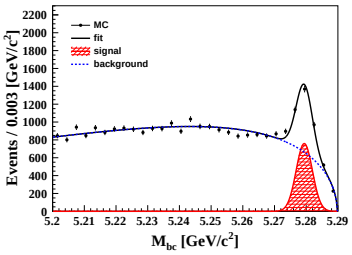
(d) Data with BDT selection

Figure H.1: K_S^0 mass distributions and fitted functions

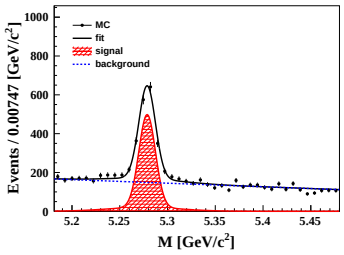
Appendix I

Fit results from validation using generic MC sample

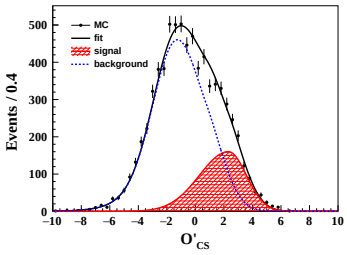
We show the fitted distributions from the validation study using the generic MC sample in Sec. 5.5.3.



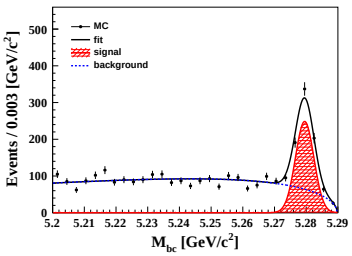
(a) $B^+ \to K^0_S K^0_S K^+$, $M_{\rm bc}$ in signal region of M



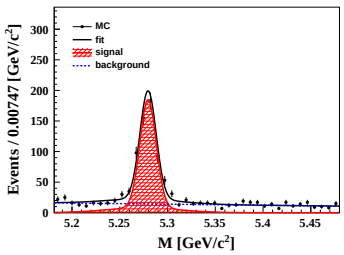
(b) $B^+ \to K^0_S K^0_S K^+$, M in signal region of $M_{\rm bc}$



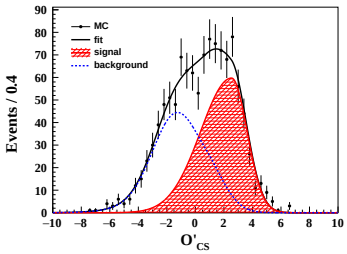
(c) $B^+ \to K^0_S K^0_S K^+$, $\mathcal{O}'_{CS}$ in signal region of $M_{\rm bc}$



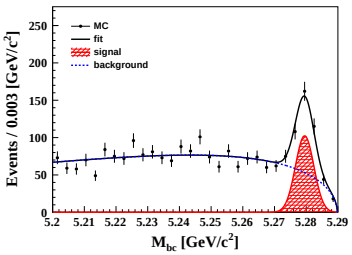
(d) $B^0 \to K^0_S K^0_S K^0_S$ TD event, $M_{\rm bc}$ in signal region of M



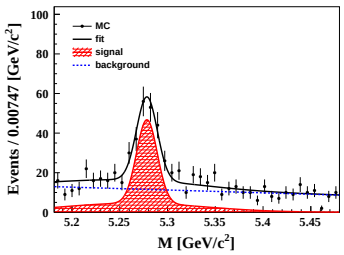
(e) $B^0 \to K^0_S K^0_S K^0_S$ TD event, M in signal region of $M_{\rm bc}$



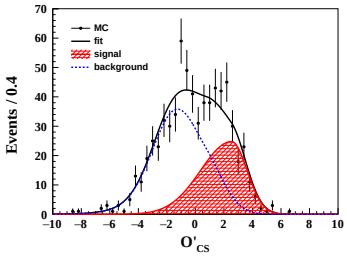
(f) $B^0 \to K^0_S K^0_S K^0_S$ TD event, $\mathcal{O}'_{CS}$ in signal region of $M_{\rm bc}$



(g) $B^0 \to K^0_S K^0_S K^0_S$ TI event, $M_{\rm bc}$ in signal region of M

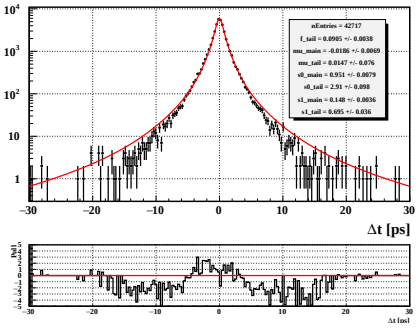


(h) $B^0 \to K^0_S K^0_S K^0_S$ TI event, M in signal region of $M_{\rm bc}$

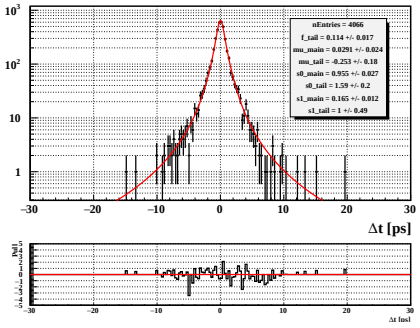


(i) $B^0 \to K^0_S K^0_S K^0_S$ TI event, $\mathcal{O}'_{CS}$ in signal region of $M_{\rm bc}$

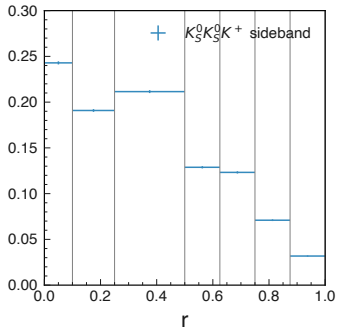
Figure I.1: Results of signal extraction fits to generic MC



(a) $B^+ \to K^0_S K^0_S K^+$ sideband Δt distribution

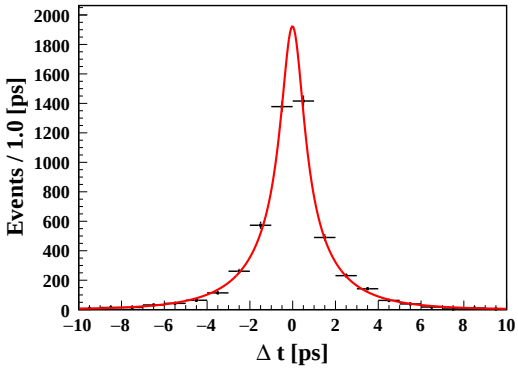


(b) $B^0 \to K^0_S K^0_S K^0_S$ sideband Δt distribution

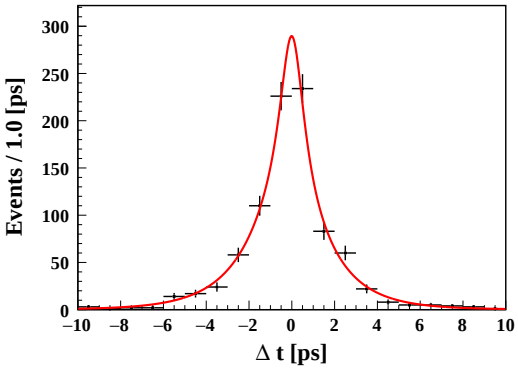


(c) $B^+ \to K^0_S K^0_S K^+$ sideband r bin distribution

Figure I.2: Sideband distributions in generic MC

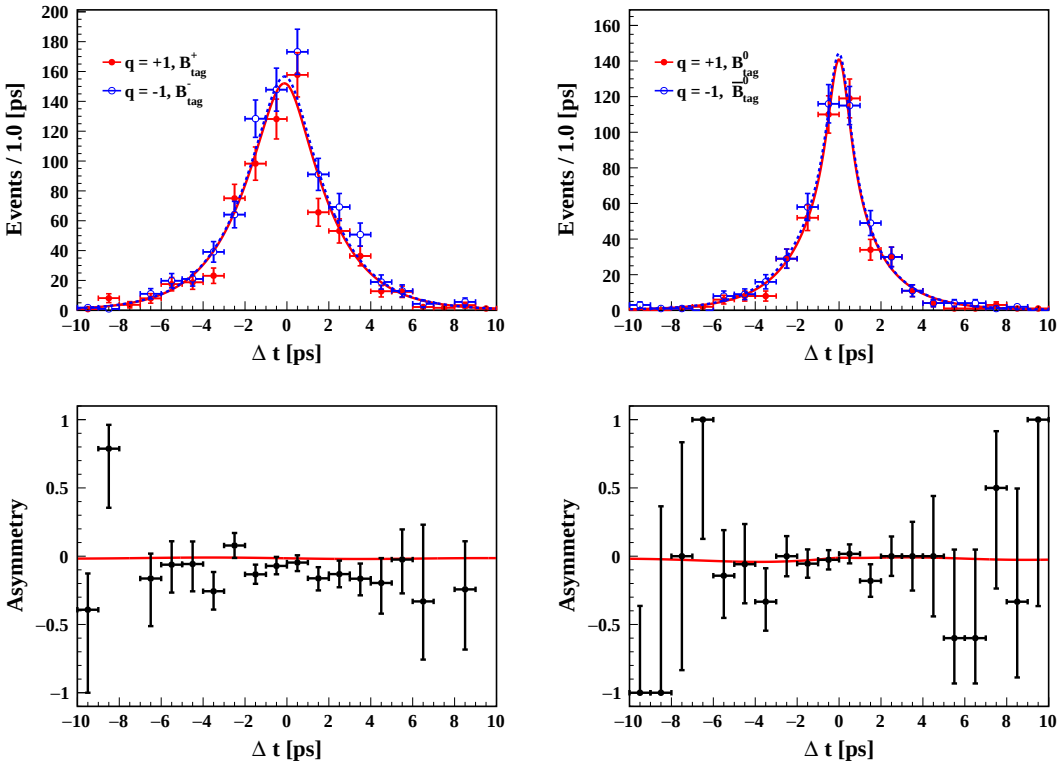


(a) $B^+ \to K^0_S K^0_S K^+$



(b) $B^0 \to K^0_S K^0_S K^0_S$

Figure I.3: Results of lifetime fit to generic MC



(a) $B^+ \to K^0_S K^0_S K^+$ (b) $B^0 \to K^0_S K^0_S K^0_S$

Figure I.4: Results of *CP* fit to generic MC

Appendix J

Invariant mass of $K^0_S K^0_S$

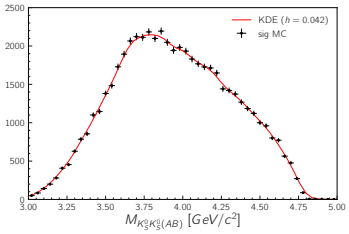
We use the distribution of invariant mass of two K^0_S's from B_{CP}, $M_{K^0_S K^0_S}$, to determine the yield of $B^0 \to \chi_{c0} K^0_S$ events and to draw the Dalitz plot.

J.1 Yield determination of $B^0 \to \chi_{c0} K^0_S$ events

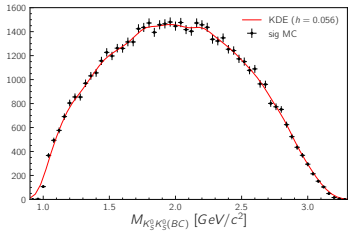
We veto $B^0 \to \chi_{c0} K^0_S$ events in our analysis. Here, we do not veto the decays and try to estimate its yield by fits to the $M_{K^0_S K^0_S}$ distributions. We name the K^0_S's A, B, and C in descending order of momentum and label the $M_{K^0_S K^0_S}$ by the K^0_S names. We define templates of $M_{K^0_S K^0_S}$ distributions using signal and MC samples and fit the sum of templates and a Gaussian function corresponding to χ_{c0} resonance to the data to determine the yield of each component. It should be noted the signal MC only includes non-resonant $B^0 \to K^0_S K^0_S K^0_S$ decays so the signal MC template does not necessarily represent the real distribution. Signal and background MC distributions are smoothed by kernel density estimation (KDE) using a Gaussian kernel. The mean and standard deviation of the Gaussian function for χ_{c0} are fixed to be $3420\,\mathrm{MeV}/c^2$ and $21\,\mathrm{MeV}/c^2$ based on generic MC. Figure J.1 shows the distributions. In the fit, we use only the events within the signal region and impose a Gaussian constraint to the yield of background component, 86 ± 12, based on a result of signal extraction fit without χ_{c0} veto. The sum of $\chi_{c0} K^0_S$ yields are estimated to be $2.0^{+4.7}_{-3.7}$, which is consistent to an MC expectation of 5.7.

J.2 Dalitz plot

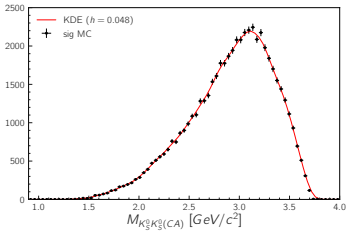
Dalitz plot describes the kinematics of three body decays and resonant structure. It is a 2D scatter plot of invariant masses of different combinations of two daughter particles, m_{ab} and m_{bc}. Because in our case the three final-state particles are identical, there are six definitions for m_{ab} and m_{bc}. We can always choose a combination for each event to let it in the shaded region in Fig. J.2, by defining the largest invariant mass as $m_{ab}(s_{\max})$ and the smallest as $m_{bc}(s_{\min})$. Such a plot is called symmetrized Dalitz plot. Figure J.3 shows



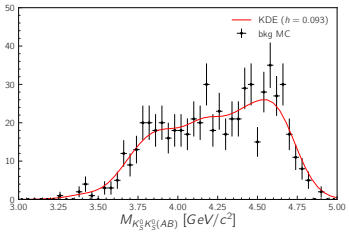
(a) $M_{K^0_S K^0_S}(AB)$, signal MC



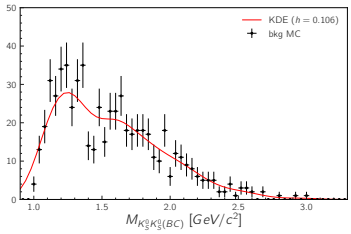
(b) $M_{K^0_S K^0_S}(BC)$, signal MC



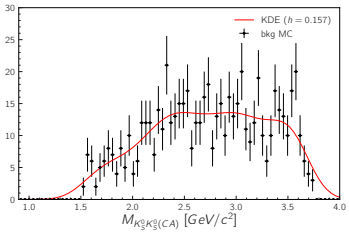
(c) $M_{K^0_S K^0_S}(CA)$, signal MC



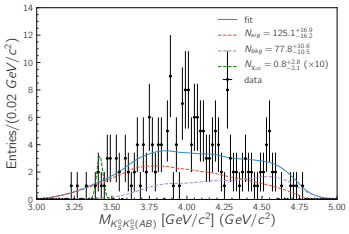
(d) $M_{K^0_S K^0_S}(AB)$, background MC



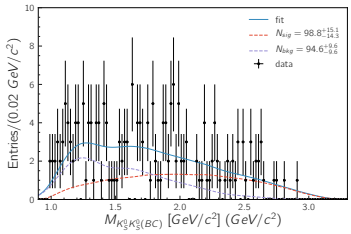
(e) $M_{K^0_S K^0_S}(BC)$, background MC



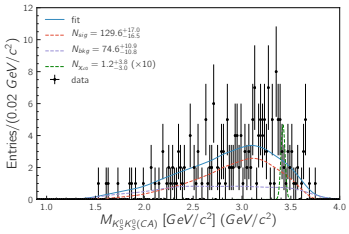
(f) $M_{K^0_S K^0_S}(CA)$, background MC



(g) $M_{K^0_S K^0_S}(AB)$, data



(h) $M_{K^0_S K^0_S}(BC)$, data



(i) $M_{K^0_S K^0_S}(CA)$, data

Figure J.1: $M_{K^0_S K^0_S}$ distribution of signal MC, background MC, and data. The signal and background MC distributions are smoothed by KDE. The χ_{c0} component fitted to the data distribution is scaled by ten for better visibility.

the symmetrized Dalitz plot using the events within the signal region without applying χ_{c0} veto. The resonant structure is not clear due to low statistics.

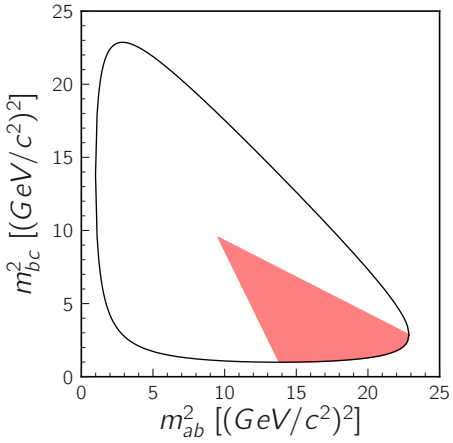


Figure J.2: Concept of symmetrized Dalitz plot

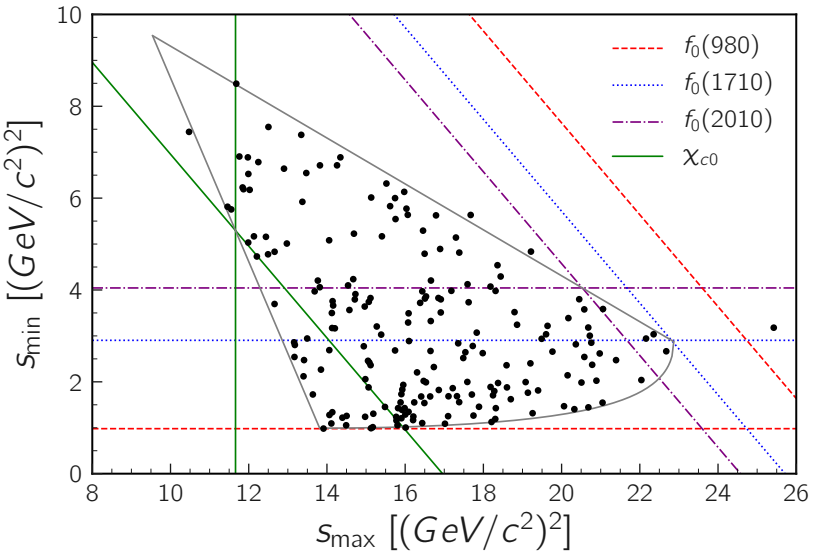


Figure J.3: Symmetrized Dalitz plot and the resonances expected to contribute to $B^0 \to K^0_S K^0_S K^0_S$ decays

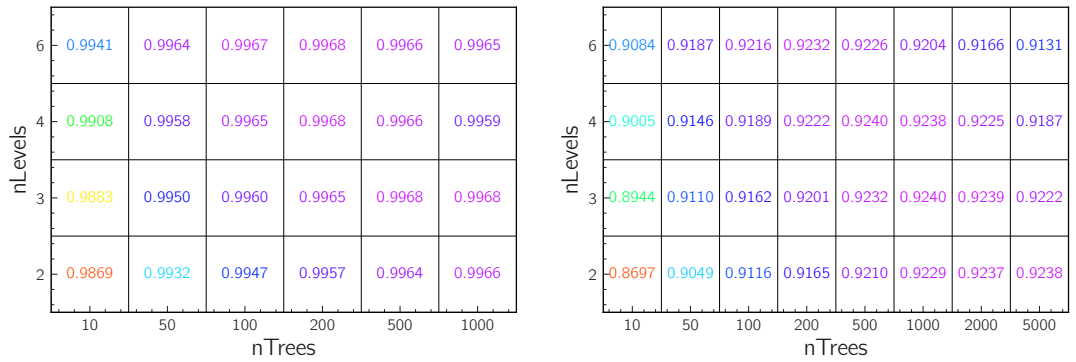
Appendix K

Hyperparameters of boosted decision trees

We set the following hyperparameters for the two BDTs used for K_S^0 selection and continuum suppression described in Sec. 4.2.1 and Sec. 4.6 (See Ref. [31] for the definition of the hyperparameters.):

- number of trees : 200
- depth of trees : 3
- learning rate : 0.1
- sampling rate : 0.5
- number of events in the training sample : 200,000
- number of features : 22 (K_S^0), 18 (continuum suppression)

Since the performance of a BDT depends largely on its number and depth of trees, we check if our choice of the parameters is proper. Figure K.1 shows the results of grid search, where we repeat training the BDTs while changing the two hyperparameters. The performance of the BDTs are evaluated as AUC, which stands for normalized area under the ROC curve. It ranges from zero to one and a larger value means better separation. We confirm that the current hyperparameters are close to the optimal choice.



(a) K_S^0 selection BDT

(b) continuum suppression BDT

Figure K.1: AUC of the K_S^0 (left) and continuum suppression (right) BDTs as a function of the number and depth of trees

Appendix L

Correlation between CP-side and tag-side vertex positions

We use the Btube constraint in the $B_{\rm tag}$ vertex fit (see Sec. 4.4.2). As the constraint relies on the result of the B_{CP} vertex fit, the fit results may be correlated with each other in principle. However, we confirm that the correlation is negligible in the signal MC as described in the following.

We naively expect a possible positive correlation between the vertex position residuals, $\delta\ell^{CP}$ and $\delta\ell^{\rm tag}$. Such a correlation should be clearer in the events where only one track is available for $B_{\rm tag}$ vertex fit. Figure L.1 shows the distribution of $\delta\ell^{\rm tag}$ in the signal MC events with and without the single-track selection on the tag side. The distributions are divided according to the sign of $\delta\ell^{CP}$. We do not observe a correlation between $\delta\ell^{CP}$ and $\delta\ell^{\rm tag}$.

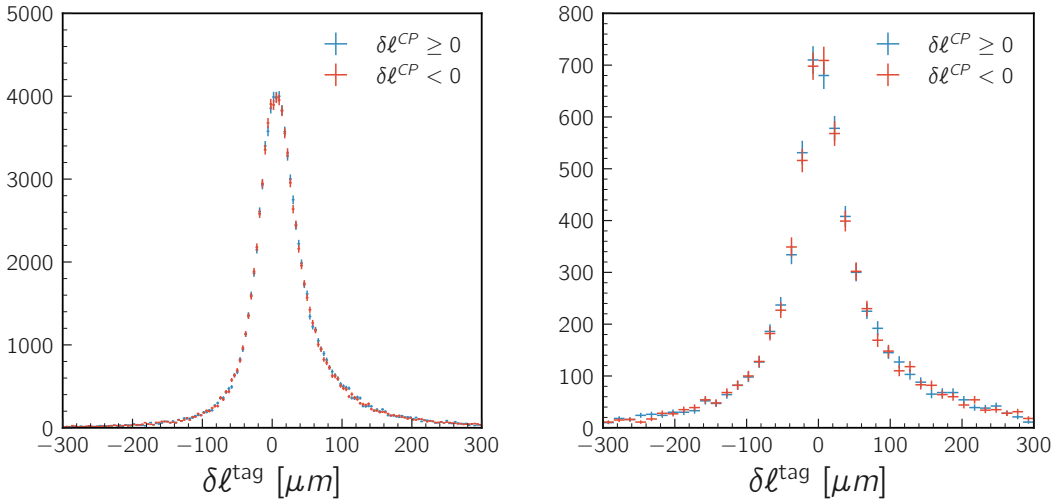


Figure L.1: Distribution of $\delta\ell^{\text{tag}}$ for the signal MC events divided according to the sign of $\delta\ell^{CP}$. The left plot shows all events passing the reconstruction criteria and the right plot shows the events where only one track is available for B_{tag} vertex fit.

Bibliography

[1] F. Abudinén et al. B-flavor tagging at Belle II. *Eur. Phys. J. C*, 82(4):283, 2022.

[2] Yuval Grossman and Mihir P. Worah. CP asymmetries in B decays with new physics in decay amplitudes. *Phys. Lett. B*, 395:241–249, 1997.

[3] Lincoln Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51:1945, 1983.

[4] Andrzej J. Buras, Markus E. Lautenbacher, and Gaby Ostermaier. Waiting for the top quark mass, K+ —> pi+ neutrino anti-neutrino, B(s)0 - anti-B(s)0 mixing and CP asymmetries in B decays. *Phys. Rev. D*, 50:3433–3446, 1994.

[5] J. Charles, Andreas Hocker, H. Lacker, S. Laplace, F. R. Le Diberder, J. Malcles, J. Ocariz, M. Pivk, and L. Roos. CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories. *Eur. Phys. J. C*, 41(1):1–131, 2005.

[6] I. I. Bigi and A. I. Sanda. *Quantum mechanics of neutral particles*, page 90–114. Cambridge University Press, 2 edition, 2009.

[7] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[8] Owen Long, Max Baak, Robert N. Cahn, and David P. Kirkby. Impact of tag side interference on time dependent CP asymmetry measurements using coherent B0 anti-B0 pairs. *Phys. Rev. D*, 68:034010, 2003.

[9] Hai-Yang Cheng, Chun-Khiang Chua, and Amarjit Soni. *cp*-violating asymmetries in B^0 decays to $K^+K^-K^0_{S(L)}$ and $K^0_SK^0_SK^0_{S(L)}$. *Phys. Rev. D*, 72:094003, Nov 2005.

[10] Yasmine Sara Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018. *Eur. Phys. J.*, C81:226, 2021. updated results and plots available at `https://hflav.web.cern.ch/`.

[11] Mikolaj Misiak, Stefan Pokorski, and Janusz Rosiek. Supersymmetry and FCNC effects. *Adv. Ser. Direct. High Energy Phys.*, 15:795–828, 1998.

[12] Luca Silvestrini. Searching for new physics in $b \to s$ hadronic penguin decays. *Ann. Rev. Nucl. Part. Sci.*, 57:405–440, 2007.

[13] SUSY Summary Plots March 2022. Technical report, CERN, Geneva, Apr 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-013.

[14] Yusuke Shimizu, Morimitsu Tanimoto, and Kei Yamamoto. Supersymmetry contributions to CP violations in b→s and b→d transitions taking account of new data. *Phys. Rev. D*, 87(5):056004, 2013.

[15] K. H. Kang et al. Measurement of time-dependent CP violation parameters in $B^0 \to K^0_S K^0_S K^0_S$ decays at Belle. *Phys. Rev. D*, 103(3):032003, 2021.

[16] J. P. Lees et al. Amplitude analysis and measurement of the time-dependent CP asymmetr y of $B^0 \to K^0_S K^0_S K^0_S$ decays. *Phys. Rev. D*, 85:054023, 2012.

[17] Wan Kun. *Time Dependent Charge-Parity Violation in $B^0 \to K^0_S K^0_S K^0_S$ in Belle II early operation*. PhD thesis, Tokyo U., 2021.

[18] W. Altmannshofer et al. The Belle II Physics Book. *PTEP*, 2019(12):123C01, 2019. [erratum: PTEP2020,no.2,029201(2020)].

[19] Z. Natkaniec et al. Status of the Belle silicon vertex detector. *Nucl. Instrum. Meth. A*, 560:1–4, 2006.

[20] Q. Liu et al. Operational Experience and Performance of the Belle II Pixel Detector. *JPS Conf. Proc.*, 34:010002, 2021.

[21] K. Adamczyk et al. The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector. 1 2022.

[22] B. Wang et al. Operational experience of the Belle II pixel detector. *Nucl. Instrum. Meth. A*, 1032:166631, 2022.

[23] T. V. Dong, S. Uno, M. Uchida, H. Ozaki, N. Taniguchi, E. Nakano, and K. Trabelsi. Calibration and alignment of the Belle II central drift chamber. *Nucl. Instrum. Meth. A*, 930:132–141, 2019.

[24] Kazuki Kojima. The operation and performance of the TOP detector at the Belle II experiment. *PoS*, EPS-HEP2021:803, 2022.

[25] Martin Bessner. Performance of the Belle II imaging Time-Of-Propagation (iTOP) detector in first collisions. *Nucl. Instrum. Meth. A*, 958:162318, 2020.

[26] M. Yonenaga et al. Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data. *PTEP*, 2020(9):093H01, 2020.

[27] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J.*, C74:3026, 2014.

[28] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II Core Software. *Comput. Softw. Big Sci.*, 3(1):1, 2019.

[29] The Belle II Collaboration. Belle ii analysis software framework (basf2).

[30] Mikihiko Nakao et al. Performance of the Unified Readout System of Belle II. *IEEE Trans. Nucl. Sci.*, 68(8):1826–1832, 2021.

[31] Thomas Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, 1(1):2, 2017.

[32] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[33] S. Jadach, B. F. L. Ward, and Z. Was. The Precision Monte Carlo event generator K K for two fermion final states in e+ e- collisions. *Comput. Phys. Commun.*, 130:260–325, 2000.

[34] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[35] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003.

[36] J. F. Krohn et al. Global decay chain vertex fitting at Belle II. *Nucl. Instrum. Meth. A*, 976:164269, 2020.

[37] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[38] H. Albrecht et al. Search for Hadronic $b \to u$ Decays. *Phys. Lett. B*, 241:278–282, 1990.

[39] H. Tajima et al. Proper time resolution function for measurement of time evolution of B mesons at the KEK B factory. *Nucl. Instrum. Meth. A*, 533:370–386, 2004.

[40] Giovanni Punzi. Comments on likelihood fits with variable resolution. *eConf*, C030908:WELT002, 2003.

[41] Gary J. Feldman and Robert D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, 1998.

[42] Kazuo Abe et al. Evidence for CP violating asymmetries B0 —> pi+ pi- decays and constraints on the CKM angle phi(2). *Phys. Rev. D*, 68:012001, 2003.

[43] Muriel Pivk and Francois R. Le Diberder. SPlot: A Statistical tool to unfold data distributions. *Nucl. Instrum. Meth. A*, 555:356–369, 2005.

[44] Thibaud Humair, Justin Skorupa, Jakub Kandra, Radek Zlebcik, Caspar Schmitt, and Vladimir Chekelian. BELLE2-NOTE-PH-2021-046Measurement of the B0 lifetime and mixing frequency using hadronic decays with the Moriond 2022 dataset, Oct 2021.

[45] K. Hara. Belle Note 872 - Study of tag side interference effect in time-dependent CP violating asymmetry measurements, 2006.

[46] Y. Yusa, K. Miyabayashi, and ICPV group. Belle Note 1149 - Measurement of time-dependent CP violation in $B^0 \to (c\bar{c})K^0$ decays with 772 M $B\overline{B}$, 2011.

[47] K. Kang, Park H. Higuchi, T., and K. Sumisawa. Belle Note 1493 - Measurement of Time-dependent Charge Parity Violation in $B^0 \to K^0_S K^0_S K^0_S$ Using the Full Data Sample, 2020.

[48] K. Sumisawa et al. Measurement of time-dependent CP-violating asymmetries in B0 —> K0(S) K0(S) KO(S) decay. *Phys. Rev. Lett.*, 95:061801, 2005.

[49] K. F. Chen et al. Observation of time-dependent CP violation in B0 —> eta-prime K0 decays and improved measurements of CP asymmetries in B0 —> phi K0, K0(s) K0(s) K0(s) and B0 —> J/psi K0 decays. *Conf. Proc. C*, 060726:823–826, 2006.

[50] Atsushi Hayakawa, Yusuke Shimizu, Morimitsu Tanimoto, and Kei Yamamoto. Squark flavor mixing and CP asymmetry of neutral B mesons at LHCb. *Phys. Lett. B*, 710:446–453, 2012.