博士論文

Doctoral Dissertation

Time Dependent Charge-Parity Violation in $B^0 \to K_S^0 K_S^0 K_S^0$ in Belle II early operation

(Belle II 初期データを使った$B^0 \to K_S^0 K_S^0 K_S^0$ 崩壊の時間に依存する荷電・パリティ非保存の研究)

令和2年12月博士(理学)申請

A Dissertation Submitted for the Degree of Doctor of Philosophy

December 22nd, 2020

東京大学大学院理学系研究科物理学専攻

Department of Physics, Graduate School of Science,

The University of Tokyo

万 琨

Wan Kun

Time Dependent Charge-Parity Violation in $B^0 \to K_S^0 K_S^0 K_S^0$ in Belle II early operation

by

Kun Wan

Submitted to the Department of Physics
on December 22nd, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The Belle II experiment is a next-generation super B-factory experiment. The targeted instantaneous luminosity is 8×10^{35} cm^{-2}s^{-1} and the expected integrated luminosity is 50 ab^{-1} by 2030 with the majority of data collected at the $\Upsilon(4S)$ resonance using SuperKEKB accelerator.

The thesis is based on the time-dependent CP violation study of $B^0 \to K_S^0 K_S^0 K_S^0$ decay to precisely measure the CP parameters $\mathcal{S}$ and $\mathcal{A}$ in penguin-dominated $b \to s$ transition, which is sensitive to New Physics effects. Such a precise measurement mainly depends on determination of the distance between two vertices of two neutral B mesons. The blind analysis and fit by a unbinned maximum likelihood method are performed using about 62.8 fb^{-1} recorded experiment data from Belle II detector 2019 and 2020 (spring and summer) operation. The measurement results: $\mathcal{S} = -\sin(2\phi_1) = -0.82 \pm 0.85$ (stat) ± 0.07 (syst) and $\mathcal{A} = -0.21 \pm 0.28$ (stat) ± 0.06 (syst) are obtained. The result is dominated by the statistical uncertainty and currently consistent with the Standard Model, as well as the previous results in the Belle and BaBar.

At the 50 ab^{-1} future Belle II full luminosity, we estimate the total uncertainty of $\mathcal{S}$ to be $0.040 \sim 0.048$ in $B^0 \to K_S^0 K_S^0 K_S^0$. This estimation is a preliminary result based on the limited data at present and conservative assumptions on the reduction of the uncertainties, which is subject to change in future. Overall, such total uncertainty can provide a much better probe for searching the New Physics effects in future.

Thesis Supervisor: Hiroaki Aihara
Title: Professor

Contents

Chapter 1

Introduction

1.1 The Standard Model

The Standard Model (SM) was built in the late 70th of 20th century to describe the matter compositions and interactions using a group of fundamental particles - fermions and bosons. In the Standard Model, there are three generations of quarks and leptons, along with their anti-particles, which are all fermions. On the other hand, the bosons in the Standard Model consist of gluons, photons, $W^{\pm}$ and Z^0 bosons that are all gauge bosons and one Higgs boson that is a scalar boson. This group of particles is summarized in Figure 1-1. The Standard Model depicts the interactions between elementary particles as the exchange of the bosons. The strong interaction requires the exchange of gluons. Photons, $W^{\pm}$ and Z^0 bosons carry the electromagnetic force and weak force, which are unified as the electroweak interaction in the Standard Model. Higgs boson is responsible for the generation of masses for the gauge bosons through electroweak symmetry breaking [1]. The Standard Model has been proved to be an excellent theoretical model that can be used to explain many experimental observations, but sadly not all of them. For instance, neutrino mass is expected to be zero in the Standard Model but the flavor oscillation indicates non-zero mass of neutrinos. The observation of Charge-Parity (CP) asymmetry in universe presented by the absence of antimatter can not be fully explained by the CP violation sources within the Standard Model. These experimental observations require further researches beyond the Standard Model, which is called New Physics (NP) studies.

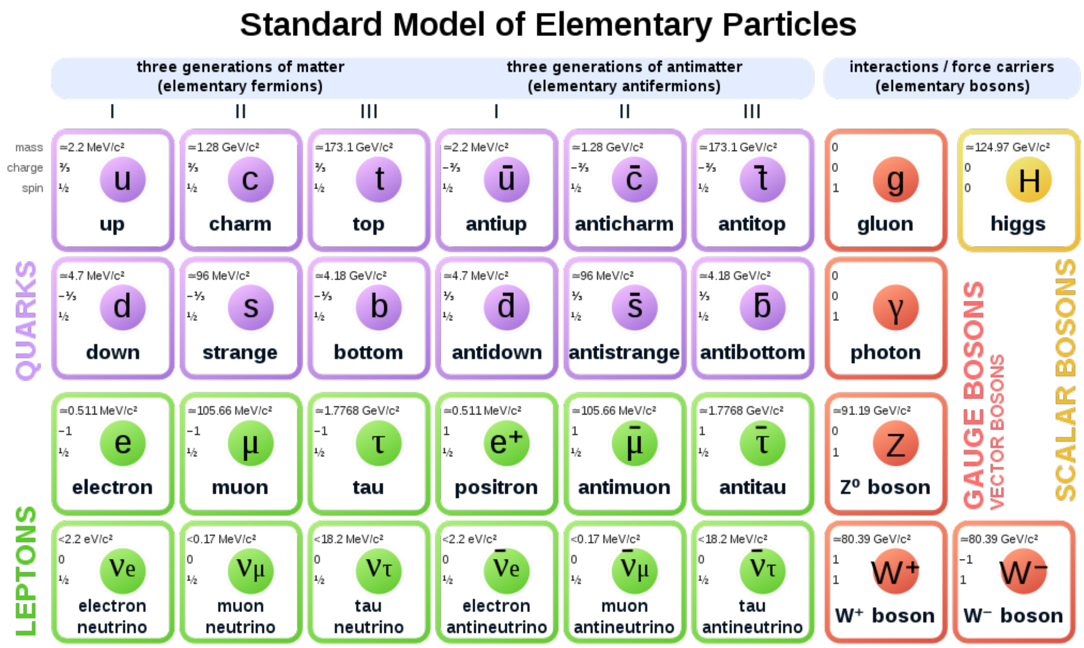


Figure 1-1: Elementary particles in the Standard Model [2].

1.2 Symmetry Violation

Symmetry violation has been one of the focuses in NP studies due to the internal link between symmetries and conservation laws, which makes it a good probe for possible NP theories beyond the SM. When a known symmetry is found to be broken, it usually leads to the discovery of a new theory.

There are three types of discrete symmetric operations which play important roles in particle physics. Charge-conjugation C is the operation that turns particle to its anti-particle. Parity transformation P is the one that puts a negative sign before all the spatial related vector such as $\overrightarrow{r} \rightarrow -\overrightarrow{r}$. The time-reversing operation T is to reversely proceed a physical process backward time. Physicists were convinced that each of these three symmetric operations makes no change to any physics system. However, in 1950s, Lee and Yang [3] first questioned that parity symmetry might be broken in the weak interaction. They offered a few possible ways to test it and then by Wu [4], an observation on the β decay of ^{60}Co was presented that the electrons emitted from ^{60}Co decay prefers the direction of nuclear spin that can be controlled by the external magnetic field. The violation of P symmetry was discovered by this clear evidence.

The first evidence of CP violation was discovered in neutral K^0 system by Cronin and Fitch's experiment [5]. The neutral K^0 mesons can be observed as two states that have significantly different lifetime (called as "K^0_S" and "K^0_L" for short and long lifetime particles) with opposite CP eigenvalues. The experiment measured the decay products at 57 foot of a neutral K^0 beamline assuming all the particle at the end of the beam should be long lifetime K^0_L, nearly no K^0_S. But 0.002% of K^0_L were found to decay into $\pi^+\pi^-$ which is the main decay process of K^0_S (CP eigenvalue $= 1$ in $\pi^+\pi^-$ final states, while K_L has CP eigenvalue $= -1$). Given that the expected distance to have 0.002% of K^0_S at about speed of light is no more than 1 meter in the beamline, such a deviation at 57 foot is an obvious evidence that $K^0_L \to \pi^+\pi^-$ exists and therefore CP symmetry is violated in the neutral K^0 system.

In 1973, Kobayashi and Maskawa introduced a quark mixing matrix called CKM matrix for three or more generations of quarks before the discovery of the third generation of the quark family [6]. The theory naturally explained an irreducible complex phase in CKM matrix and it accounts for the origin of CP asymmetries of the weak interaction in the Standard Model. The experimental evidence of CP violation in B meson system was observed in 2001 by Belle and BaBar experiments [7; 8]. They measured the time-dependent decay time difference of B and $\bar{B}$ in the decay of $B^0 \to J/\psi K^0_S$. This channel provides a good clearness in theoretical prediction and has relatively large branching fraction, hence it is called the "golden mode" [9]. In 2008, Kobayashi and Maskawa were rewarded the Nobel Prize to highly value their contribution to CP violation mechanism in the SM, to which Belle experiment contributes greatly. Later in 2010, the upgrade of Belle, Belle II and the upgrade of KEK accelerator, SuperKEKB, were approved to further push the understanding of CP violation along with other topics in New Physics researches.

1.3 CKM mechanism

The source of the flavor violation in the SM is the Yukawa interaction between fermions and the Higgs doublet,

$$\Phi = \begin{pmatrix} \phi^+ \\ v + \frac{H+i\chi}{\sqrt{2}} \end{pmatrix}. \tag{1.1}$$

where the value of H is 174 GeV as the expected Higgs potential for vacuum [10]. The ϕ^+ and χ are the pseduo-Goldstone fields related to longitudinally polarized W^+ and Z bosons. The Lagrangian for Yukawa interaction of the quark fields [11] can be presented as:

$$\mathcal{L}^q_{\text{Yuk}} = -\bar{Q}_j Y^d_{jk} \Phi d'_{Rk} - \bar{Q}_j Y^u_{jk} \epsilon \Phi^* u'_{Rk} + h.c., \tag{1.2}$$

where $j,\ k$ label the generations of quarks (repeated indices are summed over). The u'_{Rk} and d'_{Rk} are the singlets of the electroweak gauge group SU(2) which present the weak eigenstates of quarks. The left-handed quarks ensemble the SU(2) doublets as Q_j.

Yukawa matrix is an arbitrary 3×3 complex matrix $Y^{u,d}$ which gives the rise of up and down type massive quark field $M^{u,d} = Y^{u,d} v$ according to Equation 1.2. The representation of the quark fields using weak eigenstates can be transformed to mass eigenstates by Equation 1.3 and 1.4.

$$S^u_{L,R} \begin{pmatrix} u' \\ c' \\ t' \end{pmatrix}_{L,R} = \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} \tag{1.3}$$

$$S^d_{L,R} \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{L,R} = \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R} \tag{1.4}$$

In Equation 1.3 and 1.4, $S^{u,d}_{L,R}$ are all unitary matrices since they are generated by the normalized eigenstates of Yukawa matrix. The mass item in the Lagrangian can be presented as Equation 1.5

$$\mathcal{L}_m = -\sum_{q=u,c,t,d,s,b} M_q q^\dagger q \tag{1.5}$$

where the $q = (q_L + q_R)$ is four-component Dirac field, and $q_L^\dagger q_L = q_R^\dagger q_R = 0$. As a result of diagonalizing $Y^{u,d}$, the charged current $W^\pm$ interactions coupe to the physical quarks and the Lagrangian is written as Equation 1.6, where $V_{CKM} \equiv S_L^u S_L^{d\dagger}$.

$$\mathcal{L}_W^q = \frac{g}{\sqrt{2}} \left[\begin{pmatrix} \overline{u} & \overline{c} & \overline{t} \end{pmatrix}_L \gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \begin{pmatrix} \bar{d} & \overline{s} & \bar{b} \end{pmatrix}_L \gamma^\mu W_\mu^- V_{CKM}^* \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L \right] \tag{1.6}$$

The Lagrangian hereby clearly declares the transition of different quarks through the coupling of charged current $W^\pm$, where such a coupling only applies for the left-handed quarks. For example, a left-handed charm quark only transits to left-handed strange quark by a W boson. By only applying C or P conjugation, the Lagrangian is not invariant, indicating the non-conservation of C or P individually. However, if the CP conjugation is applied, the first term in Equation 1.6 transits under CP conjugation as Equation 1.7 shows.

$$\begin{pmatrix} \overline{u} & \overline{c} & \overline{t} \end{pmatrix}_L \gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \overset{CP}{\Leftrightarrow} \begin{pmatrix} \bar{d} & \overline{s} & \bar{b} \end{pmatrix}_L \gamma^\mu W_\mu^- V_{CKM} \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L \tag{1.7}$$

It is clear that the right term in Equation 1.7 resembles the second term in Equation 1.6 by only the difference in the V_{CKM}, indicating that the CP invariance is related to the CKM matrix. If considering the case of $N < 3$ where N is the number of the generations of quarks, CP invariance can be conserved by rendering non-physical phases to the quark fields. However, for $N \geq 3$, there is at least one irreducible complex phase in the CKM matrix that gives:

$$V_{CKM} \neq V^*_{CKM} \tag{1.8}$$

thereby accounts for the dominant mechanism of CP violation in flavor-changing transitions. The 3×3 unitary CKM matrix can be written as Equation 1.9 based on the quark fields using Equation 1.3 and 1.4.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.9}$$

It can be parameterized into the form of Equation 1.10.

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

where the $c_{jk} = \cos(\theta_{jk})$ and $s_{jk} = \sin(\theta_{jk})$, and δ is the irreducible complex phase. By the experimental evidence from the branching fraction of $b \to c$, $s \to u$ and $b \to u$ processes, the following relation can be determined,

$$|V_{ub}| \ll |V_{cb}| \ll |V_{us}| \ . \tag{1.11}$$

Furthermore, the relations in Equation 1.12 are often used to simplify CKM matrix presentation.

$$s_{12} = \lambda, s_{23} = A\lambda^2, s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta) \tag{1.12}$$

By using Equation 1.12, CKM matrix is parameterized as Equation 1.13.

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

Using the unitary condition, Equation1.14 can be obtained.

$$1 + \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0 \tag{1.14}$$

Based on Equation 1.14, 1.15 and 1.16, the shape of CKM triangle can be determined on the complex plane as shown in Figure 1-2.

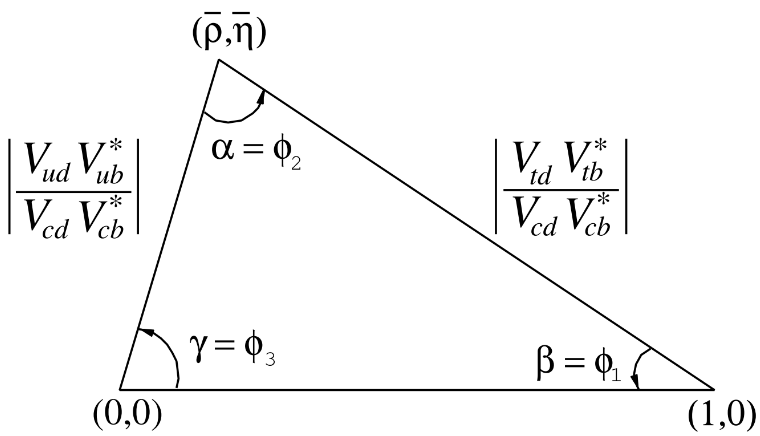


Figure 1-2: The unitary triangles of CKM [11].

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \tag{1.15}$$

$$1 - (\bar{\rho} + i\bar{\eta}) = -\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \tag{1.16}$$

These angles are obtained by drawing the $(\bar{\rho}, \bar{\eta})$ on the complex coordinates, and they are also well-known in the names as: $\phi_1 = \beta, \phi_2 = \alpha, \phi_3 = \gamma$. The results presenting the measurement of CKM angles or $(\bar{\rho}, \bar{\eta})$ in 2019 are shown in Figure 1-3.

The measurement of ϕ_1 and ϕ_2 are mainly obtained from the time-dependent CP violations (TDCPV) measurement. The ϕ_1 in the tree-level dominated decays has

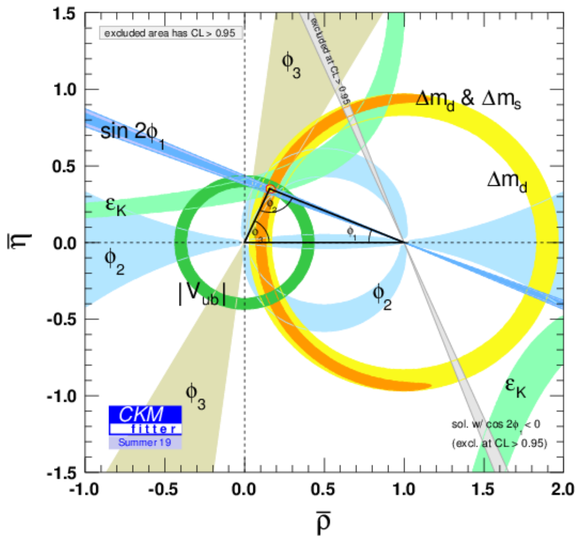


Figure 1-3: The CKM triangle fit in the complex plane of $\bar{\rho} - \bar{\eta}$ [12].

been precisely measured due to the small hadronic uncertainties. Flavor-Changing-Neutral-Current (FCNC) processes can rise through the $B^0 - \bar{B}^0$ mixing in box diagram, and it is believed that potential NP processes might contribute to the difference between results of CKM angles measured from experiments, such as ϕ_1 value in tree-dominated processes and penguin-dominated processes, where both involve $b \to s$ transition. It requires the precise measurements on multiple decay channels to search for the potential NP effects. The prospective large Belle II data and improved detector performance will be much useful to help the discovery of NP in future.

1.4 Time Dependent CP violation

1.4.1 CP violation in neutral B system

The measurements of ϕ_1, ϕ_2 and ϕ_3 essentially access the CKM CP violating phase since there is only one complex phase in the CKM matrix and it can be determined by these three angles. For determining the value of ϕ_1, TDCPV measurements provide a good experimental environment. From Figure 1-2, one can obtain ϕ_1 and ϕ_2 by

Equation 1.17 and 1.18.

$$\phi_1 = \mathrm{Arg}(-\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}) \tag{1.17}$$

$$\phi_2 = \mathrm{Arg}(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}) \tag{1.18}$$

The time-dependent CP violation comes from the interference of neutral B mixing phase and the weak phase in the decay amplitude. The mass eigenstates which are driving the propagation of neutral B meson states with mixing are: $|B\rangle_{H,L} = p|B\rangle \pm q|\overline{B}\rangle$, where H and L stand for the heavier and lighter mass eigenvalues. The $|B\rangle$ and $|\overline{B}\rangle$ present the flavor eigenstates of neutral B mesons. The Hamiltonian matrix can be written using flavor eigenstates as shown in Equation 1.19.

$$M_\Gamma = \begin{bmatrix} m - i/2\Gamma & M_{12} - i/2\Gamma_{12} \\ M_{12}^* - i/2\Gamma_{12}^* & m - i/2\Gamma \end{bmatrix} \tag{1.19}$$

Considering the time evolution of mass eigenstates, the time-dependent states can be shown as Equation 1.20 and 1.21 by using the notation of $B_{H,L}$ as physical states at $t = 0$:

$$B_H(t) = e^{-im_H t} e^{-\frac{\Gamma_H}{2}t} B_H \ , \tag{1.20}$$

$$B_L(t) = e^{-im_L t} e^{-\frac{\Gamma_L}{2}t} B_L \ , \tag{1.21}$$

where $m_{H,L}$ and $\Gamma_{H,L}$ are the masses and decay widths of two mass eigenstates. We can express the mass eigenstates using flavor eigenstates, which are shown in Equation 1.22 and 1.23.

$$B(t) = (1/2p)e^{-im_H t}e^{-\frac{\Gamma_H}{2}t}(pB + q\bar{B}) + (1/2p)e^{-im_L t}e^{-\frac{\Gamma_L}{2}t}(pB - q\bar{B}) \tag{1.22}$$

$$\bar{B}(t) = (1/2q)e^{-im_H t}e^{-\frac{\Gamma_H}{2}t}(pB + q\bar{B}) - (1/2q)e^{-im_L t}e^{-\frac{\Gamma_L}{2}t}(pB - q\bar{B}) \tag{1.23}$$

Replacing $g_\pm(t) = \frac{1}{2}(e^{-im_H t - \frac{\Gamma_H}{2}t} \pm e^{-im_L t - \frac{\Gamma_L}{2}t})$, Equation 1.22 and 1.23 become Equation 1.24 and 1.25.

$$B(t) = g_+(t)B + \frac{q}{p}g_-(t)\bar{B} \tag{1.24}$$

$$\bar{B}(t) = g_+(t)\bar{B} + \frac{p}{q}g_-(t)B \tag{1.25}$$

The q/p is introduced by the coefficient of mass eigenstates from weak eigenstates. Using the Hamiltonian matrix, q/p can be presented as

$$q/p = \frac{\Delta M - i/2\Delta\Gamma}{2(M_{12} - i/2\Gamma_{12})} \ , \tag{1.26}$$

where the M_{12} and Γ_{12} stand for the contribution of non-diagnosed term in the Hamiltonian matirx. $\Delta M = m_H - m_L$ and $\Delta\Gamma = \Gamma_H - \Gamma_L$ are the difference of mass and decay width for two mass eigenstates, respectively. Using the Hamiltonian and the decay amplitude A_f ($\bar{A}_{\bar{f}}$), the time dependent decay rate of $B^0/\bar{B}^0$ into final states $f/\bar{f}$ can be expressed as

$$\Gamma(B \to f, t) = |A_f|^2 \cdot [|g_+(t)|^2 + |\lambda_f|^2|g_-(t)|^2 + 2\mathrm{Re}(\lambda_f g_+^*(t)g_-(t))] \ , \tag{1.27}$$

$$\Gamma(\bar{B} \to \bar{f}, t) = |\bar{A}_{\bar{f}}|^2 \cdot [|g_+(t)|^2 + |\bar{\lambda}_{\bar{f}}|^2|g_-(t)|^2 + 2\mathrm{Re}(\bar{\lambda}_{\bar{f}} g_+^*(t)g_-(t))] \ , \tag{1.28}$$

$$\Gamma(\bar{B} \to f, t) = |A_f|^2 \cdot |\frac{p}{q}|^2 \cdot [|g_-(t)|^2 + |\lambda_f|^2|g_+(t)|^2 + 2\mathrm{Re}(\lambda_f g_+(t)g_-^*(t))] \ , \tag{1.29}$$

$$\Gamma(B \to \bar{f}, t) = |\bar{A}_{\bar{f}}|^2 \cdot |\frac{q}{p}|^2 \cdot [|g_-(t)|^2 + |\bar{\lambda}_{\bar{f}}|^2|g_+(t)|^2 + 2\mathrm{Re}(\bar{\lambda}_{\bar{f}} g_+(t)g_-^*(t))] \ . \tag{1.30}$$

where the parameter λ_f and $\bar{\lambda}_{\bar{f}}$ are defined as Equation 1.31 and 1.32.

$$\lambda_f \equiv (q/p)(\bar{A}_f/A_f) \tag{1.31}$$

$$\bar{\lambda}_{\bar{f}} \equiv (p/q)(A_{\bar{f}}/\bar{A}_{\bar{f}}) \tag{1.32}$$

It is obvious that if $|A_f| \neq |\bar{A}_{\bar{f}}|$, direct CP violation will occur. The time-dependent decay rate difference is defined as Equation 1.33.

$$A_{CP}(t) \equiv \frac{\Gamma(B \to f, t) - \Gamma(\bar{B} \to \bar{f}, t)}{\Gamma(B \to f, t) + \Gamma(\bar{B} \to \bar{f}, t)} \\ = \frac{\mathcal{S}\,\sin(\Delta M t) - \mathcal{A}\,\cos(\Delta M t)}{\cosh(\Delta\Gamma t/2) + A^{f}_{\Delta\Gamma}\sinh(\Delta\Gamma t/2)} \tag{1.33}$$

where

$$\mathcal{S} = \frac{2\mathrm{Im}(\lambda_f)}{1 + |\lambda_f|^2} \tag{1.34}$$

$$\mathcal{A} = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \tag{1.35}$$

$$A^{f}_{\Delta\Gamma} = -\frac{2\mathrm{Re}(\lambda_f)}{1 + |\lambda_f|^2}\ . \tag{1.36}$$

From Equation 1.34 and 1.35, the time-dependent CP violation parameters $\mathcal{S}$ and $\mathcal{A}$ are dependent on the parameter λ_f.

1.4.2 ϕ_1 from $B^0 \to J/\psi K^0_S$

If final states are CP eigenstates, the amplitudes are obtained by $\mathcal{A}_f \equiv \langle f|H|B\rangle$ and $\bar{\mathcal{A}}_f \equiv \langle f|H|\bar{B}\rangle$. In $B^0_d - \bar{B}^0_d$ mixing system, the q/p can be treated as $e^{i\phi_d}$ as a pure phase term. This relative phase accounts the transition from b to up-type quarks to strange quark s in mixing, so it can be presented as $\phi_d = \mathrm{Arg}[(V^*_{td}V_{tb})/(V^*_{tb}V_{td})] \approx 2\phi_1$ based on negligible correction to the SM. In mode $B^0 \to J/\psi K^0_S$, considering $\Delta\Gamma$ can be treated as zero in the SM in this case [13], Equation 1.33 can be reduced to

$$A_{CP}(t) = \mathcal{S}\sin(\Delta M t) - \mathcal{A}\cos(\Delta M t)\ , \tag{1.37}$$

which receives contributions from both tree-level and loop-level processes as shown in Figure 1-4 , which receives contributions from tree-level and loop-level processes shown in Figure 1-4 ,

Using the relation $|V_{ub}| \ll |V_{cb}| \ll |V_{us}| < |V_{cs}|$, it is obvious that $V^*_{ub}V_{us} \ll V^*_{cb}V_{cs}$, so the penguin-mode is suppressed in the Standard Model. The η_f is defined as the

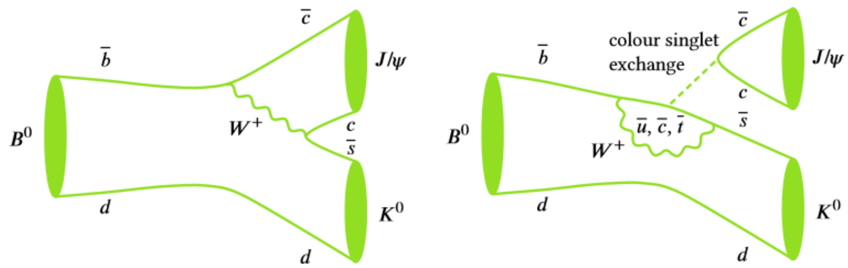


Figure 1-4: The dominated tree-level (left) and the suppressed loop-level (right) of $B \to J/\psi K^0$, in which K^0 particles are detected as K_S^0 [14].

CP eigenvalue of the final states. Given $\eta_f = -1$ and $|\lambda_f| = 1$ in $B^0 \to J/\psi K_S^0$, from Equation 1.34, CP violation parameters can be presented as shown in Equation 1.38.

$$\mathcal{S} = \mathrm{Im}(\lambda_f) = -\sin(\phi_d)\eta_f = \sin(2\phi_1);\ \mathcal{A} = 0 \tag{1.38}$$

From Equation 1.38 , ϕ_1 can be obtained precisely in the measurement of time-dependent CP violation in $B^0 \to J/\psi K_S^0$.

1.4.3 ϕ_1 from penguin-dominated mode $b \to q\bar{q}s$

Compared to $B^0 \to J/\psi K_S^0$ channel, the measurement of $\mathcal{S}$ and $\mathcal{A}$ from penguin-dominated channels through $b \to q\bar{q}s$ where q is u, d, s can be different due to the varied tree-to-penguin amplitude ratio. Furthermore, they are quite sensitive to NP effects for the following reasons [10]. First, they can probe $B^0-\bar{B}^0$ mixing through different short-distance vertices compared to the tree-level dominated decays. Second, the tree-level decay amplitude is suppressed and penguin-level amplitude is dominated, while the overall non-NP amplitude is relatively small so the NP effects may show up easier. Last but not least, they comprise a large number of different final states, which can help disentangling non-perturbation long-distance physics from short-distance information, such as ϕ_1 or NP contributions to the weak Hamiltonian.

Considering possible New Physics contribution besides the tree-level and penguin-level processes as A_f^{NP}, the decay amplitude can be rendered as Equation 1.39

$$A_f = \lambda_u^s T_f + \lambda_c^s P_f + A_f^{NP} \tag{1.39}$$

where T_f and P_f are tree-level and penguin-level amplitudes. The coefficients λ_u^s and λ_c^s are determined from CKM matrix elements by $\lambda_i^q \equiv V_{ib}^* V_{iq}$. Note that compared to the $B^0 \to J/\psi K_S^0$, the tree level amplitude T_f is suppressed and penguin amplitude P_f is dominated in $b \to q\bar{q}s$. It is also worth noting that T_f contains tree-level $W^{\pm}$ exchange, QCD and electroweak penguin contributions. These carry the combination of CKM matrix elements $\lambda_t^s = V_{ts}V_{tb}^* = -(1+\epsilon_{uc})\lambda_c^s$ where $\epsilon_{uc} \equiv \lambda_u^s/\lambda_c^s = \mathcal{O}(\lambda^2)$. In the SM with neglected ϵ, $b \to q\bar{q}s$ modes are pure penguin with the same weak phase as $B^0 \to J/\psi K_S^0$ has. Thus, the direct CP violation vanishes and the time-dependent CP violation reflects $\mathcal{S}$ in the same way as $B^0 \to J/\psi K_S^0$ does.

Departures from this limit, non-neglected tree amplitude T_f (often called "tree pollution"), as well as possible NP effects, could give different results on ϕ_1. The ϕ_1 differences can be reflected by the $\mathcal{S}$ difference, hence ΔS is defined as Equation 1.40.

$$\Delta S = \mathcal{S}_f - \mathcal{S}_{J/\psi K_s^0} \tag{1.40}$$

Introducing the tree-penguin ratio $r_f^T = T_f/P_f$, NP-to-SM ratio $r_f^{NP} = \mathcal{A}_f^{NP}/(\lambda_c^s P_f)$, the following statements are usually used [10]:

- Branching ratios are affected at $\mathcal{O}(|\epsilon_{uc} r_f^T|, |r_f^{NP}|)$
- Direct CP violations in the SM are of $\mathcal{O}(\epsilon_{uc}\mathrm{Im}(r_f^T))$
- $-n_f^{CP}\mathcal{S} = \sin(2\phi_1) + \Delta\mathcal{S}$, where $\Delta\mathcal{S} = 2\cos 2\phi_1 \sin\phi_3 |\epsilon_{uc}| \mathrm{Re}(r_f^T) + \Delta\mathcal{S}^{\mathrm{NP}}$.

The third statement suggests that non-zero ΔS without the NP effect is still allowed in a small scale within the SM. Therefore, for the precise measurement using the future Belle II data, it is important to understand ΔS theoretically within the SM correction to properly explain the experiment results. The ΔS can be larger than the SM allowed value by 5σ to be called as the evidence of the NP effects, where σ is the total uncertainty of ΔS.

1.4.4 ϕ_1 from $B^0 \to K_S^0 K_S^0 K_S^0$

Since the Belle experiment has reported the time-dependent CP analysis on various $b \to q\bar{q}s$ which experimentally shows that the difference on ϕ_1 has a margin for NP effects [15], the improved measurements with a larger data collection is popularly discussed in order to reduce the impact of uncertainties and clear the tension between results. The decay channel $B^0 \to K_S^0 K_S^0 K_S^0$, shown in the left side of Figure 1-5, is one of the most promising modes for this purpose. The CP eigenvalue of $B^0 \to K_S^0 K_S^0 K_S^0$ is $+1$ called CP-even. There is no up-quark appearing in the final states, so the potential contribution of $b \to u\bar{u}s$ re-scattered into $b \to s\bar{s}s$ is almost of absence, which makes $B^0 \to K_S^0 K_S^0 K_S^0$ a much cleaner channel compared to $B^0 \to K^+ K^- K_S^0$ [16]. In all final states with three K_S^0 from a neutral B decay, the phase-space based decay process and the resonant decay process such as $B^0 \to f_0(980) K_S^0 (f_0(980) \to K_S^0 K_S^0)$ are shown in the left and middle of Figure 1-5, which all yield CP-even states treated as signal events. In the meanwhile, $b \to c \to s$ can also produce the final states with three K_S^0 through a tree-level process like $B^0 \to \chi_{c0} K_S^0 (\chi_{c0} \to K_S^0 K_S^0)$ with a different weak phase and CP-odd states, as shown in the right of Figure 1-5. Such a tree-level process is treated as background and can be rejected by applying veto on two K_S^0 invariant mass within χ_{c0} mass window, which is considered as a minor background at the current luminosity. Due to the same weak phase in the decay amplitudes, any potential NP effects expected in the $B^0 \to \phi K_S^0$ and $B^0 \to \eta^{'} K_S^0$ should also affect $B^0 \to K_S^0 K_S^0 K_S^0$ and the absence of NP effects will lead to the close CP violation as $J/\psi K_S^0$ [16]. As discussed in the previous section that the SM correction due to the different tree-penguin ratio, $B^0 \to \eta^{'} K_S^0$, $B^0 \to \phi K_S^0$ and $B^0 \to K_S^0 K_S^0 K_S^0$ modes could create non-zero $\Delta\mathcal{S}$ in a slightly different level. There were attempted theoretical calculation based on the QCD model for the $\Delta\mathcal{S}$ in these modes. For $B^0 \to \eta^{'} K_S^0$ and $B^0 \to \phi K_S^0$, the details about such calculation on $\Delta\mathcal{S}$ within the SM can be find in Ref. [17]. As for $B^0 \to K_S^0 K_S^0 K_S^0$, the calculation of $\Delta\mathcal{S}$ within the SM can be find in Ref. [18]. From these references, the expected SM-allowed $\Delta\mathcal{S}$ from QCD model suggests the upper limit at ~ 0.05 for $B^0 \to \phi K_S^0$, 0.03 for

$B^0 \to \eta' K_S^0$, and 0.06 for $B^0 \to K_S^0 K_S^0 K_S^0$. For $B^0 \to K_S^0 K_S^0 K_S^0$, the theoretical uncertainty is expected to be $\mathcal{O}(0.05)$, which is small compared with the present experimental uncertainties. The theoretical calculation suggests the positive sign of $\Delta\mathcal{S}_{3K_S^0}$. We take $\Delta\mathcal{S} \sim 0.05$ as the SM-allowed upper limit predicted by the QCD model. Thus, in the comparison of SM-allowed ΔS and the experimental results, we should currently focus on the statistical and systematic uncertainties.

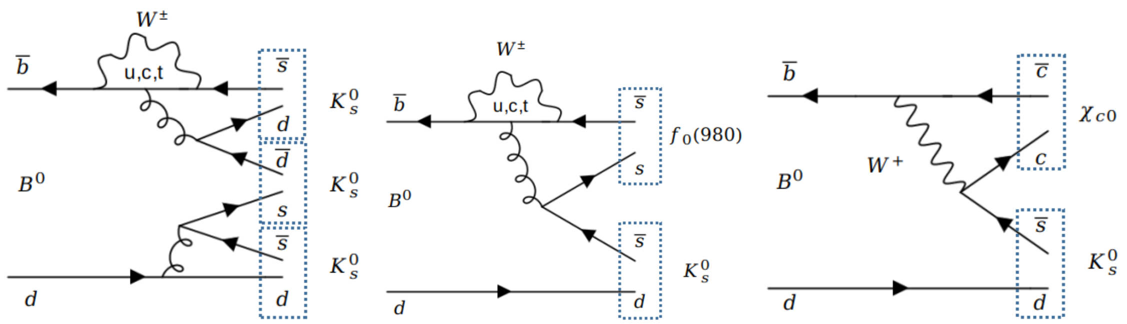


Figure 1-5: Feynman diagrams for $B^0 \to K_S^0 K_S^0 K_S^0$. The left is the non-resonant signal, the middle is the resonant signal and the right is resonant background.

The current result of $B^0 \to J/\psi K_S^0$ using the full Belle data is presented as $\mathcal{S}_{J/\psi K_S^0} = +0.670 \pm 0.029$ (stat) ± 0.013 (syst) [10]. In the meantime, the latest result from $B^0 \to K_S^0 K_S^0 K_S^0$ using the full Belle data [19] is presented as: $\mathcal{S}_{3K_S^0} = -0.71 \pm 0.23$ (stat) ± 0.05 (syst), and the result from BaBar [20] is: $\mathcal{S}_{3K_S^0} = -0.94^{+0.21}_{-0.24}$ (stat) $\pm$ 0.06 (syst). Both results have shown a small deviation from the result in $B^0 \to J/\psi K_S^0$ while the statistical uncertainties are much dominated which prevents the claim about whether NP effects are existed. For $\Delta\mathcal{S}$ from $B^0 \to K_S^0 K_S^0 K_S^0$, the experimental sensitivity of $\Delta\mathcal{S}$ will be dominated by $\mathcal{S}_{3K_S^0}$ uncertainty because the total uncertainty from $J/\psi K_S^0$ will be reduced to about 0.005 at 50 ab^{-1} Belle II data [10], which is negligible. The Figure 1-6 shows the expected $\Delta\mathcal{S}$ uncertainty from the Belle II technical design report with respect to the luminosity in future Belle II [21], which requires that the Belle II results should be prepared for $\Delta\mathcal{S} < 0.2$ in which we can start to discuss about the NP effects. The curves are extrapolated based on the Belle results from 492 fb^{-1} data and take into account the reducible systematic and statistical uncertainties [22]. Table 1.1 shows the corresponding total uncertainties of $\Delta\mathcal{S}$ in Figure 1-6 at 0.5 ab^{-1}, 5 ab^{-1} and 50 ab^{-1} luminosity. For $B^0 \to K_S^0 K_S^0 K_S^0$

at 50 ab^{-1}, if $\Delta\mathcal{S}_{3K_S^0}$ could be 0.25, the total uncertainty should be less than 0.04 to claim a 5σ deviation from the SM upper limit of 0.05 based on the QCD models for the appearance of the NP effects, which is close to the expected total uncertainty in Table 1.1. Considering the current Belle measurement, $\Delta\mathcal{S}_{3K_S^0} \sim 0.25$ is possible. The estimation shown in Table 1.1 was performed before the Belle II operation. In this thesis, we discuss the reduction of total uncertainties including results with the Belle II early data.

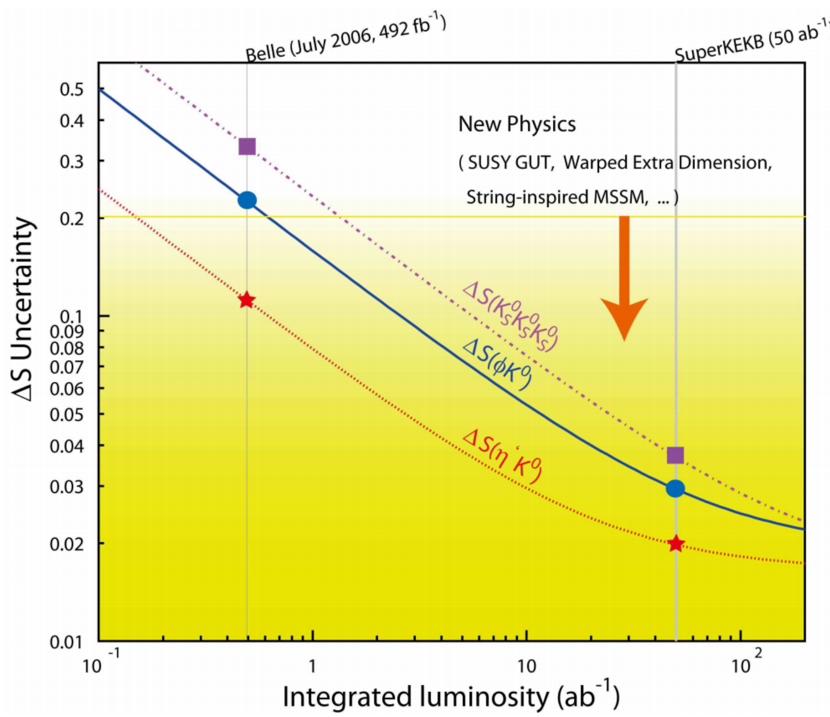


Figure 1-6: Expected sensitivity of $\Delta\mathcal{S}$ with respect to the integrated luminosity of the Belle II future data from the Belle II technical design report [21].

Table 1.1: Estimated total uncertainties on $\Delta\mathcal{S}$ with respect to the integral luminosities in the Belle II technical design report [21]. Three decay modes receive the same potential NP effects due to the same weak phases involved in the decay processes [16].

Observable	Belle (0.5 ab^{-1})	Belle II (5 ab^{-1})	Belle II (50 ab^{-1})
$\Delta\mathcal{S}_{\phi K_S^0}$	0.22	0.073	0.029
$\Delta\mathcal{S}_{\eta' K_S^0}$	0.11	0.038	0.020
$\Delta\mathcal{S}_{K_S^0K_S^0K_S^0}$	0.33	0.105	0.037

Chapter 2

Belle II experiment

2.1 Belle II and SuperKEKB overview

The goal of the Belle II experiment is to search for evidence of New Physics, and the expected operation period is from 2019 to the end of 2030. The facilities are located in KEK, Tsukuba City, around 70 km in the north of Tokyo, Japan. The SuperKEKB accelerator enables electron-positron collision at the center-of-mass energy on the region of $\Upsilon(4S)$ resonance which is just above the mass of two B mesons. The electron and positron beams are designed at 7 GeV and 4 GeV, respectively, with a boost factor of 0.28, providing an environment for measuring time-dependent CP violation by displacing the decay vertices of a B meson pair in a measurable distance along the boosted direction. The SuperKEKB has a targeted luminosity of 8×10^{35} cm^{-2}s^{-1}, a factor of 40 times higher than its predecessor, the KEKB. Some key parameters of the SuperKEKB are listed in Table 2.1. The schematic view of SuperKEKB and Belle II are shown in Figure 2-1.

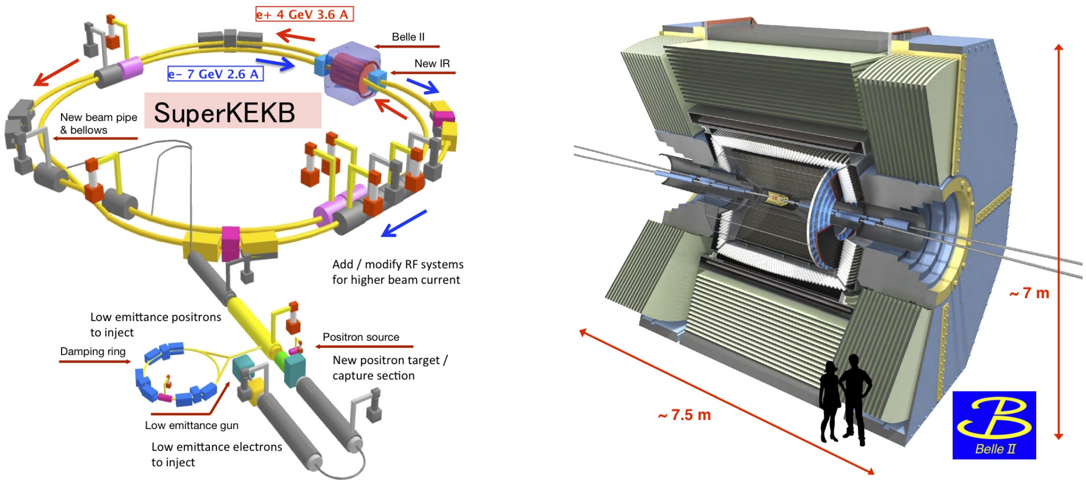


Figure 2-1: The schematic view of SuperKEKB and Belle II detector [21].

Table 2.1: SuperKEKB parameters for low energy (LER) and high energy (HER) rings [10].

Parameters	LER (e^+)	HER (e^-)	Unit
Energy	4.0	7.0	GeV
Half crossing angle	41.5		mrad
Horizontal emittance	3.2	4.6	nm
Emittance ratio	0.27	0.25	%
Beta functions at IP (x/y)	32/0.27	25/0.30	mm
Beam currents	3.6	2.6	A
Beam–beam parameter	0.0881	0.0807	
Luminosity	8×10^{35}		$\mathrm{cm}^{-2}\mathrm{s}^{-1}$
Perimeter of ring	3		km

The Belle II detector has a close size as the Belle detector so that it is placed in the same shell, but all sub-detectors and electronic systems have been either newly built or considerably upgraded. The advantage of the SuperKEKB requires that the

Belle II has to be able to stably operate at a 40 times higher event rates as well as 10 to 20 times higher beam background compared to that in the Belle. The mitigation of the effects caused by such high beam background is essential to the success of the Belle II. Higher background level leads to higher occupancy and radiation damage to the detectors, along with more fake hits in the vertex detectors and central drift chamber, pile-up backgrounds in electromagnetic calorimeter and neutron-induced hits in muon detector. Data-acquisition system (DAQ) and trigger are also upgraded not only to adapt to higher luminosity but also for a better low-multiplicity event sensitivity. The Belle II detector in the top view is shown in Figure 2-2.

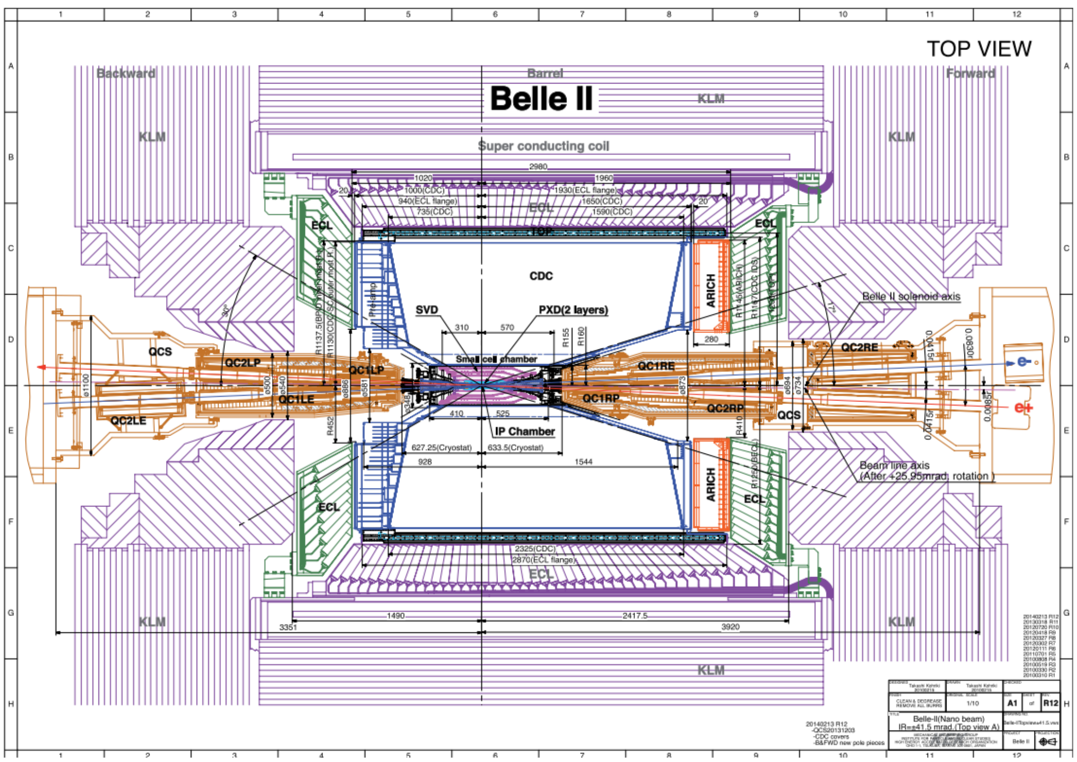


Figure 2-2: The Belle II detector top view [10].

The success of the Belle II detector depends on the complex of sub-detectors where each of them is design for specific purposes. The critical components and features are explained in the following sections.

2.2 Vertex detector (VXD)

The vertex detector is composed of two detectors, the silicon based pixel detector (PXD) and silicon based vertex detector (SVD), where total 6 layers are placed in the inner-most region from interaction point (IP). The geometry of VXD is shown in Figure 2-3. The PXD is placed at a radii of $r = 14$ mm and $r = 22$ mm with DEPFET [21] type pixel sensors, which is designed to provide two dimensional hit position information. The inner layer leaves a sufficient space for possible variations of the beampipe layout. The size of two layers are determined by the required acceptance angle from 17 degrees (forward) to 150 degrees (backward). The pixel sensor is a monolithic structure with current-digitizing electronics at the end of the senor that makes a very thin layer at about 50 microns. The schematic view of sensors on PXD is shown in Figure 2-4. As the very close range the PXD is, the sensors are exposed to

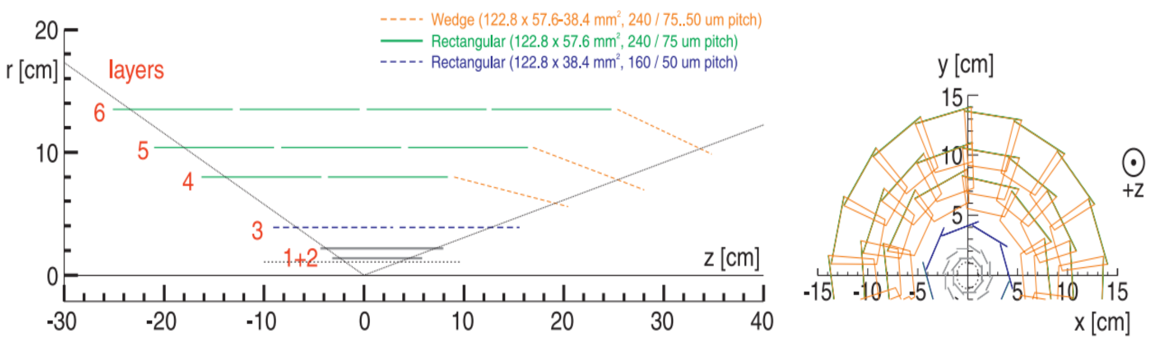


Figure 2-3: A schematic view of of PXD and SVD. Two PXD layers are in grey, SVD layer 3 is in blue and layer 4, 5, 6 are in green [21].

a very high event rate and very high beam background environment. The large data flow from PXD without any data reduction scheme is problematic for data acquisition system. In order to reduce the data that is not interested by physics analysis such as beam backgrounds, a fast online tracking system is built up for searching a "region of interest" (ROI) on the PXD sensors. To be specific, the data from PXD will be first readout to a system called "ONSEN" which can store large size temporary data up to 5 seconds. In this timing window, a fast online tracking system will perform a track fitting using vertext detector and central drift chamber to extrapolate the fitted

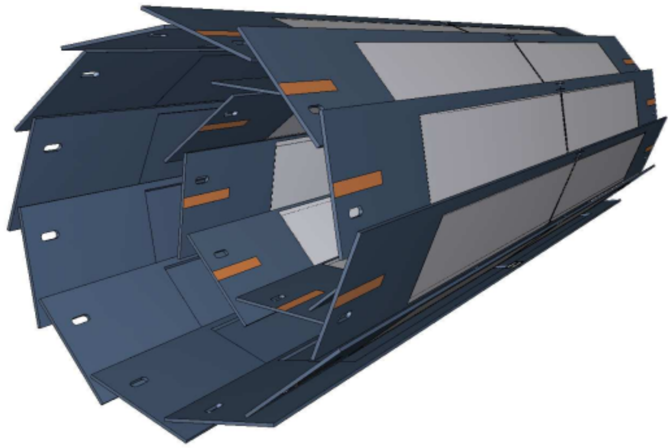


Figure 2-4: The geometry of sensors on PXD where the light grey surfaces are DEPFET sensors with a thinness of 50 microns. The full length including the out modules is 174 mm [21].

tracks backward to PXD plane so the ROI on the PXD sensors can be defined. The data from PXD outside of the ROI is not read out to external tapes where offline data is written.

SVD detector consists of 4 layers of detectors called “double-sided silicon strip detectors” (DSSDs) at 39 mm, 80 mm, 104 mm, and 135 mm away from IP, respectively. The two sides of the sensors are called p-side and n-side, where the former is for the strips on $r-\phi$ direction (transverse direction) and the latter is for the strips on the z direction (beamline direction). To suppress the background hits, a readout chip with a fast shaping time of $\mathcal{O}(50\text{ ns})$ is indispensable. The APV25 chip [23] is chosen as the readout chip that was originally developed for CMS silicon tracker, with total 128 identical channels of low-noise preamplifiers followed by a 50 ns peaking time shaper stage. The polar angular acceptance ranges from 17 degrees to 150 degrees, which is asymmetric to account for the forward boost of the center-of-mass frame. The combination between sensors, electronics and the supporting structure uses so-called “Origami” concept that stands for a chip-on-sensor design, as shown in Figure 2-5. In the Origami scheme, the readout chips APV25 are placed on a single flexible circuit mounted on the n-side of the sensors. The channels of p-side are attached by small flexible fan-outs wrapped around the edge of the sensors. All connections between flex pieces, sensor, and APV25 chips are made by wire bonds.

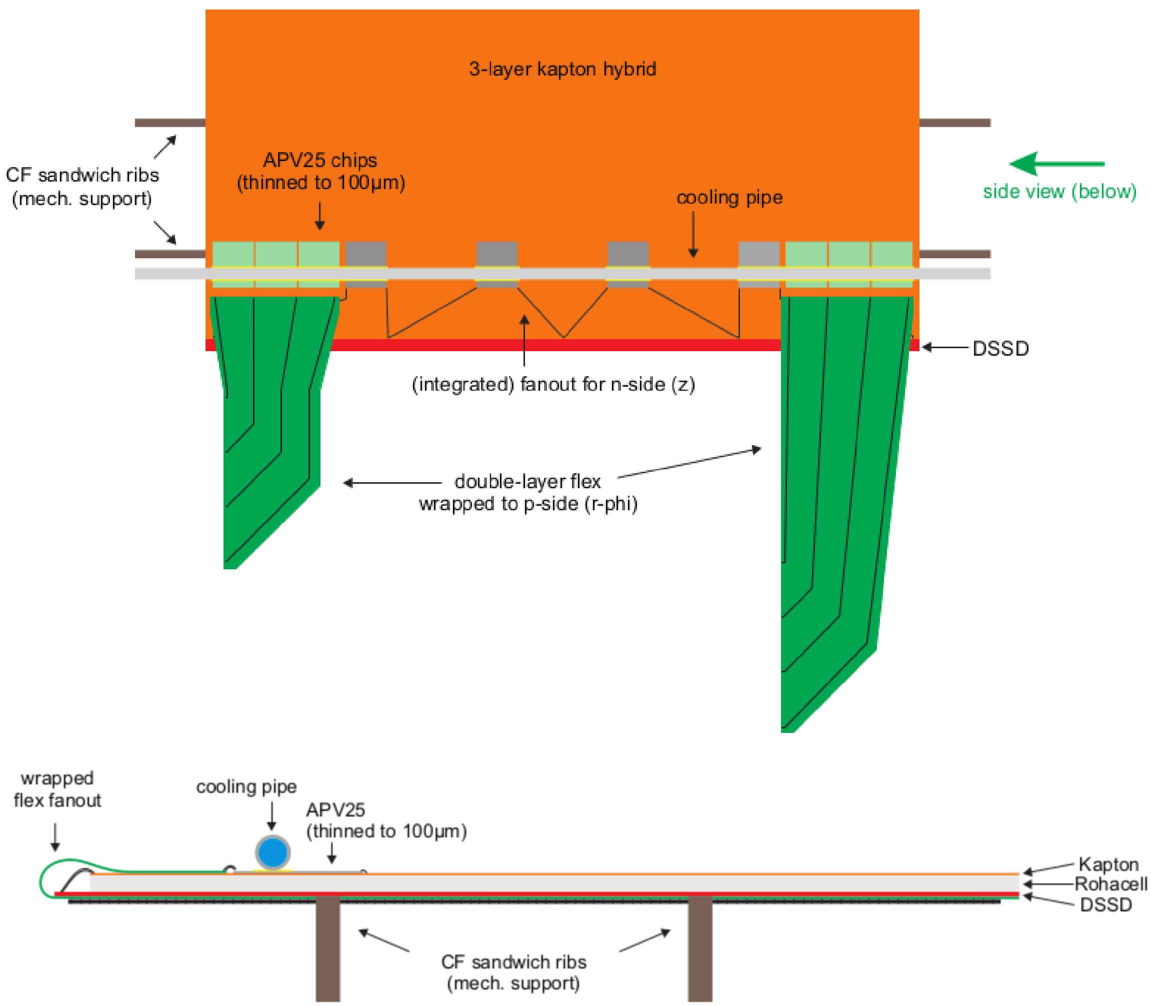


Figure 2-5: The top and side views of Origami chip-on-sensor design for DSSDs of SVD. Top: the APV25 chips in grep read out the same side sensors channel while chips in green read out the sensors on the opposite side using wrapped-around flex pieces. Bottom: side view of the Origami design shows the location of wrapped flex which connects the strips of the bottom sides which are placed at the left edge [21].

2.3 Central drift chamber (CDC)

The central drift chamber (CDC) is the core component of spectrometer in the Belle II that consists of a fairly big drift chamber made of many small drift cells filled with gas. The chamber gas is comprised of a He–C_2H_6 50%:50% mixture with an average drift velocity of 3.3 cm μs^{-1} and a maximum drift time of about 350 ns for a 17 mm cell size. The outer radius of CDC has been extended to 1130 mm from 880 mm of Belle, owed to a new thinner particle identification detector which will be introduced in the next section. The whole CDC contains 14336 sense wires in 56 layers, placed in the axial direction and the stereo direction [10][21]. Such a design can utilize the information from axial and stereo wires to construct a full 3 dimensional hits which reflects helix tracks in the CDC volume. Thus, CDC is one of the key components for measuring the helix parameters for tracking, providing precise information on the charged tracks momentum. Also, it provides particle identification information using measurements of energy loss within its gas volume. Low-momentum tracks, which do not reach the particle identification device, can be identified using the CDC alone. Last but not least, it provides efficient and reliable trigger signals for charged particles.

The Belle II CDC is expected to handle higher trigger rates with less dead time. The front-end electronics are located near the backward end-plate and send digital signals to the electronics hut through optical fibers. Due to the higher radiation and higher beam background in the Belle II, also to create more space for SVD volume, the inner radius of CDC in Belle II is 160 mm. CDC can also create three dimensional trigger information from a dedicated trigger type called z-trigger [21] based on the 3D tracking achieved by an FPGA using axial and stereo wires.

The structure of CDC consists of three main components which are a thin carbon-fiber reinforced plastic (CFRP) inner cylinder, two aluminum endplates, and a CFRP outer cylinder, as shown in Figure 2-6. The outer cylinder is a thickness of 5 mm structure supporting most of the wire tension of 4 tonnes. The inner cylinder is as thin as 0.5 mm to minimize the material and support small cell chamber such as the layers in the inner most region.

2.4 TOP and ARICH detectors

The particle identification (PID) system of the Belle II mainly consists of two parts, time-of-propagation counter (TOP) and aerogal based Cherenkov radiation imaging

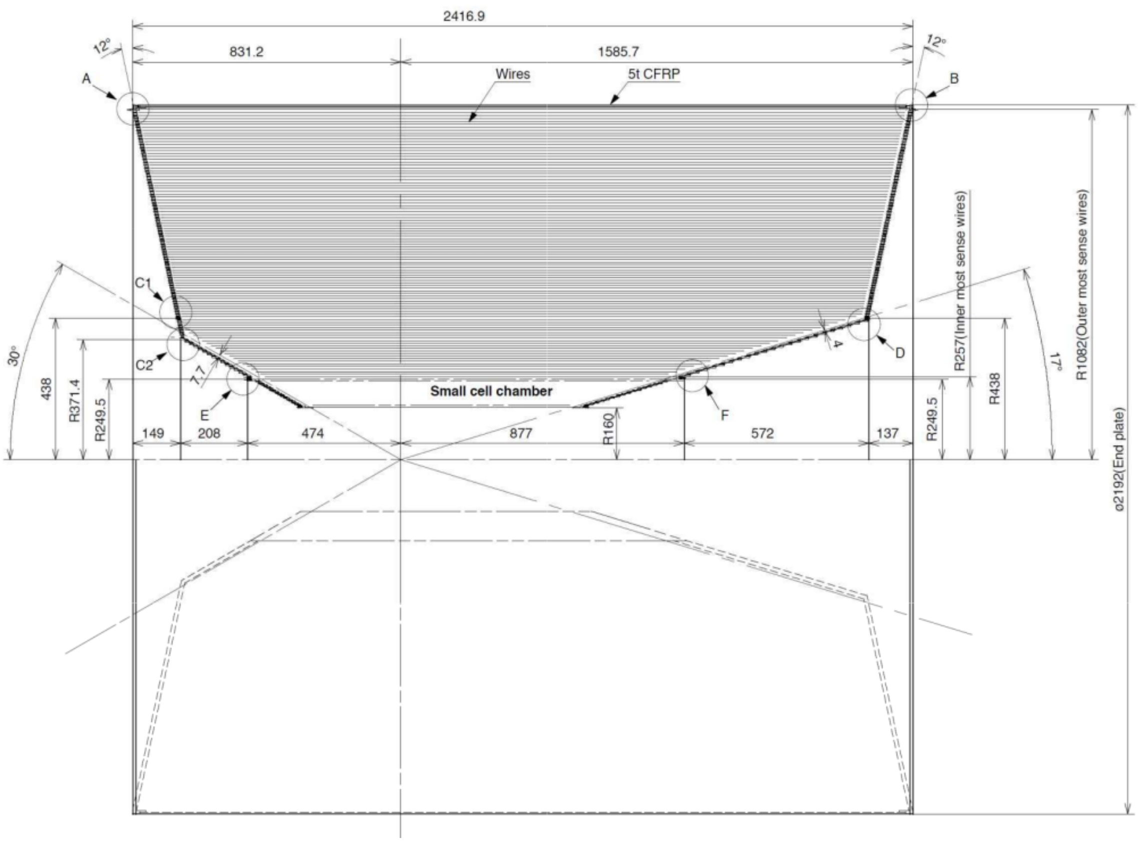


Figure 2-6: CDC structure schematic view [21].

ring (ARICH).

TOP is the specialized detector that can reconstruct Cherenkov radiation time of arrival and generated position by a photon detector placed at the end of a 2.6 cm quartz bar. The TOP is placed at the barrel region of the spectrometer, as shown in Figure 2-2. The conceptional view and the working principle of TOP counter are shown in Figure 2-7. In this counter, the time of propagation of the Cherenkov photons that are internally reflected inside a quartz radiator is measured. The quartz radiator is composed of three components. The first is a long bar for radiating Cherenkov photons. The photons then propagate via total internal reflection towards the bar end, where the MCP-PMTs are mounted. The second is a spherical mirror installed on the forward end of the bar for focusing the photons. The third is a prism that attaches to the backward end of the bar which allows the Cherenkov ring image to expand before the photons are recorded by the PMTs. By this structure, a 3-dimentional information with $x - y$ position and a timing information are obtained by micro-channel plate (MCP) PMTs at the end surfaces of the quartz bar. The resolution of starting time is achieved about 50 ps [21]. As the key component of the photon detector, the squared shape MCP PMTs, donated as SL-10 [24], have been developed with a 4×4 anode array, a multi-alkali photocathode, two MCP plates with 10 μm pore size, and an aluminum layer on the second MCP to protect against ion feedback. The image of a SL-10 MCP PMT and an anode schematic view are shown in Figure 2-8.

Aerogel Ring-Imaging Cherenkov detector (ARICH) is located at the forward endcap in Figure 2-2 to separate charged particles in a momentum range from 0.5 GeV/c to 4 GeV/c, which requires a single-photon-sensitive high-granularity sensor to reconstruct the Cherenkov angle with small photon yield. Hamamatsu company and the hardware experts from the Belle II collaboration have developed a hybrid avalanche photon detector (HAPD) to meet the requirements. Each sensor is 73×73 mm^2 embedded with 144 channels to accelerate emitted electrons in a 8 kV field. Avalanche photo-diodes (APD) are used for the detection of electrons at the end of electron acceleration, see Figure 2-9. The ARICH detector outlook and the ring

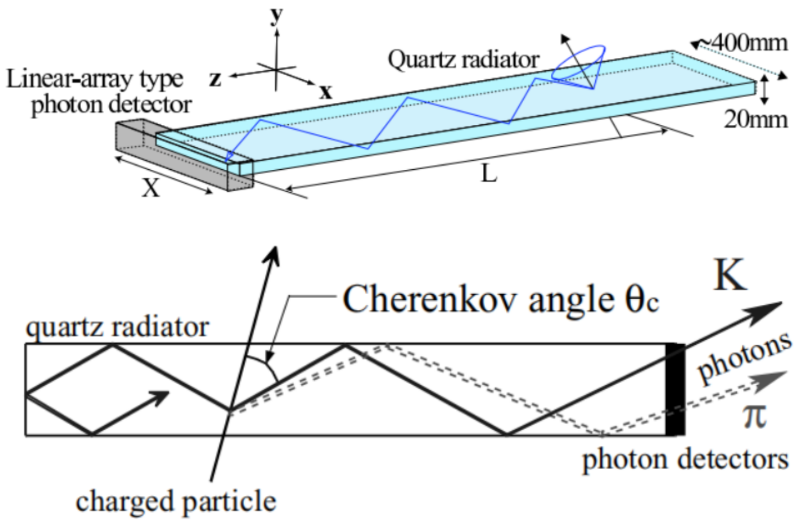


Figure 2-7: Conceptional view of TOP counter (up) and its imaging process of $K^{\pm}$ and $\pi^{\pm}$ (down) [21] for PID purpose.

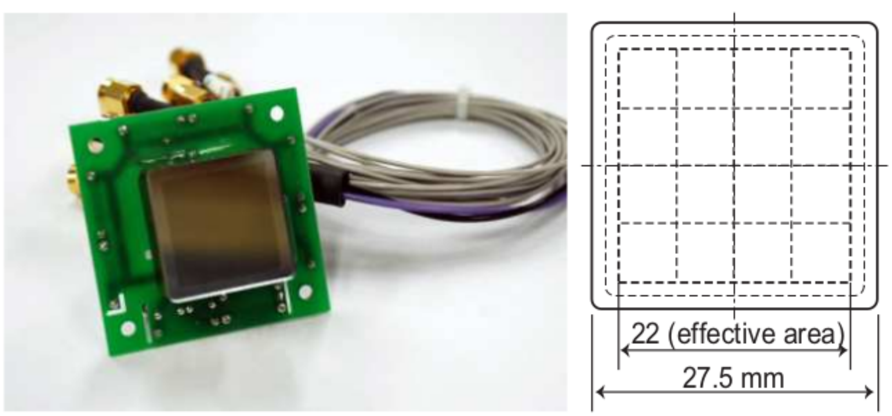


Figure 2-8: SL-10 MCP PMT (left) and the schematic view of 4×4 anode (right) [21]

image of cosmic muon on the HAPD sensors are shown in Figure 2-10.

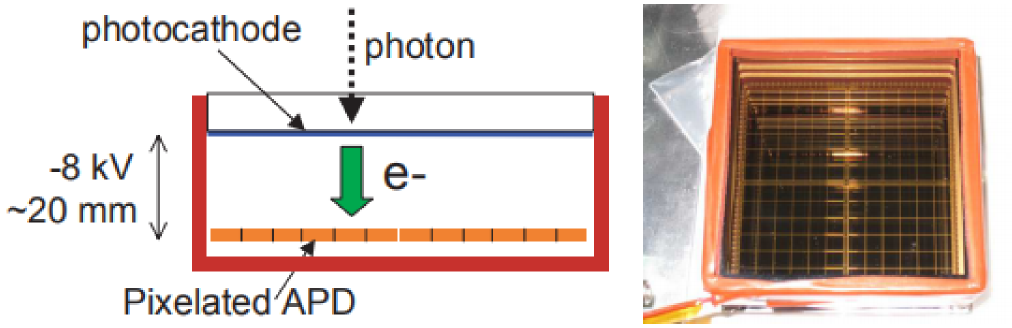


Figure 2-9: Photon-electrons acceleration (left) and pixelated APD (right) at the end [21].

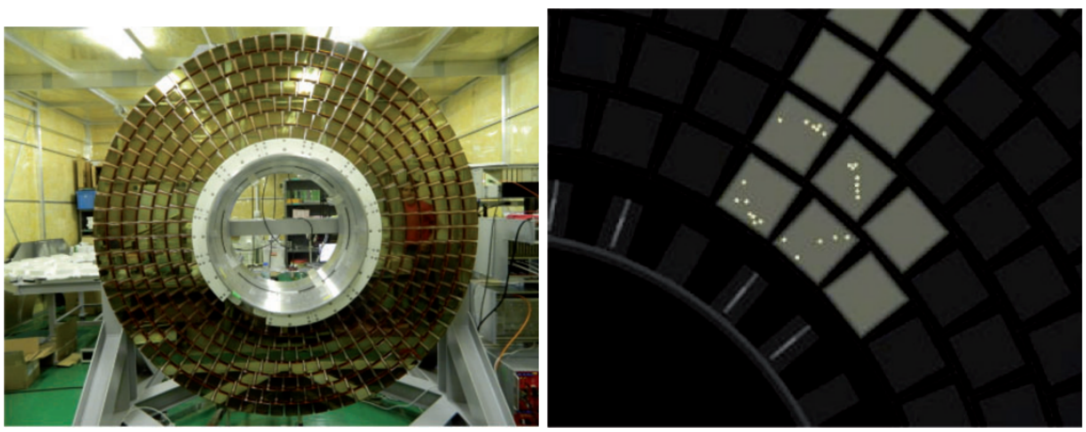


Figure 2-10: ARICH detector (left) and the ring image of cosmic muon on the HAPD sensors[10].

2.5 Electromagnetic calorimeter (ECL)

The electromagnetic calorimeter (ECL) in the Belle II is mainly responsible for the detection of γ radiation and electrons, providing energy deposition information for trigger, particle reconstruction and PID. ECL consists of three sections as shown in Figure 2-2: a 3 m long barrel section with an inner radius 1.25 m, and two annular endcaps at $z = 1.96$ m (forward) and $z = -1.02$ m (backward) from the IP. The barrel section contains 6624 CsI(TI) crystals of 29 distinct shapes and each crystal is a pyramid shape with about 6×6 in cross section and 30 cm in length. The endcaps section contains 2112 CsI crystals of 69 shapes and the total number of crystals is 8736, with a total mass of about 43 tons [21].

As the basic component of ECL, the thallium doped caesium iodide CsI(TI) crystals are assembled tightly in end-caps and barrel sections. Compared to the previous ECL in Belle, the pre-amplifiers and the structures remain unchanged, while the readout electronics have been upgraded. The estimated background level in Belle II ECL will cause the much longer decay time in the scintillation of CsI(TI). This will lead to the pile-up effect of readout noise. To compensate this effect, wave-form sampling electronics are embedded with the photon detectors (PMT).

2.6 K_L^0 muon detector (KLM)

The K_L^0 and muon detector (KLM) system of the Belle II consists of a sandwich stacked iron plates at outside of the superconducting solenoid and it acts as a return york of the magnet. The iron plates serve as the interaction materials with > 3.9 times the interacting length of material ($\sim$ 132.1 g/cm^2) compared to the ECL, allowing K_L^0 particles to shower through. The octagonal barrel covers the polar angle range from 45 degrees to 125 degrees, while the endcaps extend this coverage from 20 degrees to 155 degrees. There are 15 detector layers and 14 iron plates in the barrel and 14 detector layers and 14 iron plates in each endcap. The side view of KLM is shown in Figure 2-11. The Belle KLM material uses the glass-electrode resistivity plate chambers (RPC) which is not suitable for the Belle II due to high background level. Neutrons dose is significantly larger due to the much more electromagnetic radiation reaction on detector materials. The long dead time of RPC under such dose rate will reduce the efficiency of KLM. To mitigate this problem, the RPCs are replaced by the layers of scintillator strips with wavelength-shifting fibers, read out by silicon photomultipliers (called “SiPMs”, Geiger mode operated APDs) as light sensors, which is proven to be able to reliably operate by setting up the discrimination threshold [10].

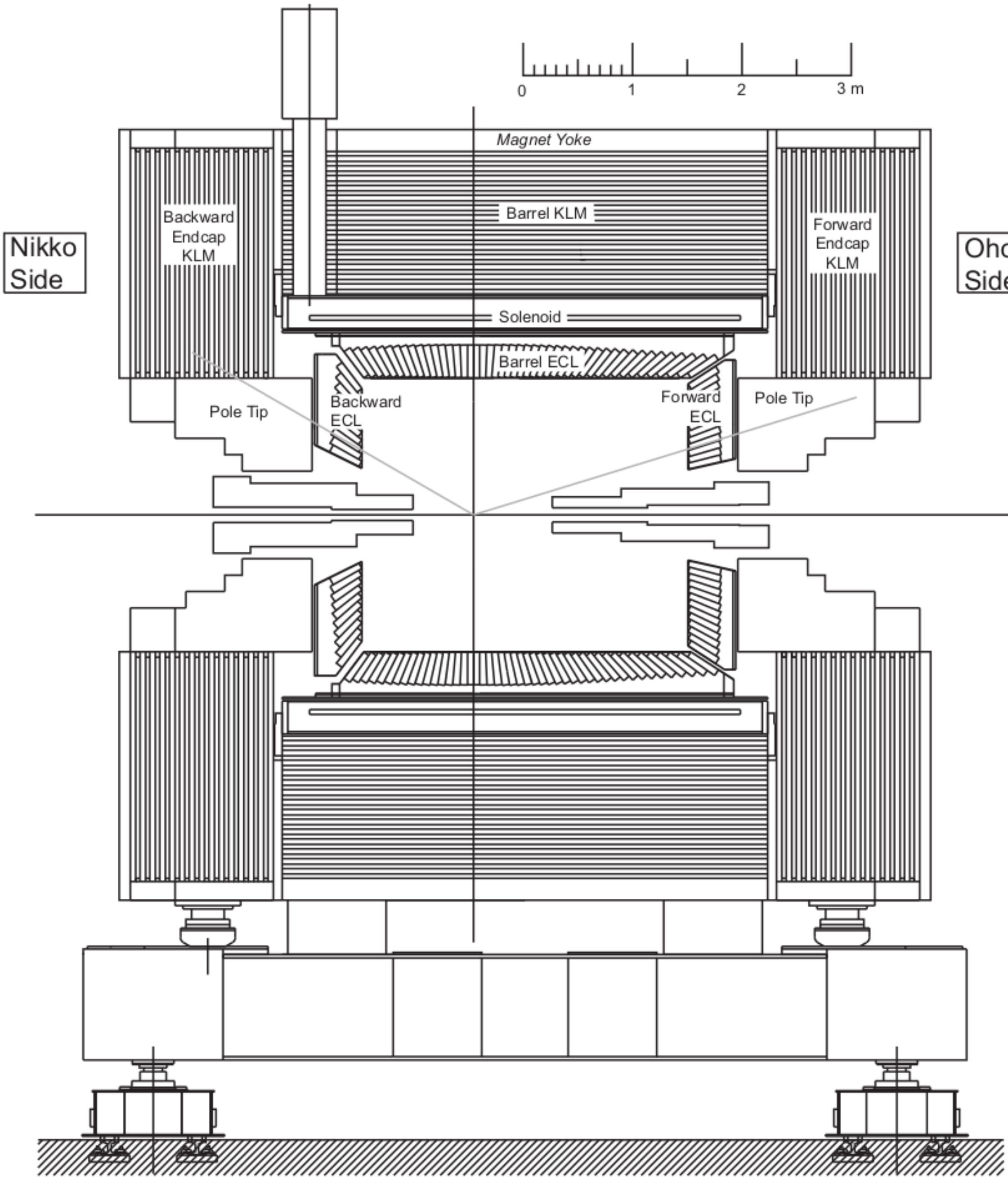


Figure 2-11: The side view of KLM in between the ECL and the solenoid, which the grey lines presents the nominal acceptance angle of the Belle II [21].

2.7 Trigger and DAQ system

The interesting topics in Belle II physics analysis highly depend on the trigger system. The Belle II trigger system is composed of two levels: a hardware-based, low-level trigger called "L1" trigger, and a software-based high-level trigger (HLT). The L1 trigger has a latency of $\sim 5\ \mu$s and the maximum trigger output rate is 30 kHz, which is limited by the read-in rate of data acquisition system (DAQ). Considered the high event rate and background level from future Belle II luminosity, a series of upgrades have been implemented for L1 trigger. The key improvements of L1 come from the firmware-based reconstruction algorithm and trigger logic.

The HLT, as the second level of Belle II trigger systema, plays an important role in DAQ. As discussed in the section of PXD, the data size in PXD is huge at high luminosity and the ROI selection must be applied to reduce it. The HLT will first use fast online tracking by CDC and ECL information to further reject the residual beam background not found by L1 trigger. Only the events passing this step are considered for the full event reconstruction. Then the information from all detectors except for PXD are fed into the first event builder for full event reconstruction. The event rate is reduced to about 6 kHz by HLT which uses the full reconstruction information to find track-associated hits on PXD, introduced as ROI before. The workflow of DAQ with HLT is demonstrated in Figure 2-12. The reduced event rate by applying ROI finding on PXD and other detector read-out systems are combined into the second event builder and eventually written to the offline storage.

Since the primary goal of the Belle II is focusing on B physics studies, it is natural that the trigger system should be able to operate over all of the interesting B physics conditions, with normally 3 or more CDC tracks and large energy deposition in ECL. By studying the efficiency using the simulated events, close to 100% B decays are recorded by Belle II trigger system. Besides, the Belle II detector is expected to capture many other physics events such as searching for leptonic flavor violation using τ decays or dark matter particles, of which the performance is highly affected by the beam background level and trigger efficiency. Therefore, the control

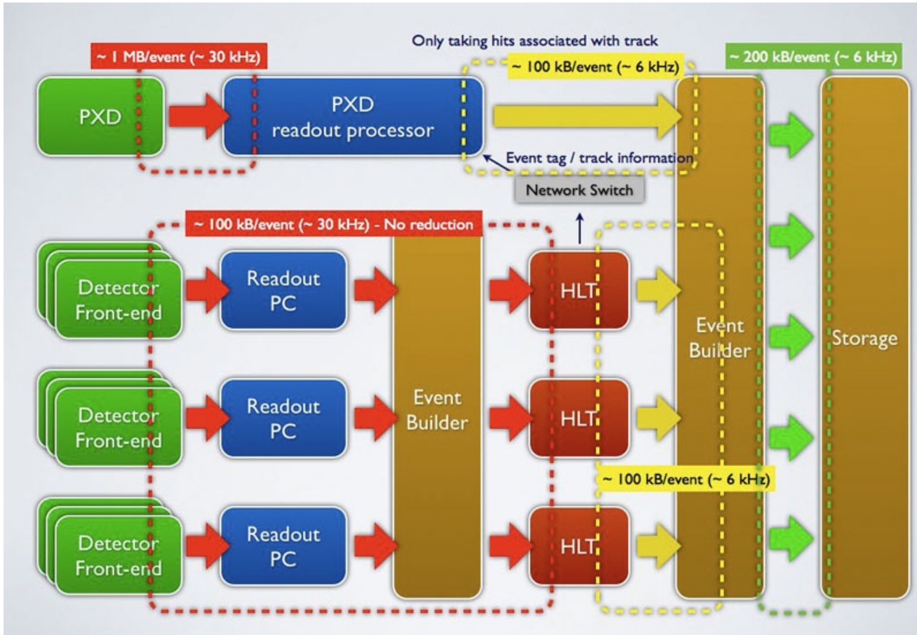


Figure 2-12: The Belle II DAQ workflow with HLT between two event builder to reduce the original 30 kHz event rate down to about 6 kHz for offline storage.

of beam background becomes essential, which mainly consists of beam-gas scattering, synchrotron radiation, the radioactive Bhabha scattering, the two-photon process, beam-beam effects, and Touschek effect. Their impacts depend on many factors such as beam current, luminosity and vacuum conditions, etc. One of the featured topology of these beam background events is the combination of two charged tracks in CDC and one or two clusters in ECL. The sources of the main beam backgrounds and their event rates in simulation are listed in Table 2.2.

2.8 Analysis software framework

The data acquired by the Belle II experiment or simulation can be processed by the Belle II Analysis Software Framework, called BASF2. It has a good capability to handle multiple tasks for the Belle II data analysis, from the simulated data production to physics events reconstruction. The BASF2 takes the advantage of good efficiency and reliability of C++ as the programming language, but the use of Python is also encouraged when it shows clear advantages, such as steering the analysis workflow.

Table 2.2: Simulated beam background rate [10]

Type	Source	Rate (MHz)
Radiative Bhabha	HER	1320
Radiative Bhabha	LER	1294
Radiative Bhabha(wide angle)	HER	40
Radiative Bhabha (wide angle)	LER	85
Touschek scattering	HER	31
Touschek scattering	LER	83
Beam–gas interactions	HER	1
Beam–gas interactions	LER	156
Two-photon QED	-	206

2.8.1 BASF2 Core Structure

The core structure of BASF2 contains three major parts: the analysis packages required by the needs of analyzing the Belle II data such as finding tracks and combining particles, the external libraries as the third-party software such as ROOT [25], and the tools for configuring and installing BASF2 which are mostly Python and shell scripts. Data analysis is supported by providing a series of modules belonged to BASF2 for appropriate reconstruction based on their specific needs. To realize this, a modular analysis workflow, where each module can handle the event data through an unified method such as ROOT I/O based object persistency, is desired. Other processes, such as data summary table (DST) processing, simulation of each sub-detectors, and data skimming, are done with the packages built for sub-detectors.

The packages are categorized based on the different levels of Belle II detector components, like the packages of base-level system control called "framework", the package that provides the simulation of each sub-detectors like "svd", the package for track reconstruction called "tracking", and the package for post-reconstruction data analysis called "analysis", etc. Users can work either with compiled binary version of BASF2 installed centrally on working servers, or build from the source based on their own need. Furthermore, the distributed computing is also supported by the installations of BASF2 through the managment service provided by DIRAC system [26]. The detail information about the core structure of BASF2 can be found in Ref. [27].

2.8.2 Event processing workflow

The data from Belle II detector or from the simulation, are organized into a set of runs that are defined by either experimental conditions or simulation conditions. For instance, the simulation data from the condition of a certain detector is packed together, marked with the condition database index used during the simulation. Such data sample then is divided into different runs based on estimated luminosity from experiment, which can contain the different number of events in each run. This scheme is used for categorizing experimental data as well, so that users can easily know which experiment conditions are used. Thus, when BASF2 processes a data set, the functions are called for every event based on different configurations that are corresponding to the different experiment conditions. For example, in a data set where events are recorded with the different magnetic fields, BASF2 can automatically change the configurations of the magnetic fields event-by-event to provide a better track measurement. Based on this idea, all BASF2 functions (called "modules") are developed based on a python module class which contains following embedded functions to be called at event-based level:

- initialize: called at the start of processing events to prepare this run, including how many events will be processed and declaration of the buffer space and memory required by this module.

- beginRun: called after the initialization is finished and before the event read-in starts, including setting up database conditions used in this run (run-dependent configurations) or event (event-based configurations).

- event: called when each event is read and start to process. This is the actual processing step, such as perform tracking or combining all daughters to find a mother particle.

- endRun: called at the end of a run, usually to register all processed information to the storage, such as physics variables from all reconstructed particles.

• terminate: called at the end of the processing of all events, release the buffered space and memory.

BASF2 executes a series of modules loaded dynamically to process the data set for analysis purposes, which is shown as Figure 2-13. The selection, configuration and executed order of the modules are defined by a file called "steering file" written in Python. The modules parameters are attributes which can be set during the runtime using the steering file.

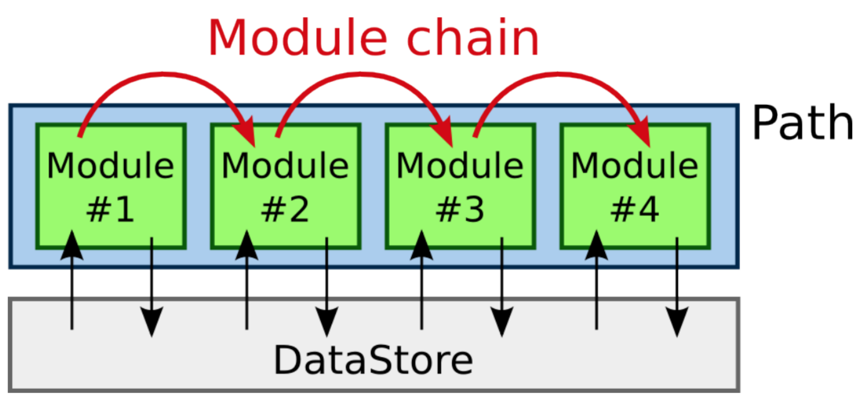


Figure 2-13: The module-based analysis workflow in BASF2.

The object that interacts with BASF2 I/O is called "DataStore", as shown in Figure 2-13. This implementation doesn't depend on the event data model. The only mandatory component is called "EventMetaData" which presents the experiment, run and event number of a event. "Unpacker" module converts the raw digits into digits-based object in BASF2. In simulation, digitization is done by module called "digitizer". The digits-based objects are further processed to form hits or clusters depending on detector types. Higher level functions such as tracking and decay reconstructions are implemented based on these basic information by their packages. Eventually, BASF2 writes out the information based on users' needs, like kinematics variables, to ROOT format files, or simply prints out processing statistics to the standard output.

2.8.3 mDST structure

The output of BASF2 processing from the online data contains several detector-specific objects, which are restored as mini data summary table (mDST) type ROOT file. For a mDST level analysis, the goal is usually aimed to find particles from physics processes and reconstruct decay information. A output mDST ROOT file contains the reconstructed objects from each sub-detectors, and the following items are required for $B^0 \to K_S^0 K_S^0 K_S^0$ analysis.

- Track: object presenting any charged particle trajectory. It is linked to multiple track fit results using different nominal mass hypotheses as well as their track fit quality to help select good tracks.

- TrackFitResult: the fitting result of tracks with different mass hypotheses. It consists of five helix parameters, their covariance matrix and p-value from the fit. It also stores the information of hit pattern on VXD and CDC.

- V0: object for the relative long-lived neutral particles that fly out of interaction region but mostly decay or interact inside detector region. In Belle II, these are mostly K_S^0, Λ and photon converted to a electron-positron pair. V0 also stores their relation to the charged daughter tracks and track fit results for further selections.

- PIDLikelihood: it presents for the possiblity of a charged track to be an electron, muon, charged kaon and pion, proton and deuteron provided by particle identification system.

- MCParticle: simulated particles and particle-detectors relations are created if simulated particles are correctly reconstructed as tracks or clusters.

2.8.4 Conditional Database

In addition to the physics data, analysis relies on various conditional data that are different calibration of detector, weight files for multi-variate analysis usage like PID and so on. This data is stored in a central database server called central Conditional Database (CDB) [28]. Conditions are made of payloads and each payload has its own "Intervals of Validity" (IoV) which defines in which runs the payload is valid. A collection of the payloads that are produced based on a certain stage of the experiment is packed together and called as a global tag (GT).

Users can create a GT, add objects of payloads to it and commit the GT to the configured database with a user-supplied IoV. This includes the support for run dependency as well. The capability to use a local file-based database allows for easy preparation and validation of new payloads before they are uploaded to the CDB. Only the creator of the payload objects has the right to add, recall, replace and remove the GT from CDB, which guarantees the stability.

2.9 Belle II simulation

This section briefly describes simulation (MC) used in the studies presented in this thesis. As this analysis is based on neutral B meson, which is from the $\Upsilon(4S)$ events, the simulation is based on the electron -positron collisions at center-of-mass (CMS) energy $\sqrt{s} = 10.58$ GeV.

In the previous section, it is shown that external packages and functionalities have been integrated with BASF2, including the core components of Belle II simulation in B decay: *evtgen* as event generator [29] and *GEANT4* as the simulator of detectors [30]. For the simulation and the reconstruction used in this analysis, the latest release of BASF2 (release-05-01-01) was used. Based on the CDB management, BASF2 can utilize the same constants such as the magnetic field distribution for the consistence between simulation and reconstruction.

All simulations start with at least one event generator that configures the physics processes. The *evtgen* requires a decay file that describes the decay chain from a certain mother particle, branching fraction for all processes and decay-related information such as flavor mixing or CP violation information. MC sample is centrally produced using Belle II grid computing service by DIRAC system and skimmed, of which the output is for physics analysis to create ROOT files. Each round of MC sample is packed and marked by their production index, such as *MC13*, which is the latest MC sample with improvements in PID. In the following content of this thesis,

Table 2.3: Production cross section for different hadronic flavors from collision at $\sqrt{s} = 10.58$ GeV used in Belle II *generic MC* [10].

Processes	$\Upsilon(4S)$	$u\bar{u}(\gamma)$	$d\bar{d}(\gamma)$	$s\bar{s}(\gamma)$	$c\bar{c}(\gamma)$
Cross section [nb]	1.110 ± 0.008	1.61	0.40	0.38	1.30

all MC samples are produced in *MC13* if not specifically stated.

For the analysis in this thesis, there are two MC samples included, where one is called *signal MC* and the other is called *generic MC*. *Signal MC*, as its name suggests, is the MC sample that describes the whole decay chain of $B^0 \to K_S^0 K_S^0 K_S^0$. The mother particle of the decay chain is $\Upsilon(4S)$, then it decays into a pair of $B^0 - \bar{B}^0$ at branching fraction of 100%, with the model *EvtVSSMix* [29] describing the decay model. Then, one of the B^0 meson is set to decay into three K_S^0 based on phase-space model ($PHSP$) at 100% branching fraction. The default configuration of evtgen can not handle multi-bodies charmless B decay with TDCPV. A modified decay model profile is under-development and not fully validated yet. Thus, MC sample of $B^0 \to K_S^0 K_S^0 K_S^0$ yields zero CP violation by default. As for the other B meson, it decays into all possible final states that are described by the Belle II generic decay file.

As for *generic MC*, all hadronic processes in a $\sqrt{s} = 10.58$ GeV collision are simulated. The total production cross section receives contributions from not only $\Upsilon(4S)$ (b-flavor decay dominated), but also u, d, s, c. Their relative branching fractions are taken from cross sections at $\sqrt{s} = 10.58$ GeV as shown in Table 2.3. *Generic MC* sample contains 6 types of MC samples due to this production arrangement, where $\Upsilon(4S)$ produces *mixed* (neutral) and *charged* B meson pairs and the rest are other flavor mesons possibly with one extra photon emission named as $u\bar{u}(\gamma)$, $d\bar{d}(\gamma)$, $s\bar{s}(\gamma)$, and $c\bar{c}(\gamma)$, respectively. In this thesis, the latter 4 types of MC samples are combined and called $q\bar{q}$ for simplicity. In the mixed MC sample, the branching fraction of $B^0 \to K_S^0 K_S^0 K_S^0$ is set at 6×10^{-6} and the branching fraction of $K_S^0 \to \pi^+\pi^-$ is set at 0.692. Both values are taken from Particle Data Group (PDG) [31]. As the same as *signal MC*, *CP* violation is set to zero for signal events in *generic MC* since they use the same model at generator level.

In addition to the simulation of physics processes, simulated data is produced with at least two beam background conditions, called *BG0* without beam background and *BG1* with one overlay of beam background. The components of them have been discussed briefly in section 2.7. The mixing of simulated beam background to simulated physics events is done by adding simulated hits on each sub-detector output. Possible pile-up of hits is therefore inherently included. The average number of background events of a given type to be added to a single simulated event is determined from the rate R_{BG} of beam background sample and the time window Δt in which the background is mixed shown in Equation 2.1:

$$\bar{N} = sR_{BG}\Delta t \tag{2.1}$$

where s is an optional scaling factor. The injected background events are based on a Poisson distribution with mean $\bar{N}$. Within the timing window, the background events are shifted randomly to simulate contributions from different bunches. To use real experiment background events (data-based beam background), the random triggered events are measured and added to simulated $BG0$ MC sample for a more precised background configuration. This method can give a more realistic description of actual beam background but with a possibility to introduce bias due to the pile-up effect of multiple background events in a short timing window. In the early stage of the Belle II, the level of background is not high and the background pile-up effect is small.

In total, there are 2 million events generated in *signal MC*. Half of the *signal MC* (1 million) is produced without beam background for cross-checking the reconstruction performance. For *generic MC*, 1 ab^{-1} sample including mixed, charged and $q\bar{q}$ events are produced with beam background at $\sqrt{s}$ = 10.58 GeV. The MC sample used in this analysis is summarized in Table 2.4

2.10 Belle II data taking

The Belle II beam test operation started in 2016 which was focused on the commissioning and test of the SuperKEKB accelerator. Later in 2018, the commissioning

Table 2.4: MC samples with and without beam background used in $B^0 \to K_S^0 K_S^0 K_S^0$ analysis.

Events number	BG0	BG1
signal MC	10^6	10^6
generic MC	None	1 ab^{-1}

of the Belle II detector was accomplished, with partial installation of PXD and full installation of SVD. From 2019 April, the physics run operation has officially started. The rest of the PXD is scheduled to be installed in 2023. By the end of 2020, Belle II has been operating for 4 total run seasons. The integrated luminosity collected during this period of time is about 84.73 fb^{-1}, shown in Figure 2-14. The indices of physics runs are labeled which are experiment 7,8,10 for 2019 data taking and experiment 12 and 14 for 2020 data taking, as shown in Figure 2-14. The data processing is regularly performed along with the data taking. For the analysis reported in this thesis, the experimental data collection from experiment 7, 8, 10 and 12 is used. Correspondingly, the integrated luminosity for offline reconstruction used in this thesis is about 62.8 fb^{-1} [32]. The experiment 14 is not used due to the unfinished processing of the latest experiment data by the time this thesis is composed.

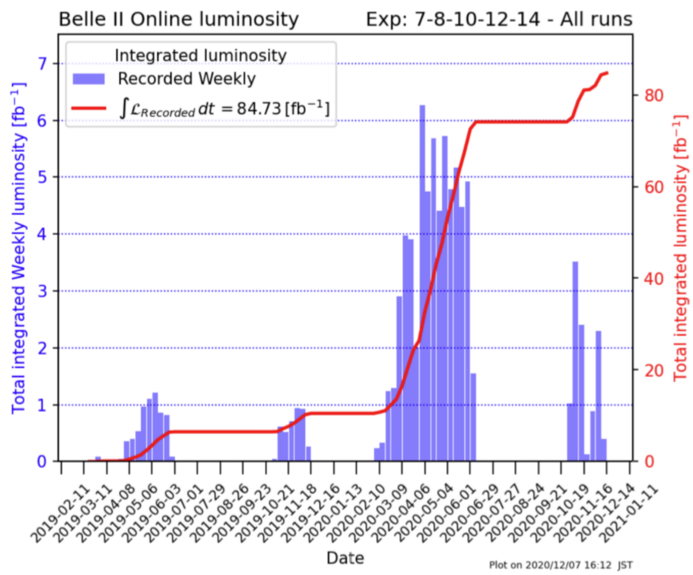


Figure 2-14: Belle II online luminosity from 2019 April to the end of 2020. The experiment 7 and 8 were conducted during 2019 March to June. The experiment 10 was conducted during 2019 October to 2019 November. The experiment 12 was conducted during 2020 February to June. The experiment 14 was conducted during 2020 September to November.

Chapter 3

K_S^0 reconstruction study

The final states of $B^0 \to K_S^0 K_S^0 K_S^0$ only depends on the decay of K_S^0. The main decay channels of K_S^0 is to either $\pi^+\pi^-$ at branching fraction of about 0.692, or to $\pi^0\pi^0$ at branching fraction of 0.307, referenced from PDG [31]. The characteristics of these two decays are much different in terms of the response from the Belle II detector. The charged decay that yields $\pi^+\pi^-$ leaves two tracks originating from VXD or CDC volumes with the opposite charges. On the other hand, the π^0 main decay channel is $\pi^0 \to \gamma\gamma$ which typically results in the photon clusters on the ECL. There are mainly two reasons for not selecting π^0 to be used as final states for reconstructing B^0. First, $\pi^0 \to \gamma\gamma$ can yield a large fraction of fake K_S^0. The reconstruction of two photons using ECL clusters provides no constrain on K_S^0 vertex so it is almost impossible to suppress the combinatorial background using vertexing information in this case. The photons could be originating from many other resources, such as beam background and charged particles radiation. Besides, the most useful selection is the invariant mass of K_S^0 which is typically distributed around its nominal mass with a few hundred of keV. However, using the mass window of K_S^0 could not effectively reject the noticeable fraction of fake K_S^0, especially when using photons. Second, B^0 that decays to one or more K_S^0 reconstructed from neutral pions have poorly reconstructed vertices. Even with $B^0 \to K_S^0 K_S^0 K_S^0$ which only uses K_S^0 from charged pions in the final states, there is no direct charged tracks from IP,leading to a worse resolution of vertex position compared to the channel like a $B^0 \to J/\psi K_S^0$ that has two direct charged tracks of e^+e^- or $\mu^+\mu^-$ from J/ψ. If one (or more) of K_S^0 has the poor vertexing quality from its decay products, it can further reduce the precision of vertex positions of B^0. Such a degradation of precision of the vertex position eventually results in the large uncertainties in decay time difference Δt as the key observable in the time dependent CP violation (TDCPV) study. Therefore, only K_S^0 reconstructed using charged pions is considered in this analysis.

3.1 Cut-based K_S^0 Reconstruction

The average life time of K_S^0 is $(8.954 \pm 0.004) \times 10^{-11}$ s according to the PDG, which corresponds to the average flight length of a few centimeter. Therefore, the flight length of K_S^0 is comparable with the scale of VXD size. In the Belle II energy scale, the range of the flight length of K_S^0 could be from a few μm away from B vertex to more than 13.5 cm that is further than the outmost layer of SVD ladders, see Figure 3-1. Due to the different topology of B^0 decay, the average momentum of K_S^0 in *generic MC* is different from the ones of the *signal MC*.

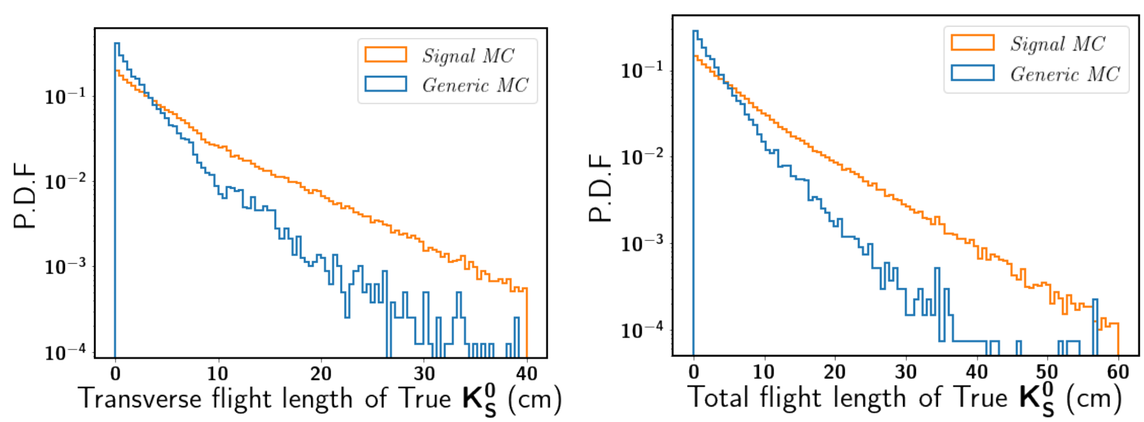


Figure 3-1: The left is the transverse flight length distribution and the right is the total flight length distribution from true K_S^0. The blue is from *generic MC* and the orange is from *signal MC*. Both plots are normalized.

The cut-based reconstruction for K_S^0 is first performed by the selection of invariant mass from its decay products. After the selection on invariant mass is applied, a vertex fit for each K_S^0 using two reconstructed charged pions is done without IP constraint.

This reconstruction is mainly achieved by using standard a BASF2 particle list, in which two K_S^0 collections are first reconstructed and then merged. We first take all the V0 objects from BASF2 which use 2 online reconstructed charged tracks with opposite charges and a converged fitted vertex. In this step, charged particles with mass hypothesis of $\pi^{\pm}$ are used, where the tracks and PID of charged pions are pre-selected by the criteria in Table 3.1. Then the K_S^0 candidates with invariant mass M between $0.45 < M < 0.55$ GeV are selected. In addition to these K_S^0 from V0 objects, another K_S^0 collection from offline reconstruction is also formed by using the same selection criteria for pions and K_S^0 invariant mass. The V0 based K_S^0 and offline reconstructed K_S^0 are merged and the vertex fit is performed using *TreeFit* [33]. After the vertex fit, the invariant mass of K_S^0 of the fitted pions is required to be in between $0.3 \sim 0.7$ GeV to further rejected fake candidates. The duplication of K_S^0 between two K_S^0 collections is possible so that the object indices of two charged pion tracks in BASF2 are compared, from which the identical combinations are removed to avoid duplication. The B^0 reconstruction efficiency is highly sensitive to the efficiency of charged pions because the final state particles are three identical K_S^0 decaying to six charged pions. That is why a very loose selection on $\pi^{\pm}$ is applied. The selected K_S^0 collection using cut-based method contains many fake candidates. The distribution of the invariant mass using *signal MC* is shown in Figure 3-2, which shows 39% true K_S^0 and 61% fake K_S^0.

Table 3.1: Pre-selection criteria of $\pi^+\pi^-$ for K_S^0 reconstruction.

Selection	θ	CDC Hits Number	PID
Criteria	CDC acceptance	> 20	pionID > 0.1

The reconstruction quality of K_S^0 also depends on the flight distance. K_S^0 that decay in the inner region of VXD yields more hits on the SVD layers associated with the charged tracks of pions, which is critical for providing tracking information together with CDC hits. The Belle II track fitting quality becomes much worse for those without inner detector hits association, especially SVD hits information. To further study the reconstruction of K_S^0 based on their SVD hits, they are categorized

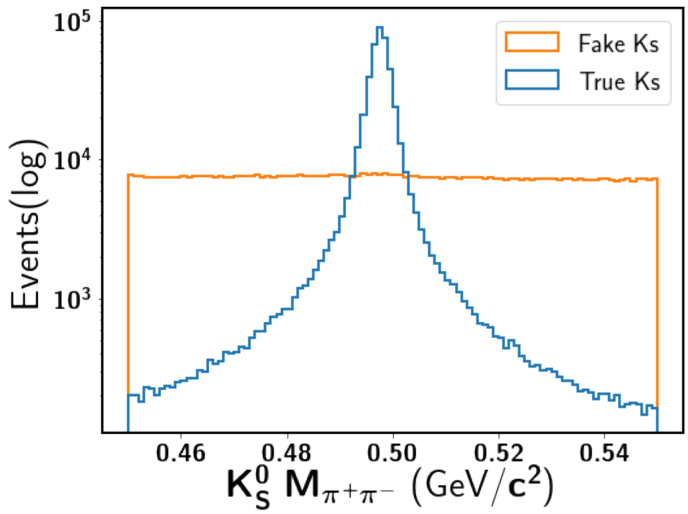


Figure 3-2: M of K_S^0 from cut-based selection in *signal MC*. The blue line is the true K_S^0 and the orange is the fake K_S^0. 200000 candidates are used in total.

by how many SVD hits their daughter tracks are associated with, in which *SVD10* and *SVD01* stands for K_S^0 that only π^+ and π^- has non-zero SVD hits number, *SVD11* and *SVD00* stands for K_S^0 that both or neither charged pions have SVD hit non-zero SVD hits number. The K_S^0 fraction of each category are listed in Table 3.2.

If we compare the distribution of the invariant mass before and after the K_S^0 vertex fit in each category, *SVD00* K_S^0 shows a large dispersion from the $0.45 \sim 0.55$ GeV to $0.3 \sim 0.7$ GeV, while *SVD11* K_S^0 shows a much smaller dispersion in Figure 3-3. This indicates that the absence of SVD information leads to the inaccurate K_S^0 reconstruction. Therefore, considering the K_S^0 candidates with different SVD hits, a series of different cuts on invariant mass $M_{\pi^+\pi^-}$ are applied to improve the purity for well-reconstructed K_S^0 candidates. As shown in Figure 3-4, the sideband regions, where fake K_S^0 is much higher than true K_S^0, are excluded. The cut windows are listed in Table 3.3.

Fake K_S^0 candidates can cost a large extra processing time and the number of combinatorial backgrounds in $B^0 \to K_S^0 K_S^0 K_S^0$ becomes high so that it can significantly reduce the signal significance and introduce bias to the CP parameters measurement. Thus, a multi-variate analysis (MVA) based K_S^0 classification package, *KsFinder*, is developed to further reject the fake K_S^0 from cut-based selected candidates.

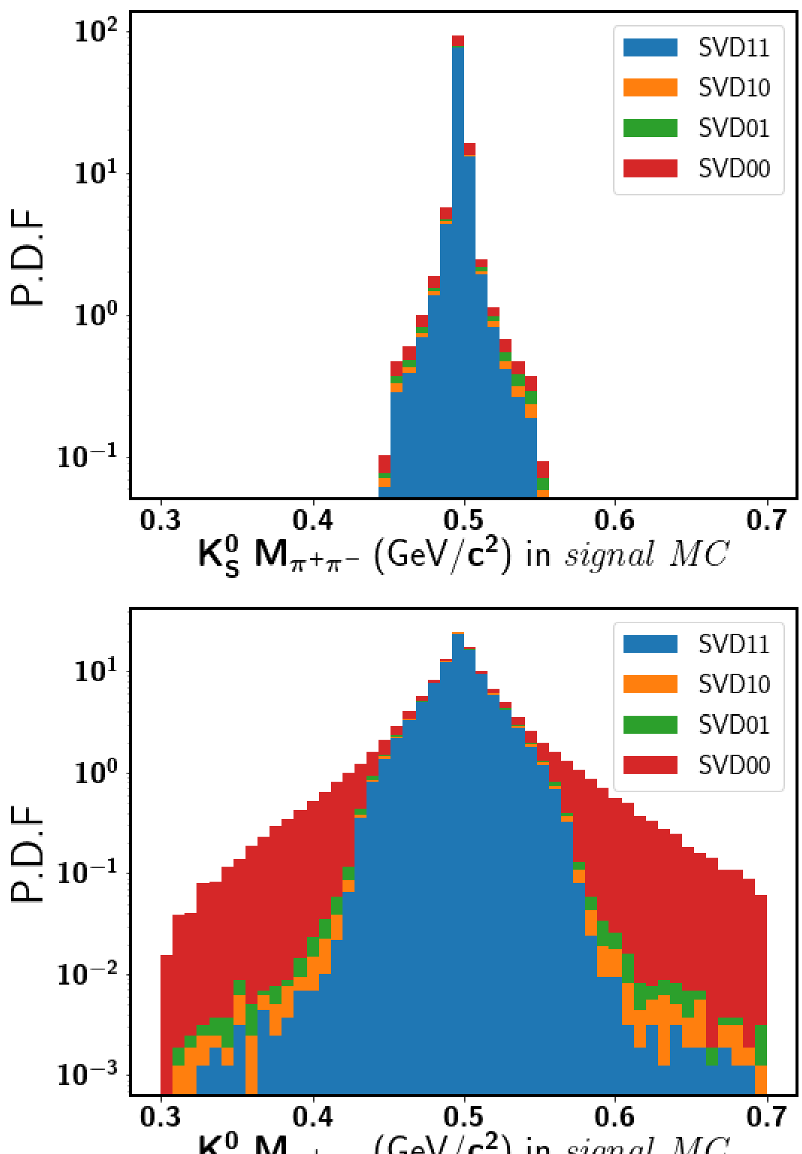


Figure 3-3: The invariant mass before (top) and after (bottom) vertex fit distribution based on SVD types, which shows a clear dispersion in *SVD00* particularly, indicating the inaccurate reconstruction of K_S^0 masses without SVD information.

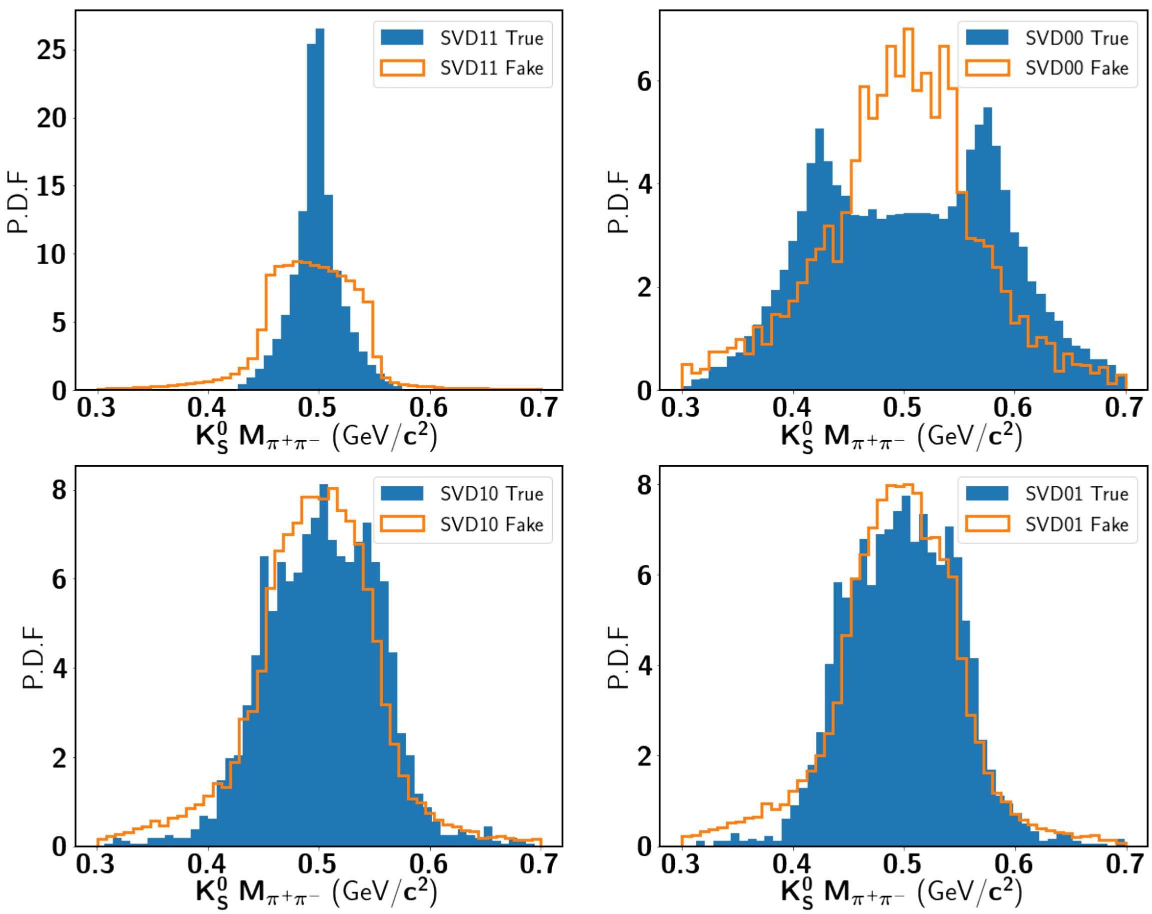


Figure 3-4: K_S^0 invariant mass after vertex fit, where the sideband regions are excluded in these distributions to further reject fake K_S^0

K_S^0 type	*SVD11*	*SVD00*	*SVD10*	*SVD01*
% in *signal MC*	52%	38%	5%	5%

Table 3.2: The fraction of each category of K_S^0 based on pions SVD hits in $B^0 \to K_S^0 K_S^0 K_S^0$ *signal MC*.

K_S^0 type	SVD11	SVD10	SVD01	SVD00
$M_{\pi^+\pi^-}$ window (GeV/c^2)	(0.45,0.55)	(0.38,0.7)	(0.38,0.7)	(0.3,0.7)

Table 3.3: The invariant mass windows after K_S^0 vertex fit based on the number of SVD hits in Figure 3-4. The K_S^0 outside these regions are rejected.

3.2 MVA-based K_S^0 selection

3.2.1 Belle II K_S^0 classification

The reconstruction of K_S^0 can be treated as a typical classification problem. The input is a set of variables that describes the characteristics of $K_S^0 \to \pi^+\pi^-$ decay. The training target is the true or fake flag from the MC truth-matching variable called *isSignal* where *isSignal* = 1 (0) stands for being a true (fake) K_S^0. The new Belle II K_S^0 classification tool aims to improve the limitations from the similar tool used in Belle.

In Belle, the K_S^0 reconstruction was first done by using cut-based method to select primary candidates, then a MVA-based classifier was implemented by assigning two likelihood indicators to each K_S^0 candidates. The package used by Belle is called *nisKsFinder* [10] which outputs the two likelihood variables based on NeuroBayes algorithm [34], called *nb_nolam* and *nb_vlike*. The Belle tool use them to define the goodness of a K_S^0 candidate. As their names suggest, *nb_nolam* is the likelihood of not being a Λ particle and *nb_vlike* is the likelihood of being a V0-like particle. A good K_S^0 candidate from *nisKsFinder* is the one with a low likelihood of being Λ particle and a high likelihood of being a V0-like particle, assuming the major backgrounds for K_S^0 are the mis-identified Λ among V0-like particles. By putting cuts on these two variables, a purification of K_S^0 can be made, shown in Figure 3-5. It can effectively reduce fake K_S^0 from cut-based selected candidates, however, there are a couple of disadvantages about this method. First, NeuroBayes is a commercial product that was developed over 10 years ago. The official support and update is stopped nowadays, so it is not an ideal method for an experiment like the Belle II that has a quite long prospective in operation. Second, the classification is based on a joint cut on two variables, which might make the cut values hard to choose. For example, two different cuts might have very close purity. Besides, the computation speed of NeuroBayes algorithm is not optimized in training large data set.

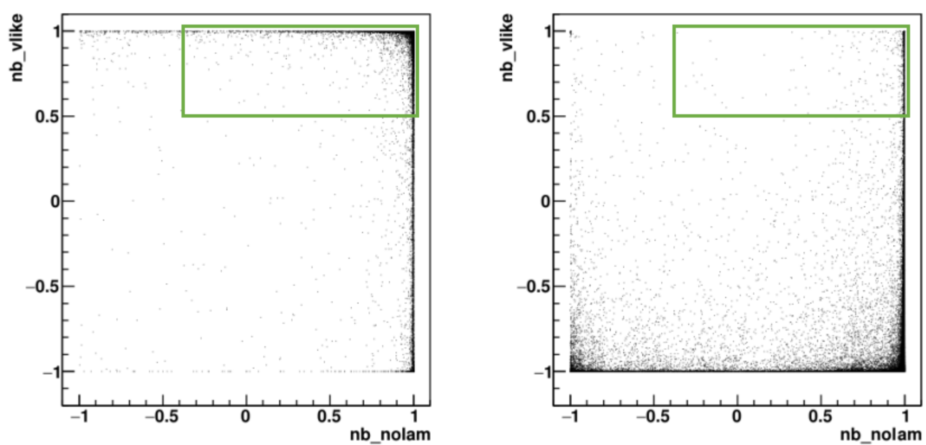


Figure 3-5: The distribution of two variables outputs: *nb_nolam* and *nb_vlike* for K_S^0 candidates from Belle *signal MC*. The left is from true K_S^0 and the right is from the fake K_S^0. In Belle, the standard cuts for K_S^0 is *nb_vlike* > 0.5 and *nb_nolam* > -0.4, which is shown as the green boxes [19].

Such a dedicated K_S^0 classification tool is not implemented yet in BASF2 framework until 2019. Considered the limitation of NeuroBayes, the development of K_S^0 classifier demands another algorithm and structure. The *Boosted Decision Trees* (BDT) is widely employed for multivariate classification and regression tasks in high energy physics field. Particularly, a speed-optimized and cache-friendly implementation of such a method called FastBDT (FBDT) is popularly used [35]. Compared to other popular classification algorithms such as TMVA [36], scikit-learn [37] and XGBoost [38], FastBDT method is proven to be one order of magnitude faster during the training and applying phases [35]. By using FastBDT algorithm, *KsFinder* in

Belle II is expected to give a single output which directly presents the goodness of a candidate of being a true K_S^0. Since the FastBDT algorithm depends on the variables that are differently distributed in signal and backgrounds, a set of training variables are selected based on K_S^0 decay topology. The K_S^0 variables used in the training of *KsFinder* might be differently distributed in different decay channels, therefore a *KsFinder* trained using MC sample from one channel may not be able to perform a good classification on the other. Thus, *KsFinder* is designed as a general package that provides a mode-dependent K_S^0 classification which mainly consists of four components: *KsFinderSampler*, *KsFinderTeacher*, *KsFinderApplier* and *KsFinderTest*. *KsFinderSampler* is a function that automatically generates training and/or testing sample from mDST files where the cut-based reconstruction is already implemented as explained in Section 3.1. *KsFinderTeacher* is responsible for extracting variables to perform training of the FastBDT model and generate a weight file containing all the node information in ROOT format, which also provides a function to communicate with BASF2 CDB so that users can share or download other weight file in their own analysis. *KsFinderApplier* can apply the weight file generated by *KsFinderTeacher* (or downloaded from BASF2 CDB) to the independent data sample and assign each K_S^0 candidate a goodness index used as a single cut value in the further analysis. *KsFinderTest* is the evaluation function that can use a test sample to check for over-training, efficiency, and purity. By providing MC samples from certain decay modes, users can easily generate their own weight files of K_S^0 classification that suit different decay modes. Such a design largely improves the flexibility of *KsFinder* compared to Belle MVA tool which indirectly classify K_S^0 with two outputs.

3.2.2 Decay Topology of $K_S^0 \to \pi^+\pi^-$

As introduced in Section 3.2.1, the first step for developing K_S^0 MVA classification is to determine the input variables for FastBDT algorithm that can represent the decay features of K_S^0 against possible backgrounds. The remaining background of $K_S^0 \to \pi^+\pi^-$ after the cut-based reconstruction comes from different sources, mainly including the false combination of tracks (including $\pi^\pm$ misidentification), V0-like particle misidentification and self-looped tracks. For instance, a D^0/D^* from a B decaying to $K\pi$ with K misidentified as π, could give a false combination of tracks. On the other hand, it is also possible that both of two tracks are correctly identified as $\pi^{\pm}$ but they are not from the same mother particle, or the mother is not a K_S^0 particle due to the missing of other daughters, such as $D^+ \to K_S^0(\to \pi^+\pi^-)\pi^+$. These two cases as the fake K_S^0 are demonstrated in Figure 3-6.

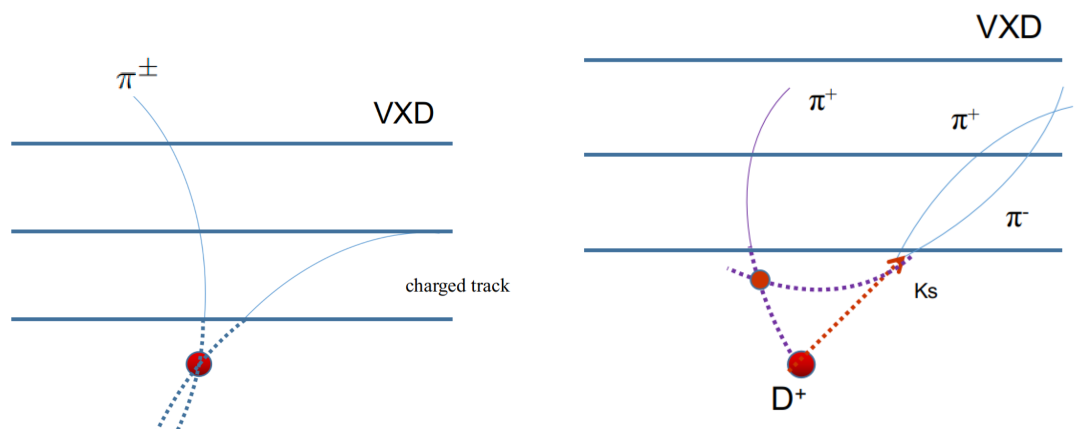


Figure 3-6: The left shows the case when a charged track (not $\pi^{\pm}$) combined with a charged pion to form a fake K_S^0, the right shows the case when two daughters are correctly reconstructed as pions but not from the correct mother particle, which is falsely taken as a K_S^0.

The V0-like particles mainly refer to K_S^0, Λ and γ. $\gamma \to e^+e^-$ yield is significantly lower than the other two types and the mass difference between pion and electron is very large, so the PID values can be used to well-distinguish them. As for the contribution of $\Lambda \to p^+\pi^-$, it happens when the positive charged tracks (proton track) is wrongly identified as π^+, see Figure 3-7 left. The key observable to distinguish this background is the invariant mass of mother particle, which is 1.115 GeV for Λ, much larger than the K_S^0. The number of left-over Λ after the cut-based reconstruction in section 3.1 is small, and can be further reduced by rejecting the candidates whose positive charged daughter has PID($\pi^{\pm}$) smaller than PID(p).

When a charged pion only carries a minimal of its mother's transverse momentum p_T, the curvature of its track may form a self-loop of which radius is comparable with the size of Belle II detector (mainly VXD and CDC). In this case, one charge pion could leave two charged tracks candidates with the opposite charge and similar p_T, with a possibility to form a converged vertex to form a fake K^0_S, see the right of Figure 3-7.

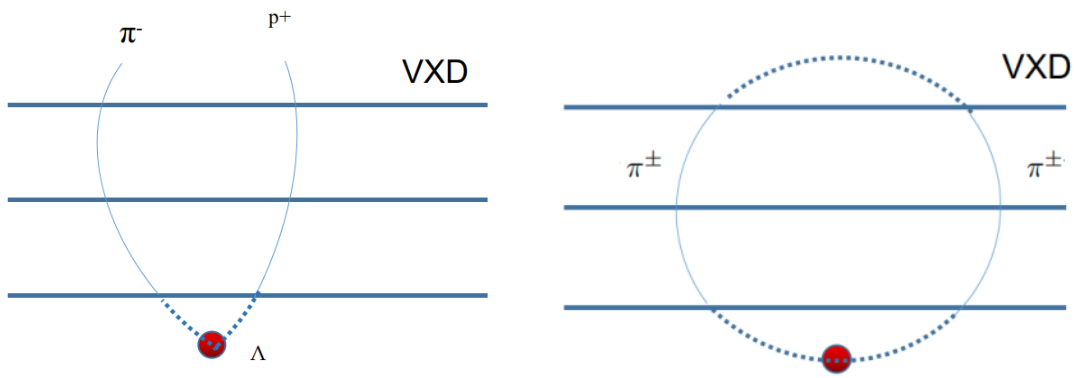


Figure 3-7: The left shows the $\Lambda \to p^+\pi^-$ decay shape that can be treated as K^0_S, the right shows a self-loop formed by a low p_T charged pion reconstructed as two separated tracks with a vertex.

3.2.3 Determination of training variables from K^0_S decay

Given the characteristics of $K^0_S \to \pi^+\pi^-$ discussed in the previous section, a set of variables as training features of *KsFinder* can be selected. The set includes variables related to K^0_S kinematics, decay shape parameters, particle identifications and detector hits information. The summarized information of training variables is listed in Table 3.4.

The cosine between K^0_S vertex and momentum direction (named *cosVertexMomentum*) is regarded as the most useful variable to separate true and fake K^0_S, which is originally used as an extra cut in the first measurement of $B^0 \to K^0_S K^0_S K^0_S$ before MVA based K^0_S classification tool was developed [39]. For instance, if a falsely reconstructed K^0_S is made of two tracks, it is likely that the momentum direction of the fake K^0_S is not aligned with the its vertex direction from IP. So the projection of vertex position of K^0_S on the reconstructed momentum direction could be negative

[1]The decay angle of two daughters are essentially the equivalent variables because they are defined in the mother's rest frame, which in future will be replaced by only one variable of the positive charged pion decay angle.

Table 3.4: Summary of *KsFinder* input variables, where "lab" means angles in lab frame and "K_S^0 CMS" means in K_S^0 rest frame. Other variables are calculated in lab frame by default. The last column shows the number of the variables correspondingly.

K_S^0 variables(#)	Meaning	#
cosVertexMomentum	cosine of vertex and momentum direction (lab)	1
flight distance	K_S^0 flight distance along its momentum direction	1
significanceOfDistance	flight length from IP divided by relative error	1
cosHelicityAngleMomentum	cosine between $\pi^{\pm}$ and K_S^0 (lab)	1
ImpactXY	Impact parameters in transverse plane for K_S^0	1
x, y, z, px, py, pz	K_S^0 vertex position and momentum	6
p_D1(p_D2)	momentum magnitude for $\pi^+(\pi^-)$	2
pionID, muonID	PID values of π^+	2
decayAngle_D1(D2)[1]	angle between $\pi^+(\pi^-)$ and K_S^0 (K_S^0 CMS)	2
daughterAngle2body	angle between $\pi^{\pm}$ (lab)	1
daughtersDeltaZ	Z-direction distance of two tracks helix	1
nSVDHits_D1(D2)	SVD detector hits of $\pi^+(\pi^-)$	2
nPXDHits_D1(D2)	PXD detector hits of $\pi^+(\pi^-)$	2
M(InvM)	K_S^0 invariant mass before (after) vertex fit	2

value for fake K_S^0. While in case of a true K_S^0, such projection is almost always a positive value, shown in Figure 3-8. This often happens when the two tracks taken as $\pi^{\pm}$ are accidentally crossed, or due to the misidentified tracks. The distribution of *cosVertexMomentum* using *signal MC* is shown in the Figure 3-9. By requiring the cut *cosVertexMomentum* > 0.9, fake K_S^0 fraction can be reduced to about 20%, which is still not good enough.

The other variables in Table 3.4 are not as contributive as *cosVertexMomentum* in selecting true K_S^0 and reject fake ones at the same time, but still important in increasing the discriminating power of FastBDT model. For instance, the significance of flight distance distribution is shown in Figure 3-10. The fake K_S^0 can have relatively smaller significance because of the larger error of the vertex fit. FastBDT algorithm can give the importance of each variable after the training, and the total classification ability of the model depends on the combined power of all input variables. By combining these variables, the rejection of fake K_S^0 is targeted to be as good as the Belle *nisKsFinder*, which should exceed 95%.

Because FastBDT method relies on the distribution of variables to calculate signal and background separation, there are a few points to be checked before feeding the

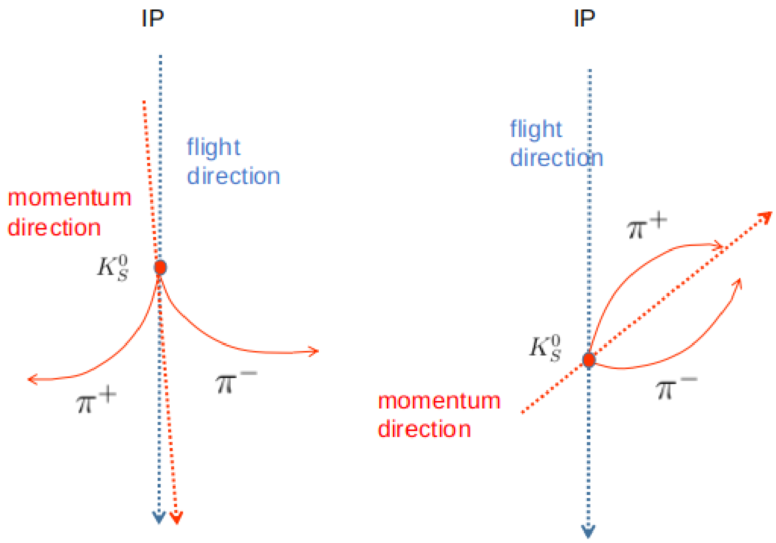


Figure 3-8: The left shows a true K_S^0 decay shape where the cosine angle of K_S^0 vertex position (blue dashed arrow) against reconstructed momentum direction (red dashed arrow) is positive. While the right shows a fake K_S^0 decay shape where the cosine angle of K_S^0 vertex position against the reconstructed momentum direction can be negative.

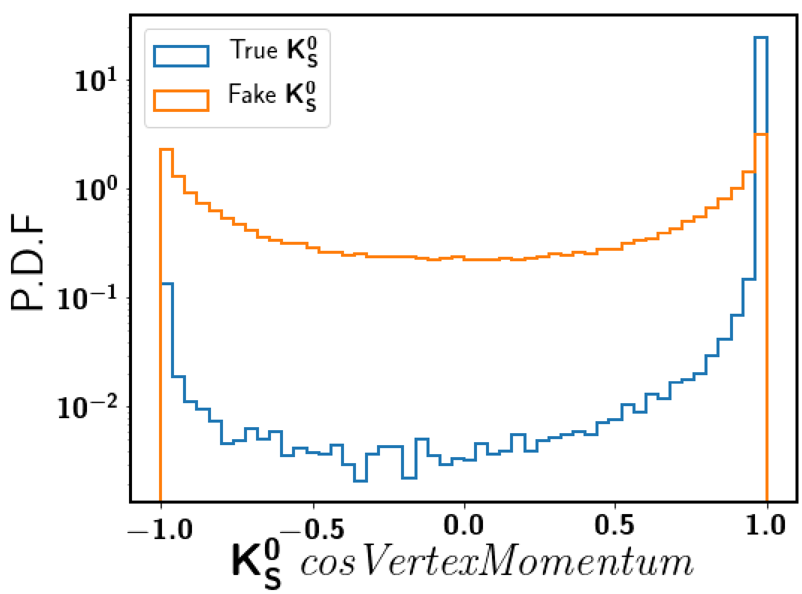


Figure 3-9: The distribution of *cosVertexMomentum* using *signal MC*. The true and fake K_S^0 ratio is set to be 1:1, where the most true K_S^0 gives *cosVertexMomentum* > 0.9.

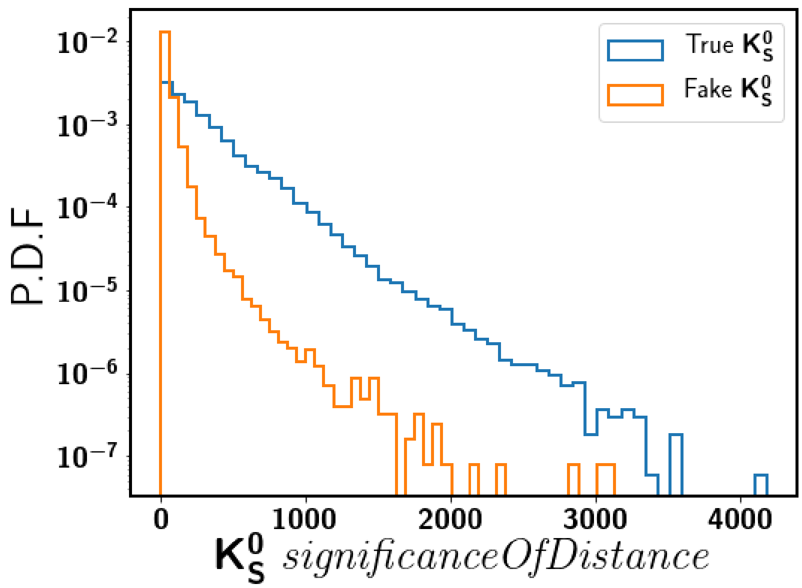


Figure 3-10: The distribution of *significanceOfDistance* using *signal MC*. The true and fake K_S^0 ratio is set to be 1:1, where the most fake K_S^0 are distributed in *significanceOfDistance* < 1000 region.

training sample to the model or applying the classification on real data. First, the distribution of the observables should be different in true K_S^0 and the fake ones, so the FastBDT classifier can effectively separate the true and the fake K_S^0 at each node to maximize the separation gain. Second, there will a correlation among the training observables and they should also be different in signal and background. The boosting step will create a sequence of shallow decision trees (DT) whose structures are not the same. Different correlations helps improve the performance of decision trees in tuning of structure. For instance, a true K_S^0 flights longer (flight distance) due to larger momentum in general, so the number of daughter detector hits (nSVDHits_D1 or nSVDHits_D2) becomes fewer. Then these two observables have negative correlations in true K_S^0. In case of a fake K_S^0, the flight length could be a deep outside of VXD but daughters may have full hits on SVD, without strong correlation, see Figure 3-11 . At last, one should also avoid using many observables with too strong correlations, since in this case, many DTs might have a potentially equivalent structure in the boosting step. Therefore, the separation power of many DTs doesn't gain any improvement and the collection of observables might be redundant. The correlation between variables are shown in Figure 3-11.

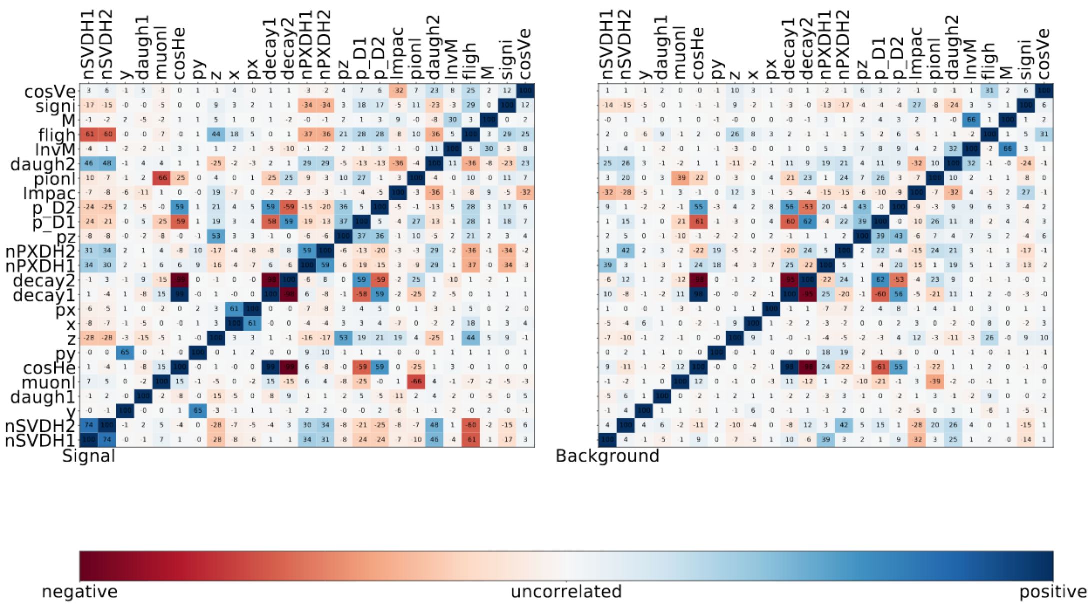


Figure 3-11: The correlation between input variables for *KsFinder*. As the given example, flight length has negative correlation with SVD hits in signal while uncorrelated in background.

3.2.4 Training, Applying and Testing of *KsFinder*

The variables are internally registered inside the *KsFinder* so it can automatically retrieve their values from a mDST file in BASF2. The Table 3.5 shows the abbreviations of the input variables used in the *KsFinder*. The first step of using *KsFinder* is to call *KsFinderSampler* on a MC sample to generate training and testing data sample. To show the flexibility and stability of *KsFinder* on different modes, *KsFinderSampler* extracts MC data points from both *signal MC* and *generic MC* (see MC definition in section 2.9), respectively. Also, it can separately sample true and fake K_S^0. In this analysis, we are looking for the K_S^0 specifically from $B^0 \to K_S^0 K_S^0 K_S^0$ decay, of which the branching fraction is 6×10^{-6} according to the PDG value. In the *generic MC*, the fraction of K_S^0 from $B^0 \to K_S^0 K_S^0 K_S^0$ is quite low even among the true K_S^0. Most of the K_S^0 particles in the *generic MC* are from the decays related to $c \to s$ transition. If the true and fake K_S^0 are unbalanced, it may not be optimized for the training of *KsFinder*. Therefore, *KsFinderSampler* is configured to extract true K_S^0 from *signal*

Table 3.5: The abbreviations of the input variables used in the training of *KsFinder*.

Observables	Abbreviations
cosVertexMomentum	cosVe
flight distance	fligh
significanceOfDistance	signi
cosHelicityAngleMomentum	cosHe
ImpactXY	Impac
x	x
y	y
z	z
px	px
py	py
pz	pz
p_D1	p_D1
p_D2	p_D2
muonID_pi	muonI
pionID_pi	pionI
decayAngle_D1	decay1
decayAngle_D2	decay2
daughterAngle2body	daugh2
daughtersDeltaZ	daugh1
nSVDHits_D1	nSVDH1
nSVDHits_D2	nSVDH2
nPXDHits_D1	nPXDH1
nPXDHits_D2	nPXDH2
M	M
InvM	InvM

MC, where the majority of K_S^0 from $B^0 \to K_S^0 K_S^0 K_S^0$. The true K_S^0 in *signal MC* may contain some candidates from tag-side generic decay, where the fraction is estimated below 5% on average. Hence, there are two types of training samples prepared. First, (a) a true K_S^0 sample is composed of 95000 true K_S^0 from $B^0 \to K_S^0 K_S^0 K_S^0$ and 5000 true K_S^0 from *generic MC*, to allow the *KsFinder* learn from both cases. Second, (b) we directly sample 100000 true K_S^0 only from *generic MC*. For both cases, the fake K_S^0 are sampled from *generic MC* with the same number as the true K_S^0. The testing samples corresponding to these two training samples are prepared in the same way. As for the FastBDT training options, the *KsFinder* configures that the depth of each DT is 3, learning rate is 0.3 and the boosting steps is 200.

To train the *KsFinder*, *KsFinderTeacher* function is called and weight files are saved. To apply the classification of K_S^0, *KsFinderApplier* reads in the testing samples and calculate the output using saved weight files, so that each K_S^0 candidate is assigned with a goodness index named *FBDT_Ks*. It ranges from 0 to 1 where 1 stands for the best goodness. To evaluate the performance of *KsFinder*, *KsFinderTest* is called which compares the results between the training sample and the testing sample, including the over-training check.

3.2.5 The Performance and Over-training check

To evaluate the performance of *KsFinder* on both (a) and (b) samples, signal efficiency, background rejection and purity are calculated by cutting on the different values on *FBDT_Ks*, as defined:

$$\text{signal efficency} = \frac{\text{Number of true } K_S^0 \text{ with } \mathit{FBDT\_Ks} > \text{cut value}}{\text{Number of all true } K_S^0}, \quad (3.1)$$

$$\text{background rejection} = \frac{\text{Number of fake } K_S^0 \text{ with } \mathit{FBDT\_Ks} < \text{cut value}}{\text{Number of fake true } K_S^0}, \quad (3.2)$$

$$\text{purity} = \frac{\text{Number of true } K_S^0 \text{ with } \mathit{FBDT\_Ks} > \text{cut value}}{\text{Number of all } K_S^0 \text{ with } \mathit{FBDT\_Ks} > \text{cut value}}. \quad (3.3)$$

The ROC (receiver operating characteristics) curve is usually taken as an indicator

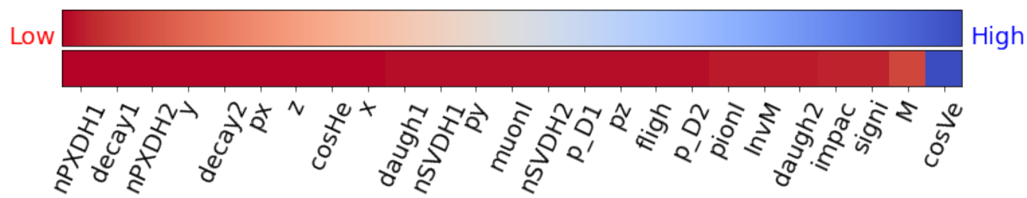


Figure 3-12: The importance rank of the input variables for *KsFinder*. The *cosVertexMomentum* is the most important variable.

of the performance where the curve shows the dependence of background rejection power with respect to the signal efficiency. The larger area under a ROC curve means that the better performance is achieved. The ROC curves as well as the efficiency & purity with respect to the *KsFinder* cut are shown in Figure 3-13 and Figure 3-14, where the former is for sample (a) and the latter is for sample (b). With increasing the efficiency, the cut on the output of *KsFinder* is getting loose. The background rejection only starts to drop when the efficiency exceeds about 90% in both training and testing sample.

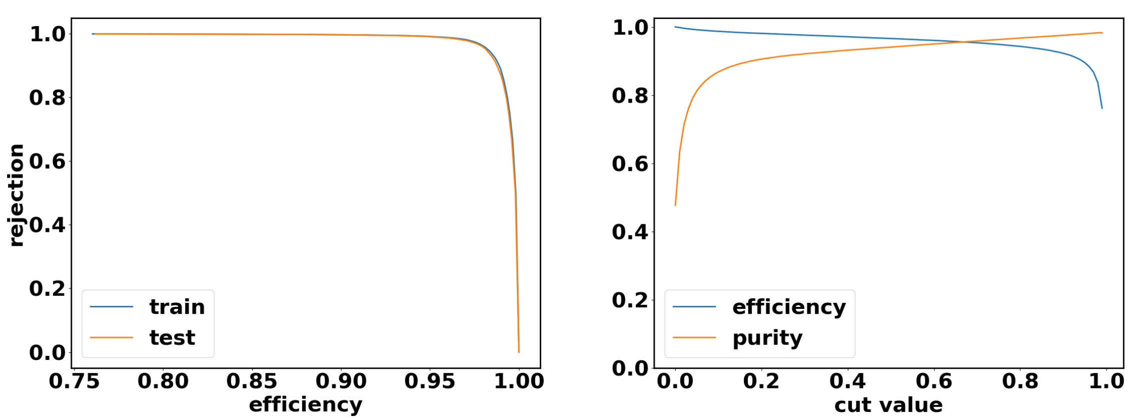


Figure 3-13: The left is ROC curve(blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of *KsFinder* output. The results are obtained by applying the weight file from training sample (a) to testing sample (a).

Because the ROC curves are consistent in the training and testing samples, it proves the absence of noticeable over-training in classification, however, the detailed check can be made by comparing the distributions of *KsFinder* output on true and

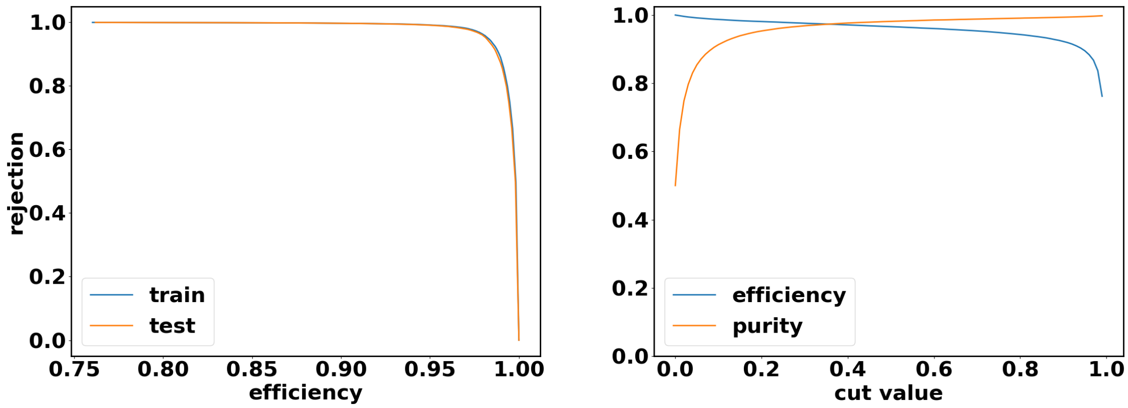


Figure 3-14: The left is ROC curve (blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of *KsFinder* output. The results are obtained by applying weight file from training sample (b) on testing sample (b).

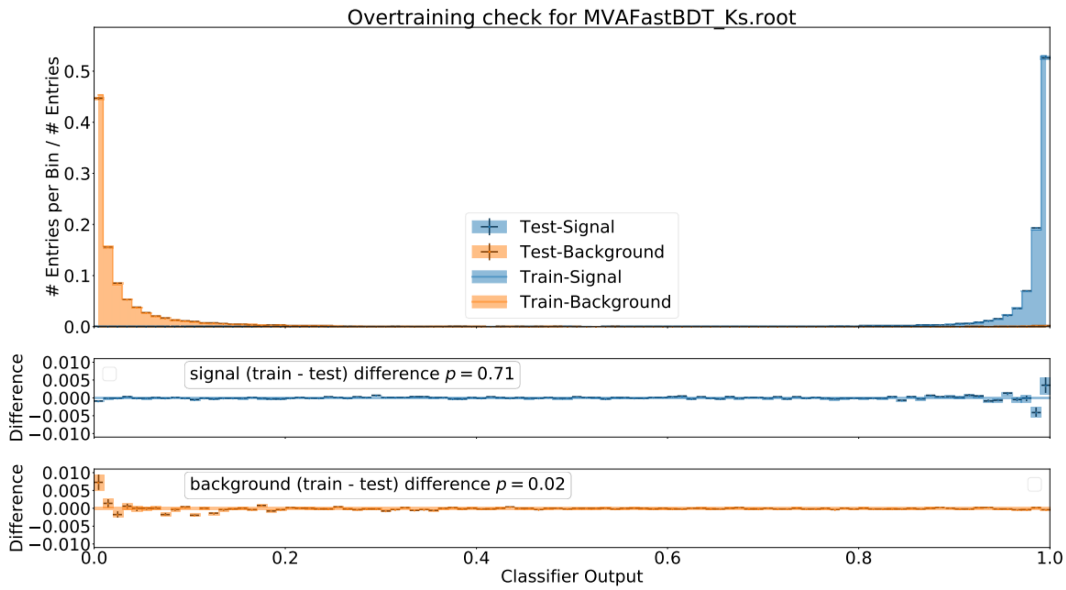
fake K_S^0 in training and testing samples. Therefore, the distribution of input variables in the true and fake K_S^0 in training and testing sample with respect to the *KsFinder* output are plotted, where a distinctive separation for both sample (a) and sample (b) are shown and no over-training is found, as shown in Figure 3-15.

The best cut value for *FBDT_Ks* is determined by maximizing the "Figure of Merit" (FOM), as shown Equation 3.4, where S and B is the number of true and fake K_S^0 after the cut, respectively.

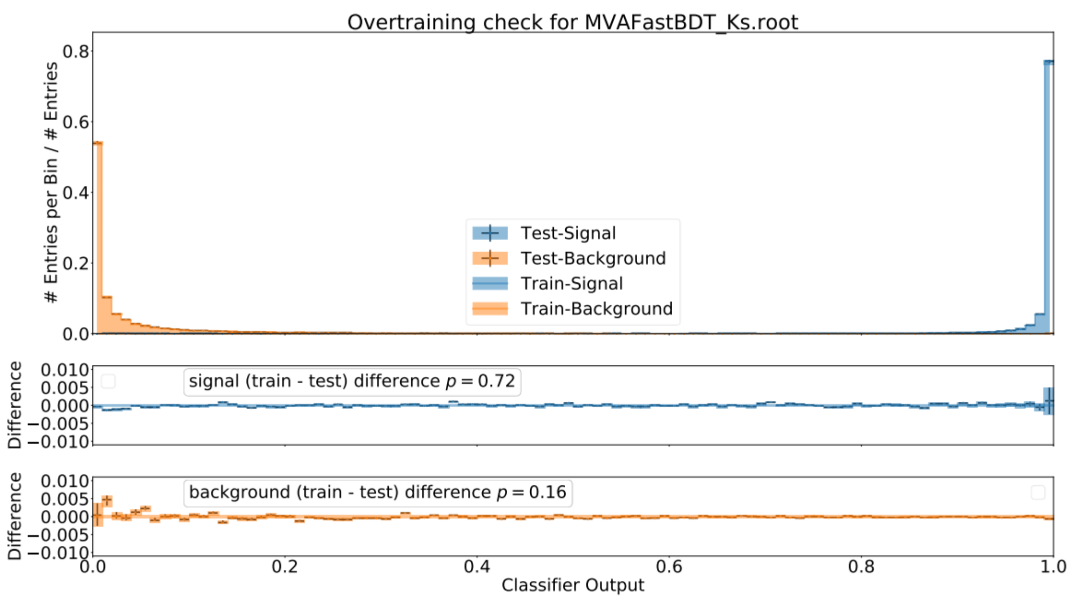
$$\text{FOM} = \frac{S}{\sqrt{S+B}} \tag{3.4}$$

The FOM distribution depending on the cut value of *FBDT_Ks* is shown in Figure 3-16. In this analysis, we primarily focus on the K_S^0 from $B^0 \to K_S^0 K_S^0 K_S^0$, thus the weight file from (a) is chosen to perform K_S^0 classification since it is supposed to learn more on K_S^0 from our analysis channel. The maximum FOM is achieved at *FBDT_Ks* $= 0.74$, which is going to be used as the cut value[2]to further reject fake K_S^0. The FOM curve is not sensitive to the cut value in between $0.5 \sim 0.9$, which achieves similar performance on average.

[2]The cut value of 0.74 is obtained by using 1:1 ratio of true and fake candidates in the FOM calculation, while the maximum FOM is achieved at 0.76 if using the ratio from *generic MC* which



a) Over-fitting check for sample (a).



b) Over-fitting check for sample (b).

Figure 3-15: The over-training check based on the comparison between training/testing data points in both *signal* and *generic MC*.

To demonstrate the improvement from using *KsFinder*, the *KsFinder* cut is applied to the cut-based selected K_S^0 sample additionally based on the sample used in Figure 3-2. The true K_S^0 fraction before applying *KsFinder* cut is 39%, and 95% of them are kept after the cut is applied. In the meantime, the fake K_S^0 fraction before applying the cut is 61%, and 98% of them are rejected after the cut is applied. The purity of the K_S^0 candidates is largely improved as shown in Figure 3-17.

The comparison of K_S^0 reconstruction performance using different cuts and approaches is summarized in Table 3.6. It is clear that *KsFinder* can provide a better K_S^0 reconstruction performance using 25 input variables compared to only use cuts on one or two important variables. The importance to the output of *KsFinder* of each variable is obtained from the *KsFinderTest*, where *cosVertexMomentum* has the highest rank, shown in Figure 3-12. From Figure 3-13 and 3-14, the purity can exceed 95% by choosing proper cut value. Instead, by only applying *cosVertexMomentum*>0.9 on the cut-based selections, purity can only reach about 80%, demonstrating the necessity of including more variables to improve the power of classification despite that each variable may only weakly discriminate the true and fake K_S^0.

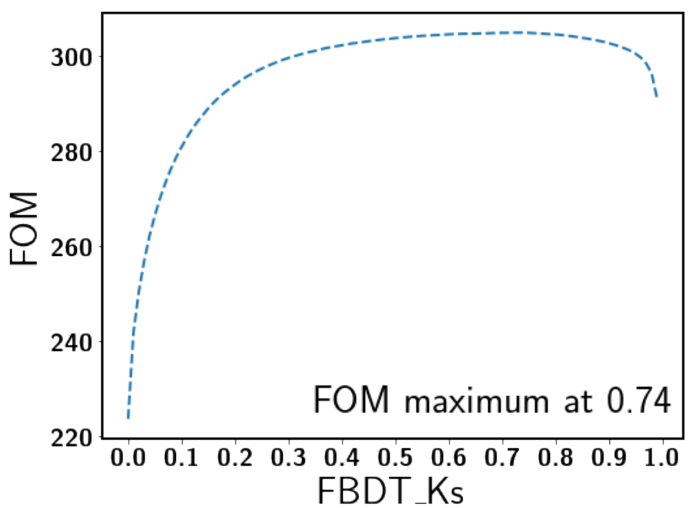


Figure 3-16: FOM of classifier output (*FBDT_Ks*) in *signal MC*, the maximum value is achieved at 0.74. The FOM curve is almost flat between $0.5 \sim 0.9$, which is insensitive to the cut value in this region.

is about 1:10. Since the difference is fairly small and FOM is almost flat in this range, we use 0.74 as the cut value in this analysis.

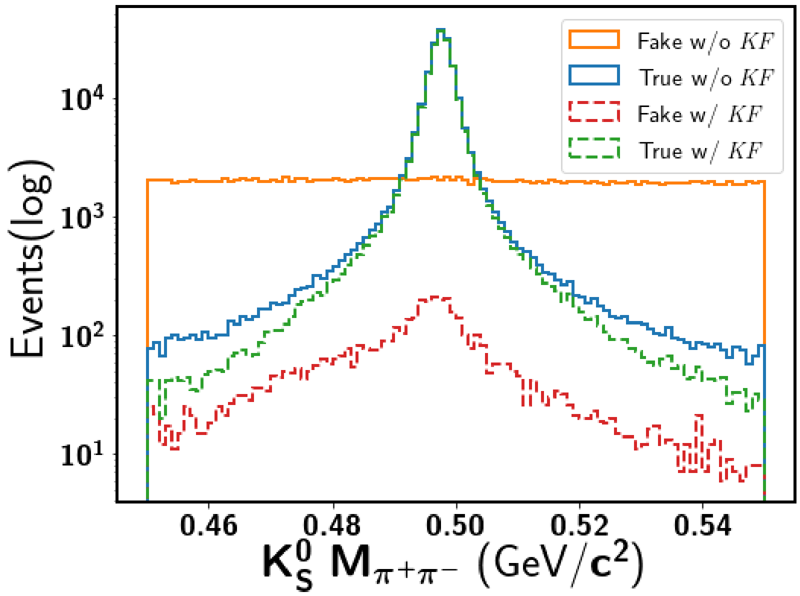


Figure 3-17: K_S^0 purity improvement with *KsFinder*(KF) cut at 0.74. The blue solid line is true K_S^0 without *KsFinder* and green dashed line is the true K_S^0 with the cut applied. The orange solid line is fake K_S^0 without the cut and the red dashed line is fake K_S^0 with the cut. About 95% of true K_S^0 are kept while 98% of the fake ones are rejected by applying the cut.

Table 3.6: The summarized performance of K_S^0 reconstruction in different approaches. The Belle II results are based on the cut-based reconstruction, with extra cuts from *cosVertexMomentum*(cosVe), *significanceOfDistance*(signi) and *KsFinder*.

Cuts	Efficiency	purity	BKG rejection
Belle(default): nb_vlike > 0.5 & nb_nolam > −0.4	90%	95%	95%
Belle II cut: cosVe > 0.9	96%	82%	80%
Belle II cut: cosVe> 0.9 & signi > 50	92%	89%	89%
Belle II *KsFinder* cut: FBDT_Ks > 0.74	95%	97%	97%

3.2.6 Data Validation for ***KsFinder***

The results from MC studies show an excellent performance of *KsFinder*. However, the validation of such a tool on the real experiment data is necessary. Since there is no MC truth on target variable in real data, the FastBDT method is based on variables in MC samples. If these variables shows close distributions among MC and data, the classifier obtained using MC can be applied to the real data with the expectation of the close performance. In addtion, due to the fact that K_S^0 candidates are used for the further reconstruction of B^0, the mass and energy distributions may change after applying the cut, thus it is also required to show that no clear bias on B^0 for signal extraction.

For comparison between MC and data, a small data sample from Belle II experiment 7 and 8 is used. The integral luminosity at $\Upsilon(4S)$ resonance for this data sample is 5.17 fb^{-1}. The MC sample is extracted from *generic MC* with equivalent luminosity. Data and MC events are filled in the binned histogram of each variables to check the consistency. The number of events in each bin is assumed to follow the *Poisson* Distribution, which is approximately equivalent to the *Gaussian* distribution when the number of events is large enough according to the *Central Limit Theorem*. Therefore the standard deviation of each bin is calculated as $\sqrt{N_i}$ using *Poisson* distribution property, where N_i is the event number in the i-th bin. We use three times the standard deviation in each bin as a conventional reference in the drawing error bars. The Figure 3-18 shows the invariant mass and momentum distributions from data and MC samples. The *generic MC* is shown in blue solid lines with no *KsFinder* cut used. Similarly, data without using *KsFinder* is shown in yellow dots, which are closely distributed as the *generic MC*, indicating a good data-MC consistency. The purple solid lines are presenting the K_S^0 distribution in *generic MC* with *KsFinder* cut at 0.74, while the red dots are the K_S^0 in data after using the same *KsFinder* cut. To compare the number of events ratio between data and MC, a ratio distribution is also produced for each variable below the main comparison plot, where the blue circle dots are the data/MC ratio without *KsFinder* cut and red reverse-triangle dots are the ones with *KsFinder*. The distribution plots of invariant mass and momentum are shown in Figure 3-18. After applying *KsFinder* cut, it shows about 20% difference on the data/MC ratio, indicating the data is reduced more than the *generic MC* on the *KsFinder* cut. The distribution plots of *cosVertexMomentum* is shown in Figure 3-19. The ratio plot of *cosVertexMomentum* shows a discrepancy in the background dominated region (*cosVertexMomentum*< 0.0), while remains close to 1 for the signal dominated region (*cosVertexMomentum*> 0.5) after the *KsFinder* cut. These discrepancies will be monitored and improved in future by the better data and MC match-up. The full distributions comparison between data and *generic MC* by using *KsFinder* cut are included in the Appendix A.

Not only the similar distributions of the input variables are important for applying *KsFinder* trained from MC to the real data, the correlations between these variables should also be similar between data and MC. This requires the comparison of the correlation matrices using true and fake K_S^0. However, because it is hard to precisely know whether a K_S^0 from data is definitely a true candidate or not, it is difficult to directly compare the correlation among data and MC in true and fake K_S^0 separately. One possible solution to obtain the true K_S^0 sample in real data is to use other control channels which contain at least one K_S^0 in the decay chain and have very high purity even without *KsFinder*. In the early stage of the Belle II, the preparation of such clean control samples is not ready, especially for the multi-body hadronic B decays that are similar to $B^0 \to K_S^0 K_S^0 K_S^0$ with more than one K_S^0. This will be an advanced topic for widely validating *KsFinder* in the future and it is not implemented in this thesis. Hence, instead of using clean control samples containing very high purity K_S^0 collections, we just compare the correlation matrices for both true and fake K_S^0 together in data and *generic MC*, where the correlation in data is divided by the correlation in MC, as shown in Figure 3-20. The most of the variables present the close correlation factor between data and MC, while the z momentum presents a largely different correlation with the daughter tracks distance in z direction. The *cosVertexMomentum* only shows a different correlation with the y-direction momentum, which is acceptable due to the low importance of the y-direction momentum. The differently

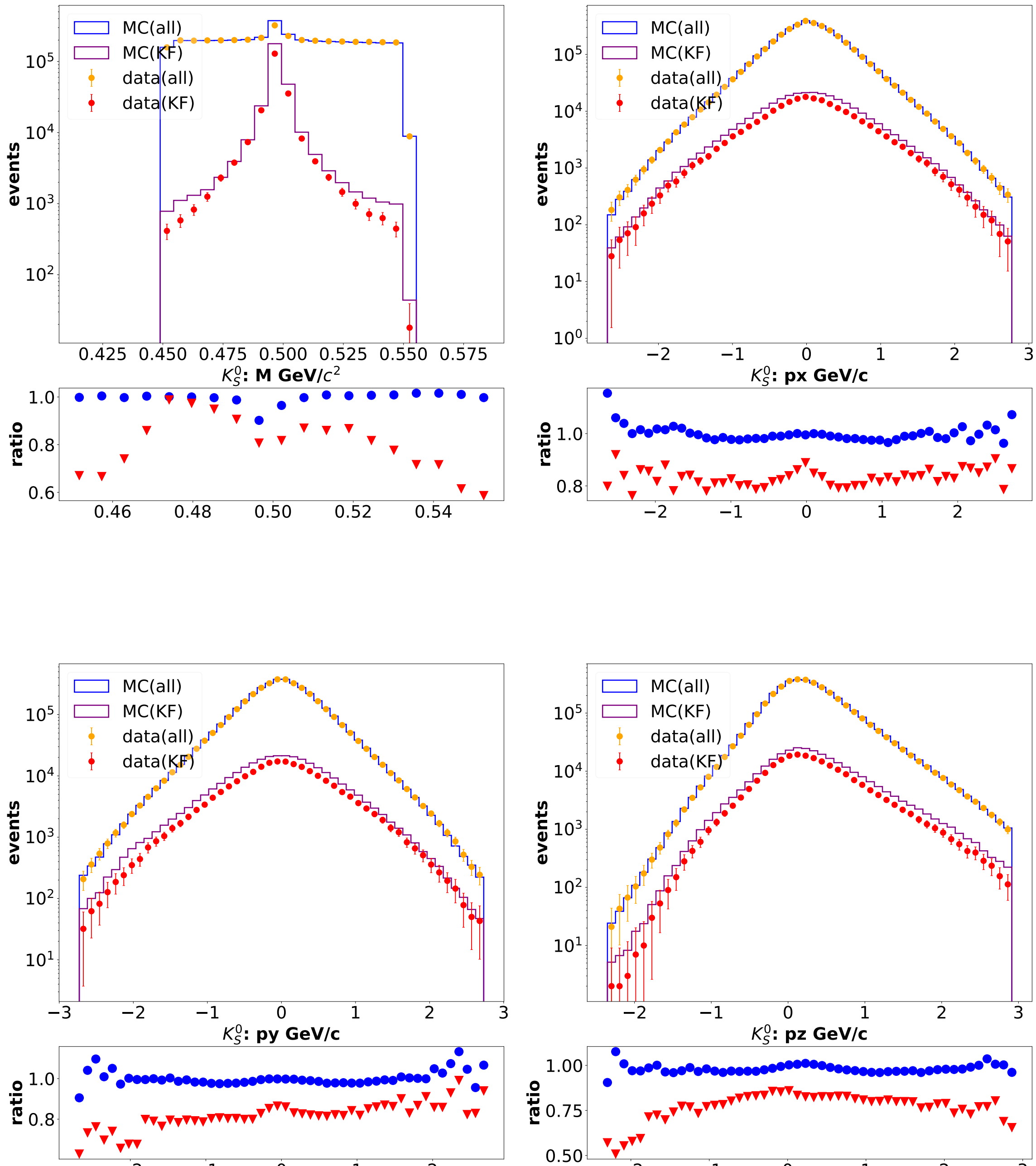


Figure 3-18: The distribution of invariant mass from charged pions and the momentum of K_S^0 in x, y, z directions. The blue line is from all *generic MC* and the purple line is the K_S^0 after *KsFinder* (KF). The yellow dots are data with no *KsFinder* (KF) cut applied and the solid red dots are data after applying the same cut. The bottom sub-plots are the data/MC ratio before and after applying *KsFinder* in data (blue) and *generic MC* (red).

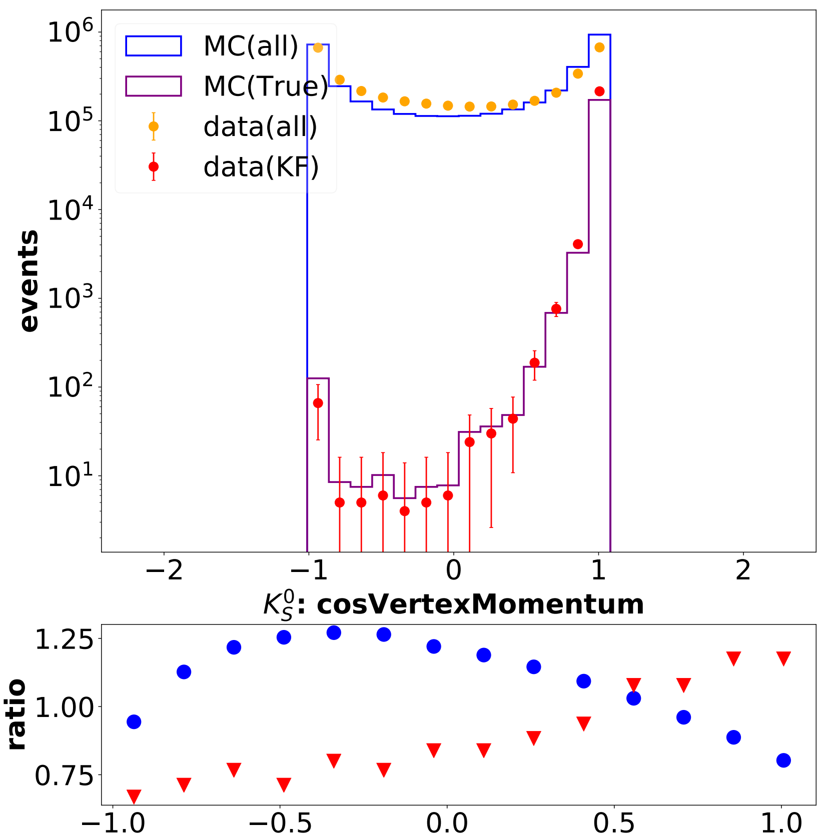


Figure 3-19: The distribution of $cos VertexMomentum$ in data and MC with or without $KsFinder$ cut applied.

correlated variables are mostly low ranked, and no appearance of the large difference among the high ranked variables is observed. In the future, such discrepancies will be monitored with improved simulation along with the data recording.

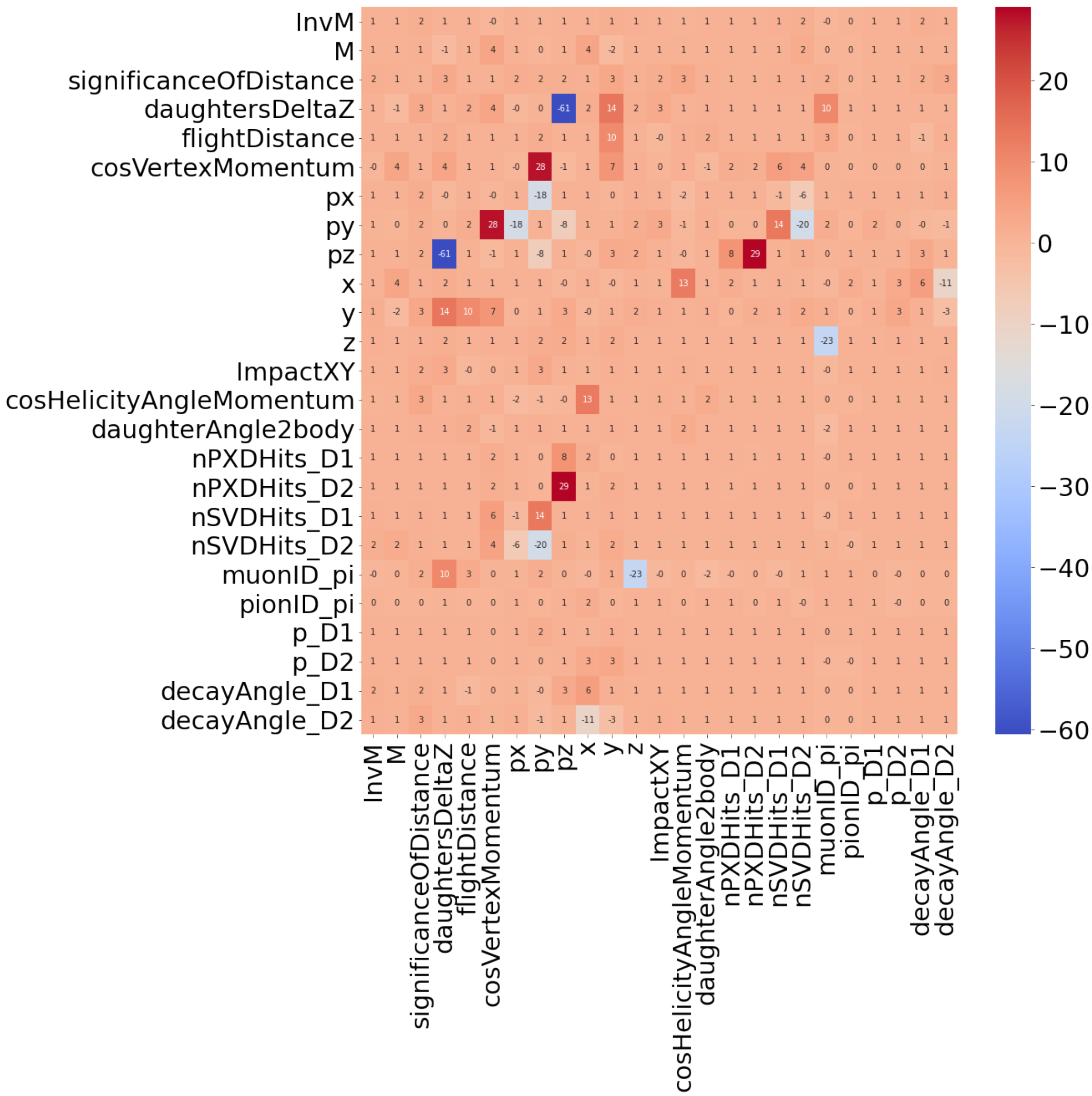


Figure 3-20: The correlation ratio between *KsFinder* input variables, where the value is calculated by using the one from data divided by that from *generic MC*.

3.2.7 Data and MC correction by *KsFinder*

As the previous section shows, implementing *KsFinder* cut on data may induce bias on the event numbers for K_S^0 because the training set of *KsFinder* is extracted from

MC. To compensate such potential effect, a ratio as the data and MC correction is calculated based on the expected signal yield after using *KsFinder*. A maximum likelihood fit on invariant mass for K_S^0 with *FBDT_Ks* > 0.74 is performed, where signal shape is modeled as a triple-Gaussian and background shape is modeled as a Chebyshev polynomial. The signal yield fraction is defined as:

$$f_{K_S} = \frac{N_{sig}}{N_{tot}} \,, \tag{3.5}$$

where N_{sig} is the signal number from the fit result and N_{tot} is the total events number. The fit is performed on both *generic MC* and data to obtained f_{K_S}, respectively, as shown in Figure 3-21, where the left is for data and the right is for *generic MC*. The fit results are 0.933 ± 0.008 for *generic MC* and 0.924 ± 0.006 for data. The fraction of the true K_S^0 in *generic MC* sample is 0.939, consistent with the fit result. The $\mathcal{R}_{K_S}$ is defined as the ratio of signal yield fraction f_{K_S} from *generic MC* and data as shown in Equation 3.6.

$$\mathcal{R}_{K_S} = \frac{f_{K_S}^{MC}}{f_{K_S}^{data}} \tag{3.6}$$

By applying the *KsFinder* cut at 0.74, the $\mathcal{R}_{K_S}$ is calculated to be 1.009 ± 0.011 from the f_{K_S}, Similarly, the ratio for data and *generic MC* in terms of the number of B^0 can be defined as:

$$\mathcal{R}_{B^0} = \frac{f_{B^0}^{MC}}{f_{B^0}^{data}} \simeq \mathcal{R}_{K_S}^3 \,, \tag{3.7}$$

Since the final state consists of three K_S^0, the $\mathcal{R}_{B^0}$ is expected to be the cube of $\mathcal{R}_{K_S}$, with the uncertainty propagated from the uncertainty of $\mathcal{R}_{K_S}$. By using $\mathcal{R}_{K_S} = 1.009 \pm 0.011$, the result of $\mathcal{R}_{B^0}$ is 1.027 ± 0.033, which is close to one within its uncertainty. Hence, the correction $\mathcal{R}_{B^0}$ is not applied in signal extraction of B^0, but the impact is taken into account as a possible source of systematic uncertainty. We take the larger one among the center value shift of $\mathcal{R}_{B^0}$ (0.027) and the uncertainty of $\mathcal{R}_{B^0}$ (0.033) as the systematic uncertainty source in the *CP* measurement. Thus, the $\mathcal{R}_{B^0} = 1.033$ and $\mathcal{R}_{B^0} = 0.967$ are applied to the calculation of the signal fraction when performing the systematic uncertainty evaluation in *CP* violation measurement.

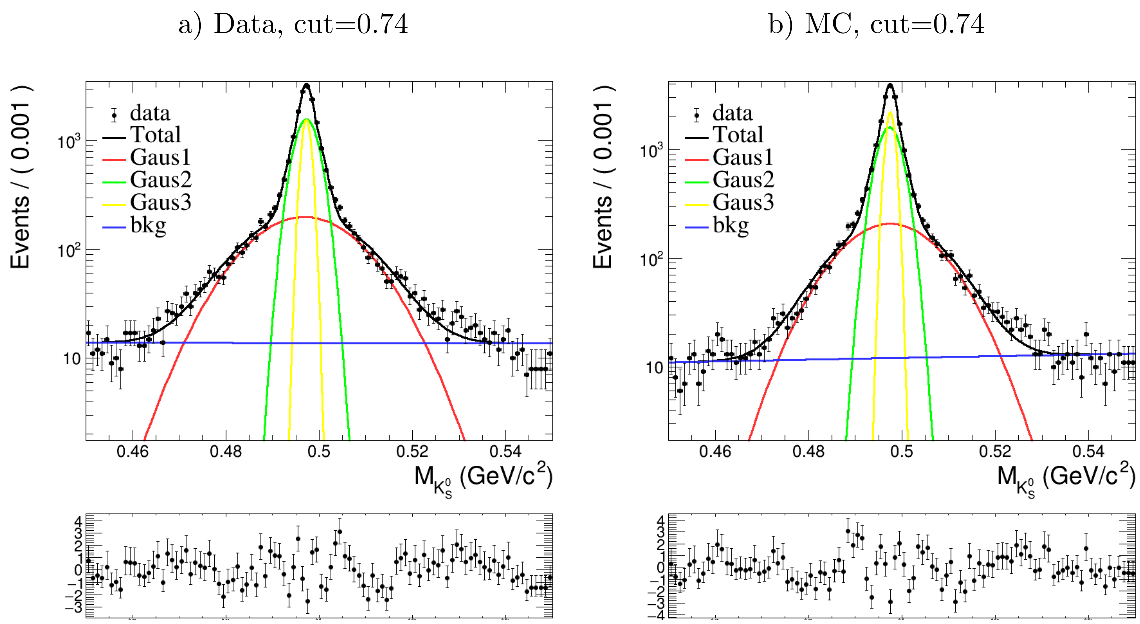


Figure 3-21: The fit on invariant mass $M_{K_S^0}$ where signal component is modeled as a triple-Gaussian and background component is modeled as a Chebyshev polynomial. The bottom plots are the pull of the fitted lines and data points. The signal fraction is slight higher in *generic MC* compared to that in data.

Chapter 4

B^0 reconstruction and event selection

As introduced in Section 2.9, the branching fraction of $B^0 \to K_S^0 K_S^0 K_S^0$ is 6.0×10^{-6}. The simulation sets the $\Upsilon(4S)$ as the mother particle then $\Upsilon(4S)$ decays into to two scalar B^0 mesons with mixing. $B^0 \to K_S^0 K_S^0 K_S^0$ decay process is simulated based on the possible phase-space the final state particles could obtain, where no CP violation is implemented in the generator level, meaning that the input of $\mathcal{S}(\sin 2\phi_1)$ and $\mathcal{A}$ are both zero.

4.1 K_S^0 Selection

K_S^0 is first reconstructed by the cut-based method using two charged pions which contains a large fraction of fake candidates, as discussed in Chapter 3 and Table 3.1. In addition, a cut on K_S^0 is used considering momentum distribution of $B^0 \to K_S^0 K_S^0 K_S^0$, where the huge fake K_S^0 appear in low momentum region. Only the K_S^0 candidates with momentum larger than 0.05 GeV are selected, as shown in Figure 4-1. To further reduce the fake candidates in K_S^0 using *KsFinder*, only K_S^0 with *FBDT_Ks* larger than 0.74 are kept.

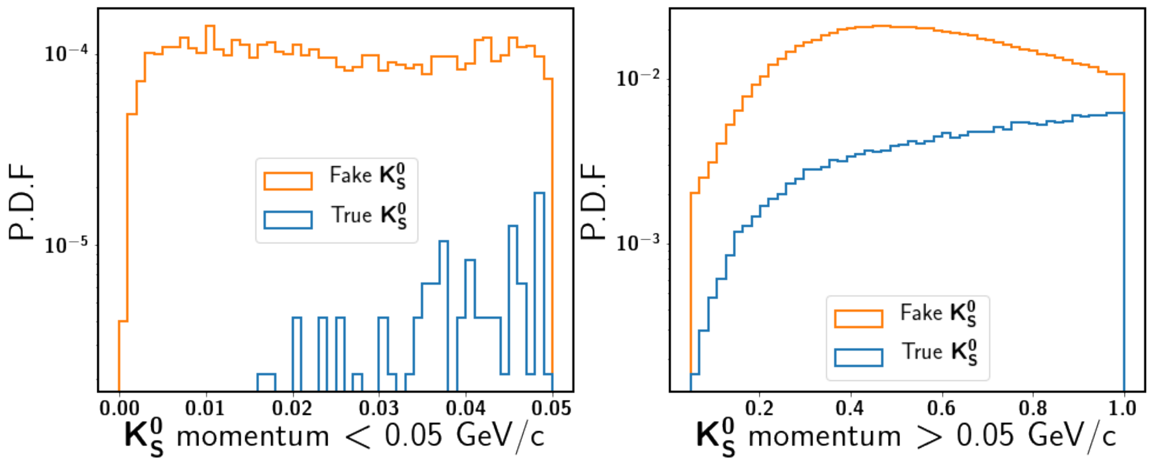


Figure 4-1: The distribution of K_S^0 momentum. Candidates smaller than 0.05GeV/c are rejected.

4.2 B^0 Reconstruction

By combining three K_S^0 particles from selected K_S^0 candidates, B^0 candidates can be reconstructed. The beam-constraint mass M_{bc} and energy difference ΔE are used to extract signal, as defined in Equation 4.1 and 4.2, respectively.

$$M_{bc} = \sqrt{\frac{s}{4} - p_B^{*2}} \tag{4.1}$$

$$\Delta E = E_B^* - \frac{\sqrt{s}}{2} \tag{4.2}$$

For M_{bc}, $\sqrt{s}$ is defined as the invariant mass of the center-of-mass which is calculated from the beam energies and p_B^* is the reconstructed B momentum in the center-of-mass frame. For ΔE, E_B^* is the reconstructed energy in the center-of-mass frame. These two variables are quite useful for discriminating signal and background events for hadronic B decay with fully reconstructed final states. In Belle II, the B^0 candidates with $M_{bc} > 5.2$ GeV and $|\Delta E| < 0.2$ GeV are required.

The vertex information of the fully reconstructed $B^0 \to K_S^0 K_S^0 K_S^0$, called as *CP*-side, is obtained by the vertex fit using *TreeFit* and the χ^2 probability of the fit is calculated. Only B^0 candidates with converged vertex fit results are kept by a very loose cut of $P(\chi^2) > 0.001$. On the other hand, the *CP* violation measurement does not require the certain decay mode of the other B^0 meson in the Rest-Of-Event which includes all particles except for the ones used on the CP-side reconstruction. The B meson in the Rest-Of-Event is called as tag-side B since this B is used to tag the flavor of CP-side B. Thus, no full B^0 reconstruction on the tag-side is performed, meaning that the vertex information can not be obtained by the specific final states. Considering this strategy, the vertex fit on the tag-side is done by *KFit* that only takes advantage of well-reconstructed charged particle tracks. The vertex on the tag-side are required to be located inside the standard PXD region, despite that PXD is not fully installed yet. In future, such requirement is subjected to be modified by the PXD hits requirements.

After performing the vertex fit for both CP-side and tag-side, we check the potential impact of applying *KsFinder* on the reconstructed vertex positions, as well as the impact on the M_{bc} and ΔE. It is necessary for mainly two reasons. First, the *KsFinder* might change the original distributions of M_{bc} and ΔE of B^0 candidates, which are used for the signal extraction. The signal extraction will provide the signal fraction information that is used during the CP parameter measurement. Second, the *KsFinder* might introduce the bias on the distribution of the vertex positions of B^0. The K_S^0 candidates with less SVD hits on their daughter pion tracks usually have poorer reconstruction quality and more likely to be rejected as fake candidates. Therefore, we check the distribution of M_{bc} and ΔE before and after the applying *KsFinder*, as well as the distribution of vertex positions on the z-axis. The details about the comparison can be found in the Appendix E. In conclusion, applying *KsFinder* has a negligible impact on M_{bc}, ΔE and the vertex positions on the z-axis. The contribution of *KsFinder* as a possible systematic uncertainty source mainly comes from the different data and MC responses of the K_S^0 classification, as discussed in the Section 3.2.7.

When multiple B^0 candidates are obtained in a single event, the best candidates selection (BCS) is performed by ranking their χ^2 of the CP-side vertex fit. Since the BCS is based on the χ^2 that might introduce bias in the vertex positions for CP fit, we check the distribution of the vertex χ^2, as shown in Figure 4-2 top left where the data and *generic MC* present a good consistence within 1σ on average. The distribution of the candidate number per event without BCS is shown in top right of Figure 4-2 as well, showing an agreement between data and *generic MC* within around 1σ. The distribution of the candidate number per event from the *signal MC* is also in the bottom left of Figure 4-2. The 2D distribution of M_{bc} and ΔE from $B^0 \to K_S^0 K_S^0 K_S^0$ *signal MC* is shown in Figure 4-2 bottom right, where the correlation factor is about 15% between two observables.

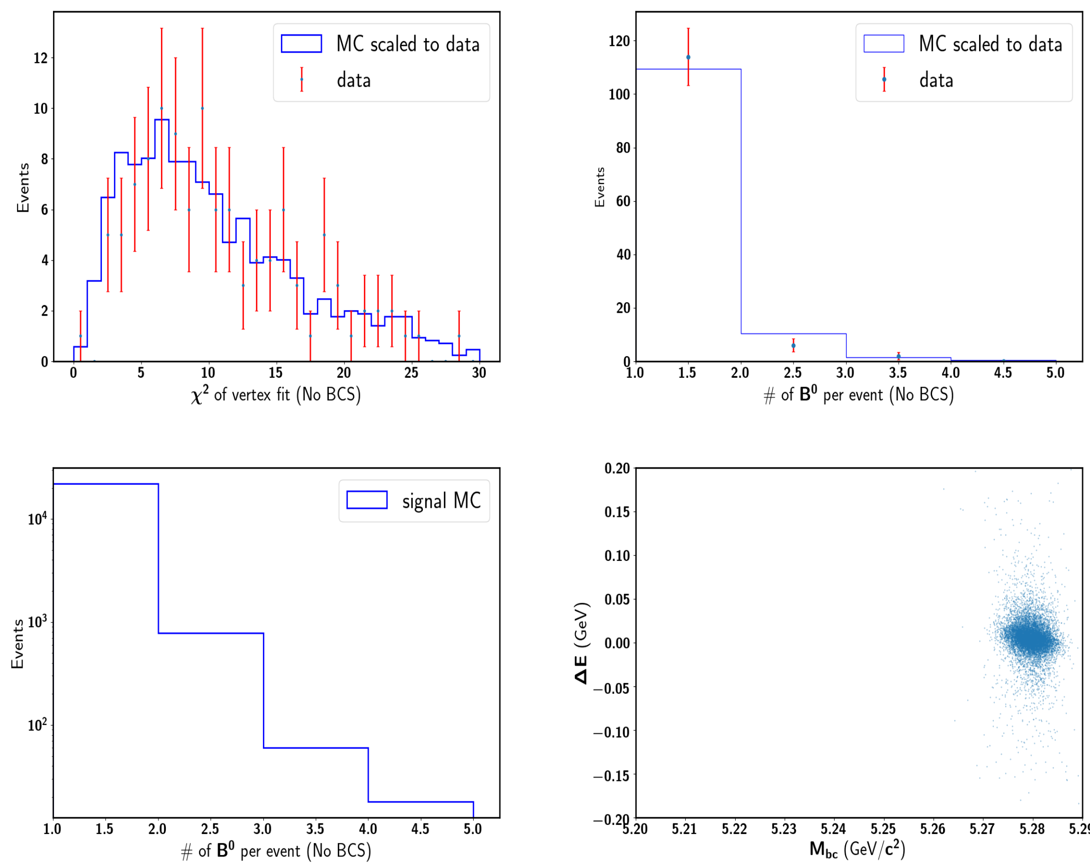


Figure 4-2: Top left is the χ^2 for data and *generic MC* before BCS. Top Right is the B^0 candidates per event in data and *generic MC* before the BCS. Bottom left is the number of B^0 candidates per event from *signal MC*. Bottom right is the 2D M_{bc} and ΔE distribution from *signal MC*.

4.3 Continuum Suppression

The generic decay of Υ(4S) produces neutral and charged B mesons, as well as other flavor mesons $q\bar{q}$. Since the branching fraction of $B^0 \to K_S^0 K_S^0 K_S^0$ is relatively low, the B^0 candidates after the reconstruction contain a large fraction of fake ones if no special reduction on $q\bar{q}$ is applied. The $q\bar{q}$ background events are distributed as a continuum-like shape in the distribution of M_{bc} and ΔE. This calls a demand to distinguish $B\bar{B}$ decay events from $q\bar{q}$ events, which is called as continuum suppression (CS). The rejection is essential because it is the dominated background in this analysis. The most useful information to reject $q\bar{q}$ events is to use the event shape information. In a $B\bar{B}$ event, two mesons are produced almost at rest in the CMS frame since the resonance state $\Upsilon(4S)$ is just slightly lighter than the beam energy. As a result, decay products are emitted more isotropically compared to continuum background events which are more jet-like, back-to-back flying out from the interaction region. The ARGUS and CLEO collaboration developed a set of variables to suppress the continuum background [40], which has also been implemented into BASF2 framework. The two major sets of variables are the CLEO cone momentum variables and the modified Super Fox-wolfram momentum variables.

CLEO cone momentum can be presented as Equation 4.3, where p_i is momentum of i-th particle in the Rest-Of-Event (ROE). The particles used in a reconstructed CP-side B^0 are therefore excluded. The θ_i is an angle between $\vec{p_i}$ and the momentum thrust axis of the reconstructed CP-side B meson. The angle is divided into the binned intervals in nine cones of 10 degrees around the thrust. The L_n stands for the combination that includes particles in a certain cone. The demonstration of the intervals of the CLEO cone is shown in Figure 4-3. The distribution of the first CLEO cone momentum variable in the signal and continuum background events is shown in Figure 4-4.

$$L_n = \sum_{i \in ROE} p_i \times |\cos\theta_i| \tag{4.3}$$

The modified Super Fox-wolfram momentum (KSFW momentum) can be defined

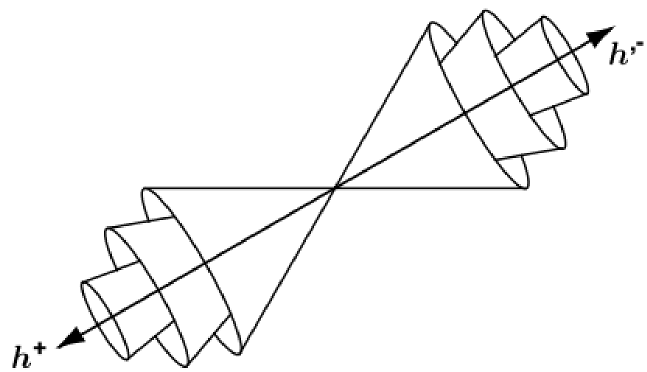


Figure 4-3: A graphical illustration of the CLEO cone. The h^{+} and $h^{'-}$ present the hadronic tracks from a B decay. The first three cones are drawn [40]. The nine CLEO cone momentum can be calculated using the particles in each cone.

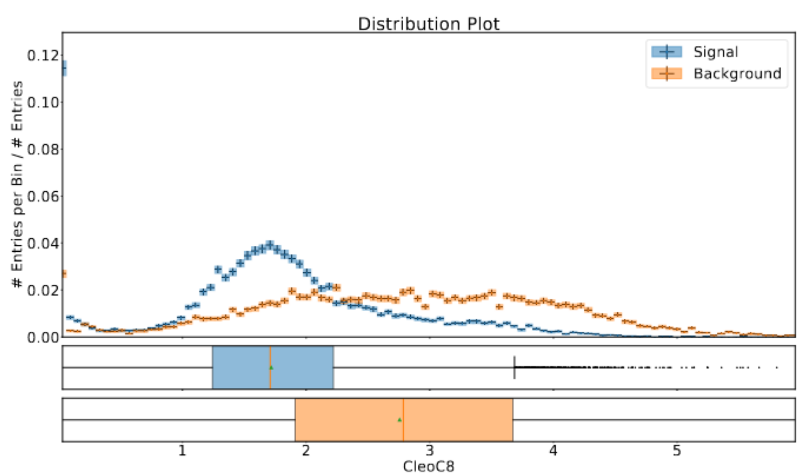


Figure 4-4: The distribution of the first CLEO cone momentum variable in the signal and continuum background events.

as shown in Equation 4.4.

$$KSFW = \sum_{l=0}^{4}(R_l^{so} + R_l^{oo}) + \gamma \sum_{n=1}^{N_t} |P(t)_n| \tag{4.4}$$

The R_l^{so} and R_l^{oo} are the functions which depend on both CP and tag-side particles. Their values are also affected by whether l is even or odd. The $P(t)_n$ is the scalar sum of the transverse momentum of each particle multiplied by a free parameter γ and N_t is the total number of particles. The detailed definition for each KSFW momentum is described in Ref. [40]. As an example, the distribution of one of the KSFW momentum variables is shown in Figure 4-5.

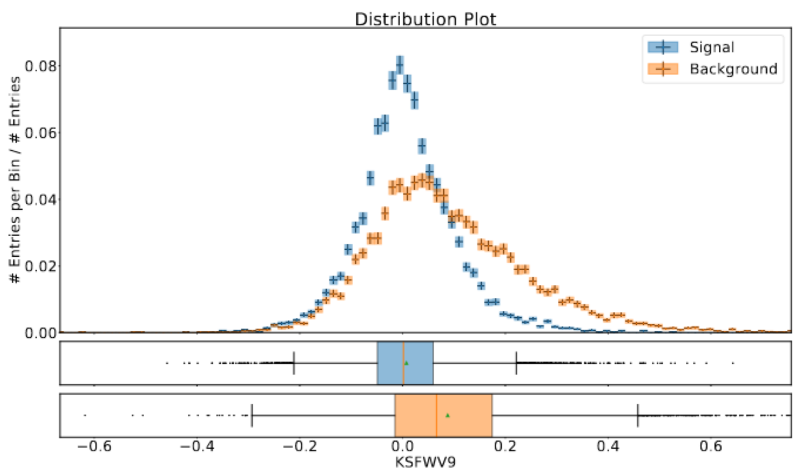


Figure 4-5: The distribution of the variable *KSFWV9* in the signal and continuum background events.

In addition to the CLEO cone and KSFW momentum, a few other variables that are related to the event-shape topology are also used in the Belle II CS framework in order to obtain a better continuum background rejection performance. This includes R_2, $cosTBz$, $cosTBTO$, $thrustOm$, and $thrustBm$. R_2 is defined as the normalized second Fox-Wolfram moment ratio, which is widely used in the decay shape studies. The distribution of R_2 is shown in Figure 4-6. The $cosTBTO$ is cosine of angle between thrust axis of the CP-side B meson and thrust axis of Rest-Of-Event. The $cosTBz$ is cosine of angle between thrust axis of the CP-side B meson and z-axis. The $thrustOm$ and $thrustBm$ are the magnitude of the CP-side B thrust axis and

Rest-Of-Event thrust axis, respectively.

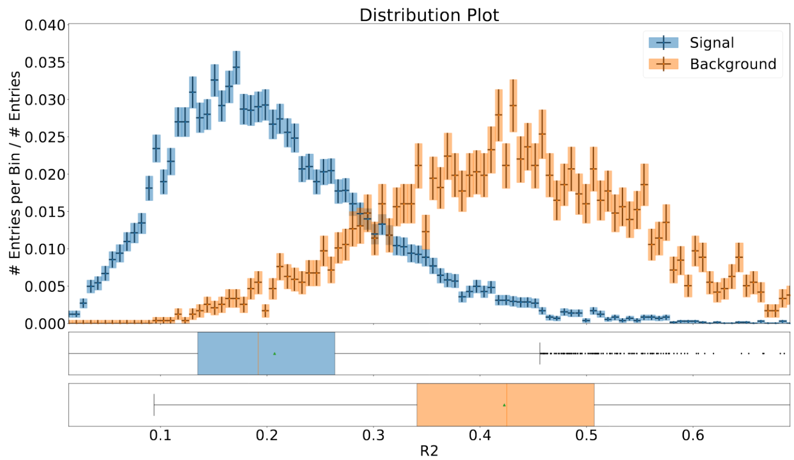


Figure 4-6: R_2 is the ratio of the second to the zeroth KSFW momentum in Equation 4.4 of which the distribution in *signal MC* sample which serves as the highest weight as a variable in discriminating the continuum events, having a quite different distribution between signal and background.

For the Belle II CS strategy, the default method is to use the above variables as an input for a FastBDT classifier to discriminating the signal and continuum background. For the specific decay mode $B^0 \to K_S^0 K_S^0 K_S^0$, the training and testing samples prepared from *signal MC* and *generic MC* are used. The target variable of the training is the continuum event truth named $NotContinuumEvent$, where for signal (continuum background) the value is 1 (0). The same event-reconstruction procedure for B^0 is applied for both MC samples. Events passing the selection using M_{bc} and ΔE are used for training the CS classifier. The fraction of signal and background is set to 1:1. The output of CS classifier is called *FBDT_CS*. We determine the cut value at 0.66 based on the maximum of *FOM* curve, as shown in Figure 4-7. The input variables are listed in Table 4.1 with their abbreviations in the training. After the training, the importance of the input variables can be evaluated, shown as Figure 4-8.

The correlation between these training variables are shown in Figure 4-9 which are varied between signal and continuum background events. The ROC curve and the efficiency/purity using the testing samples with respect to the classifier output are shown in Figure 4-10, yielding a close performance to the training.

The overtraining check is made by comparing the distribution of signal and back-

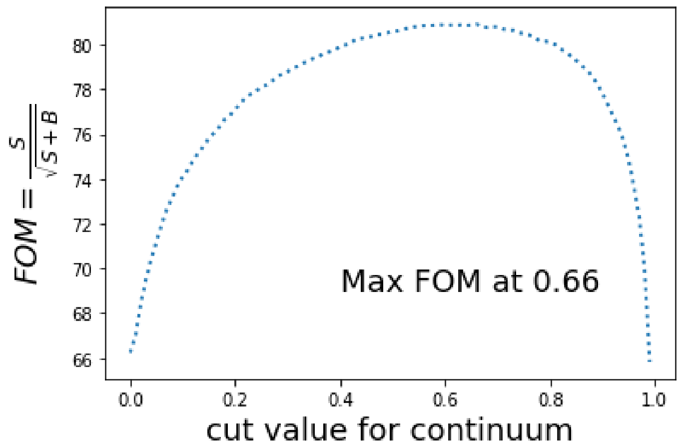
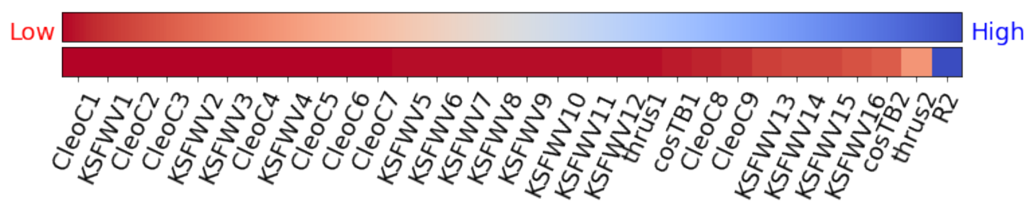


Figure 4-7: FOM depending on the cut value of continuum classifier output, cut value at 0.66 is used for continuum suppression.

Figure 4-8: The importance rank of the input variables for the Belle II CS framework.



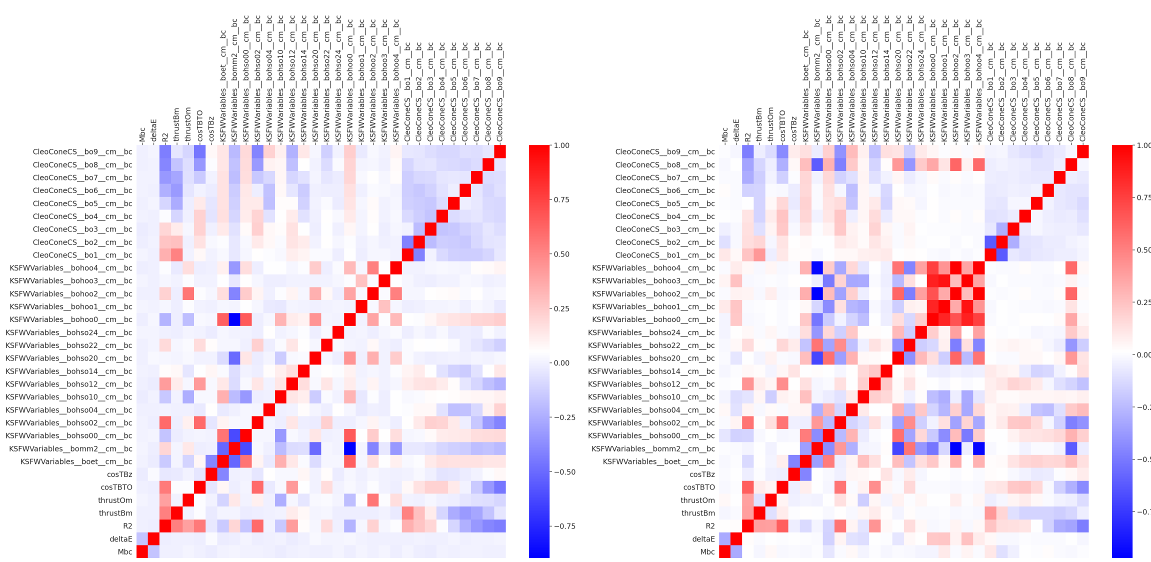
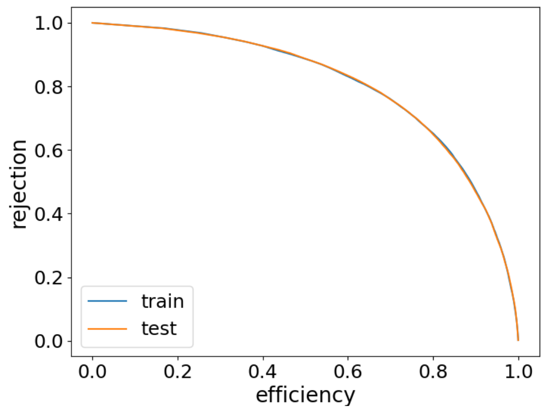


Figure 4-9: The correlation in variables for continuum suppression. The left is for signal and the right is for background.

Table 4.1: Input variables and the abbreviations in the continuum suppression framework of BASF2.

Observables	Abbreviations
CleoConeCS(9)	CleoC1
KSFWVariables(hoo1,)	KSFWV1
CleoConeCS(7)	CleoC2
CleoConeCS(5)	CleoC3
KSFWVariables(hso22)	KSFWV2
KSFWVariables(hoo3)	KSFWV3
CleoConeCS(4)	CleoC4
KSFWVariables(hoo4,)	KSFWV4
CleoConeCS(3)	CleoC5
CleoConeCS(6)	CleoC6
CleoConeCS(8)	CleoC7
KSFWVariables(hso14)	KSFWV5
KSFWVariables(hso00)	KSFWV6
KSFWVariables(et)	KSFWV7
KSFWVariables(hso24)	KSFWV8
KSFWVariables(hso04)	KSFWV9
KSFWVariables(hso20)	KSFWV10
KSFWVariables(mm2)	KSFWV11
KSFWVariables(hoo2)	KSFWV12
thrustOm	thrus1
cosTBz	cosTB1
CleoConeCS(1)	CleoC8
CleoConeCS(2)	CleoC9
KSFWVariables(hso02)	KSFWV13
KSFWVariables(hoo0)	KSFWV14
KSFWVariables(hso12)	KSFWV15
KSFWVariables(hso10)	KSFWV16
cosTBTO	cosTB2
thrustBm	thrus2
R2	R2



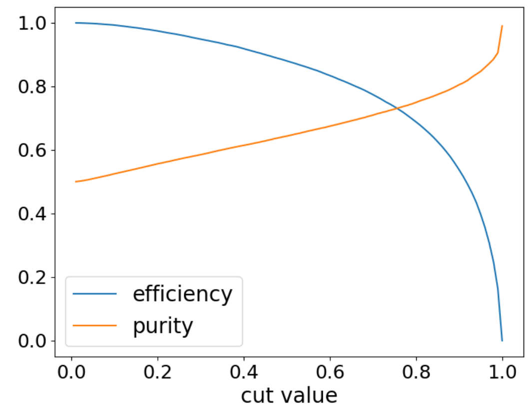


Figure 4-10: The left is the ROC curve (blue for training and orange for testing) and the right is the efficiency (blue) and purity (orange) regarding the classifier output *FBDT_CS*.

ground depending on the classifier output in both training and testing samples. The testing sample shows about 1% lower in each bin for both signal and background events, which is within the acceptable range.

4.3.1 Event selection summary

The summary of Event selections is listed in Table 4.2, including the application of *KsFinder* (by *FBDT_Ks*) and continuum suppression (by *FBDT_CS*).

B^0	$M_{bc}(\text{GeV}/c^2)$	$\Delta E(\text{GeV})$	$P(\chi^2)$	*Rank*	*FBDT_CS*	*FBDT_Ks*
Criteria	> 5.20 & < 5.29	$\|\Delta E\| < 0.2$	> 0.001	$= 1$	> 0.66	> 0.74

Table 4.2: B^0 selection criteria, $P(\chi^2)$ is from B^0 *CP*-side vertex fit and *Rank* is from the BCS

Combined with the previous paragraph, the reconstruction performance of B^0 is summarized in Table 4.3. The efficiency, purity, fraction of multiplicity events and best candidates fraction of B^0 are slightly improved in the Belle II compared to the ones referenced from Belle [19].

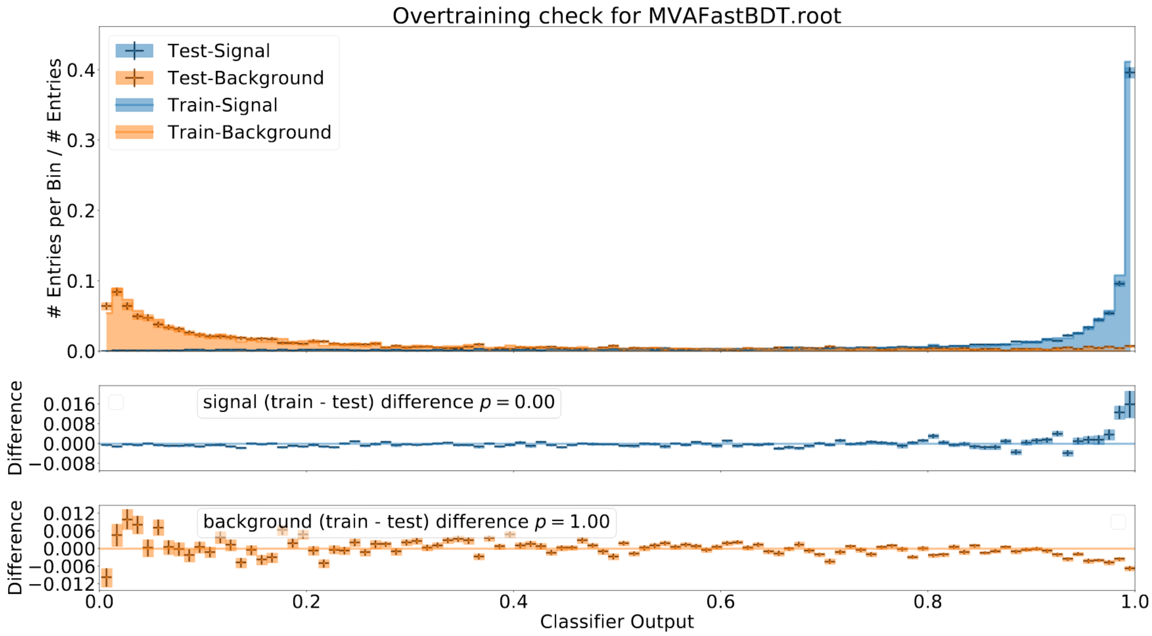


Figure 4-11: Over training check of continuum classifier, where a very small difference in training and testing (1%) is shown.

event selection	efficiency	purity	f_{MB}	BCS
Belle Standard	35%(33%)	96%(99%)	6%(6%)	83%(96%)
Belle II (*BG1*)	36%(34%)	96%(98%)	4%(4%)	95%(96%)
Belle II (*BG0*)	40%(36%)	96%(99%)	3%(3%)	97%(97%)

Table 4.3: The efficiency is defined by the fraction of best candidates among the MC input number. Purity is the fraction of true B^0 in best candidates. f_{MB} stands for multiple B^0 events fraction in true signal events. BCS is the fraction of best candidates being a true signal. All values in the parenthesis are calculated in $|M_{bc}| - 5.28 < 0.1$ and $|\Delta E| < 0.1$, called as "signal region" where efficiency is lower but purity is higher, compared to the full range of M_{bc} and ΔE in Table 4.2. *BG1* and *BG0* indicate that the corresponding values are obtained using *signal MC* with or without the beam background respectively, which the *BG0* sample yields a slightly better reconstruction performance.

4.4 Resonance Background

The CP eigenvalue of $B^0 \to K_S^0 K_S^0 K_S^0$ is $\eta_f = +1$ if it is a loop-level $b \to s$ transition, called CP-even. However, the charmonium resonances from $b \to c$ tree-level transition can give an odd CP eigenvalue ($\eta_f = -1$) while producing the same final states as $B^0 \to K_S^0 K_S^0 K_S^0$. This would cause the contamination to the CP measurement due to the fact that $\mathcal{S} = -\eta_f \sin(2\phi_1)$. The loop-level CP-even processes are demonstrated in the left and middle diagrams of Figure 1-5, including the phase-space based decay and resonance decays. Both are considered as the signal events. The tree-level CP-odd process is shown in the right diagram of Figure 1-5, which is usually referred as the resonance background and can be rejected by applying cut on the invariant mass of two K_S^0 around the possible charmonium resonant states. Therefore, to check wether a resonance decay is a signal or background event, a list of possible $B^0 \to X(\to K_S^0 K_S^0) K_S^0$ processes are considered. For the resonance decays of signal events, X could be $f_2(1270)$, $f_0(1500)$, $f_2^{'}(1525)$, $f_0(980)$, $f_0(1710)$ and $f_2(2010)$. For the resonance background, X could be D^0, η, J/ψ, $\psi(2S)$, χ_{c0}, χ_{c1}, and χ_{c2}. The branching fraction of these decay modes in the PDG and the Belle II decay profile are both checked and compared, where some of them are not yet measured or not implemented in the Belle II simulation. Thus it is hard to precisely evaluate all the contributions from these decay modes. The expected contributions are calculated by using $2.14 \times 10^8 B\bar{B}$ pairs corresponding to 400 fb^{-1} as shown in the Table 4.4. The main contribution from the resonance background is $B^0 \to \chi_{c0} K_S^0$ and the main contribution from the resonant signal is $B^0 \to f_0(980) K_S^0$. Given the very limited statistics of data accumulation we used in this analysis, the contribution of CP-odd resonance background should be smaller than one event. Besides, the comparison of two K_S^0 invariant mass using data and *generic MC* is performed, while the small number of reconstructed events in data can not provide reasonable information on the distribution, see Appendix C. Hence, currently no veto of two K_S^0 invariant mass is applied to avoid potential bias in this low statistics scenario. In the future with more data collected, the veto will be first checked by the proper comparison between data and MC, then applied according to mass of the main contributive resonance states.

Table 4.4: Expected yield for signal and background resonances $2.14 \times 10^8 B\bar{B}$ in *generic MC*. The branching fraction of $B \to XK_S$ and $X \to 2K_S$ are listed from both the PDG values and the values used in Belle II generic decay profile. The number of events from *CP*-odd resonance background is expected to be smaller than one at current luminosity 62.8 fb^{-1}.

Resonances	$\mathrm{Br}(B \to XK_S)$PDG	$\mathrm{Br}(X \to 2K_S)$	$\mathrm{Br}(B \to XK_S)$Dec.	$\mathrm{Br}(X \to 2K_S)$Dec.	$B\bar{B}$ pairs	Expected yields
D^0K_S	2.6×10^{-5}	1.7×10^{-4}	2.6×10^{-5}	1.8×10^{-4}	2.14×10^8	0.134
ηK_S	3.45×10^{-4}	$< 3.1 \times 10^{-4}$	4×10^{-4}	No Value	2.14×10^8	No Value
$J/\psi K_S$	4.35×10^{-4}	$< 1.4 \times 10^{-8}$	4.35×10^{-4}	0	2.14×10^8	0
$\psi(2S)K_S$	2.9×10^{-4}	$< 4.6 \times 10^{-6}$	2.9×10^{-4}	0	2.14×10^8	0
$\chi_{c0}K_S$	7.3×10^{-5}	3.16×10^{-3}	7.35×10^{-5}	3.1×10^{-3}	2.14×10^8	6.21
$\chi_{c1}K_S$	1.96×10^{-4}	6×10^{-5}	1.96×10^{-4}	1×10^{-5}	2.14×10^8	0.05
$\chi_{c2}K_S$	7.5×10^{-6}	2.6×10^{-4}	7.5×10^{-6}	$5.5 \times 10 - 4$	2.14×10^8	0.11
$f_2(1270)K_S$	1.35×10^{-6}	1.15×10^{-2}	1.35×10^{-6}	1.15×10^{-2}	2.14×10^8	0.42
$f_2'(1525)K_S$	1.5×10^{-7}	2.22×10^{-2}	No value	0.22	2.14×10^8	No Value
$f_2(2010)K_S$	5×10^{-7}	No Value	No Value	No Value	2.14×10^8	No Value
$f_0(980)K_S$	2.7×10^{-6}	No Value	2.75×10^{-6}	No Value	2.14×10^8	43.3
$f_0(1710)K_S$	5×10^{-7}	No Value	No Value	No Value	2.14×10^8	No Value
$f_0(1500)K_S$	6.5×10^{-5}	0.022	No Value	0.022	2.14×10^8	No Value
Total	-	-	-	-	-	$\simeq$50

4.5 $B\bar{B}$ background

Another possible contribution of background comes from $B\bar{B}$ events. Both neutral and charged $B\bar{B}$ pairs could produce the background events. Compared to the event number from the continuum background, the number of $B\bar{B}$ background is much fewer. By counting the background event number from $B\bar{B}$ and $q\bar{q}$ using *generic MC*, the fraction of $B\bar{B}$ takes about 3% among all background events, and no special treatment is implemented.

4.6 Signal Extraction

The event selections defined in Table 4.2 is applied to *signal MC*, *generic MC* and experiment data for extracting signal events. The integral luminosity in *generic MC* is 1 ab^{-1} and experiment data used in this analysis is 62.8 fb^{-1} from the latest official processing. The distribution of M_{bc} and ΔE from *generic MC* are shown in Figure 4-12 which contains about 21 signal events, 44 continuum background events and 2 $B\bar{B}$ background events if it is normalized to the data luminosity. The distribution of M_{bc} and ΔE from data are also shown in Figure 4-13, where 119 events are observed.

The unbinned maximum likelihood fit using RooFit is performed to extract the signal. The 2D fit using both M_{bc} and ΔE are done by taking the probability density function:

$$\mathcal{P}(M_{bc}, \Delta E) = f_{sig} \times \mathcal{P}_{sig}^{M_{bc}} \times \mathcal{P}_{sig}^{\Delta E} + (1 - f_{sig})\mathcal{P}_{bkg}^{M_{bc}} \times \mathcal{P}_{bkg}^{\Delta E} \ , \tag{4.5}$$

where $\mathcal{P}_{sig}^{M_{bc}}$ and $\mathcal{P}_{sig}^{\Delta E}$ are the probability density function (P.D.F) for the distributions of M_{bc} and ΔE. The f_{sig} is the fraction of signal events. We use the single Gaussian function as $\mathcal{P}_{sig}^{M_{bc}}$ in the distribution of M_{bc} and triple Gaussian functions as $\mathcal{P}_{sig}^{\Delta E}$ in the distribution of ΔE to model the signal component.

On the other hand, the dominated background comes from the continuum events,

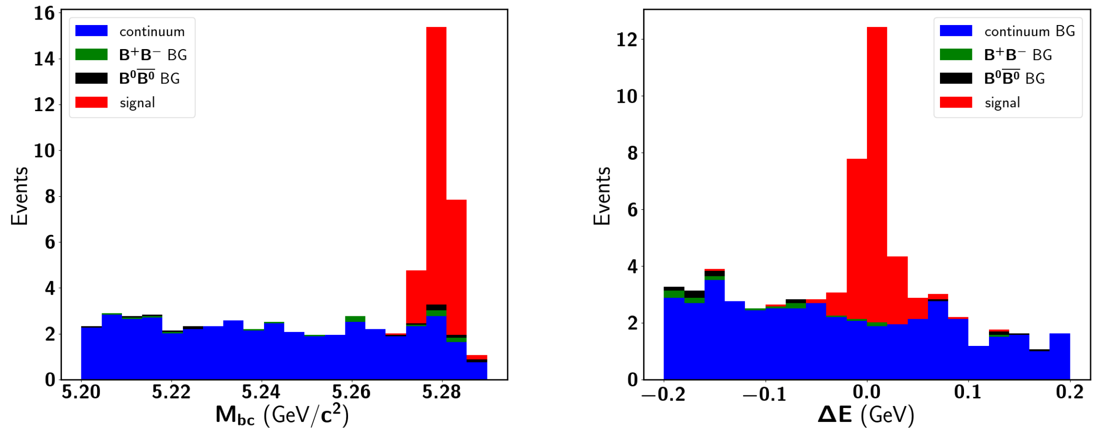


Figure 4-12: The distribution of M_{bc} and ΔE for selected events from *generic MC*, where each background components are stacked with signal normalized to 62.8 fb^{-1}. The blue component is the continuum background. The green and the black are the background from charged and neutral $B\bar{B}$ events. The red component is the signal component by checking MC truth matching.

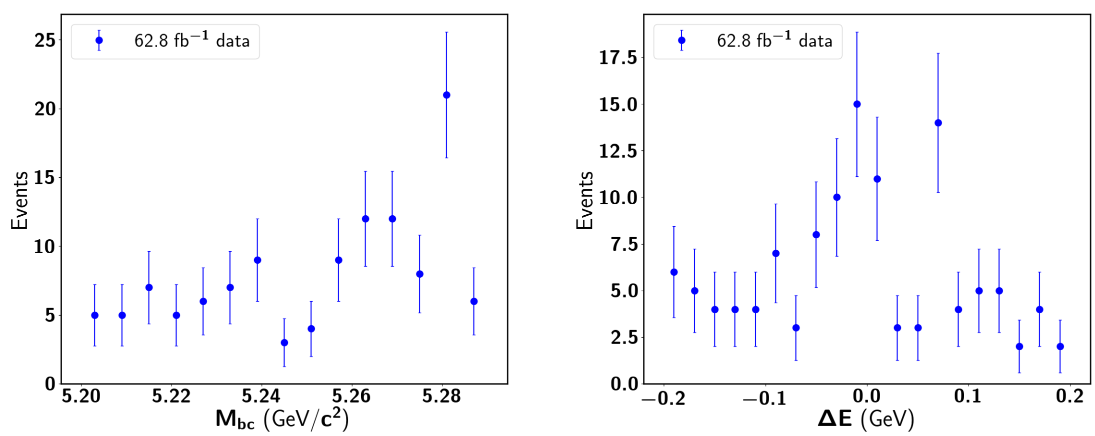


Figure 4-13: The distribution of M_{bc} and ΔE for selected events from the experiment data of 62.8 fb^{-1}.

which is modeled as the Argus function [41] in the distribution of M_{bc}:

$$\mathcal{P}_{bkg}^{M_{bc}}(x; c, \chi) = \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} \cdot \frac{x}{c^2}\sqrt{1 - \frac{x^2}{c^2}} \cdot \exp\left\{-\tfrac{1}{2}\chi^2(1 - \tfrac{x^2}{c^2})\right\} , \tag{4.6}$$

where x presenting M_{bc} is defined in $0 < x < c$ with a preset mass threshold at $c = 5.29$ GeV. The χ is a parameter of the distribution. The $\Psi(\chi) = \Phi(\chi) - \chi\phi(\chi) - \frac{1}{2}$ where $\Phi(\chi)$ and $\phi(\chi)$ are cumulative distribution and probability density function of the standard normal distribution, respectively. The ΔE distribution of continuum events is modeled by the first order Chebyshev polynomial function. The scattered distribution of M_{bc} and ΔE in the 2D plane is shown in Figure 4-14.

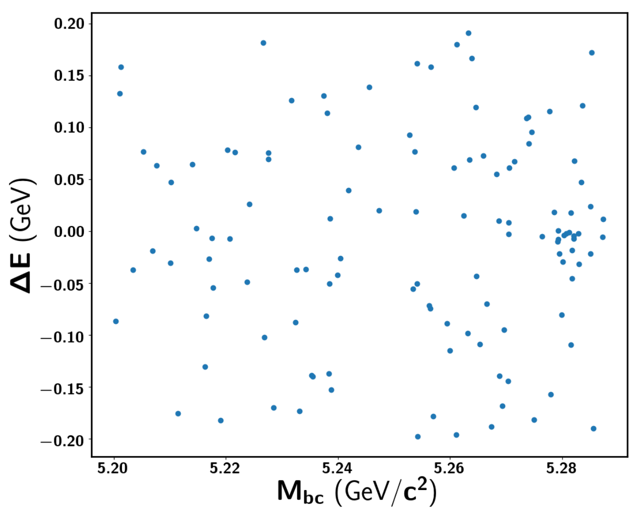


Figure 4-14: The 2D distribution of M_{bc} and ΔE in the 62.8 fb^{-1} experiment data.

The unbinned maximum likelihood fit is first performed to obtain the parameters for signal functions using *signal MC*, and then fixed them as the constants latter for 2D fit. The fit results on *signal MC* are shown in Figure 4-15. The continuum background is fitted by using $q\bar{q}$ events from *generic MC* to determine the shapes then fix them as the constants latter for 2D fit. The fit results on $q\bar{q}$ events are shown in Figure 4-16.

Then we set the number of signal and background events as floating parameters and use Equation 4.5 as the 2D model to fit on 1 ab^{-1} *generic MC* as shown in the

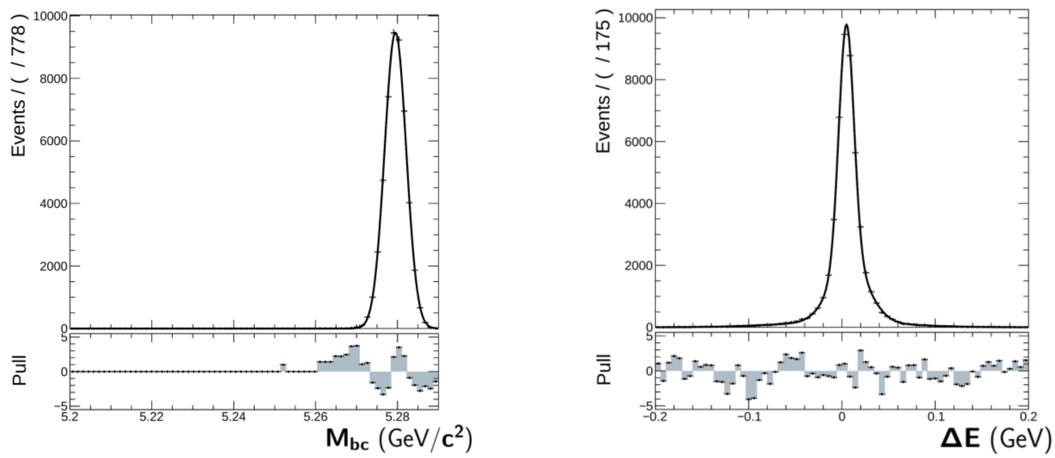


Figure 4-15: The distribution of M_{bc} and ΔE of *signal MC* of $B^0 \to K_S^0 K_S^0 K_S^0$ fitted with single and triple Gaussian functions, respectively. The bottom plots are the pull of the data points and the fit result.

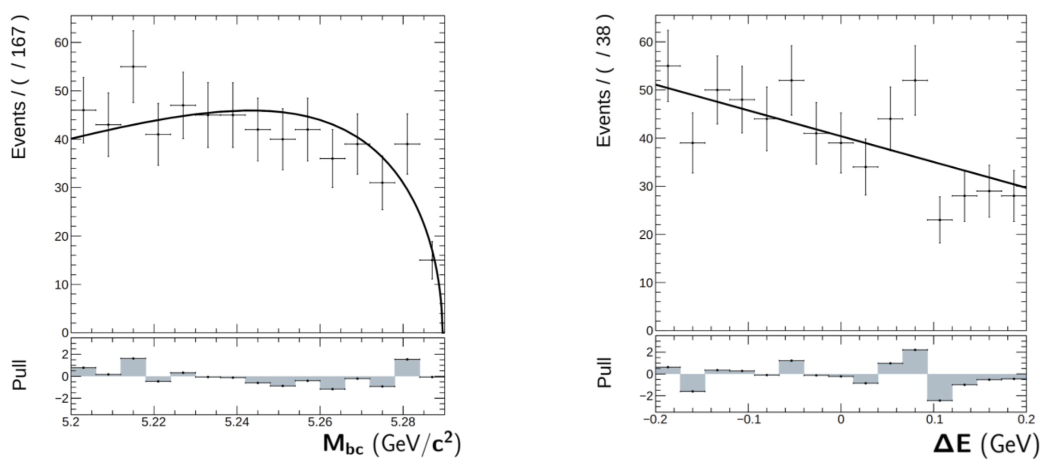


Figure 4-16: The distribution of M_{bc} and ΔE of continuum events in *generic MC* fitted with Argus and Chebyshev polynomial functions, respectively. The bottom plots are the pull of the data points and the fit result.

Figure 4-17.

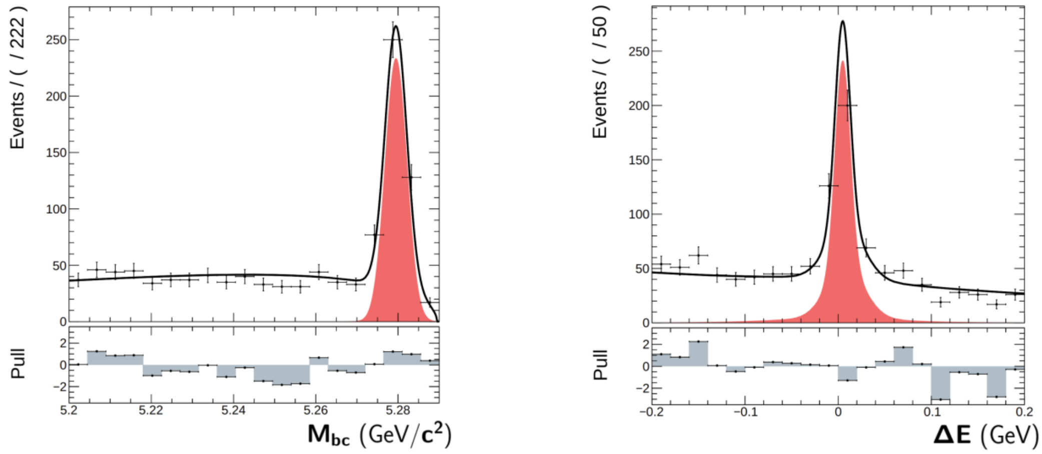


Figure 4-17: The fit result for *generic MC* projected on M_{bc} and ΔE are shown, where the red is signal component from the fit result in both plots. The bottom subplots are the pull of the fit and the distribution which is defined as the difference between the fit line and the points divided by points in each bin.

Before performing 2D fit on experiment data, the distribution of K_S^0 invariant mass from the reconstructed B^0 candidates is compared between *generic MC* and experiment data. The selection criteria in Table 4.2 are applied to both samples. The distributions are shown in Figure 4-18, where the *generic MC* is scaled to the luminosity of experiment data and an agreement within $\sim 1\sigma$ is observed.

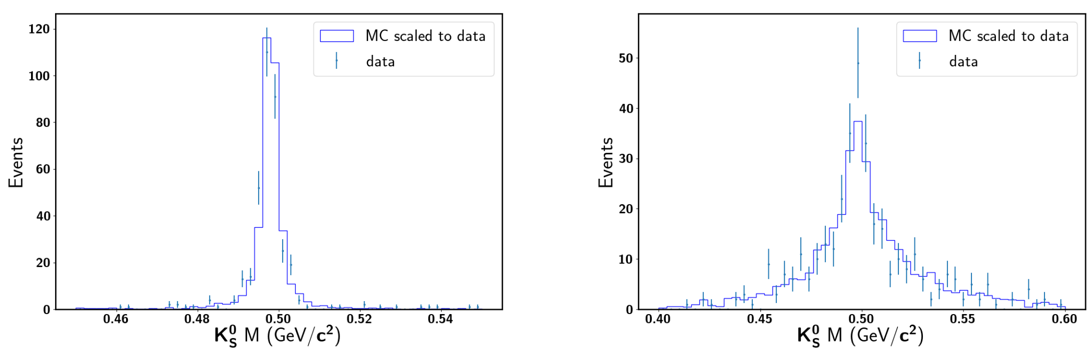


Figure 4-18: Invariant mass before (left) and after (right) B^0 vertex fit from *generic MC* and the 62.8 fb^{-1} experiment data.

As the fit procedure for the *generic MC*, the 2D fit of the experiment data is done and the distributions projected on M_{bc} and ΔE are shown in Figure 4-19. The number

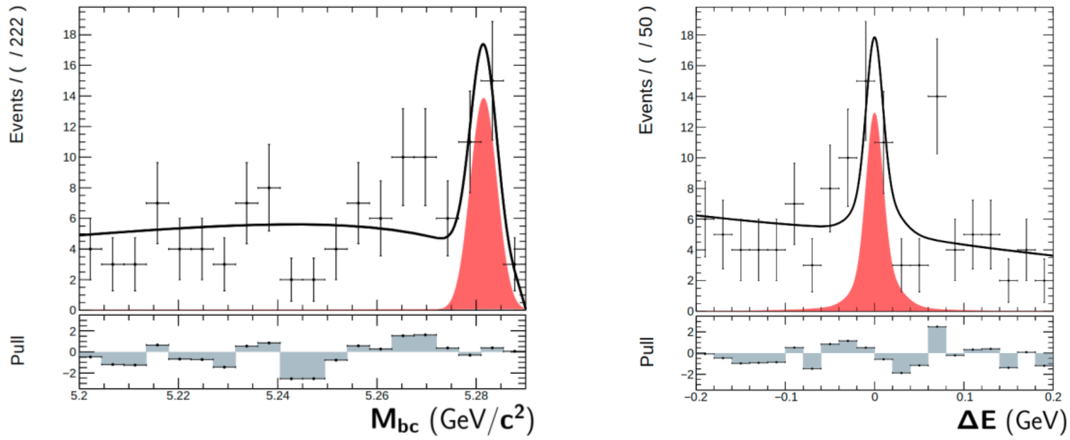


Figure 4-19: M_{bc}(GeV) and ΔE(GeV) 2D fit on 62.8 fb^{-1}data, the red is the signal component.

of signal events is extracted by the integral of fit model over the signal region which is defined as $5.27 < M_{bc} < 5.29$ GeV and $-0.1 < \Delta E < 0.1$ GeV. Using 62.8 fb^{-1} data, we extract the number of signal events to be $N_{sig} = 17.4 \pm 4.2$ from the signal region. The number of background events in the signal region is 7.2 ± 3.6. When we directly count the number of events from the data, there are 30 events in the signal region and 60 events in the sideband region which is defined as $M_{bc} < 5.26$ GeV/c^2. The events in the signal region will be used as the input data points for the *CP* parameters measurement later.

In the meantime, the number of events can also be estimated by using Equation 4.7:

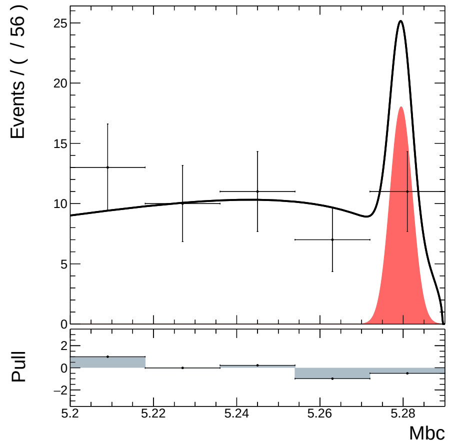
$$N_{sig} = \mathcal{B}(B^0 \to K_S^0 K_S^0 K_S^0) \times \mathcal{B}(K_S^0 \to \pi^+\pi^-)^3 \times \epsilon_{rec} \times N_{B\bar{B}} \times 2 \ , \qquad (4.7)$$

where the $N_{B\bar{B}}$ is the number of neutral B meson pairs calculated from the integrated luminosity and ϵ_{rec} is the reconstruction efficiency. The factor 2 accounts for the fact that both B^0 and $\overline{B^0}$ can decay to three K_S^0 in the final states. Thus, the number of signal and background events in 62.8 fb^{-1} *generic MC* is calculated to be ~ 20.6 and ~ 4.1, respectively. The calculated results are in an agreement with the normalized event number from Figure 4-12, as well as the ones from the 2D fit of data. The number of signal and background events inside the signal region from the 2D fit using data and *generic MC* are summarized in Table 4.5.

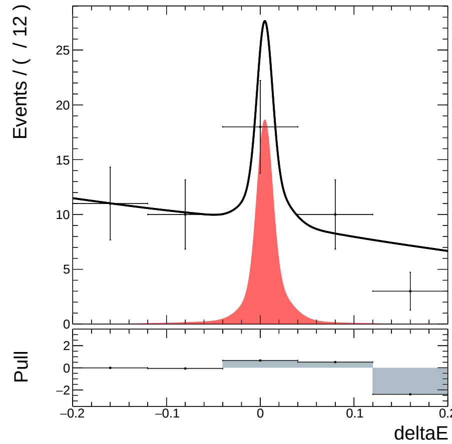
Table 4.5: Reconstructed signal and background events using M_{bc} and ΔE 2D fit, compared with expected numbers from *generic MC*.

Events (signal region)	Signal	Background
62.8 fb^{-1} *generic MC*	~ 20.6	~ 4.1
62.8 fb^{-1} data	17.4 ± 4.2	7.2 ± 3.6

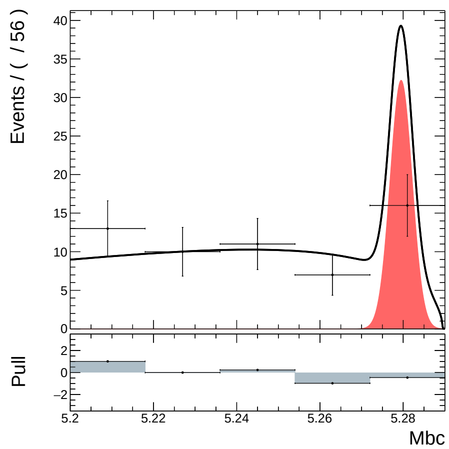
To check reliability of the number of events fitted from the M_{bc} and ΔE in this low statistics case, we test the 2D fit method using MC, by merging the background events and the different number of signal events, to check the linearity of the input and output. In the full range of M_{bc} and ΔE, *generic MC* yields about 44 continuum events and 2 $B\bar{B}$ background events at the luminosity of 62.8 fb^{-1}, where both of these two background components are included in the model of Argus function and first order Chebyshev polynomial function. Therefore, 46 background events from the *generic MC* is used as a constant background component in the linearity test. Then the number of signal events from 5 to 30 with 5 events per step are injected into the background events, to perform the M_{bc} and ΔE 2D fit to obtain the output signal events number. The M_{bc} and ΔE distributions and fit results in each injection test are shown in Figure 4-20. The output signal and background events depending on the injected number of the signal events are presented in Figure 4-21, where the dependence on both signal and background events are fitted with a linear function $y = ax + b$. The error bar on each data point is taken from the statistical uncertainty of the number of signal or background events of the 2D fit results. The fit results show a good linearity on the input and output of the number of signal events while the number of continuum events remain close the constant 46 as the input number. Due to this fact, the signal event yield from the current low luminosity is considered as a reliable result.



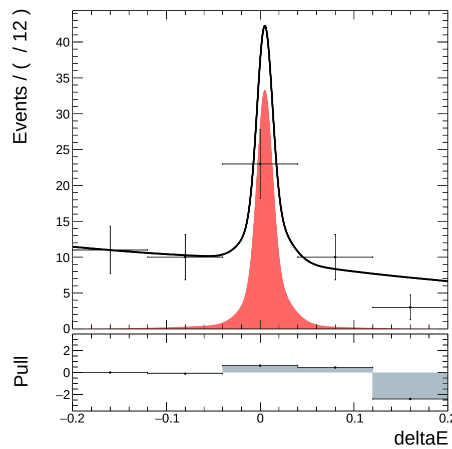
a) signal injected: 5



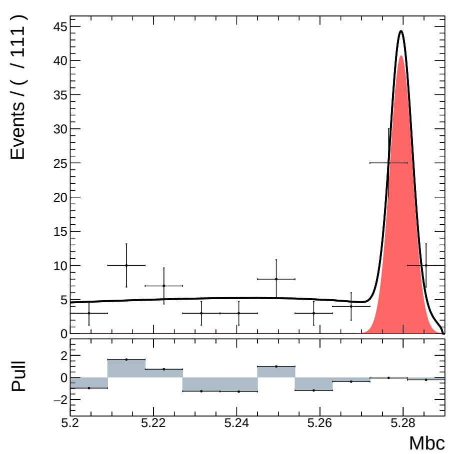
b) signal injected: 5



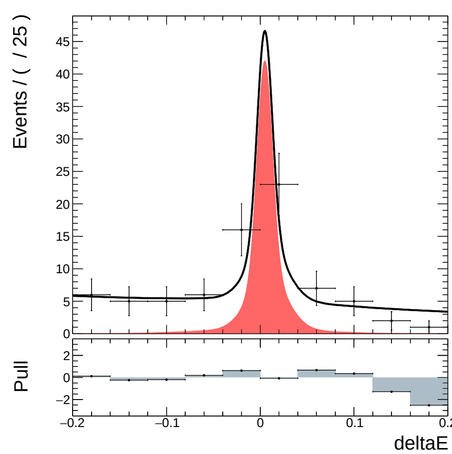
c) signal injected: 10



d) signal injected: 10



e) signal injected: 30



f) signal injected: 30

Figure 4-20: The fit results of M_{bc} and ΔE in signal injection test, where the number of signal events of 5, 10, and 30, injected with 46 continuum events, are shown. The full results including other values of the number of signal events are included in the Appendix D.

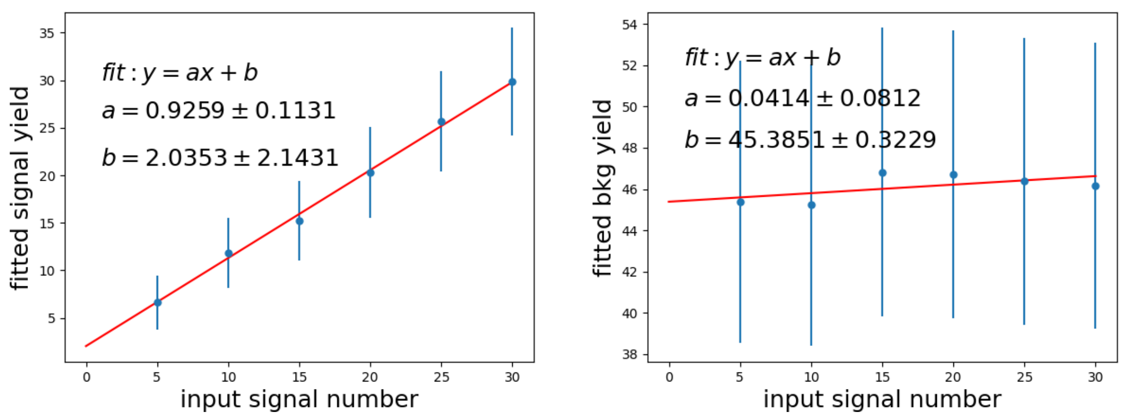


Figure 4-21: Injection test for signal extraction. The linearity is clearly observed between the input and output signal events numbers. The error bar on each point is taken from the statistical uncertainty of the number of signal or continuum events from 2D fit results.

Chapter 5

CP parameter measurement

The measurement of CP parameters $\mathcal{S}$ and $\mathcal{A}$ is perform by fitting Equation 5.1 to the distribution of the decay time difference Δt of events with flavor q, where $\Delta t = t_{CP} - t_{tag}$ and $q = +1$ (-1) when the tag-side B meson is B^0 $(\bar{B}^0)$.

$$\mathcal{P}_{sig}(\Delta t, q) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \left\{1 + q \cdot \Big[\mathcal{S} \cdot \sin(\Delta M_d \Delta t) + \mathcal{A} \cdot \cos(\Delta M_d \Delta t)\Big]\right\} \tag{5.1}$$

The Equation 5.1 describes the distribution of signal events only with τ_{B^0} and ΔM_d as physics parameters. To perform the unbinned maximum likelihood fit on data, a complete model for i-th event that takes into account background and outlier components is defined as:

$$\begin{aligned}\mathcal{P}(\Delta t_i, q_i, f_i^{sig}, \mathcal{S}, \mathcal{A}) = (1 - f_{ol}) &\left[f_i^{sig} \mathcal{P}_{sig}(\Delta t_i, q_i, \mathcal{S}, \mathcal{A}) + (1 - f_i^{sig}) \mathcal{P}_{bkg}(\Delta t_i)\right] \\ &+ f_{ol} \mathcal{P}_{ol}(\Delta t_i) \ ,\end{aligned} \tag{5.2}$$

where f_i^{sig} is the signal fraction assigned to the i-th event and f_{ol} is the fraction of the outlier components, respectively. The $\mathcal{P}_{bkg}$ and $\mathcal{P}_{ol}$ are defined as

$$\mathcal{P}_{bkg}(\Delta t_i) = f_{bkg}^{\delta} \delta(\Delta t_i - \mu_{bkg}^{\delta}) + (1 - f_{bkg}^{\delta}) \frac{1}{2\tau_{bkg}} e^{-|\Delta t_i - \mu_{\tau}^{bkg}|/\tau_{bkg}} \ , \tag{5.3}$$

$$\mathcal{P}_{ol}(\Delta t_i) = G(\Delta t_i, \sigma_{ol}) \ , \tag{5.4}$$

where $\delta(\Delta t_i - \mu^{\delta}_{bkg})$ is a Dirac δ function and G is a single Gaussian. The outlier component is to improve the fit quality with large Δt events.

5.1 Vertex Resolution Model

The Equation 5.2 presents an ideal distribution of Δt_i for each event without considering the difference between measured and the true position of the vertex. The difference can be described by introducing resolution functions, turning Equation 5.2 into Equation 5.5.

$$\begin{aligned}\mathcal{P}(\Delta t_i, q_i, f_i^{sig}, \mathcal{S}, \mathcal{A}) = (1 - f_{ol})[&f_i^{sig}\mathcal{P}_{sig}(\Delta t_i) \otimes R_{sig}(\Delta t_i) \\ &+ (1 - f_i^{sig})\mathcal{P}_{bkg}(\Delta t_i) \otimes R_{bkg}(\Delta t_i)] \\ &+ f_{ol}\mathcal{P}_{ol}(\Delta t_i) \otimes R_{ol}(\Delta t_i)\end{aligned} \tag{5.5}$$

The R_{sig} stands for the resolution function for signal events, which receives smearing effect from CP-side and tag-side separately, namely R_{cp} and R_{tag}. The treatment of CP-side and tag-side is different because of the different vertexing strategies. For CP side, the vertex of B^0 is reconstructed by fitting all the daughter particles using *TreeFit*. Instead, in tag-side, there is no full reconstruction of B^0 so that a vertex fit is applied for the selected charged tracks in the Rest-Of-Event using *KFit*. The background events have its own resolution model which is independent from CP violation parameters. The outlier component is used to smooth fit for large Δt events. In the low statistics case like the current luminosity, the outlier component is not taken into account in the fit model of Equation 5.5.

For signal events, the resolution functions are studied for CP-side and tag-side based on each possible degradation such as detector resolutions, the effect of tracks from non-primary B vertex and so on. Such approaches have been used in Belle analyses. Details are summarized in Ref. [42]. The vertex position difference Δz for signal events can be broken down to

$$\Delta z = \Delta z' + (z_{cp} - z'_{cp}) - (z_{tag} - z'_{tag}) \; , \tag{5.6}$$

where the primed ones stand for physics truth of the position and the non-primed one are the measured value. Importantly, the resolution functions on both CP- and tag-sides depend on the applied IP constraints since it will change the obtained information of vertices. Considering that the fine structure of IP profile is not yet fully understood and small discrepancies have been observed between data and simulation [43], there is no IP constraint applied for both sides in vertex fit to avoid the potential bias from IP profile under this low statistical situation. From the Equation 5.6, the total vertex resolution of signal events can be presented as the sum of the two variables which stand for the resolution effects on the both sides. Therefore, the vertex resolution function of signal events is defined as

$$R_{sig} = R_{cp} \otimes R_{tag} \tag{5.7}$$

where R_{cp} and R_{tag} are the vertex resolution functions for CP-side and tag-side, respectively. The determination of these two components are described in Section 5.1.1 and 5.1.2.

5.1.1 CP-side resolution function

A CP-side vertex position is obtained by *TreeFit* with all tracks from a reconstructed B^0, thus the resolution models only depend on the performance of the Belle II detectors, such as the vertexing and tracking accuracy affected by the detector hardware and software. Hence, the resolution effects for each event can be different based on event-by-event reconstruction quality, primarily presented by the reduced χ^2 called χ^2/N from *TreeFit*, which N is the degree of freedom of the fit. The distribution of χ^2/N in data are shown in Figure 5-1 compared with that of *generic MC*. The error bar is taken as $\sqrt{N}$ where N is the number of events in each bin.

We model the resolution function R_{cp} on CP-side by using a double Gaussian function, where the mean is fixed to zero and the standard deviation is the error of

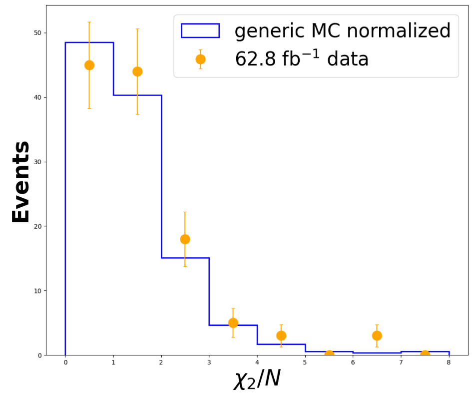


Figure 5-1: The distirbution of χ^2/N in selected events from data compared with that of *generic MC*.

reconstructed vertex $\sigma_{z_{cp}}$ scaled by χ^2/N:

$$R_{cp}(\delta z_{cp}) = (1 - f_{cp}^{tail})G(0, s_{cp}^{main}) + f_{cp}^{tail}G(0, s_{cp}^{tail}) \ , \tag{5.8}$$

where s_{cp}^{main} and s_{cp}^{tail} are

$$\begin{aligned} s_{cp}^{main} &= (s_0^{main} + s_1^{main} \cdot \chi_{cp}^2/N) \cdot \sigma_{z_{cp}} \ , \\ s_{cp}^{tail} &= (s_0^{tail} + s_1^{tail} \cdot \chi_{cp}^2/N) \cdot \sigma_{z_{cp}} \ . \end{aligned} \tag{5.9}$$

Figure 5-2 shows the distribution of the $z_{cp} - z_{cp}^{'}$ as the residual with the dependence of the χ^2/N, where the first plot is for the full range of $0 < \chi^2/N < 10$, covering all of the selected events. The rest plots are the distributions in the sliced ranges of χ^2/N and they show an small increase of the standard deviation with respect to the increasing of χ^2/N range. This validates the necessity to include the χ^2/N as a conditional parameter in R_{cp} so that for a poorly reconstructed B^0 the resolution functions should yield relatively larger deviation.

Restrictively speaking, the CP-side resolution for $B^0 \to K_S^0K_S^0K_S^0$ should be slight different from $B^0 \to J/\psi K_S^0$ due to the absence of the direct tracks of the charged particles from the B^0 vertex. The modification of the resolution function on CP-side will be further studied when more data becomes available in future. Given the current low statistics, the Equation 5.8 works well as an approximation. By fitting the resolution function using *signal MC* on CP-side, the parameters are obtained and listed in Table 5.1.

Table 5.1: Parameters in R_{cp}.

f_{cp}^{tail}	0.07420 ± 0.00080
s_0^{main}	0.9151 ± 0.0077
s_1^{main}	0.2142 ± 0.0064
s_0^{tail}	2.048 ± 0.078
s_1^{tail}	1.347 ± 0.072

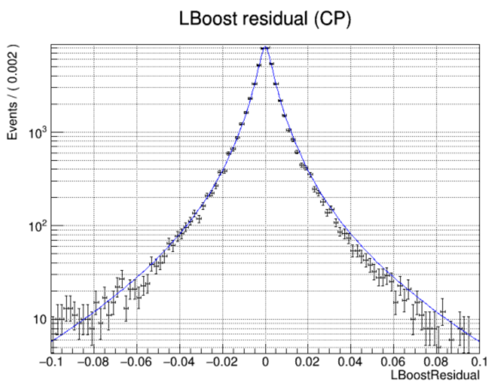
5.1.2 Tag-side resolution function

For the tag-side, the vertexing is done by using *KFit* and no IP constraint used. Since the tag-side vertex reconstruction takes into account the tracks of the charged particles which may not be directly from the tag-side B^0 decay, the vertex position could be affected by both detector effects like the CP-side and the degradation of the secondary vertex. This effect is demonstrated in Figure 5-3 where the tag-side vertex fit will give an estimated position of the vertex around the range of orange meshed area while the actual position of the primary vertex of tag-side B meson should be in the area of the blue circle.

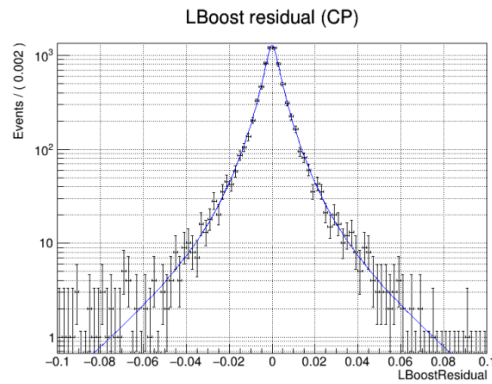
To the contrary, if all tracks that are used for tag-side vertexing are primary tracks, the resolution will only be affected by the detector effects. The vertex position difference from the MC truth can be written as Equation 5.10, where δz_{tag}^{det} and δz_{tag}^{np} are the position shifts caused by the detector effects and the vertex fit using non-primary tracks, respectively. Therefore, the effects from both detectors and non-primary tracks contributes to the total resolution on tag-side as shown in Equation 5.11.

$$\begin{aligned} z_{tag} - z_{tag}^{'} =& (z_{tag}^{'} + \delta z_{tag}^{det} + \delta z_{tag}^{np}) - z_{tag}^{'} \\ =& \delta z_{tag}^{det} + \delta z_{tag}^{np} \end{aligned} \tag{5.10}$$

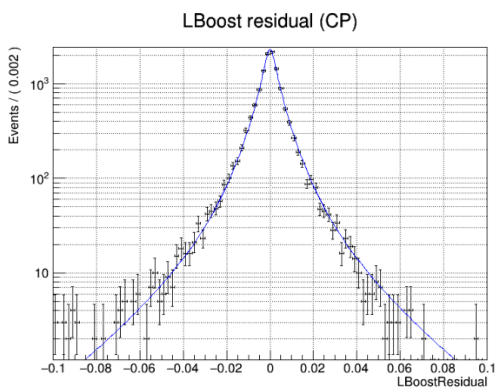
a) $0 < \chi^2/N < 10$



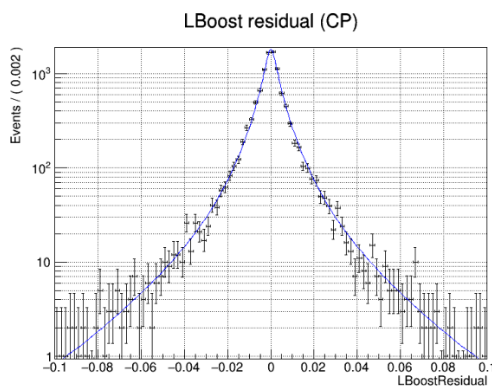
b) $0 < \chi^2/N < 0.4$



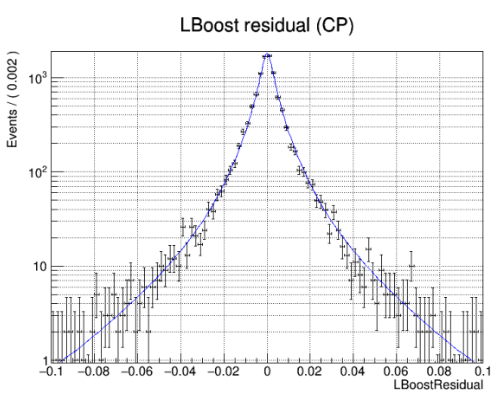
c) $0.4 < \chi^2/N < 0.8$



d) $0.8 < \chi^2/N < 1.2$



e) $1.2 < \chi^2/N < 1.6$



f) $1.6 < \chi^2/N < 10$

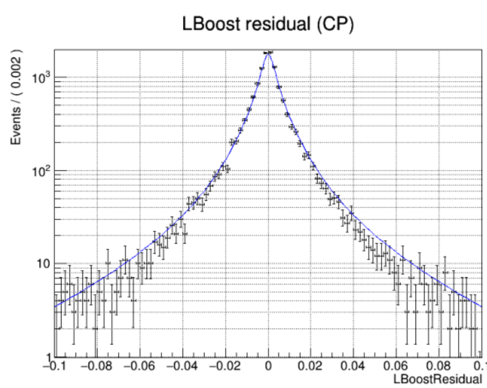


Figure 5-2: The z position residual of B^0 vertices on CP-side, which is dependent on the χ^2/N. The first plot is the fit in the full range and the rest of the plots are the fit in each sliced range of χ^2/N. The events are the true candidates taken from the *signal MC*.

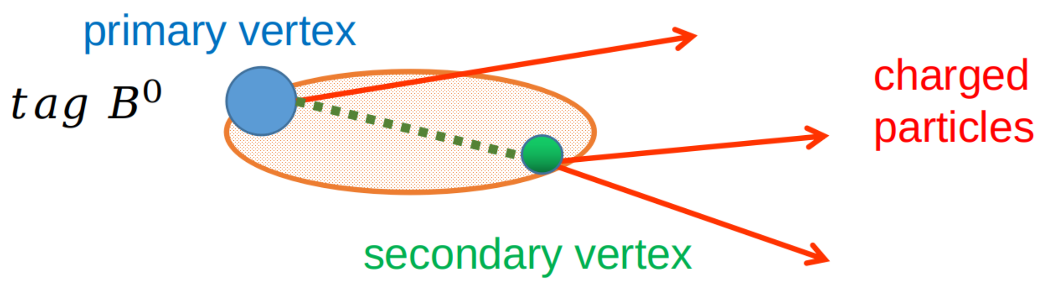


Figure 5-3: The illustration of the tag-side B^0 vertex resolution affected by the non-primary tracks from the secondary vertex. The red arrows present the charged particles used for the tag-side vertex reconstruction, where one is from primary vertex of B^0 and two are from the secondary vertex.

$$R_{tag}(z_{tag} - z'_{tag}) = R^{tag}_{det}(\delta z^{det}_{tag}) \otimes R^{tag}_{np}(\delta z^{np}_{tag}) \tag{5.11}$$

Similarly to the CP-side resolution function, detector effects are presented as

$$R^{tag}_{det}(\delta z^{det}_{tag}) = (1 - f^{tail}_{tag})G(0, s^{main}_{tag} \cdot \sigma_{z_{tag}}) + f^{tail}_{tag} G(0, s^{tail}_{tag} \cdot \sigma_{z_{tag}}) \ , \tag{5.12}$$

where the main and tail Gaussian functions have the same mean value at zero, but the standard deviation is scaled by χ^2_{tag}/N on the tag-side:

$$s^{main/tail}_{tag} = s^{main/tail}_0 + s^{main/tail}_1 \cdot \chi^2_{tag}/N \ . \tag{5.13}$$

R^{tag}_{det} can be obtained with MC samples of which tag-side tracks are all from a primary vertex.

The fit model of R^{tag}_{np} is shown in Equation 5.14 which consists of three functions, including one Dirac δ function and two single-side exponential functions E_p and E_n. The $E_p(x, y_p) = (1/y_p)e^{-x/y_p}$ when $x > 0$ and the $E_n(x, y_n) = (1/y_n)e^{x/y_n}$ when $x < 0$. The exponential factors in both positive and negative components are scaled by the tag-side vertex uncertainty $\sigma_{z_{tag}}$.

$$R^{tag}_{np}(\delta z^{np}_{tag}) = f_\delta \delta(\delta z^{np}_{tag}) + (1 - f_\delta)[f_p E_p(\delta z^{np}_{tag}, \tau_p \cdot \sigma_{z_{tag}}) + (1 - f_p)E_n(\delta z^{np}_{tag}, \tau_n \cdot \sigma_{z_{tag}})] \tag{5.14}$$

Since tag-side has no dependence on how CP-side is reconstructed, the resolution functions on tag-side are almost mode-independent. Thus these parameters are obtained by fitting to the control sample reconstructed from the MC. The control sample consists of multiple exclusive $D^{(*)}$ hadronic decays, of which the details are summarized in Appendix B. The control samples are also used in the analysis of $B^0 \to J/\psi K_S^0$ study and the values of the parameters are referenced from the $B^0 \to J/\psi K_S^0$ study [43], too. The fit plots for tag-side resolution functions are shown in Figure 5-4 and 5-5. The parameters obtained from the fit are listed in Table 5.2 and Table 5.3.

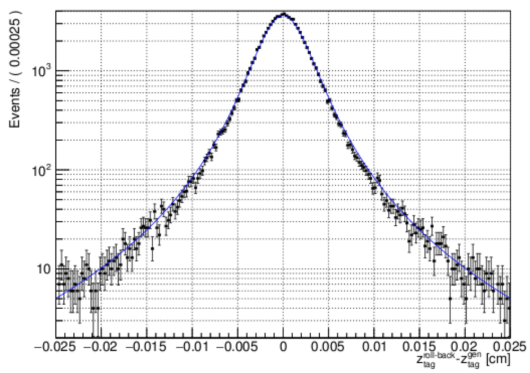


Figure 5-4: R_{det}^{tag} fit

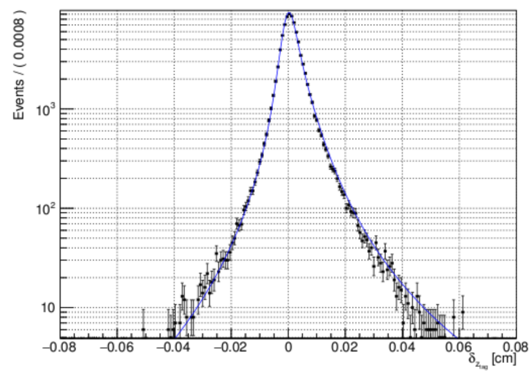


Figure 5-5: R_{np}^{tag} fit

Table 5.2: Parameters in R_{det}^{tag}

f_{tag}^{tail}	0.0523 ± 0.0025
s_0^{main}	1.1446 ± 0.0061
s_1^{main}	0.0443 ± 0.0022
s_0^{tail}	3.448 ± 0.090
s_1^{tail}	0.267 ± 0.028

Table 5.3: Parameters in R_{np}^{tag}

f_δ	0.63 ± 0.05
f_p	0.83 ± 0.01
τ_n	2.914 ± 0.076
τ_p	2.485 ± 0.027

The boost direction is not exactly same in each event, while we use $\Delta t_i = \Delta z/\beta\gamma c$ as an universal relation to calculate the decay time difference for all the events, assuming the boosted direction of the CMS frame to the laboratory frame is always precisely along the z-direction for each event. It will cause a small additional contribution to the resolution of Δt_i, called kinematic effects. This effect could be reduced mainly by two approaches. First, instead of using the vertex position difference along the z-axis, we use the the relative distance between two vertices along the boosting direction calculated for each event during the vertex reconstruction. The second approach is to still use vertex position difference along the z-axis but introduce another resolution function called R_k into Equation 5.5 for the kinematic correction of both signal and background events [44]. In the current stage of the Belle II, the dedicated study of R_k is not ready because it requires a good understanding to the beam conditions and alignment in the event basis. Thus, the first approach is used in this thesis.

5.1.3 Background events Δt distribution

The R_{bkg} model is approximately uncorrelated to vertex reconstruction method. Because the background component mainly comes from continuum events passing the selection, it is reasonable to model the vertex resolution of the background events by a Gaussian-like function. In this case, a double-Gaussian function with the standard deviations scaled by the measured uncertainties from both CP- and tag-sides is used as shown in Equation 5.15.

$$R_{bkg} = (1 - f_{tail}^{bkg})G(\Delta t_i, \sigma_{main}^{bkg}\sqrt{\sigma_{z_{cp}}^2 + \sigma_{z_{tag}}^2}) + f_{tail}^{bkg}G(\Delta t_i, \sigma_{tail}^{bkg}\sqrt{\sigma_{z_{cp}}^2 + \sigma_{z_{tag}}^2}) \quad (5.15)$$

where the $\sigma_{z_{cp}}$ and $\sigma_{z_{tag}}$ are the measured uncertainties of the vertex position in CP- and tag-sides. The σ_{main}^{bkg} and σ_{tail}^{bkg} are the non-unit scaling parameters which control the standard deviations of the main and tail Gaussian functions.

The shape of $\mathcal{P}_{bkg} \otimes R_{bkg}$ can be determined by fitting to the side-band data. The sideband region is defined as $M_{bc} < 5.26$ GeV, according to Section 4.6. The requirement of ΔE is not used in the current definition of the sideband in order to increase the number of events, which will be added in future when more data is collected. Figure 5-6 shows the fit result using 60 sideband data. There are totally seven floating parameters to be determined which are all listed in Table 5.4.

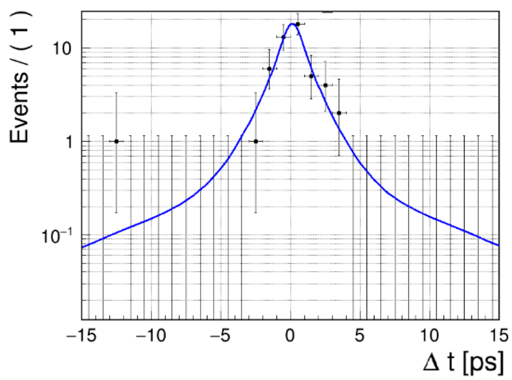


Figure 5-6: $\mathcal{P}_{bkg} \otimes R_{bkg}$ fit using 60 sideband events of the experiment data at $M_{bc} <$ 5.26 GeV.

μ_{δ}^{bkg}	0.13 ± 0.19
μ_{l}^{bkg}	0.16 ± 0.50
τ_{bkg}	1.05 ± 0.44
f_{δ}^{bkg}	0.59 ± 0.26
f_{tail}^{bkg}	0.04 ± 0.04
σ_{main}^{bkg}	1.43 ± 0.39
σ_{tail}^{bkg}	28.1 ± 8.8

Table 5.4: Parameters in Background Δt distribution.

5.2 Flavor Tagging

In order to determine the flavor of tag-side B^0, a flavor tagging algorithm has been developed. The flavor tagging uses information of charged particles including $\mu^{\pm}$, $\pi^{\pm}$,$K^{\pm}$ and Λ. The same charged particles could be originated from the different mother particles in the tag-side decay, so that their importance to the determination of the flavor of the B^0 meson can be varied. For instance, a lepton from decay of $\bar{B}^0 \to D^{*+} l^- \bar{\nu}_l$ is directly related to the flavor of the tag-side meson B, where the negative charge corresponds to a $\bar{B}^0$, that is, $q = -1$. While a lepton from a flavor-definite decay like $B^0 \to J/\psi(\to l^+ l^-) K_S^0$ on the tag-side is much less useful because the both flavor of B could produce such a decay. Thus, these charged particles are first combined into a set of total 13 categories, as demonstrated in the Figure 5-7. In this step, one particle is likely to be assigned into more than one categories, such as a K^- from $D^0 \to K^- \pi^+$, which belongs to both "Kaon" and "KaonPion" pair categories.

A event-level classifier using FastBDT algorithm is implemented to identifier the likelihood of a particle in each category. Then, for each event, the responses of the event-level FastBDT classifier of the 13 categories are combined to form an combiner-level classifier to identify the flavor of the tag-side B meson. For example, the decay $\bar{B}^0 \to D^{*+} l^- \bar{\nu}_l$, as shown in the right side of the Figure 5-7, can give high likelihood outputs in the event-level classifier in categories of "Electron", "KinLepton", "Kaon", "KaonPion", "SlowPion", "MaximumP" and "Fast-Slow-Correlation". The rest of the categories will have relatively low outputs. With combiner-level classification, these 13 outputs could give an output of this decay being a $\bar{B}^0$, so that the CP-side B is much likely to be a B^0.

Categories	Targets
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
KinLepton	ℓ^-
Intermediate KinLepton	ℓ^+
Kaon	K^-
KaonPion	K^-, π^+
SlowPion	π^+
FastHadron	π^-, K^-
MaximumP	ℓ^-, π^-
Fast-Slow-Correlation	ℓ^-, π^+
Lambda	Λ
Total= 13	

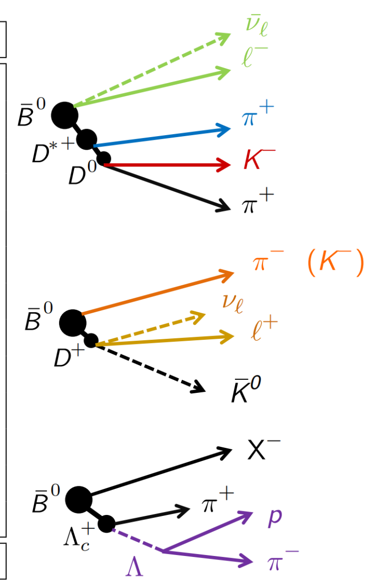


Figure 5-7: Particles and their categories used in flavor tagging algorithm [45]. The right side shows a few typical decay processes in the tag-side which are flavor definite and easy to be tagged. The color of the particles are corresponding to the possible categories they belong to in the particle-level classification.

This method is called category-based method and used in this thesis. To minimize impact of the reconstruction performance on CP-side, MC sample of $B^0 \to \nu\nu$ is used as the training sample where the final state in CP-side is completely invisible. Considering the limited power of flavor tagging accuracy, there is a certain fraction of events that are wrongly tagged. Thus, the flavor tagging efficiency ϵ and wrong tag fraction w are introduced. Taking into account of the performance of the flavor tagging, the Equation 5.1 turns into

$$\mathcal{P}_{sig}^{obs}(\Delta t, q, \epsilon, w) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \epsilon \left\{1 - q \cdot \Delta w + q \cdot (1 - 2w) \cdot \left[\mathcal{S} \cdot \sin(\Delta M_d \Delta t) + \mathcal{A} \cdot \cos(\Delta M_d \Delta t)\right]\right\} , \tag{5.16}$$

where ϵ is the flavor tagging efficiency, w is the wrong tag fraction and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ is the difference of the wrong tag fraction between B^0 and $\bar{B}^0$. Compared to the original, the term with $\mathcal{S}$ and $\mathcal{A}$ is scaled by factor $r \equiv |1 - 2w|$, called the dilution factor. The statistical uncertainty of $\mathcal{S}$ now becomes dependent to the tagging efficiency ϵ and wrong tag fraction w.

The uncertainty of w is much larger than ϵ so that w gives an important source of systematic uncertainty. The validation of flavor tagger using flavor specific tagging control samples and Belle II data in 2019 is summarized in Ref. [45]. The w for each single event is defined as a probability of being wrongly flavor tagged which can be presented by the average wrong tag fraction in a binned interval of the dilution factor. The binned values of dilution factor r is defined for the calculation of w as $[0.0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.875, 1.0]$, also named as r-bin. For all events that have been successful tagged, they are projected into histogram of r-bin, and w is calculated in each bin by the fraction of events of which the sign of the flavor tagger output is opposite to its true flavor. To easily use the information from the flavor tagger, the actual output of the flavor tagger is transformed into $q \cdot r$ where q is the true flavor. In this case, an event with a small w could give $r \simeq 1$ so that the flavor tagger output is close to its original flavor q, while $w > 0.5$ can flip the sign of the q indicating that this event is wrongly tagged with the opposite flavor. The distribution of $q \cdot r$ is shown in Figure 5-8 using *signal MC*.

The distribution of w and Δw in each r-bin are shown in the top left and bottom left of the Figure 5-9, respectively. The Δw is mostly within 3%, therefore the contribution from Δw is treated as zero in Equation 5.16 in this analysis. The distribution of ϵ and $\mu = \epsilon_{B^0} - \epsilon_{\bar{B}^0}$ in each r-bin are shown in the top right and bottom right of Figure 5-9, respectively. It is clear that the μ is mostly within a range of $\pm 2\%$, also treated as zero in this thesis. We sum the ϵ for B^0 and $\bar{B}^0$ separately in all r-bins and take the average value to be used as the efficiency in Equation 5.16 which is $(99.7 \pm 0.2)\%$, treated as 1 in this thesis. The performance of flavor tagger in $B^0 \to K_S^0 K_S^0 K_S^0$ and the ones from Ref. [45] are agreed, therefore the values of w from tagging control samples is used in this thesis.

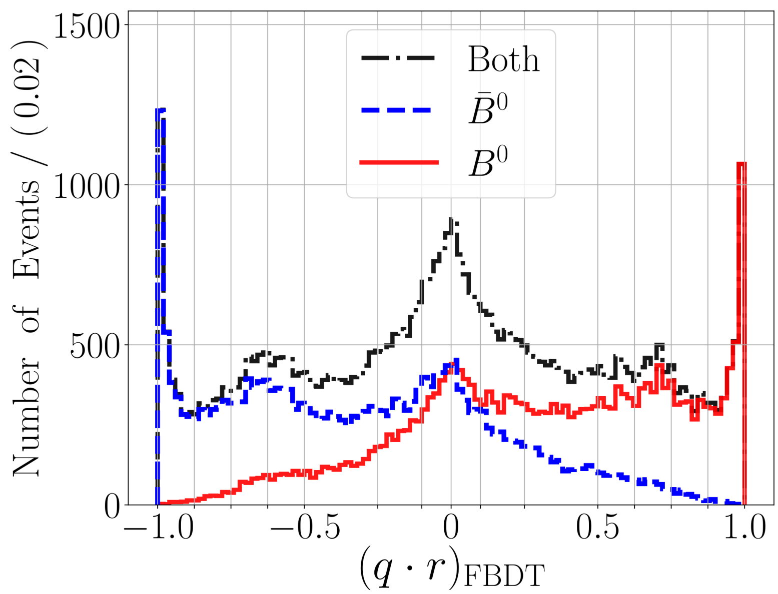


Figure 5-8: The distribution of flavor tagger output $(q \cdot r)$ for both tag-side of B^0 and $\bar{B}^0$ using *signal MC*.

5.3 *CP* Fitter

The parameters that are needed for measuring $\mathcal{S}$ and $\mathcal{A}$ are studied and obtainable. Based on the observed Δt distribution from selected events, Equation 5.5 can be fitted using the unbinned maximum likelihood method where Δt, signal fraction f^{sig}, and the flavor q are taken as observables. The signal fraction for each event is calculated based on the M_{bc} and ΔE values using the 2D fit model. In the meantime the vertexing error $\sigma_{z_{cp}}$, $\sigma_{z_{tag}}$ and χ^2/N are used as event-by-event conditional variables that are accessed during the fitting. For Belle II, a new *CP* fitter has been developed based on Python and RooFit, which is easy to use and integrated well with BASF2. The fitter requires a configuration file which contains all the parameter configurations including their ranges, initial values, floating states and uncertainties to perform the

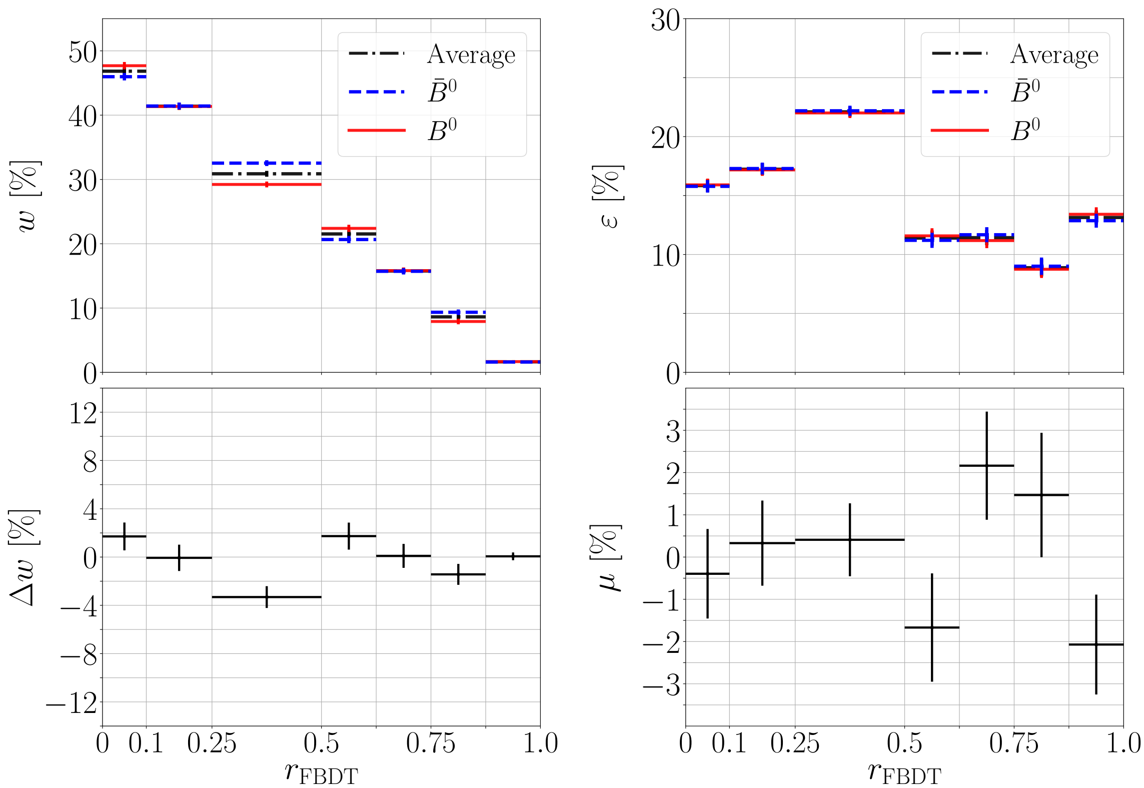


Figure 5-9: The flavor tagging efficiency, wrong tagging fraction, and their differences between different flavors sorted in each r-bin using *signal MC*.

Table 5.5: The selection criteria for events that are used for CP parameters fit.

Observables	Selections
Δt	$-70 < \Delta t < 70$ ps
CP-side χ^2/N	$0 < (\chi^2/N)_{cp} < 8$
tag-side χ^2/N	$0 < (\chi^2/N)_{tag} < 50$
$\sigma_{z_{tag}}$	$\sigma_{z_{tag}} < 0.1$ cm
signal region	$5.27 < M_{bc} < 5.29$ GeV and $\lvert\Delta E\rvert < 0.1$ GeV

CP parameter measurement.

5.4 Blind analysis and fit

As a required procedure to make sure the CP parameters are measured without any bias due to the preconceived results, a blind analysis procedure is conducted before the fit is actually performed using the experimental data. The blind fit procedure includes the CP fit on *signal MC* and *generic MC*, with different numbers of events used. To check the reliability of fit results from the CP fitter, a linearity test and toy MC studies are also performed.

5.4.1 CP fit on MC samples

We first perform the CP fit on events in *signal MC* and *generic MC*. The *signal MC* and *generic MC* are generated with the phase-space model which contains zero CP violation ($\mathcal{S} = \mathcal{A} = 0$). In addition to the event selection criteria in the Table 4.2, the cuts in Table 5.5 are also applied for rejecting the events outside of the signal region and with very poor vertex reconstruction quality. Based on Section 4.6, the number of events inside the signal region is 30, which 26 events are kept after applying cuts in Table 5.5.

We have 8873 events passing the selections from 10000 *signal MC*. For 1 ab^{-1} *generic MC*, 373 events are selected to fit CP parameters. To mimic the number of events expected in data sample, 26 events randomly sampled from *generic MC* are used to perform the fit as well, which the sampling rounds take place for 100 times to avoid the large statistical fluctuations and the fit results are averaged. The CP plots

Table 5.6: The *CP* fit results using *signal MC* and *generic MC* with only statistical uncertainties. The first two rows are the results from single time fit and the third is the average result from totally 100 times fit.

Sample (events)	$\mathcal{S}$	$\mathcal{A}$
signal MC (8873)	$\sin(2\phi_1) = 0.00 \pm 0.04$	$\mathcal{A} = -0.01 \pm 0.02$
generic MC (373)	$\sin(2\phi_1) = 0.00 \pm 0.21$	$\mathcal{A} = -0.05 \pm 0.07$
generic MC (30)	$\sin(2\phi_1) = 0.20 \pm 0.85$	$\mathcal{A} = -0.06 \pm 0.30$

are shown in Figure 5-10. The fit results of $\mathcal{S}$ and $\mathcal{A}$ are summarized in Table 5.6.

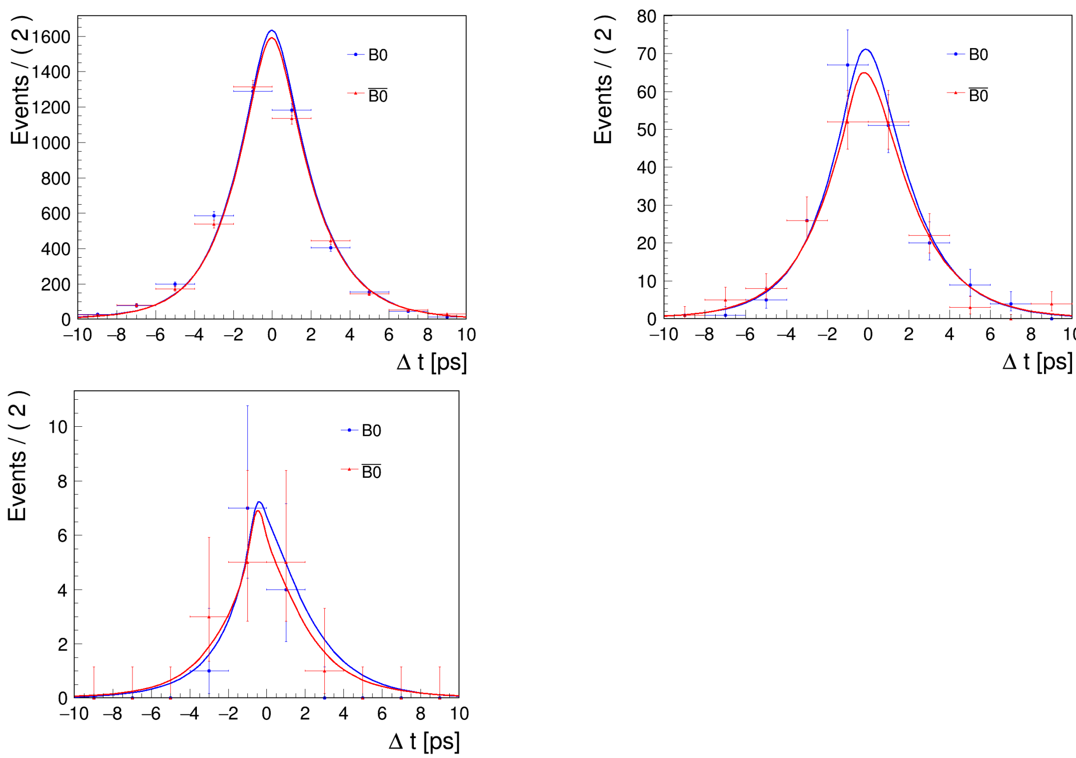


Figure 5-10: The distribution of Δt with the *CP* fit. The top left is for *signal MC* and the top right is for *generic MC*. The bottom left is the fit of the first sampling round of 30 events of *generic MC*.

The fit results are consistent with the expectation of non-*CP* violation, and the statistical uncertainties are basically proportional to the $\sim 1/\sqrt{N}$, where N is the number of events used for *CP* fit. To test a fit on non-zero *CP* violating MC, the fit on $B^0 \to J/\psi K_S^0$ *signal MC* is also done, of which the details of the event selection and the fit model determination can be found in Ref. [43]. The fit result over 10000 events is shown in Figure 5-11, which gives $\sin(2\phi_1) = 0.70 \pm 0.05$ and $\mathcal{A} = -0.01 \pm 0.02$.

The results agree with the inputs of $\mathcal{S} = sin(2\phi_1) = 0.670$ and $\mathcal{A} = 0$.

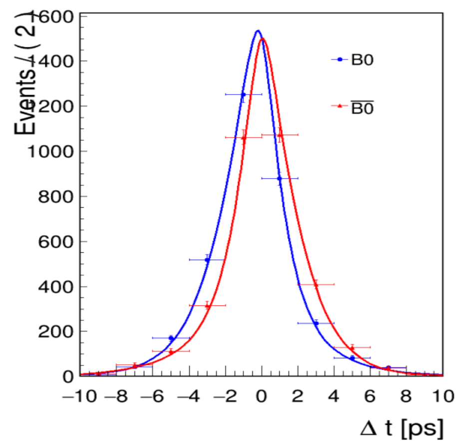


Figure 5-11: *CP* fit over 10000 $B^0 \to J/\psi K_S^0$ *signal MC*.

5.4.2 Linearity Test

To validate the *CP* fit linearity, a set of toy MC samples is generated where the χ^2 from vertex fit and the vertex errors on *CP*- and tag-side are sampled from *signal MC* as the conditional variables, while the number of events used in each toy MC sample is kept as 26 that is the same as data. The resolution function parameters are kept the same as *CP* fit on *generic MC*. The input $\mathcal{A}$ is set to zero while the input value of $\sin(2\phi_1)$ is running from 0.1 to 0.9. For each value of input $\sin(2\phi_1)$, 200 toy MC samples are generated. The center value and the error for $\sin(2\phi_1)$ and $\mathcal{A}$ are determined by the mean and its uncertainty for each input from the fit using Gaussian function. The dependence between input and output are shown in Figure 5-12 where a linear function is used to fit the dependence with the uncertainties. We fix $\sin(2\phi_1)$ at zero while floating $\mathcal{A}$ from 0.1 to 0.9 to repeat the above linearity test. The dependence between input and output are shown in Figure 5-13, where a linear function is used to fit the dependence. From the linear function fit results, the linearity tests show an agreement with the input *CP* parameters and the output values.

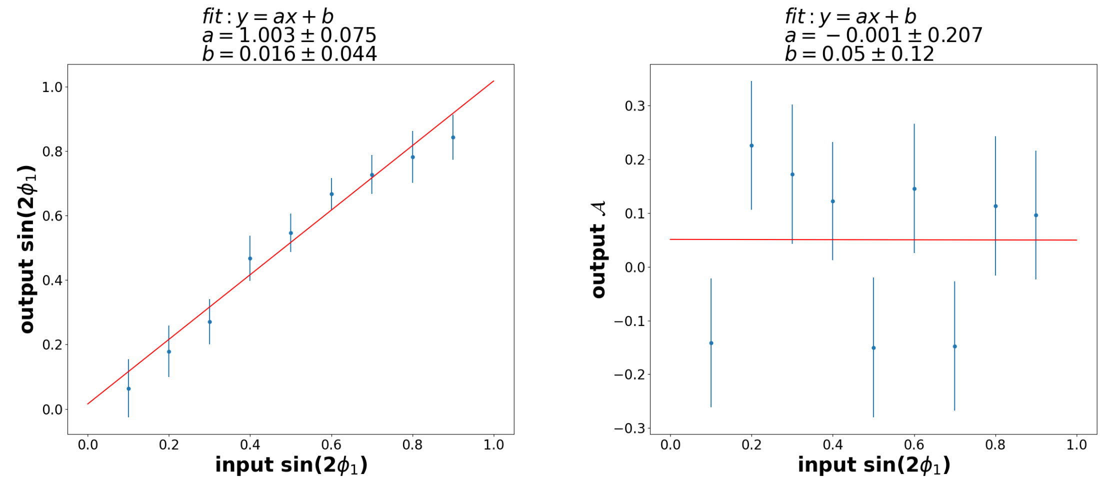


Figure 5-12: Linearity test of CP fit with a fixed input $\mathcal{A} = 0$ and floating $\sin(2\phi_1)$. The left is for the output $\sin(2\phi_1)$ and the right is for the output $\mathcal{A}$.

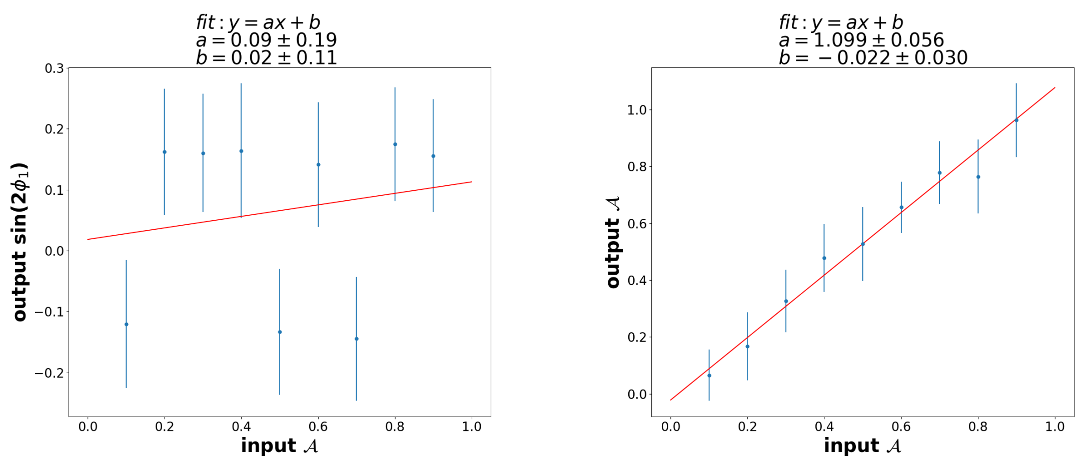


Figure 5-13: Linearity test of CP fit with a fixed input $\sin(2\phi_1) = 0$ and floating $\mathcal{A}$. The left is for the output $\sin(2\phi_1)$ and the right is for the output $\mathcal{A}$.

5.4.3 Toy MC Fit Pull

In order to check the fit bias and the size of fitted errors with input-output method, a set of 1000 toy MC samples has been created. Each toy MC sample contains 26 events to keep the same number of events as the real data. The χ^2 from vertex fit, the number of events N and vertex errors on CP- and tag-sides are sampled from the distribution of data. The input $\sin(2\phi_1)$ and $\mathcal{A}$ are both zero for the toy sets. We expect to use the standard normal distribution to fit the pull of $\sin(2\phi_1)$ and $\mathcal{A}$.

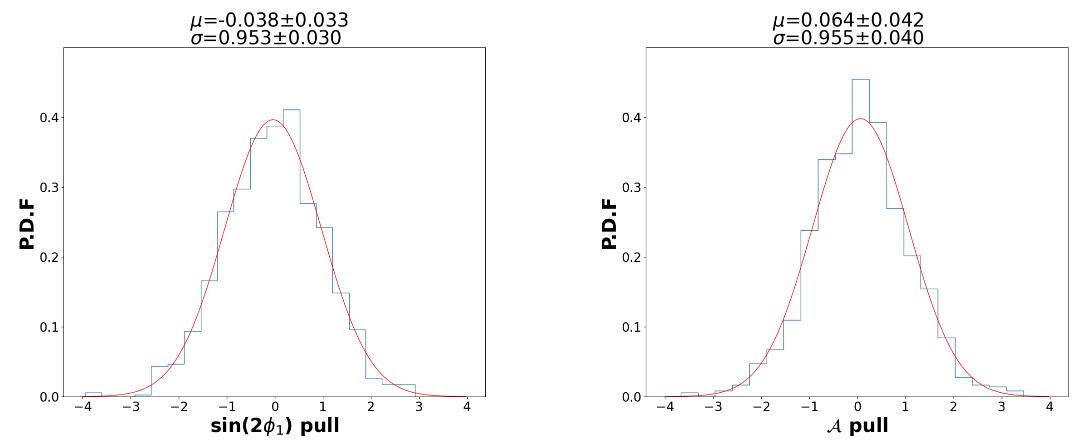


Figure 5-14: The pull distributions of $\sin(2\phi_1)$ and $\mathcal{A}$ fitted by the standard normal distribution with μ as the mean and σ as the standard deviation.

As shown in Figure 5-14, the fit results using the standard normal distribution shows a good recovery of input $\sin(2\phi_1)$ and $\mathcal{A}$ with no clear bias is spotted. The fitted errors look to be a little bit overestimated around 5% but they are still consistent with 1.5σ so that any correction is not applied.

5.4.4 Lifetime fit

Before looking at CP parameters in data, we need to check if the lifetime of B meson are consistent by configuring the CP fitter to fit it as a floating parameter. To test lifetime fit, first we use 10000 *signal MC* events which is generated by $\tau_{B^0} = 1.520$ from PDG value. The $\sin(2\phi_1)$ and $\mathcal{A}$ are fixed at zero during the fit, for which the generator level CP violation is zero. This is equivalent fit to Equation 5.17.

$$\mathcal{P}(\Delta t, \tau_{B^0}) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \tag{5.17}$$

The fit result on *signal MC* is 1.537 ± 0.024 ps which is consistent with the input. We perform the lifetime fit on data in signal region, and the CP parameters are fixed based on PDG values to: $\sin(2\phi_1) = 0.69$ and $\mathcal{A} = 0$. The fitted lifetime from $B^0 \to K_S^0 K_S^0 K_S^0$ is 1.431 ± 0.382 ps. The result is consistent with PDG value. The distribution of Δt in lifetime fit is shown as Figure 5-15. The B^0 and B^+ lifetime fit using control sample is also performed and summarized in Ref. [43]. The results are consistent with PDG values as input in MC generator.

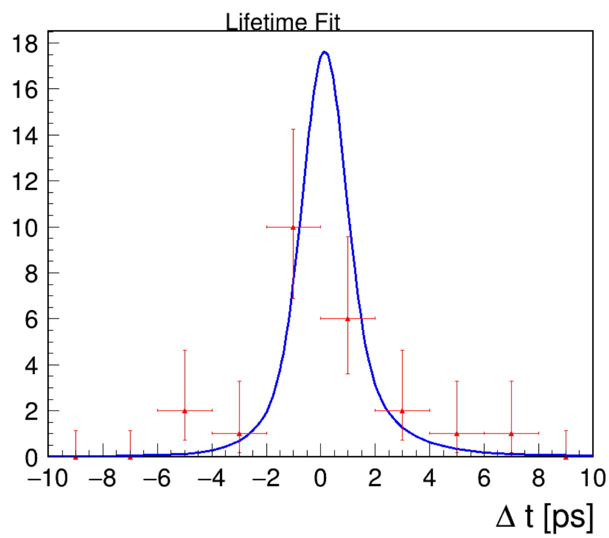


Figure 5-15: Lifetime fit on data

5.5 CP fit on data

After the CP fit procedures are reviewed by Belle II collaboration, the permission of measuring CP parameters using 62.8 fb^{-1} Belle II data is granted. The events number used for the CP fit is 26, and the fit result is shown Figure 5-16.

The results of CP parameters are obtained as

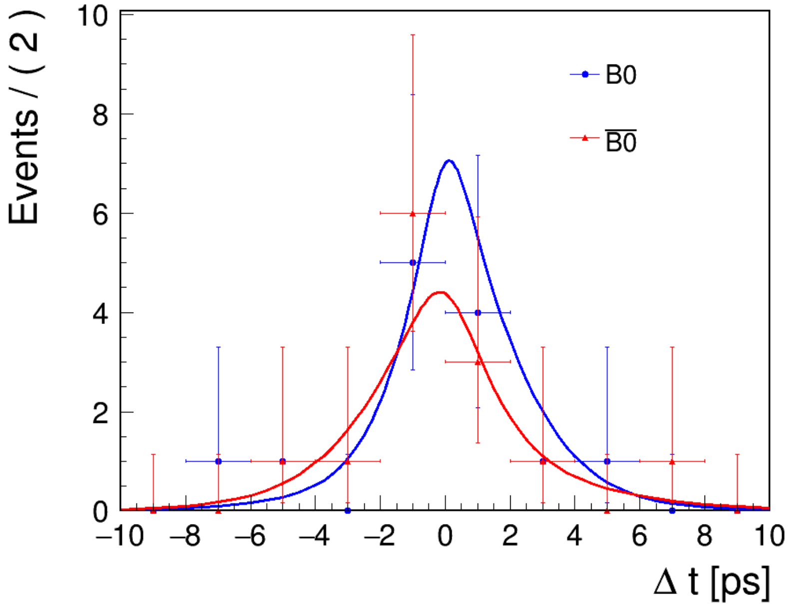


Figure 5-16: The CP fit result from data. B^0 and $\bar{B}^0$ stand for the flavor of the tag-side B meson.

$$
\begin{aligned}
sin(2\phi_1) &= 0.82 \pm 0.85 \text{ (stat) } , \\
\mathcal{A} &= -0.21 \pm 0.28 \text{ (stat) } .
\end{aligned}
\tag{5.18}
$$

5.6 Systematic Uncertainty

The systematic uncertainty that affects the CP fit results may come from many aspects of the measurement procedures. We first estimate the contribution from each source, then the total systematic uncertainty is calculated by adding-in-quadrature using each contribution. The possible systematic uncertainty from each source is summarized in Table 5.7.

The contributions from signal and background Δt shapes, wrong tag fraction w, physics parameters, and signal fraction directly come from the corresponding parameters in the CP fit model. To be specific, in each above source, if the parameters are defined with MC study, we float the value by $\pm 2\sigma$, and if the parameters are defined by data, we float the value by $\pm 1\sigma$, where σ is the uncertainty of the parameters.

The contribution from every parameter in each source is therefore marked as $\pm\delta\mathcal{S}$ and $\pm\delta\mathcal{A}$, where the signs present the difference caused by positively or negatively floating values. In the case, from Table 5.8 to Table 5.12, the total $\delta\mathcal{S}$ and $\delta\mathcal{A}$ are obtained by adding in quadrature using all $\pm\delta\mathcal{S}$ and $\pm\delta\mathcal{A}$, respectively. From Table 5.13 to Table 5.15, the total $\delta\mathcal{S}$ and $\delta\mathcal{A}$ are calculated directly by adding in quadrature of the contribution from every source.

Table 5.7: The contributions of each source and the total systematic uncertainty.

Sources	$\delta\mathcal{S}$	$\delta\mathcal{A}$
signal Δt shapes	0.034	0.0093
background Δt shapes	0.039	0.037
wrong tag fraction	0.0037	0.0038
physics parameters	0.0069	0.0013
signal fraction	0.044	0.034
fit bias	0.0098	0.0057
KsFinder impact on data	0.0036	0.0036
vertex reconstruction	0.019	0.021
tag-side interference	0.001	0.008
Total	0.072	0.056

On the other hand, for the contributions from fit bias, *KsFinder*, and vertex reconstruction, they are not directly from the parameters of the fit model, but the number of events in fit bias test, *KsFinder* correction factor, and vertex reconstruction options, respectively. We modify their values or ranges to repeat the *CP* fit to obtain their contributions as $\delta\mathcal{S}$ and $\delta\mathcal{A}$ correspondingly.

Last but not least, tag-side interference could also contribute to the total systematic uncertainty, which is caused by the interference between CKM-favored or CKM-suppressed tree-level decays. There is no estimation on systematic uncertainty from the tag-side interference in the Belle II at the current stage. Since this effect is less sensitive to the detector performance, we quote the value from the Belle $\delta\mathcal{S} \sim 0.001$ and $\delta\mathcal{A} \sim 0.008$ [46], which is a small contribution compared to the current statistical uncertainty.

In the following sections, the estimations of systematic uncertainty from each source in Table 5.7 are discussed in detail.

5.6.1 Systematic uncertainty from signal Δt shapes

The signal Δt shapes are dependent on the resolution functions of CP- and tag-sides, therefore the contributions come from the parameters used in the corresponding models, which are determined by using MC samples. The contributions from each parameters are summarized in Table 5.8, which are obtained by floating 2σ for each of them.

Table 5.8: Systematic uncertainty from signal Δt shapes

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
f_{cp}^{tail}	-0.000096	-0.000057	0.000014	0.000056
s_{0CP}^{main}	0.0054	0.0013	-0.0057	-0.0014
s_{1CP}^{main}	0.020	-0.00090	-0.020	0.00063
s_{0CP}^{tail}	-0.0032	-0.0016	0.0033	0.0016
s_{1CP}^{tail}	-0.0020	-0.000063	0.0020	0.000048
f_{tag}^{tail}	0.0031	-0.0013	-0.0031	0.0013
s_{0tag}^{main}	0.0020	-0.0014	-0.0020	0.0014
s_{1tag}^{main}	0.0051	-0.00084	-0.0050	0.00083
s_{0tag}^{tail}	-0.00014	-0.00039	0.00010	0.00044
s_{1tag}^{tail}	0.00010	0.000027	-0.00047	0.00013
f_δ	-0.0072	-0.00055	0.0072	0.00059
f_p	0.0030	0.0043	-0.0031	-0.0043
τ_n	-0.0010	-0.0028	0.00094	0.0029
τ_p	0.0045	0.0025	-0.0046	-0.0025
Total	$\delta\mathcal{S} = 0.034$		$\delta\mathcal{A} = 0.0093$	

5.6.2 Systematic uncertainty from background Δt shapes

The background Δt shapes are affected by the resolution model for background events which are determined by using the sideband data $M_{bc} < 5.26$ GeV. In this case, we float 1σ for each parameter and the results are summarized in Table 5.9.

Table 5.9: Systematic uncertainty from background Δt shapes

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
μ_δ^{bkg}	-0.014	-0.017	0.0068	0.0065
μ_l^{bkg}	-0.0028	-0.013	0.0038	0.013
τ_{bkg}	0.0014	0.0017	-0.0042	0.000085
f_δ^{bkg}	-0.011	0.0014	0.011	-0.0014
f_{tail}^{bkg}	-0.0027	0.0015	0.0025	-0.0014
σ_{main}^{bkg}	0.021	0.022	-0.024	-0.016
σ_{tail}^{bkg}	-0.00028	-0.00016	0.00018	0.00014
Total	$\delta\mathcal{S} = 0.039$		$\delta\mathcal{A} = 0.037$	

5.6.3 Systematic uncertainty from wrong tag fraction

The wrong tag fraction for each event during the CP fit is obtained by calculating the fraction of wrongly tagged events in the corresponding r-bin. As discussed in section 5.2, there are totally seven r-bins, marked as $w_1 \sim w_7$. We use the values of w obtained by the tagging control sample study [45] and float 1σ to get the contributions. The results are shown in Table 5.10.

Table 5.10: Systematic uncertainty from wrong tagging fraction

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
w_1	-0.0019	0.0019	0.0019	-0.0020
w_2	-0.0016	0.0011	0.0016	-0.0012
w_3	-0.00049	0.0013	0.00047	-0.0013
w_4	0.00066	0.00026	-0.00065	-0.00026
w_5	-0.00012	0.00020	0.00012	-0.00020
w_6	0.000095	0.000054	0.000096	-0.000045
w_7	0.00019	-0.00040	-0.00019	0.00040
Total	$\delta\mathcal{S} = 0.0037$		$\delta\mathcal{A} = 0.0038$	

5.6.4 Systematic uncertainty from physics parameters.

The physics parameters that affect the CP fit are the lifetime of B^0 (τ_{B^0}) and the mass difference of the two mass eigenstates of neutral B meson (ΔM_d), which are set as constants in the CP fit based on the PDG value. Thus, the contributions from Δm_d and τ_{B^0} are included by floating 1σ using the PDG average errors. The results are summarized in Table 5.11.

Table 5.11: Systematic uncertainty from physics parameters

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
Δm_d	-0.0018	-0.00069	0.0018	0.00070
τ_{B^0}	-0.0046	-0.00055	0.0046	0.00056
Total	$\delta\mathcal{S} = 0.0069$		$\delta\mathcal{A} = 0.0013$	

5.6.5 Systematic uncertainty from signal fraction.

The signal fraction is a event-by-event parameter based on the M_{bc} and ΔE which is determined using 2D fit results from data. Because of this, we take turns to change each parameter in the 2D fit by floating 1σ to get different 2D fit models. Then based on each model, signal fraction parameter for every event is updated and used during the CP fit. The M_{bc} and ΔE models are single and triple Gaussian functions, thus there are 8 parameters for the means and standard deviations, as well as 2 internal fraction coefficients for the triple Gaussian functions of ΔE. Besides the shape parameters in each model, the signal fraction is the floating parameter to be obtained by the 2D fit. We also directly float the signal fraction by 1σ to check its contribution. The results are summarized in Table 5.12.

Table 5.12: Systematic uncertainty from signal fraction

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
μ_{mbc}	0.00082	-0.0039	-0.00080	0.0038
σ_{mbc}	0.00048	0.0084	-0.00063	-0.0087
χ_{argus}	-0.00071	0.0041	0.0014	-0.0058
c_{argus}	-0.0055	0.0014	0.0009	-0.000078
$f^1_{\Delta E}$	0.028	0.021	-0.019	-0.0084
$f^2_{\Delta E}$	0.021	0.018	-0.016	-0.0070
$\mu^1_{\Delta E}$	-0.00044	-0.00015	0.00050	0.000088
$\mu^2_{\Delta E}$	-0.00056	0.0014	0.00059	-0.0014
$\mu^3_{\Delta E}$	-0.0032	-0.00083	0.0034	0.00098
$\sigma^1_{\Delta E}$	-0.00017	-0.00097	0.00021	0.00091
$\sigma^2_{\Delta E}$	-0.0032	0.0030	0.0026	-0.0025
$\sigma^3_{\Delta E}$	-0.0019	-0.0026	0.0025	0.0030
a_{cheb}	0.00095	0.000057	-0.00089	-0.00010
f_{sig}	-0.0046	0.0040	0.0049	-0.0035
Total	$\delta\mathcal{S} = 0.044$		$\delta\mathcal{A} = 0.034$	

5.6.6 Systematic uncertainty from fit bias.

Using maximum likelihood fit might cause bias to the fit results, which can be checked by using a large number of input events. We use 300000 *signal MC* events with zero *CP* violation to evaluate such effect. The fit result is $\mathcal{S} = 0.000127 \pm 0.009817$ and $\mathcal{A} = 0.000265 \pm 0.005702$. The difference between the output of the *CP* fit and input ($\mathcal{S} = 0$) for $\mathcal{S}$ is 0.000127 and the fit error is 0.009817, and we use the larger one among them as the fit bias contribution, as listed in Table 5.13.

Table 5.13: Systematic uncertainty from fit bias

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
fit bias	0.0098	0.0057

5.6.7 Systematic uncertainty from *KsFinder*.

Applying *KsFinder* cut at 0.74 based on MC study may introduce small bias on data due to the different response on the classifier between data and MC. Therefore the systematic uncertainty from applying *KsFinder* is considered. As discussed in section 3.2.7, the upper and lower limit of $\mathcal{R}_{B^0}$ are 1.033 and 0.967 respectively. These two ratios are applied on the data events obtained to repeat the *CP* fit, and the difference of fit results compared to the original values are used as systematic uncertainties, as shown in Table 5.14.

Table 5.14: Systematic uncertainty from *KsFinder*.

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
$\mathcal{R}_{B^0} = 1.033$	0.0025	-0.0025
$\mathcal{R}_{B^0} = 0.967$	-0.0026	0.0026
Total	0.0036	0.0036

5.6.8 Systematic uncertainty from vertex reconstruction.

For vertex reconstruction, the contributions from the options in Table 5.5 are considered. Given the fact that cut values in Table 5.5 are very loose and the number of events from data is very limited, the changing of the these values does not affect events collected from data so that systematic uncertainty can not be reflected correctly. Therefore, 1 ab^{-1} *generic MC* is used with the modified ranges to estimate the potential systematic uncertainty from vertex reconstruction. Besides, due to the absence of IP constraint in vertex fit, the contributions from the IP constraint options as well as the potential bias are not considered. However, this does not mean that the IP constraint options have no effect on the *CP* fit. Since we also do not use any IP constraint options in the study of the vertex resolution functions, the resolution observed in this case will be worse compared to the ones with good IP constraint and tighter vertex selection criteria. Thus, the systematic uncertainties from Δt shapes become larger, already indirectly reflecting the contributions from the missing IP options. Besides, the contributions due to the VXD misalignment and Δz bias are not included because the PXD is not fully installed and the beam misalignment is not well understood in the early stage of the Belle II. The results are listed in Table 5.15, where the zero values appear due to the unchanged events input under the modified ranges. This part of the systematic uncertainty will be checked with more data and *generic MC* in future to testify the possible variation.

Table 5.15: Systematic uncertainty from vertex reconstruction

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
$\sigma_{z_{tag}} < 0.05$ cm	0.0044	-0.0036
$\sigma_{z_{tag}} < 0.15$ cm	0.00	0.00
$\chi^2/N(CP) < 3$	0.018	-0.020
$\chi^2/N(CP) < 13$	0.00	0.00
$\chi^2/N(tag) < 40$	0.00	0.00
$\chi^2/N(tag) < 60$	0.00	0.00
$\|\Delta t\| < 50$ ps	0.0033	-0.00040
$\|\Delta t\| < 90$ ps	0.00	0.00
IP constraint	0.00	0.00
Total	0.019	0.021

In conclusion, the current systematic uncertainty is estimated by considering each possible source, where the total systematic uncertainty of $\delta\mathcal{S}_{syst} \simeq 0.07$ and $\delta\mathcal{A}_{syst} \simeq 0.06$ are obtained. It is clear that the dominated uncertainty at the current stage of the Belle II for time-dependent CP violation measurement in $B^0 \to K_S^0 K_S^0 K_S^0$ is still the statistical error. As a result, the current results are agreed with the previous results from Belle, as well as the SM predictions.

However, this might be changed in future when more data becomes available. In the meantime, some of the systematic uncertainty sources are considered as the reducible contributions while the others remain almost unchanged with increasing mount of data and MC samples. The prospective of the statistical and systematic uncertainties in future are fundamentally important to examine the SM predictions and search for the NP effects in this decay mode. We will discuss about the future statistical and systematic uncertainties of $\mathcal{S}$ in detail in the next chapter.

Chapter 6

Discussions, conclusions and prospects

6.1 Discussions

Based on the results and current status of the analysis, several important topics need to be discussed. In this section, we discuss the statistical and systematic uncertainty improvements in future based on the current results, as well as the importance of *KsFinder* in this analysis.

6.1.1 Statistical uncertainty in future

The precision on $\mathcal{S}$ requires the large luminosity of data as shown in Figure 1-6, which includes both statistical and systematic uncertainties. With 50 ab^{-1} luminosity from the full Belle II data sample in future, the statistical uncertainty is expected to be largely reduced. The CP fit on the MC sample with different amount of events used reflects that the statistical uncertainty is reduced proportionally around factor of $\frac{1}{\sqrt{N}}$, where N is the events used in CP fit. Also,the wrong tag fraction is contributive to the statistical uncertainty. The performance of flavor tagging algorithm could be slightly improved in future. If so, the expected statistics of correctly tagged events could be increased. The reduced statistical uncertainty from this effect is considered to be marginal. Thus, the reduction of statistical uncertainty is assumed to mainly come from the increased data sample with current reconstruction efficiency, which is shown in Figure 6-1. At 50 ab^{-1} luminosity, the statistical uncertainty is estimated to be ~ 0.03.

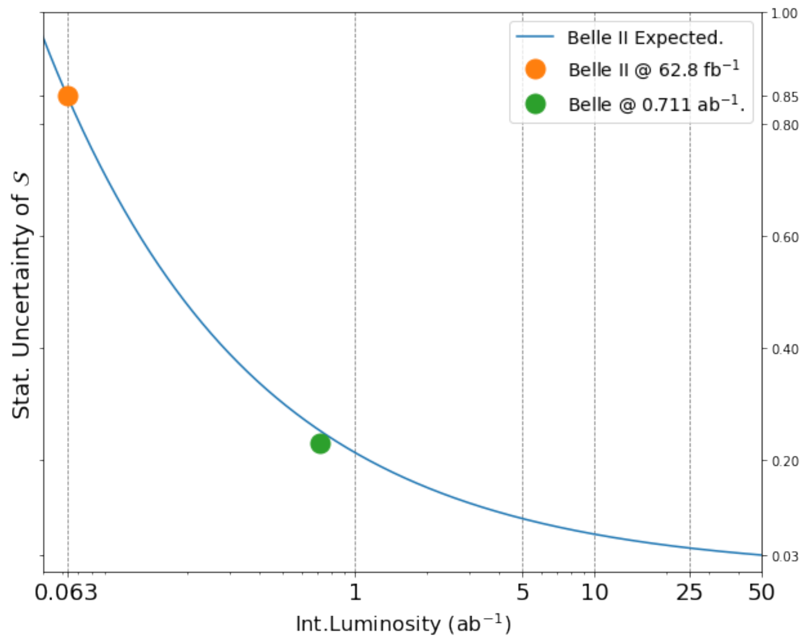


Figure 6-1: Statistical uncertainty of $\mathcal{S}$ extrapolation based on the current result of $B^0 \to K_S^0 K_S^0 K_S^0$ in Belle II, where the orange is the current value and the green is the Belle result at 0.711 ab^{-1} $\Upsilon(4S)$ data.

From Table 4.3, the current B^0 reconstruction efficiency is 34% which could be further optimized mainly by improving K_S^0 reconstruction efficiency. As discussed in Section 3, the reconstruction efficiency and vertexing quality become worse for long-flight K_S^0 particles. This is mainly due to the limitation of CDC-only tracking and the hit filters on the SVD layers. The current Belle II track finding algorithm rises a requirement for SVD hits that at least two or more SVD hits have to be associated from a same track so that they can be used together to form a track. If a K_S^0 decays outside of layer 5 at 10.4 cm, even though the daughter tracks could pass the SVD layer 6, they are much likely to become two CDC-only tracks unless they pass the overlapped regions at the edges of SVD layers.

This effect is shown in Figure 6-2, where *SVD00* type K_S^0 starts to appear at SVD layer 5 instead of layer 6. The requirement of the SVD hits in the current tracking algorithm is needed to suppress the beam background and SVD noise strips that create a large fraction of random single hits. Thus, the actual sensitive volume of SVD is reduced and the K_S^0 reconstruction efficiency is negatively affected. In future, the improvement of our tracking algorithm is expected to remove this requirement while still be able to effectively reduce single-hit background. The fraction of the *SVD00* type K_S^0 will be reduced to provide a better vertexing quality as well.

In general, the expected B^0 signal yield can be improved in future and help to further reduce the statistical uncertainty. From Figure 6-1, the extrapolated statistical uncertainty at 0.711 ab^{-1} is comparable with the Belle result. When the integrated luminosity reaches about 9 ab^{-1}, the statistical uncertainty is reduced to ~ 0.072 which is equivalent to the current systematic uncertainty. At 50 ab^{-1} integrated luminosity, the major contribution will be systematic uncertainty if no improvement is assumed. We take the extrapolated statistical uncertainty of ~ 0.030 at 50 ab^{-1} as a conservative value for estimating the total uncertainty later.

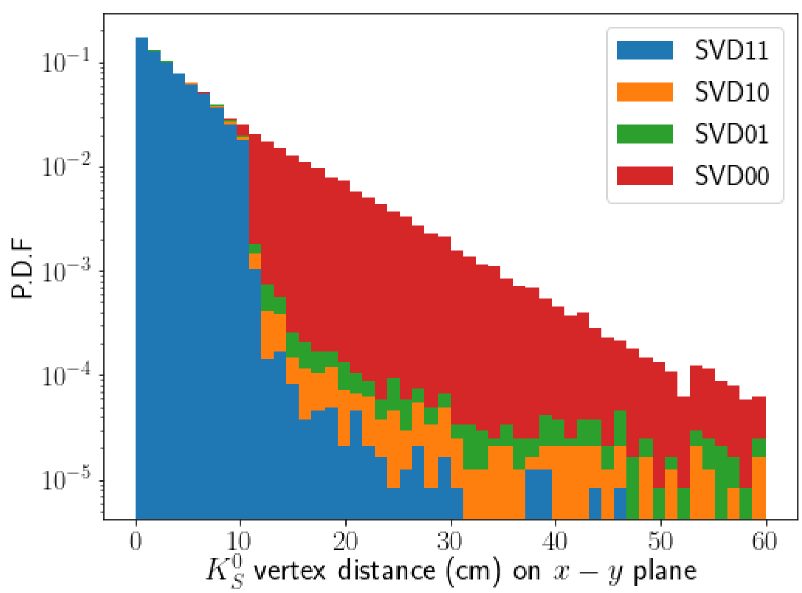


Figure 6-2: The K_S^0 are categorized into 4 types. *SVD11* (*SVD00*) are the ones whose daughter pion tracks have non-zero (zero) SVD hits. *SVD10* (*SVD01*) stands for the ones whose only positive (negative) charged pion track contains SVD hits. The distribution is the K_S^0 flight length on $x-y$ plane for each category of K_S^0, where *SVD00* type K_S^0 start to appear at about SVD layer 5.

6.1.2 Systematic uncertainty in future

When the statistical uncertainty is largely reduced to a comparable level with the systematic uncertainty in future, it is important to evaluate the reducible and irreducible systematic uncertainties in $B^0 \to K_S^0 K_S^0 K_S^0$ for future data collection based on the current measurement result. If no improvement on systematic uncertainty is expected from the current evaluation, it is still challenging to validate the evidence of the NP effects against the small theoretical predictions with the full Belle II luminosity in future.

The reducible systematic sources mainly benefit from the increased luminosity of the experiment data or the larger production of the MC samples, depending on how the corresponding parameters are determined. The systematic uncertainties from signal Δt shape and the fit bias will be reduced possibly by the increased MC sample, including both *signal MC* and *generic MC*. The systematic uncertainties from signal fraction, background Δt shape and wrong tag fraction will be reduced by the larger luminosity of data in future. As an initial study in early operation with very low statistics, it is assumed that these reducible sources are achieved by the increased number of events where the reduced uncertainties are scaled by the fraction of squared-root of increased events. To the contrary, the irreducible sources mainly refer to the ones that are not improved with increased data luminosity or MC samples, such as the irreducible vertexing related sources and tag-side interference [10].

For the signal Δt shape parameters related to CP-side resolution function, the improvement could also come from the better VXD performance and improved tracking algorithm. In the Belle II original prospects, the new PXD detector would be contributive to reduce about 50% of the systematic uncertainties caused by the vertexing resolution. This assumption is made by referencing the evaluation using $J/\psi K^0$ channel in Ref. [10] compared to the Belle result. Current, the vertexing resolution on the CP-side from our analysis is close to that of the Belle. However, such vertex-related improvement might be hard to fully achieved due to the lack of the direct tracks from the IP in our analysis, which makes the vertexing improvement less promising. In the mean time, the uncertainties of CP-side resolution parameters can be reduced if more *signal MC* samples are used. This is dependent on the actual MC production plan, where even the signal yield from the current 1 million *signal MC* is already much more than the signal events in 50 ab^{-1} data. Considering both the unclear improvements from the vertexing quality and the future MC, we keep the reduction factor of 2 as the expected value in 50 ab^{-1} data. The current contribution from CP-side resolution parameters is ~ 0.030, which can be reduced to 0.015 based on this assumption. On the other hand, if the improvement from the vertexing quality is very small and the mount of *signal MC* is not changed as a conservative case, we assume the CP-side systematic uncertainty is the same value as the current one.

For the tag-side parameters, the current contribution is ~ 0.016 determined from the MC control sample corresponding to 2 ab^{-1} luminosity. Unlike the CP-side, tag-side vertex reconstruction is almost universal for the different decay modes. We only expect that the reduced systematic uncertainties from these parameters come from the same mount of MC control sample as 50 ab^{-1}, which is going to be ~ 0.003. To be conservative, the vertexing improvement for the tag-side is not assumed because we do not perform the full reconstruction of B mesons.

In Table 6.1, the improvement where both CP- and tag-sides uncertainties are reduced is calculated to be 0.015 while a conservative case that only tag-side is improved is also considered, which is ~ 0.030.

Table 6.1: Signal Δt shape systematic uncertainties of $\mathcal{S}$ expected at 50 ab^{-1}. The second and third columns are the expected reduced systematic uncertainties for both CP/tag-side improvements or only tag-side improvement.

Luminosity (50 ab^{-1})	both (CP/tag) improved	only tag-side improved
signal Δt shape	~ 0.015	~ 0.030

For the signal fraction contribution which is the largest source at this moment, it mainly suffers from the very low statistics in data that causes the inaccurate modeling of signal shape parameters. Therefore the uncertainties of signal fraction parameters are expected to be reduced quickly with the increased experiment data in future. To be noted, the uncertainties from signal fraction parameters are not directly included in the CP fit model. The 2D signal extraction model is changed every time the parameters are floated and f^{sig} is updated for CP fit. We compare the systematic uncertainty obtained from *generic MC* with the Belle result and the luminosity scaled Belle II result. Using 1 ab^{-1} *generic MC*, the combined contribution on $\mathcal{S}$ uncertainty by floating $\pm 1\sigma$ for each parameter is calculated to be ~ 0.013. This estimation is close to the observed systematic uncertainty in the Belle data which is ~ 0.015 [19] and the Belle II data scaled value ~ 0.011. The luminosity scaling factor $1/\sqrt{N}$ still works well in this case. So we assume this source can be reduced to ~ 0.0016 by factor $\sqrt{50/0.063}$ as shown in Table 6.2 in 50 ab^{-1} Belle II data.

Table 6.2: Signal fraction systematic uncertainties of $\mathcal{S}$ expected at 50 ab^{-1}.

Luminosity (50 ab^{-1})	Improved uncertainty
Signal fraction	~ 0.0016

For the background Δt shape, the systematic uncertainties comes from the parameters determined by the sideband data. Thus, the systematic uncertainty is reduced by factor $\sqrt{50/0.063}$ to be ~ 0.001 as listed in Table 6.3.

Table 6.3: Background Δt shape systematic uncertainties of $\mathcal{S}$ expected at 50 ab^{-1}.

Luminosity (50 ab^{-1})	Improved uncertainty
Background Δt shape	~ 0.0014

For the contributions of wrong tag fraction, the uncertainties of w in each r-bin is expected to be reduced with more tagging control samples. The current wrong tag fraction is taken from 8.7 fb^{-1} early Belle II data [45]. The expected uncertainties at 50 ab^{-1} Belle II data is close to zero as listed in Table 6.4, regarded as a negligible source.

Table 6.4: Wrong tag fraction systematic uncertainties of $\mathcal{S}$ expected at 50 ab^{-1}.

Luminosity (50 ab^{-1})	Improved uncertainty
wrong tag fraction	~ 0.00049

For the fit bias contribution, currently the values are taken by the statistical fit error using 300000 events in *signal MC*. If MC sample used in future could be at least 100 times more than one million, then the fit error is possible to be smaller than input-output difference. From the current MC production plan of the Belle II, the *signal MC* sample recommended by MC production group is typically in a range of several millions. So the foreseen systematic uncertainty is still going to be the fit error, where we take a 50% reduction of the value ~ 0.0098 from Tabel 5.13 as an estimation, listed in Table 6.5.

Table 6.5: Fit bias systematic uncertainties of $\mathcal{S}$ expected at 50 ab^{-1}.

Luminosity (50 ab^{-1})	Improved uncertainty
fit bias	~ 0.005

Concluded from the above discussion, the reducible systematic uncertainties by using increased MC and data in the full Belle II luminosity are estimated and summarized in Table 6.6. It is clear that the dominated contribution in future Belle II data for systematic uncertainty is the *CP* side resolution, indicating the finer study on *CP* side resolution model for no IP-originated tracks is important.

The impact of using *KsFinder* receives contribution from the data-MC mismatch, which can be be improved by a better data-MC consistency. For physics parameters Δm_d and τ_{B^0}, the uncertainties could be reduced by the improved physics input. The vertex reconstruction options are not contributing much in this analysis mostly because they are partially reflected by the signal Δt shape contributions. Using proper IP constraint and tighter vertex cuts, the signal Δt shape will contribute less and vertex reconstruction could contribute more in future. The tag-side interference contribution is referenced to be ~ 0.001 as a small source from the Belle study [46]. The current estimations from *KsFinder* (~ 0.004), physics parameters ($\sim$0.007), the vertex reconstruction ($\sim$0.019) and tag-side interference(~ 0.001) are assumed to be unchanged in the 50 ab^{-1} Belle II luminosity to have a conservative expectation on $\mathcal{S}$ systematic uncertainty, shown in Table 6.7.

It is worth noting that these estimations on the future systematic uncertainties are very preliminary. Some assumptions and expectations are subject to change in future. The vertexing quality and IP conditions can not be precisely assessed at this moment, so as the potential improvements from them. The models used in the signal extraction, resolution functions and CP fit could be modified in future based on the new observations from data and MC, which might contain the different sources of systematic uncertainties. The misalignment of the VXD is not studied at the current stage, but should be included as one of the systematic uncertainty if it is found to be not negligible in future. Overall, we keep these estimations rather conservative to be a baseline for the future studies.

Table 6.6: Improved systematic uncertainties of $\mathcal{S}$ expected at 50 ab^{-1}. The value in the parenthesis stands for the case that only tag-side resolution is improved and no improvement on CP side resolution is implemented.

Sources	Improved uncertainty (50 ab^{-1})
signal Δt shape	$\sim$0.015($\sim$0.030)
Signal fraction	$\sim$0.002
Background Δt shape	$\sim$ 0.001
wrong tag fraction	$\sim$ 0.000
fit bias	$\sim$ 0.005
Total	$\sim$0.016 (0.030)

Table 6.7: The unchanged systematic uncertainties at 50 ab^{-1} based on the current luminosity of the Belle II data.

Sources	Unchanged uncertainty (50 ab^{-1})
KsFinder	$\sim$0.004
physics parameters	$\sim$0.007
vertex reconstruction	$\sim$0.019
tag-side interference	$\sim$ 0.001
Total	$\sim$0.021

6.1.3 Total uncertainty of $\Delta\mathcal{S}$ at 50 ab^{-1}

The total systematic uncertainty of $\mathcal{S}$ in the 50 ab^{-1} Belle II luminosity is estimated based on the improved sources from Table 6.6 and the unchanged sources from Table 6.7. If CP side resolution is not improved, the systematic uncertainty is $\sim$0.037. If the CP side resolution functions is 50% improved, the systematic uncertainty is reduced to $\sim$ 0.026. Both are shown in Table 6.8.

Table 6.8: The systematic uncertainty expected in 50 ab^{-1} Belle II luminosity. The first column is the current value of systematic uncertainty of $\mathcal{S}$ in $B^0 \to K_S^0 K_S^0 K_S^0$. The second and third columns are the systematic uncertainties for both CP/tag-side improvements or only tag-side improvement used in the combined estimation.

Luminosity(ab^{-1})	current (0.0628)	CP/tag (50)	only-tag (50)
Syst.Uncert. on $\mathcal{S}$	0.072	$\sim$0.026	$\sim$ 0.037

By adding in quadrature using estimated statistical and systematic uncertainties, the total uncertainty for $\mathcal{S}$ in $B^0 \to K_S^0 K_S^0 K_S^0$ in 50 ab^{-1} Belle II luminosity is estimated, as shown in Table 6.9. Total uncertainty of ~ 0.048 or ~ 0.040 is expected depending on if the CP-side resolution is improved or not. Considering that the total uncertainty from $B^0 \to J/\psi K_S^0$ at that time is expected to be ~ 0.005 [10], ΔS sensitivity will be dominated by the total uncertainty in $B^0 \to K_S^0 K_S^0 K_S^0$. The current Belle result from $B^0 \to K_S^0 K_S^0 K_S^0$ on ΔS is $0.71 - 0.67 \sim 0.04$ without taking into account any uncertainty. Considering the total uncertainty from the current Belle result is about 0.24, the center value of $\mathcal{S}$ is still quite likely to be 5σ away from that of $J/\psi K_S^0$ under a future uncertainty at $0.040 \sim 0.048$. In general, a total uncertainty at about $0.040 \sim 0.048$ for ΔS at Belle II full luminosity is expected to be a much better probe for addressing whether the NP effects in $B^0 \to K_S^0 K_S^0 K_S^0$ exist.

Table 6.9: The total uncertainty of $\mathcal{S}$ in $B^0 \to K_S^0 K_S^0 K_S^0$ expected in 50 ab^{-1} Belle II luminosity, calculated from the expected statistical and systematic uncertainties. The second and third columns are the total uncertainties for both CP/tag-side improvements or only tag-side improvement used in the combined estimation.

Luminosity (ab^{-1})	current (0.0628)	CP/tag (50)	only-tag (50)
Tot.Ucert. on $\mathcal{S}$	0.853	$\sim$ 0.040	$\sim$ 0.048

6.1.4 *KsFinder* importance

While monitoring the uncertainties of the CP parameters is crucial in searching the hidden NP effects, avoiding wrongly estimated CP asymmetry in the measurement is also critical. If a total uncertainty at ~ 0.03 is achieved in future, however, the center value of $\mathcal{S}_{3K_S^0}$ is wrongly shifted away from $\mathcal{S}_{J/\psi K_S^0}$, it can lead to a very wrong conclusion about the discovery of the NP effects. The *KsFinder* contributes to improve the signal purity for measuring *CP* parameters, which is essential in controlling the potential effect introduced by the large fraction of background events that yield random *CP* asymmetry due to the statistical fluctuation. The signal extraction and *CP* fit on the 1 ab^{-1} *generic MC* sample without *KsFinder* are performed. In this case, we remove the *KsFinder* cut in Table 4.2 and apply the cut $cosVertexMomentum > 0.9$ which can only achieve $\sim$ 82% purity for K_S^0 in *signal MC*. The signal fraction in the signal region defined by M_{bc} and ΔE is considerately lower than that with using *KsFinder*. The stacked histograms of M_{bc} and ΔE with much higer background are shown in Figure 6-3 where the red component is signal. There are 352 true signal events and 543 background events inside the signal region by count. In the meanwhile, the 2D fit on M_{bc} and ΔE are shown in Figure 6-4. From the 2D fit, the signal events number is 389 ± 19 and background events number is 502 ± 15. The comparison of the number of events obtained by different methods is summarized in Table 6.10.

Table 6.10: The number of the signal and background events using *KsFinder* or cut on *cosVertexMomentum* are obtained by the 2D fit, which are compared with the numbers by directly counting the events with *isSignal* = 1(0). It is clear that the number of events obtained by using *KsFinder* is closer to the MC truth.

Selection	signal	background
$FBDT\_Ks > 0.74$ (fit)	341 ± 20	61 ± 17
$FBDT\_Ks > 0.74$ (MC)	336	68
$cosVertexMomentum > 0.9$ (fit)	389 ± 19	502 ± 15
$cosVertexMomentum > 0.9$ (MC)	352	543

From Table 6.10, by using *KsFinder*, the true average signal fraction in signal region from 1 ab^{-1} *generic MC* is 83.2%, and the fit result is $(84.8 \pm 3.7)\%$. To contrary, by only using $cosVertexMomentum > 0.9$, the true average signal fraction is 39.3% and the fit result is $(43.7 \pm 1.4)\%$. In future when the instantaneous luminosity of the SuperKEKB ramps up to a much higher level, the fraction of fake K_S^0 could also rise, too. The signal extraction on a high background event collection might be hard to obtain the signal fraction correctly. Therefore, the development of *KsFinder* is particularly important in the precise *CP* measurement for $B^0 \to K_S^0 K_S^0 K_S^0$ by providing high purity signal events. The current performance of *KsFinder* is presenting a purity about 95% in K_S^0 reconstruction which means there is still a margin for the improvements. The targeted purity and background rejection power of *KsFinder* in future is $\sim$ 99% on average. The data/MC consistency should also be improved so the current correction ratio R_{B^0} is expected to be reduced to $\sim 1.00 \pm 0.01$. Thus the systematic uncertainty from different *KsFinder* responses in between data and MC is assumed to be $\mathcal{O}(0.001)$ as a negligible contribution.

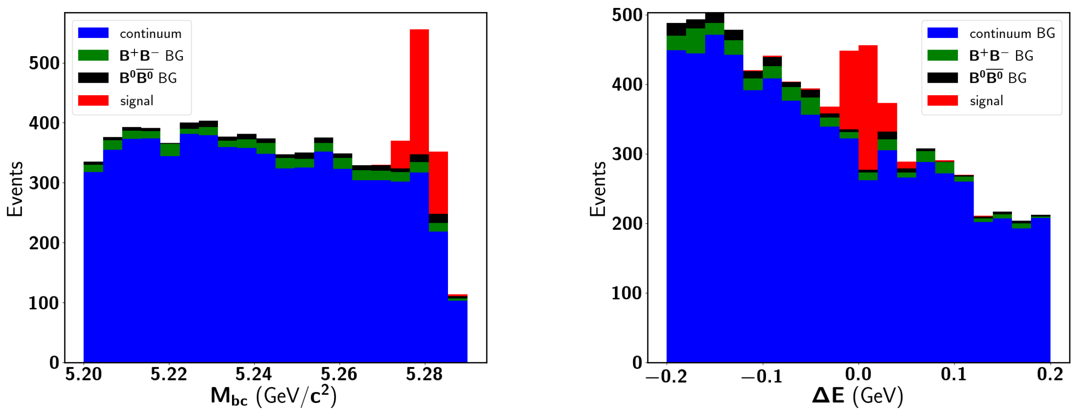


Figure 6-3: M_{bc} and ΔE stacked histogram of 1 ab^{-1} *generic MC* sample replacing $FBDT\_Ks > 0.74$ by $cosVertexMomentum > 0.9$ in Table 4.2 as a selection criteria, showing a much worse signal significance.

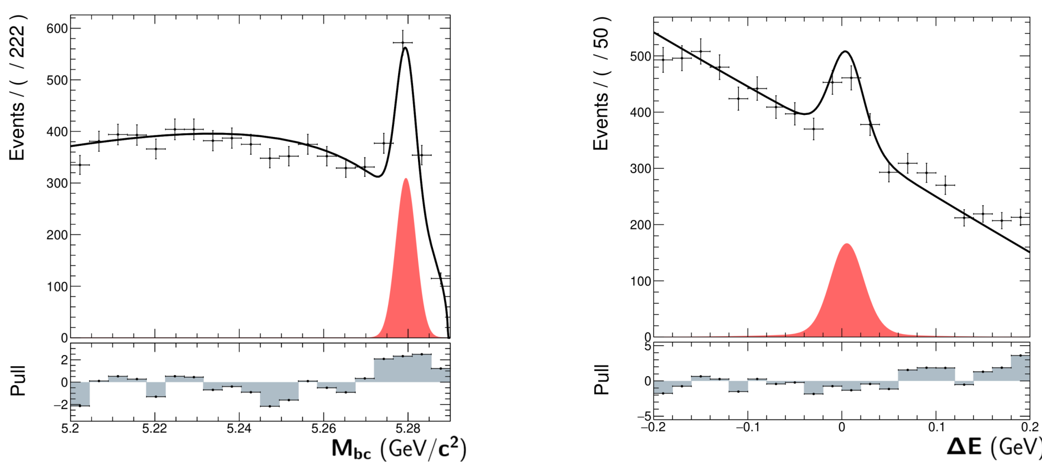


Figure 6-4: M_{bc} and ΔE 2D fit on 1 ab^{-1} *generic MC* sample, replacing $FBDT\_Ks > 0.74$ by $cosVertexMomentum > 0.9$ in Table 4.2. The red is signal component.

6.2 Conclusions

In this thesis, we perform the analysis of the CP parameter measurement in $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ using the early Belle II data, which is targeted to search for the NP effects in the penguin dominated $b \rightarrow s$ transition. The reconstruction of K_S^0 is initially done by using a traditional cut-based method with a large fraction of fake candidates. Thus a new MVA-based *KsFinder* is developed by using FastBDT algorithm which can effectively improve the signal purity. For the reconstruction of B^0, we take advantage of two variable M_{bc} and ΔE to select the events with a sufficient continuum background suppression. To obtain the signal fraction for each B^0 events, a 2D fit model is established and fitted using 62.8 fb^{-1} data. The model of the resolution of vertex positions has been studied using MC sample and sideband data based on the understanding of vertex reconstruction performance in the current Belle II detectors. To make a proper use of the reconstructed vertex information and perform a CP fit compactly, a new CP fitter is built and being validated, which will serve as a multi-functional analysis tool for the Belle II CP violation study in future. The CP parameter measurement is performed based on the validation of analysis strategies by the blind analysis that shows a consistent result for CP parameters compared to the simulation input. The linearity and pull of the CP fit are checked to demonstrate the reliability of the fit procedures. The fit result on B^0 lifetime using the experiment data is also agreed with the current value in PDG with a relatively large statistical uncertainty due to the low statistics of data.

After the CP fit procedures are validated, the CP parameters $\mathcal{S}$ and $\mathcal{A}$ using 62.8 ab^{-1} Belle II early data in 2019 and 2020 spring and summer are obtained. The result is

$$\begin{aligned} \mathcal{S} = -\sin(2\phi_1) &= -0.82 \pm 0.85 \text{ (stat)} \pm 0.07 \text{ (syst)} , \\ \mathcal{A} &= -0.21 \pm 0.28 \text{ (stat)} \pm 0.06 \text{ (syst)} . \end{aligned} \tag{6.1}$$

The result agrees with the prediction of the Standard Model and the previous results from Belle [19] and BaBar [20]. The measurement precision of CP parameters in this thesis is majorly limited by the large statistical uncertainty.

6.3 Prospects

Even though the current result on CP parameters are dominated by the large uncertainty, the previous discussions about the future results have shown a good potential of searching for the NP effects in $B^0 \to K_S^0 K_S^0 K_S^0$ based on the current analysis in this thesis. At integral luminosity at 50 ab^{-1}, the uncertainty on $\mathcal{S}$ would be reduced to a comparable value around $0.040 \sim 0.048$ realistically, as shown in Figure 6-5[1], where the statistical and reducible systematic uncertainties are assumed to be scaled by the squared root of the integrated luminosity. The expected sensitivity in full Belle II data is proven to be competitive and the analysis workflow is built which will be further improved along with the future Belle II data taking and MC production. The progress that has been made in this thesis paves a well-constructed and solid path for searching the NP effects in time dependent CP violation study of $B^0 \to K_S^0 K_S^0 K_S^0$ at the full Belle II luminosity in future.

[1]For the continuous extrapolation of the systematic uncertainties, an assumption is used that the reducible part is approximately scaled by the square-root of the future luminosity and the irreducible part is treated as a constant. Thus two Equations are formed: $0.072^2 = a/0.0628 + b$ and 0.026^2 $(0.037^2) = a/50 + b$ where a and b present the reducible and irreducible part, which are used as the coefficients of the extrapolating function.

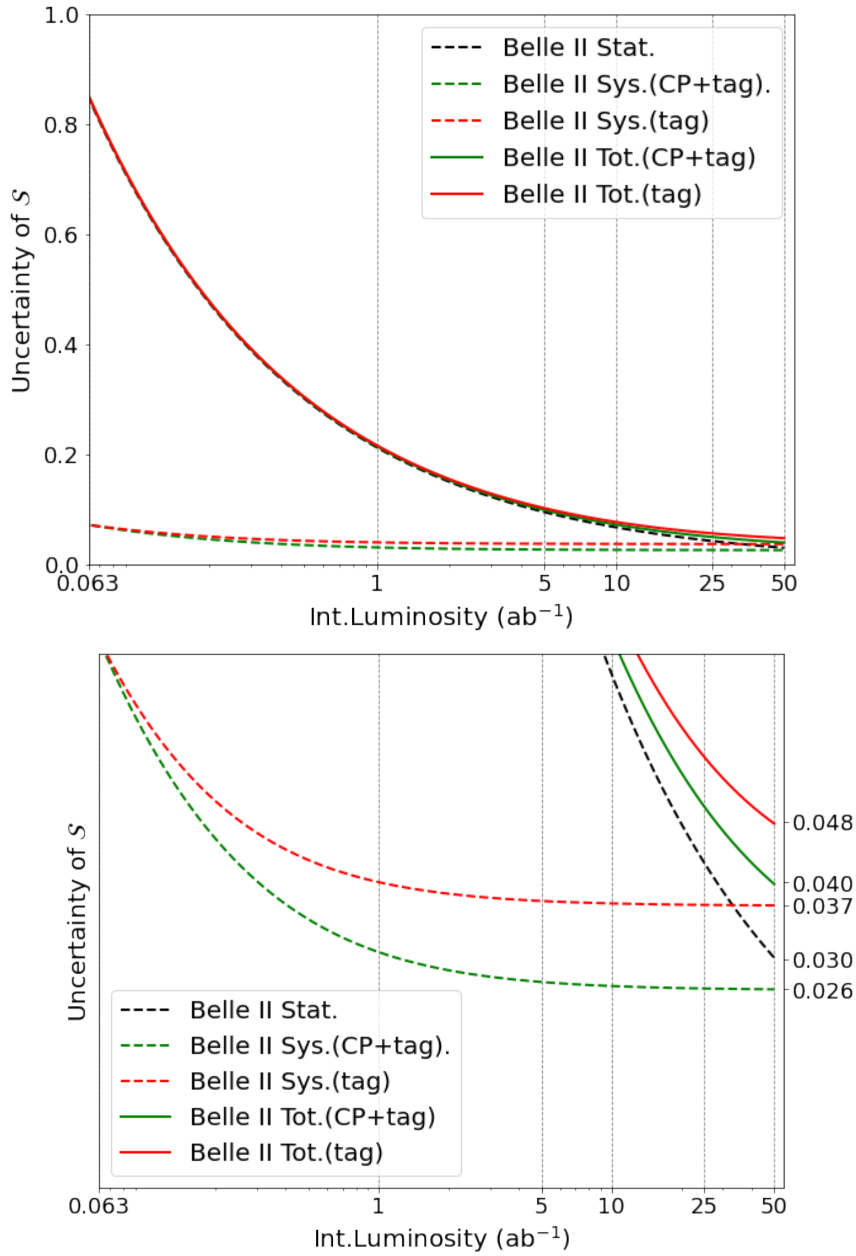


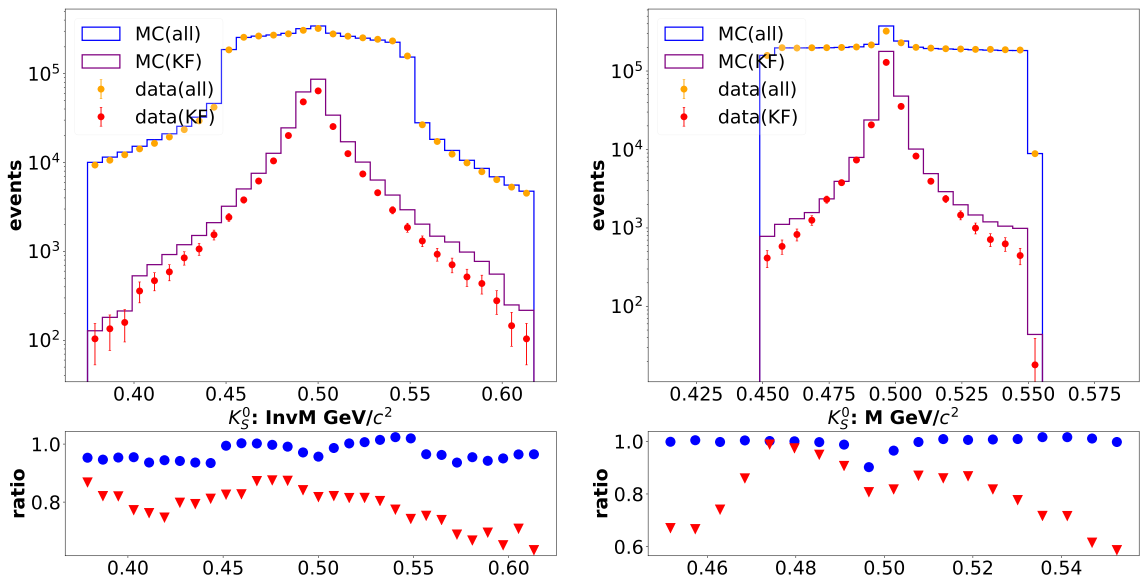
Figure 6-5: The expected total uncertainty of $\mathcal{S}$ in $B^0 \to K_S^0 K_S^0 K_S^0$, where the dashed lines are the statistical (black), CP/tag-side improved systematic (green) and only tag-side improved systematic (red) uncertainties, with the corresponding solid lines as the total uncertainties. The top is the overview for the whole Belle II luminosity range from now, and the bottom is y-axis zoom-in.

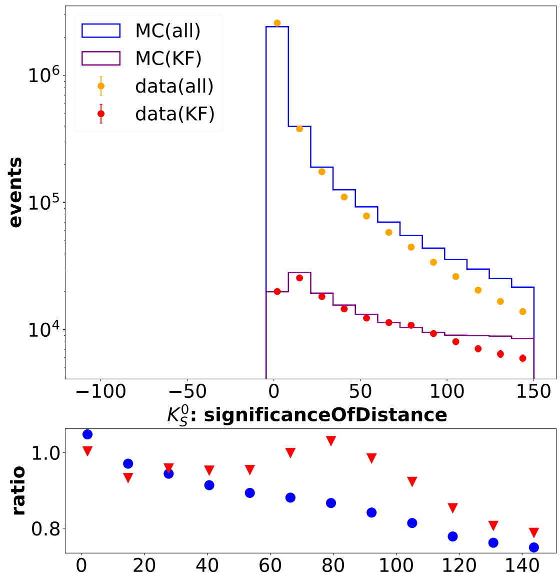
Appendix A

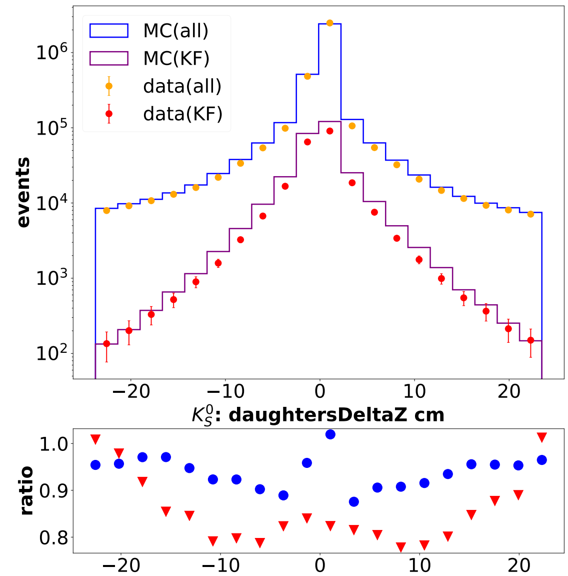
Data Validation Plots for K_S^0

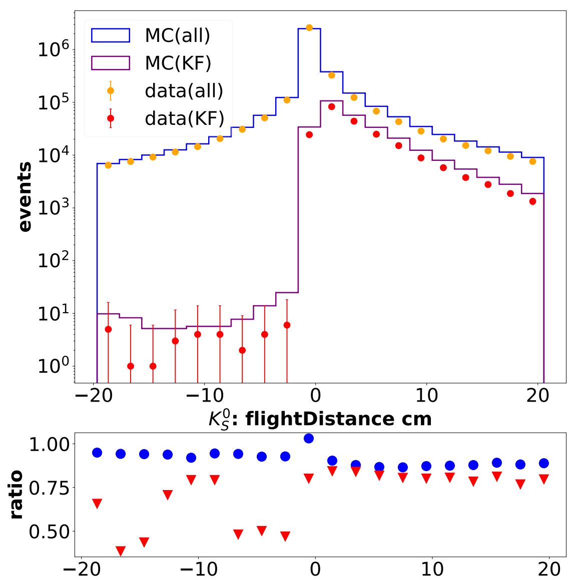
The full distributions of the input variables of *KsFinder* are shown in the Figure A-1. The small discrepancies are observed in the distributions after applying the *KsFinder* selection cut while the agreement before applying such cut is fine. The main reason is that the difference in the *cosVertexMomentum* which plays the most important role in the K_S^0 classification. Even before the use of *KsFinder*, small differences can be observed. This variable is sensitive to the K_S^0 vertexing quality and the VXD misalignment, which is still subject to be optimized in future.

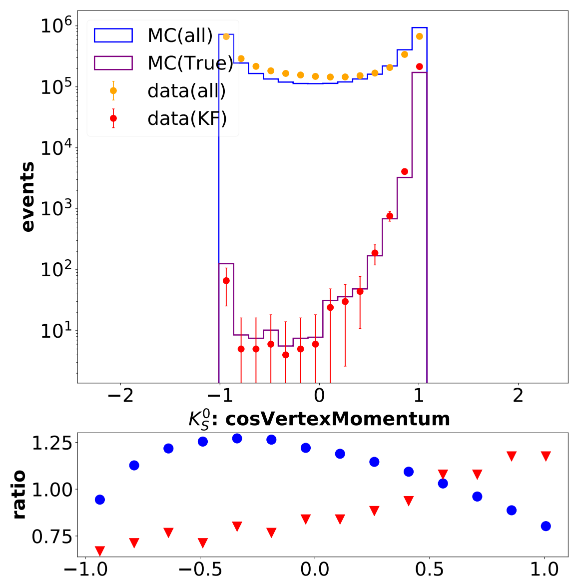
Figure A-1: The distribution of the training variables in KsFinder. The blue and purple solid lines in the top plots are the total and true K_S^0 distributions from generic MC, respectively. The bottom plots are the data-MC ratio before (blue) and after (red) applying *KsFinder* cut.

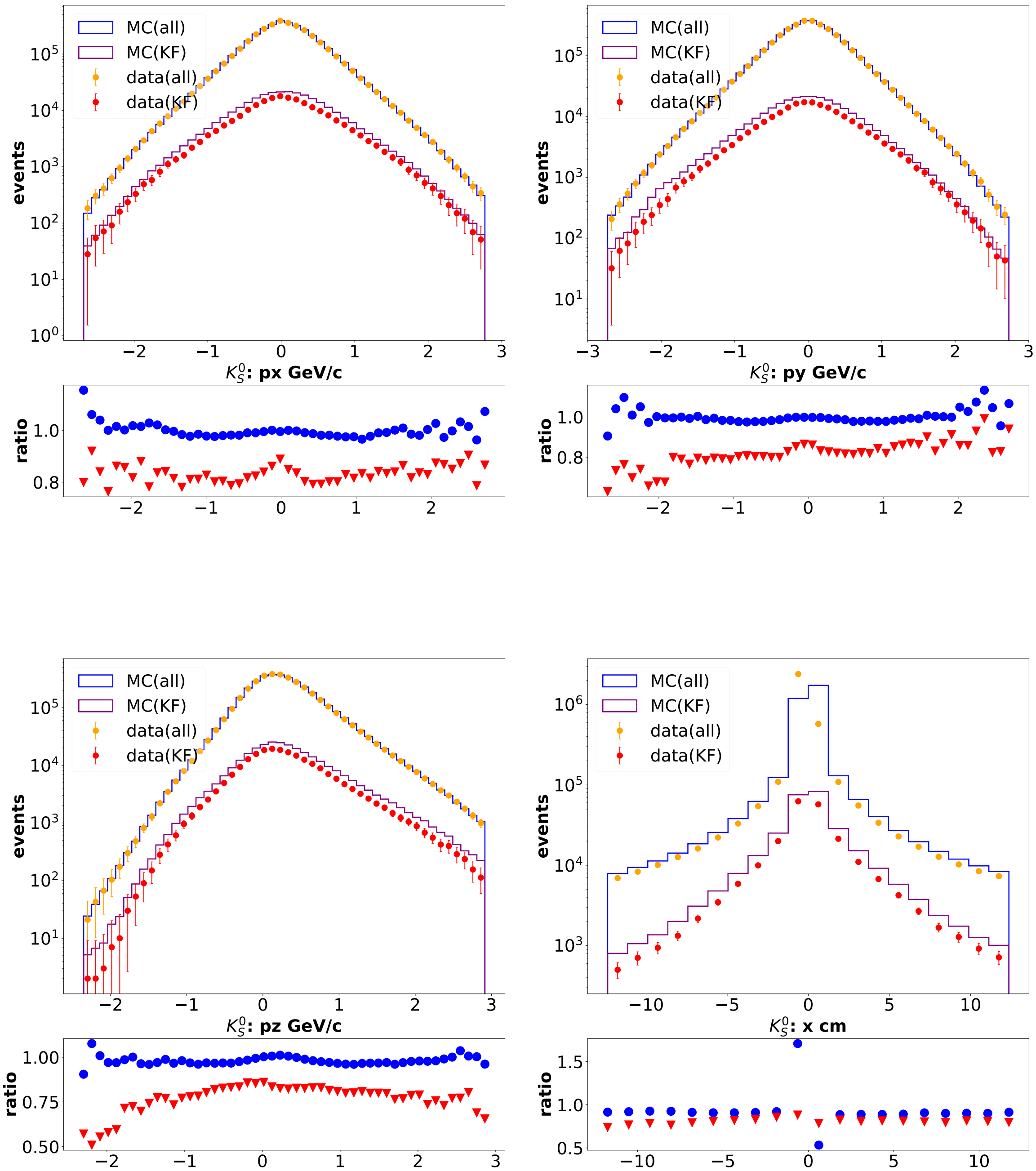


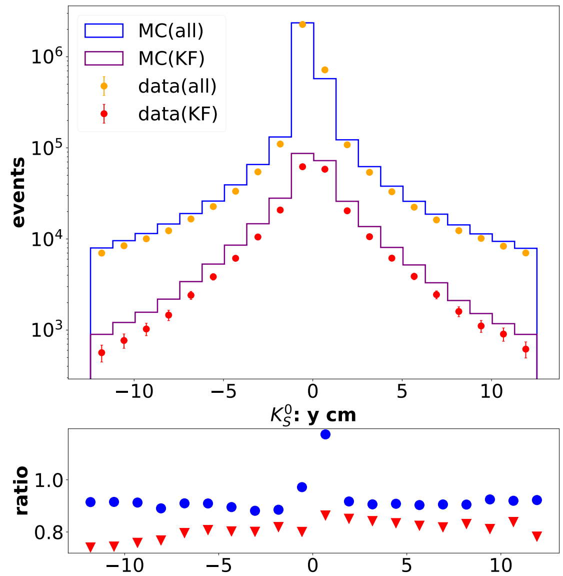

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: significanceOfDistance
ratio

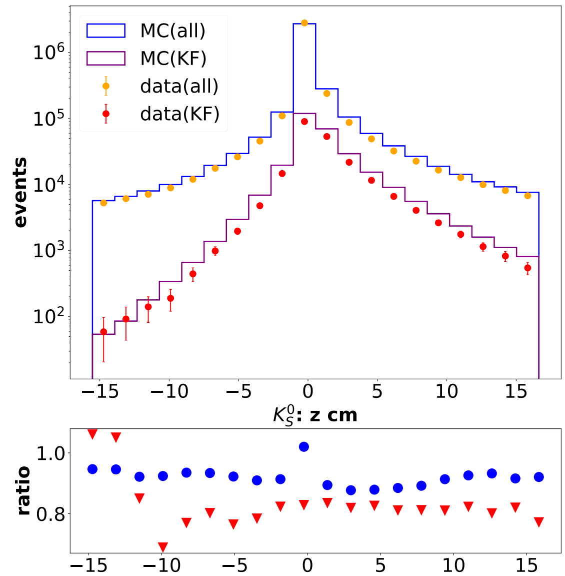

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: daughtersDeltaZ cm
ratio

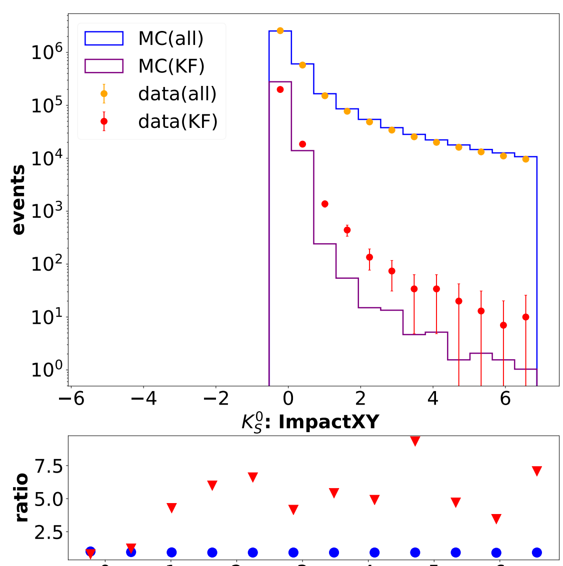

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: flightDistance cm
ratio

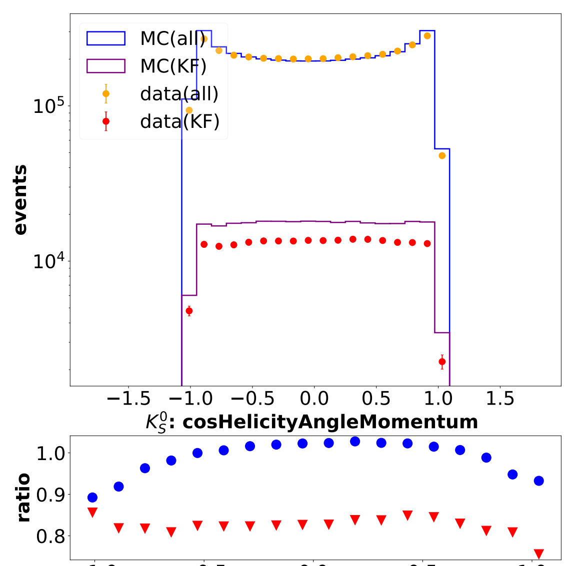

MC(all)
MC(True)
data(all)
data(KF)
events
K_S^0: cosVertexMomentum
ratio

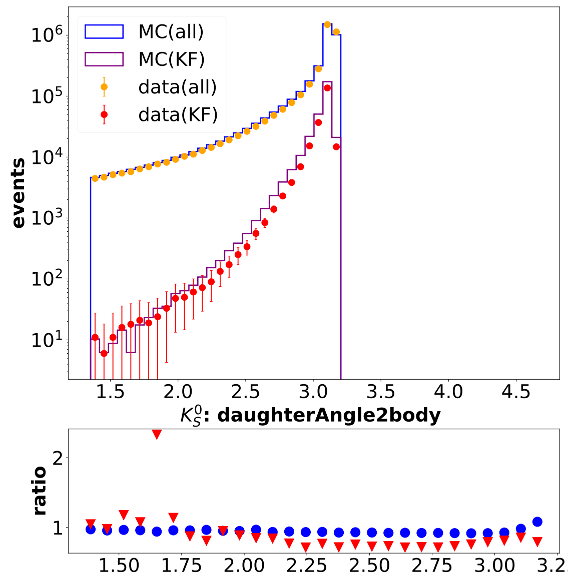

MC(all)
MC(KF)
data(all)
data(KF)
events
ratio
K_S^0: px GeV/c
K_S^0: py GeV/c
K_S^0: pz GeV/c
K_S^0: x cm

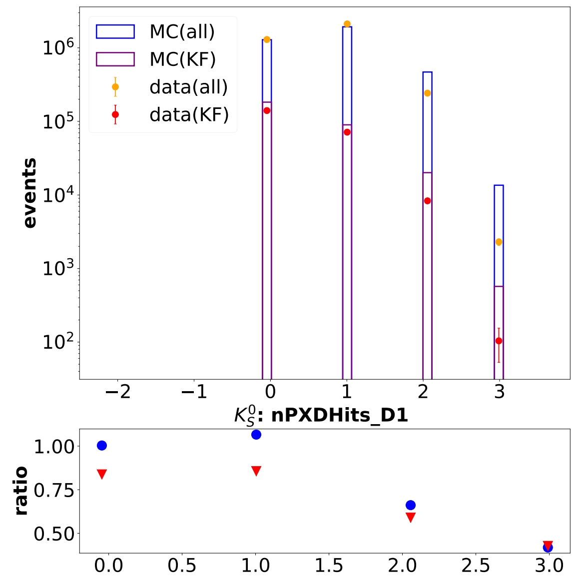

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: y cm
ratio

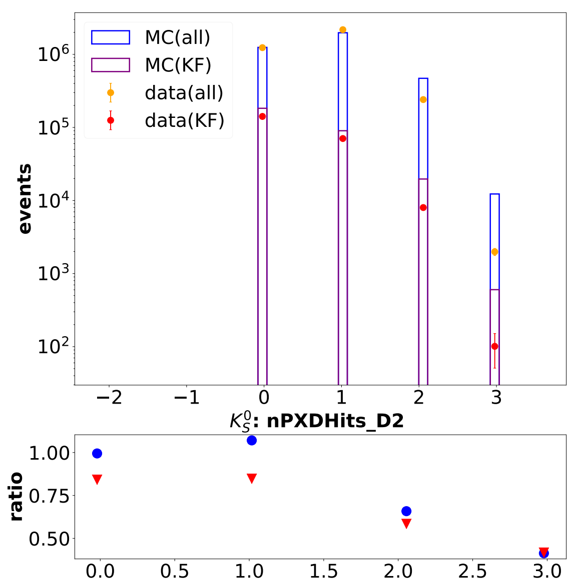

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: z cm
ratio

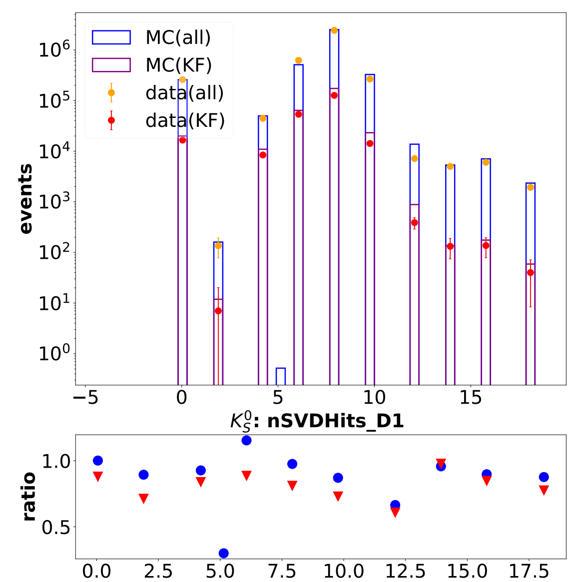

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: ImpactXY
ratio

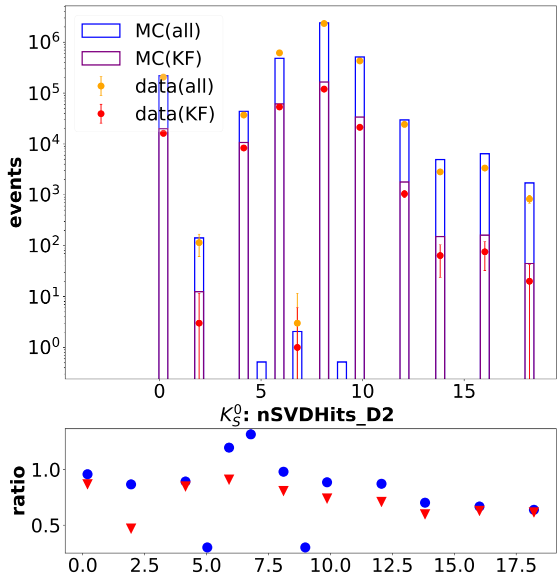

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: cosHelicityAngleMomentum
ratio

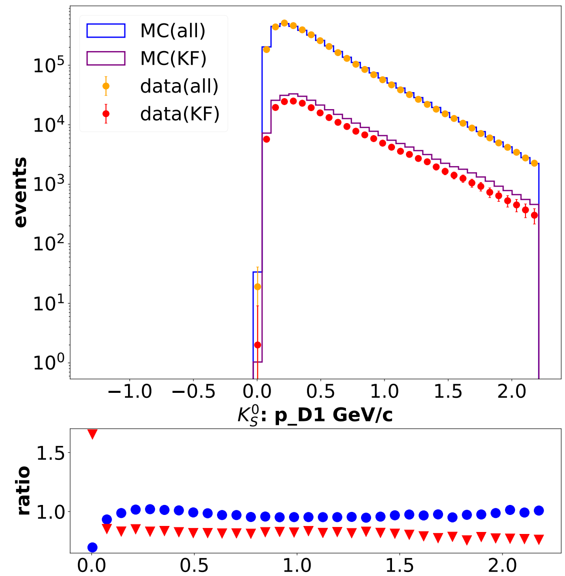

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: daughterAngle2body
ratio

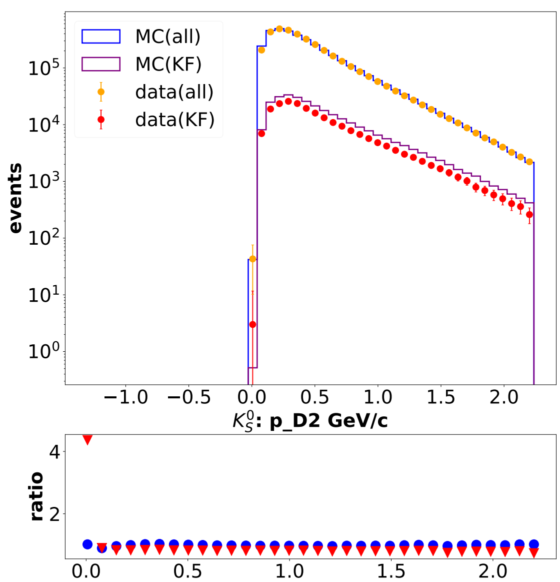

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: nPXDHits_D1
ratio

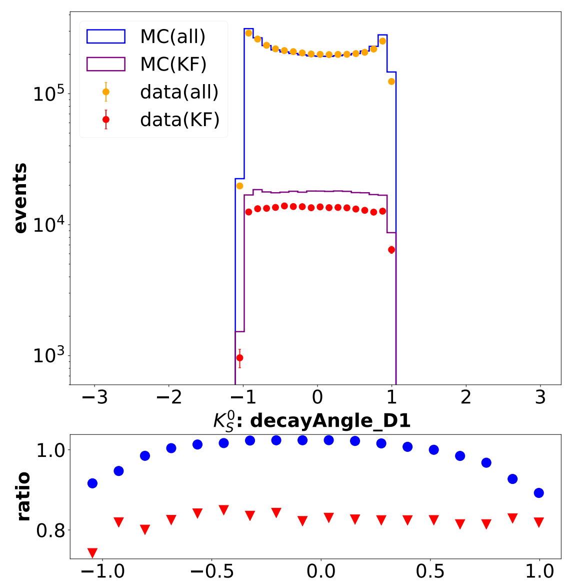

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: nPXDHits_D2
ratio

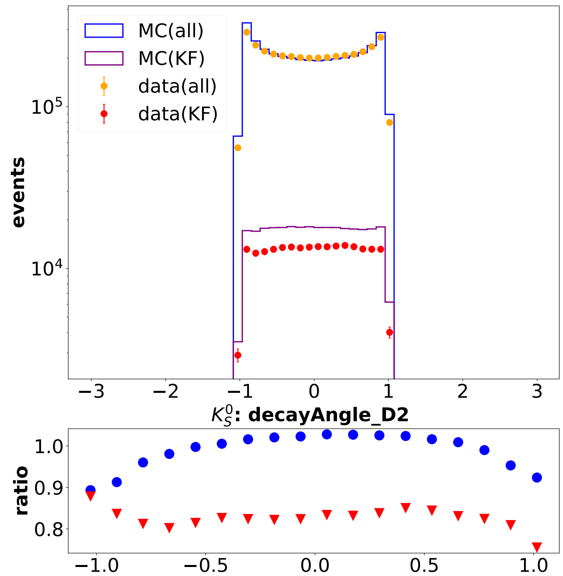

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: nSVDHits_D1
ratio


MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: nSVDHits_D2
ratio


MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: p_D1 GeV/c
ratio


MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: p_D2 GeV/c
ratio


MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: decayAngle_D1
ratio



MC(all)
MC(KF)
data(all)
data(KF)
events
10^5
10^4
−3
−2
−1
0
1
2
3
K_S^0: decayAngle_D2
ratio
1.0
0.9
0.8
−1.0
−0.5
0.0
0.5
1.0

Appendix B

Control Samples

The tag-side is almost independent from the CP-side reconstruction when it comes to a correctly reconstructed candidate. Thus, we refer to the study of the tag-side vertexing performance used in the $B^0 \to J/\psi K_S^0$ to minimize the influence of the different tag-side vertexing method. The control samples are prepared with the equivalent amount of 2 ab^{-1} luminosity, which primarily consider about the decay channels with D or D^* mesons that could give fairly misplaced secondary vertices. The summary of the reconstruction criteria of the control samples are listed in Table B.1 and the vertex fit is performed by *KFit* without IP constraint to keep it same as the tag-side vertexing method.

Table B.1: Hadronic control sample reconstruction criteria, which is used for the tag-side resolution study. The first column stands for the B decay to neutral and charged D or D^*. The second column stands for the D decay as intermediate states of the B decay, which includes $D^{*+} \to D^0\pi^+,\ D^0 \to K^-\pi^+, K^-\pi^+\pi^0, K^-\pi^+\pi^-\pi^+$ and $\rho^+ \to \pi^+\pi^0$ [43].

B decay	D decay	$\lvert M_{Kn\pi} - M_D \rvert$	ΔM_{D^*}	R_2	$cos\theta_{th}$
$B^+ \to \bar{D}^0\pi^+$	$D^0 \to K^-\pi^+$	$< 4\sigma$	-	-	-
	$D^0 \to K^-\pi^+\pi^0$	$< 3\sigma$	-	< 0.45	-
	$D^0 \to K^-\pi^+\pi^-\pi^+$	$< 2\sigma$	-	< 0.45	-
$B^0 \to D^{*-}\pi^+$	$D^0 \to K^-\pi^+$	$< 10\sigma$	< 5 MeV/c^2	-	-
	$D^0 \to K^-\pi^+\pi^0$	$< 3.5\sigma$	< 3 MeV/c^2	-	< 0.98
	$D^0 \to K^-\pi^+\pi^-\pi^+$	$< 4\sigma$	< 4 MeV/c^2	< 0.6	-
$B^0 \to D^{*-}\rho^+$	$D^0 \to K^-\pi^+$	$< 7\sigma$	< 4 MeV/c^2	< 0.6	< 0.95
	$D^0 \to K^-\pi^+\pi^0$	$< 3.5\sigma$	< 12 MeV/c^2	-	< 0.98
	$D^0 \to K^-\pi^+\pi^-\pi^+$	$< 3.5\sigma$	< 3 MeV/c^2	-	< 0.92
$B^0 \to D^-\pi^+$	$D^+ \to K^-\pi^+\pi-$	$< 2\sigma$	-	< 0.5	< 0.995

Appendix C

$2K_S^0$ invariant mass distribution

The decay $B^0 \to K_S^0 K_S^0 K_S^0$ is much cleaner compared to the $B^0 \to K_S^+ K_S^- K_S^0$ in terms of the CP-odd backgrounds, due to the absence of the u quark in the final states. However, the potential small effect from the CP-odd final states mainly created by $b \to c$ transitions needs to be checked to reduce the impact on the fit of the CP parameters. Usually the CP-odd resonant backgrounds can be effectively rejected by requiring the invariant mass of two K_S^0 to be outside of mass windows of the resonant intermediate states. In the early stage of the Belle II, the data validation of the two K_S^0 mass distribution can not be well done and such veto may contain large bias if the veto ranges are only taken from the study of MC samples. Therefore, we only present the distributions of the two K_S^0 invariant masses obtained from data and *generic MC* in Figure C-3. The contributions from CP-odd backgrounds will be monitored along with the future data taking.

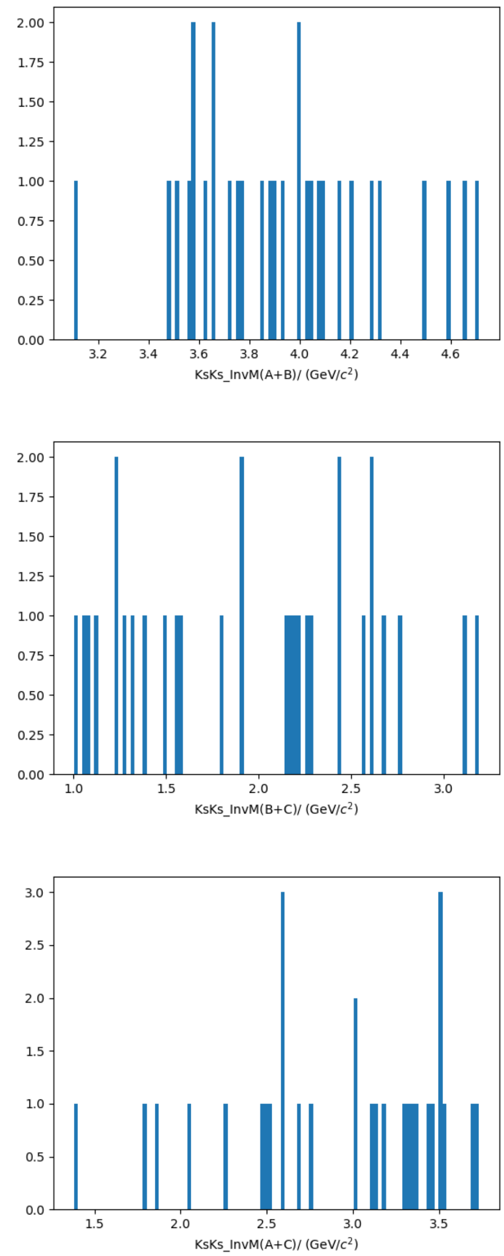


Figure C-1: Data in signal region

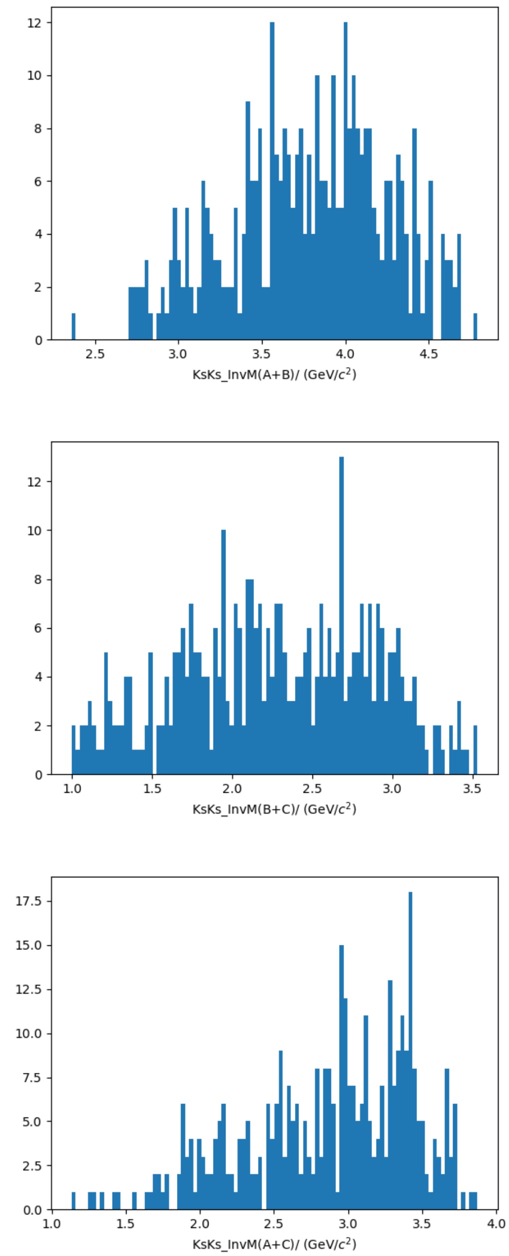


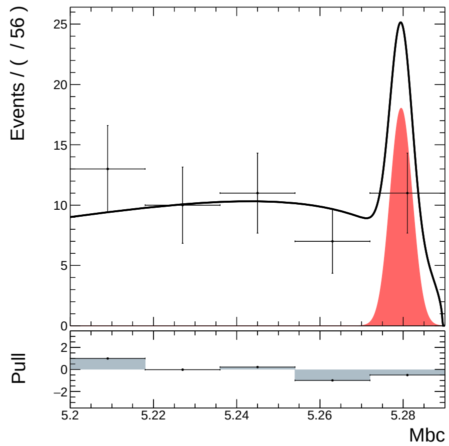
Figure C-2: Generic MC in signal region.

Figure C-3: The distribution of two K_S^0 invariant mass where the A, B and C indices are in the increasing order of K_S^0 momentum. The left is from the data and the right is from *generic MC*.

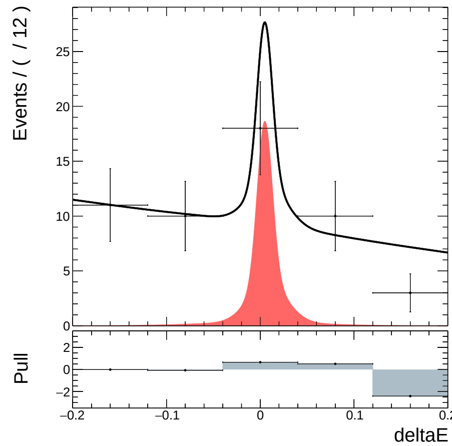
Appendix D

Injection test for B^0 signal yield

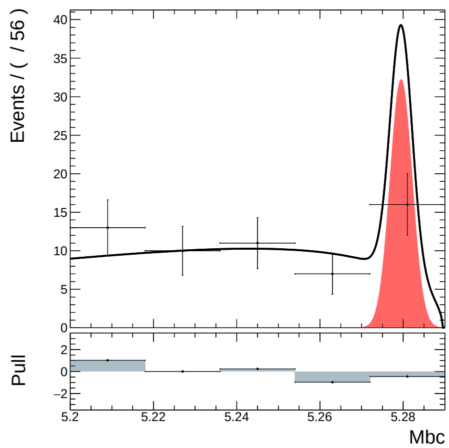
To test the reliability of the signal extraction using 2D fit in the current low statistics of experiment data, an injection test is performed where we inject 5 to 30 signal events with 46 background events randomly taken from the *generic MC*. The distribution of the each test is included in Figure D-1.



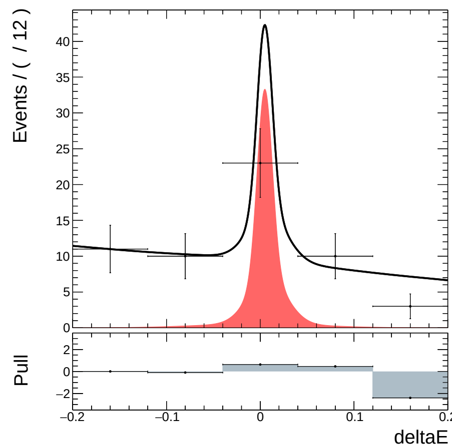
a) signal injected: 5



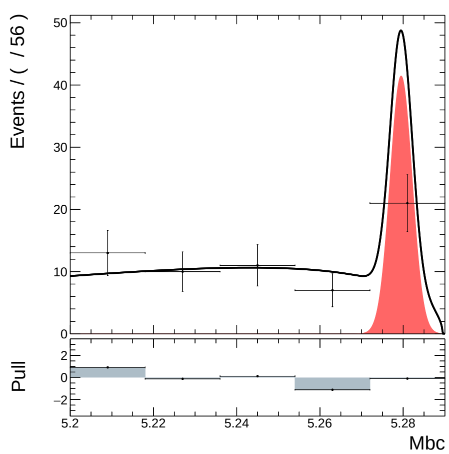
b) signal injected: 5



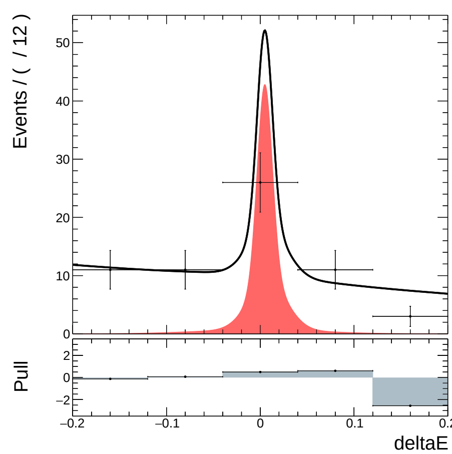
c) signal injected: 10



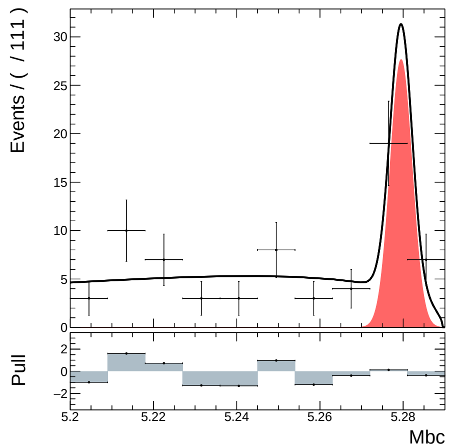
d) signal injected: 10



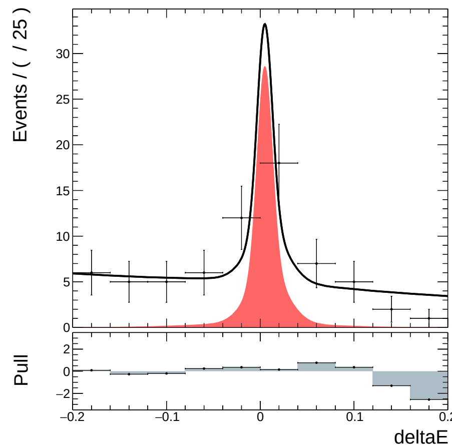
e) signal injected: 15



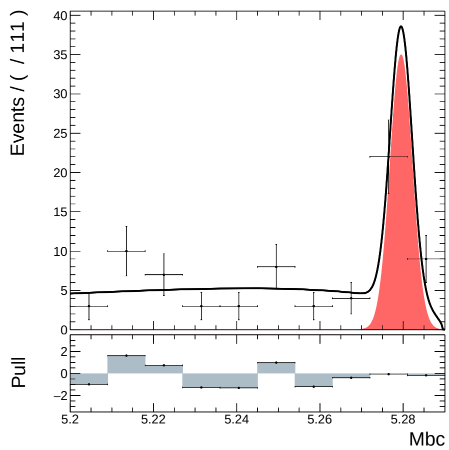
f) signal injected: 15



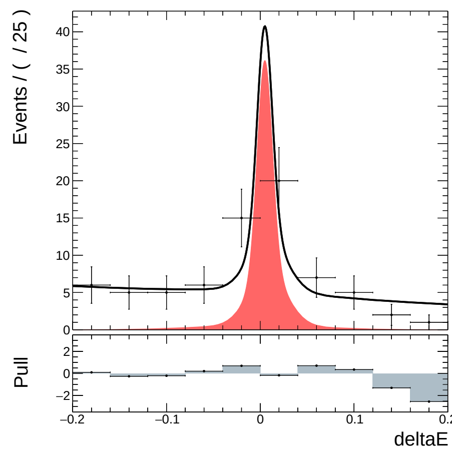
g) signal injected: 20



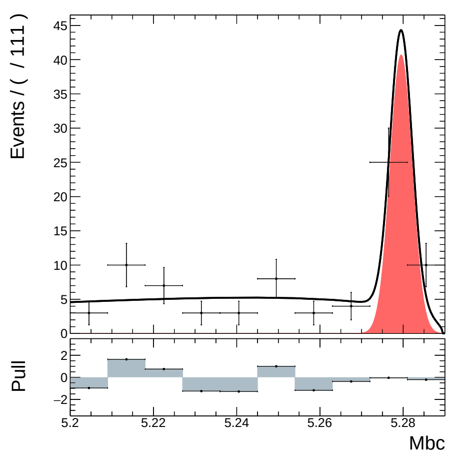
h) signal injected: 20



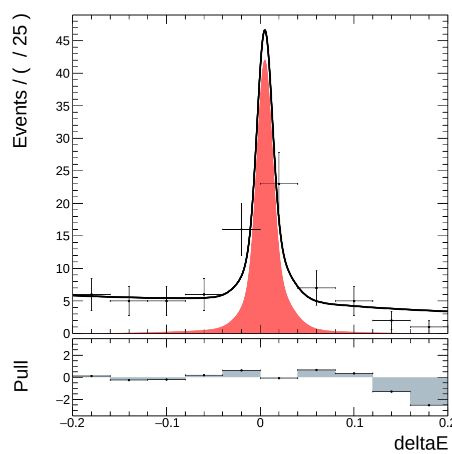
i) signal injected: 25



j) signal injected: 25



k) signal injected: 30



l) signal injected: 30

Figure D-1: The fit results of M_{bc} and ΔE in signal injection test, where signal events from 5 to 30 with 5 per step are injected with 46 continuum events.

Appendix E

The *KsFinder* impact on M_{bc}, ΔE and vertex positions

KsFinder provides a classification variable *FBDT_Ks* to be used as a cut to improve the purity of the K_S^0 and B^0. Thus, it's essential to check the potential impact on M_{bc} and ΔE, as well as vertex positions on z-axis of B^0 after applying *KsFinder*. The K_S^0 classification uses information such as invariant mass and decay vertex positions which may propagate bias into B^0 signal extraction, eventually may affect the measurement of CP parameters.

For K_S^0 candidates with full SVD hits from their daughter pion tracks, the *KsFinder* can perform a better classification on wether it's a true K_S^0 or not, compared to those K_S^0 candidates with less SVD hits information. This is because the tracking quality and invariant mass from the fit is highly dependent on the SVD hits information, which plays an important role in calculating input variables used in the *KsFinder*. For example, a true K_S^0 is more likely to be wrongly classified if all of its daughter pions have no SVD hits, so the reconstruction can only be done by using 6 CDC-only tracks. Rejecting these candidates might introduce biases. Therefore, given each type of B^0 based on how many CDC-only tracks it has in the final states, the comparison on M_{bc} and ΔE with or without *KsFinder* is performed by fitting the distribution in *signal MC*. M_{bc} and ΔE are modeled by signal and double Gaussian functions, respectively. Comparing corresponding fit results, no clear bias on M_{bc} and

ΔE is found by using *KsFinder* where fit results are agreed well within one standard deviation. The fit results are shown in Figure E-1 and E-2.

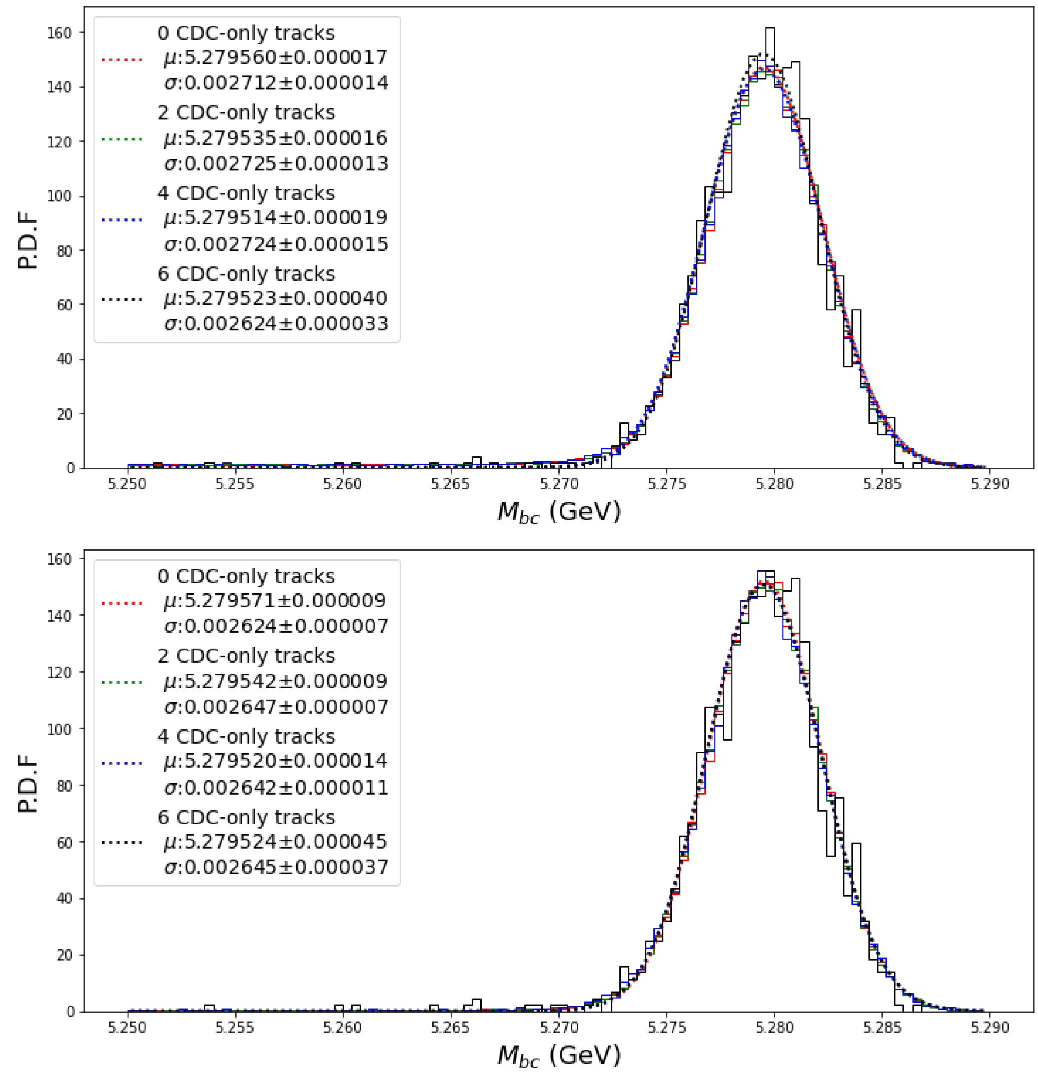


Figure E-1: M_{bc} distribution based on the number of CDC-only tracks in final states. Top: no *KsFinder* used; Bottom: *KsFinder* used. The μ and σ are the mean and standard deviation of the Gaussian function.

Similar to the comparison of M_{bc} and ΔE, the z direction vertex position and the vertex position difference Δz between CP and tag sides are also checked, in which no clear bias are found either. The z and Δz are modeled using single Gaussian with the same mean but different standard deviation. The results are shown in Figure E-3 and E-4. It is obvious that in Figure E-3, the CP-side resolution of vertex on z-axis is wider when the final states of B^0 have more CDC-only tracks, especially when all

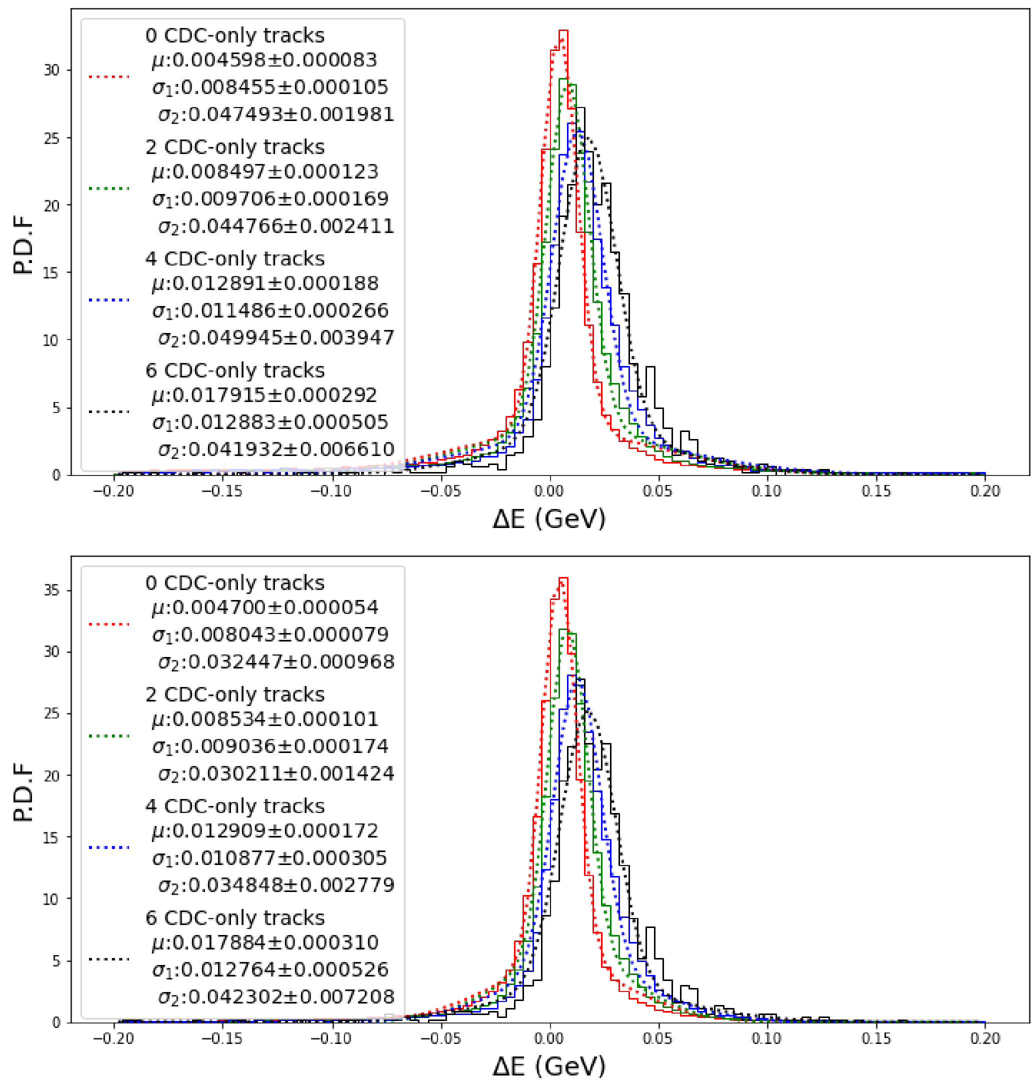


Figure E-2: ΔE distribution based on the number of CDC-only tracks in final states. Top: no *KsFinder*; Bottom: *KsFinder* used. The μ is the common mean for double Gaussian. The σ_1 and σ_2 are the standard deviations of the Gaussian function.

the tracks only contains CDC hits (6 CDC-only tracks).

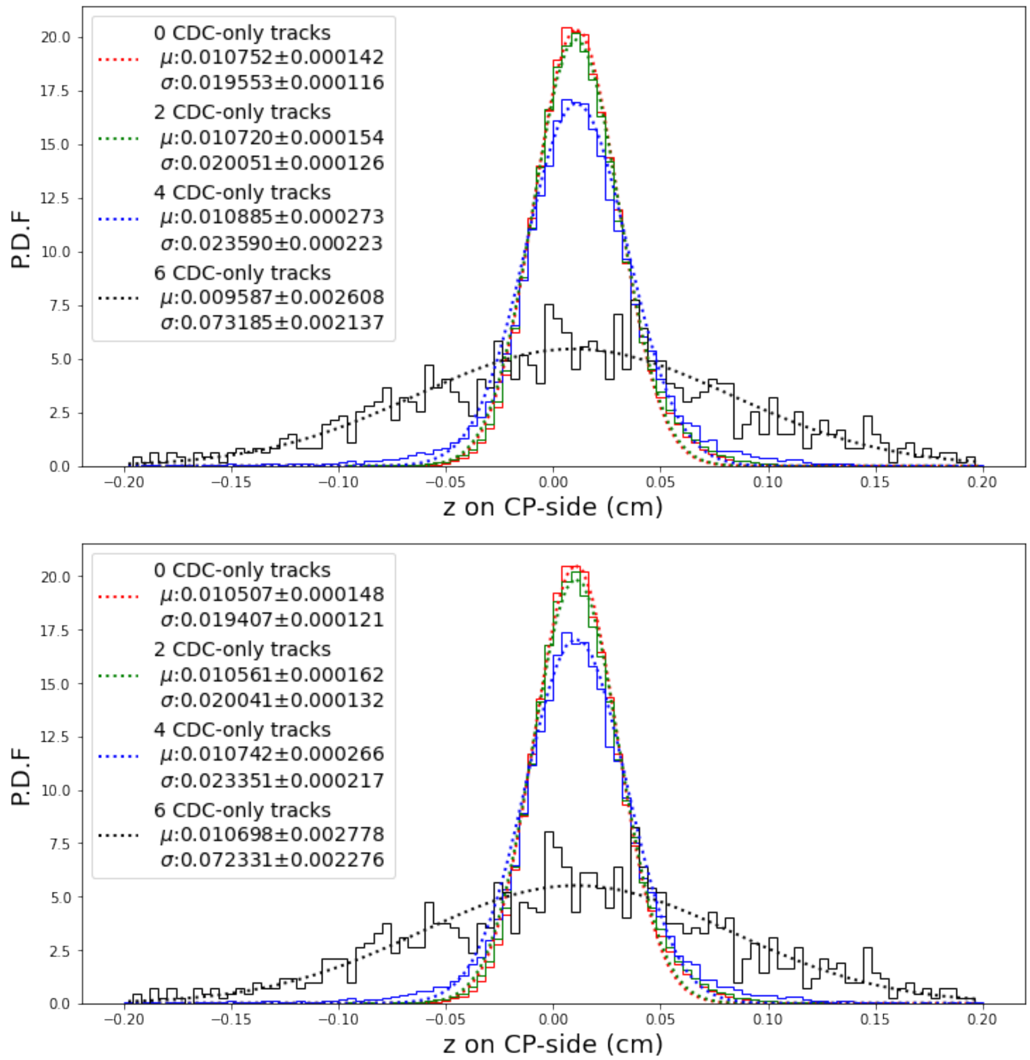


Figure E-3: Δz distribution based on the number of CDC-only tracks in final states. Top: no *KsFinder*; Bottom: *KsFinder* used. The μ and σ are the mean and standard deviation of the Gaussian function.

Above all, no clear appearance of bias on M_{bc} and ΔE distributions, as well as vertex positions from using *KsFinder* has been found, *KsFinder* may implement a small shift on the vertex position which is negligible compared to the large statistical uncertainty due to the current low luminosity. Hence, there's no correction on these observables are applied in this analysis, and the systematic uncertainty from *KsFinder* is evaluated by taking into account of R_{B^0} in signal fraction calculation which comes from the potential discrepancy of *KsFinder* responses among data and MC.

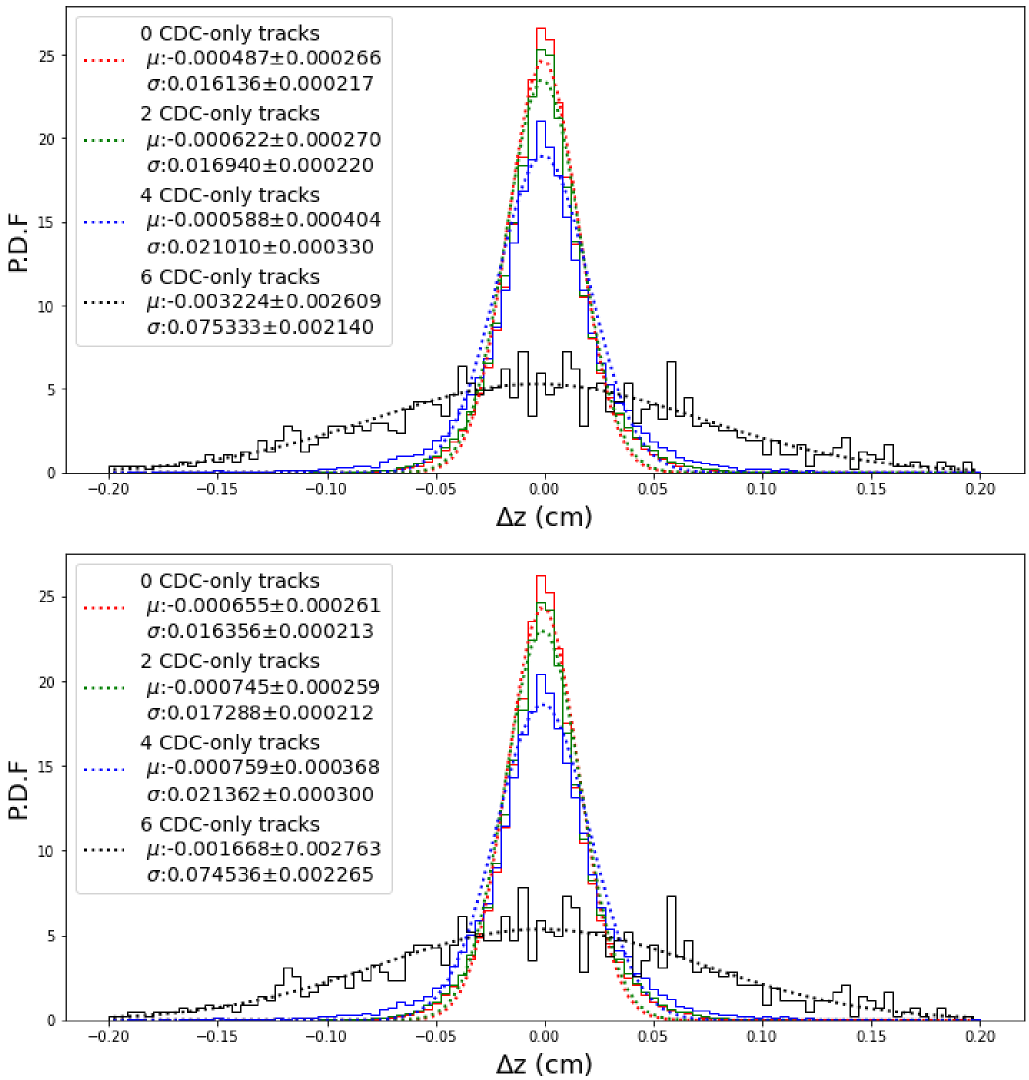


Figure E-4: Δz distribution based on th number of CDC-only tracks in final states. Top: no *KsFinder*; Bottom: *KsFinder* used. The μ and σ are the mean and standard deviation of each Gaussian function.

Bibliography

[1] Georges Aad, Tatevik Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, Ahmed Ali Abdelalim, R. Aben, B. Abi, M. Abolins, OS. AbouZeid, et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.

[2] Wikipedia on the Standard Model. Elementary particles in the Standard Model. `https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles_Anti.svg`, 2012.

[3] T. D. Lee and C. N. Yang. Question of Parity Conservation in Weak Interactions. *Phys. Rev.*, 104:254–258, Oct 1956. doi:10.1103/PhysRev.104.254. URL `https://link.aps.org/doi/10.1103/PhysRev.104.254`.

[4] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi:10.1103/PhysRev.105.1413. URL `https://link.aps.org/doi/10.1103/PhysRev.105.1413`.

[5] James H Christenson, Jeremiah W Cronin, Val L Fitch, and René Turlay. Evidence for the 2π Decay of the K_2^0 Meson. *Physical Review Letters*, 13(4):138, 1964. URL `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.13.138`.

[6] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, 02 1973. ISSN 0033-068X. doi:10.1143/PTP.49.652. URL `https://doi.org/10.1143/PTP.49.652`.

[7] Alexander Abashian, K. Abe, I. Adachi, Byoung Sup Ahn, H. Aihara, M. Akatsu, G. Alimonti, K. Aoki, K. Asai, M. Asai, et al. Measurement of the CP Violation Parameter $\sin 2\phi_1$ in B^0 Meson Decays. *Physical review letters*, 86(12):2509, 2001.

[8] Bernard Aubert et al. BaBar collaboration. *Phys. Rev. D*, 76:052005, 2007. URL `http://oldwww.unibas.it/utenti/cavallo/BaBar/Papers/NIM.4.1.pdf`.

[9] Ikaros I. Bigi, Valery A. Khoze, NG. Uraltsev, and AI. Sanda. The Question of CP noninvariance—As seen through the eyes of neutral beauty. In *CP Violation*, pages 175–248. World Scientific, 1989.

[10] E. Kou, P. Urquijo, W. Altmannshofer, F. Beaujean, G. Bell, M. Beneke, I. I. Bigi, F. Bishara, M. Blanke, C. Bobeth, and et al. The Belle II Physics Book. *Progress of Theoretical and Experimental Physics*, 2019(12), Dec 2019. ISSN 2050-3911. doi:10.1093/ptep/ptz106. URL `http://dx.doi.org/10.1093/ptep/ptz106`.

[11] Augusto Ceccucci. The CKM Quark-Mixing Matrix. 2008.

[12] The CKMfitter group. Results of the global CKM fit in the large complex plane. `http://ckmfitter.in2p3.fr/www/results/plots_summer19/ckm_res_summer19.html`, 2019.

[13] Amol Dighe, Tobias Hurth, Choong Sun Kim, and Tadashi Yoshikawa. The Width Difference of B_d Mesons, 2001.

[14] Julian Tarek Wishahi. *Measurement of CP Violation in $B^0 \to J/\psi K_S^0$ Decays with the LHCb Experiment*. PhD thesis, Dortmund U., 2014.

[15] K-F Chen, K. Hara, M. Hazumi, T. Higuchi, K. Miyabayashi, Y. Nakahama, K. Sumisawa, O. Tajima, Y. Ushiroda, Y. Yusa, et al. Observation of Time-Dependent CP Violation in $B^0 \to \eta' K^0$ Decays and Improved Measurements of CP Asymmetries in $B^0 \to \phi K^0, K_0^S K_0^S K_0^S$ and $B^0 \to J/\psi K^0$ Decays. *Physical review letters*, 98(3):031802, 2007.

[16] Tim Gershon and Masashi Hazumi. Time-dependent CP Violation in $B^0 \to P^0 P^0 X^0$ Decays. *Physics Letters B*, 596(3-4):163–172, 2004.

[17] Martin Beneke and Matthias Neubert. QCD Factorization for $B \to PP$ and $B \to PV$ Decays. *Nuclear Physics B*, 675(1-2):333–415, Dec 2003. ISSN 0550-3213. doi:10.1016/j.nuclphysb.2003.09.026. URL `http://dx.doi.org/10.1016/j.nuclphysb.2003.09.026`.

[18] Hai-Yang Cheng, Chun-Khiang Chua, and Amarjit Soni. CP-violating Asymmetries in B^0 Decays to $K^+K^-K^0_{S(L)}$ and $K^0_S K^0_S K^0_{S(L)}$. *Phys. Rev. D*, 72:094003, Nov 2005. doi:10.1103/PhysRevD.72.094003. URL `https://link.aps.org/doi/10.1103/PhysRevD.72.094003`.

[19] KH. Kang, H. Park, T. Higuchi, K. Miyabayashi, K. Sumisawa, I. Adachi, JK. Ahn, H. Aihara, S. Al. Said, DM. Asner, et al. Measurement of time-dependent CP violation parameters in $B^0 \to K_S^0 K_S^0 K_S^0$ decays at Belle. *arXiv preprint arXiv:2011.00793*, 2020.

[20] J. P. Lees et al. Amplitude analysis and measurement of the time-dependent CP asymmetry of $B^0 \to K_S^0 K_S^0 K_S^0$ decays. *Phys. Rev. D*, 85:054023, 2012. doi:10.1103/PhysRevD.85.054023.

[21] T. Abe et al. Belle II Technical Design Report. 2010. arXiv:1011.0352.

[22] T. Aushev, W. Bartel, A. Bondar, J. Brodzicka, T. E. Browder, P. Chang, Y. Chao, K. F. Chen, J. Dalseno, A. Drutskoy, Y. Enari, T. Gershon, B. Golob, T. Goto, F. Handa, K. Hara, S. Hashimoto, H. Hayashii, M. Hazumi, T. Higuchi, J. Hisano, W. S. Hou, T. Iijima, K. Ikado, K. Inami, H. Itoh, R. Itoh, H. Ishino, N. Katayama, Y. Y. Keum, K. Kinoshita, E. Kou, P. Križan, P. Krokovny, T. Kurimoto, Y. Kwon, A. Limosani, T. Matsumoto, T. Morozumi, Y. Nakahama, M. Nakao, S. Nishida, T. Ohshima, Y. Okada, K. Okumura, S. L. Olsen, T. Onogi, G. Pakhlova, H. Palka, P. Pakhlov, A. Poluektov, S. Recksiegel, H. Sagawa, M. Saigo, Y. Sakai, A. I. Sanda, C. Schwanda, A. Schwartz, K. Senyo, Y. Shimizu, T. Shindou, R. Sinha, M. Starič, K. Sumisawa, M. Tanaka, K. Trabelsi, P. Urquijo, Y. Ushiroda, E. Won, H. Yamamoto, M. Yamauchi, T. Yoshikawa, and J. Zupan. Physics at Super B Factory, 2010.

[23] MJ. French, LL. Jones, Q. Morrissey, A. Neviani, R. Turchetta, J. Fulcher, G. Hall, E. Noah, M. Raymond, G. Cervelli, et al. Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 466(2):359–365, 2001.

[24] K. Inami, T. Mori, T. Matsumura, K. Kurimoto, S. Hasegawa, Y. Suzuki, T. Murase, Y. Yurikusa, M. Akatsu, Y. Enari, et al. Cross-talk suppressed multi-anode MCP-PMT. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 592 (3):247–253, 2008.

[25] ROOT: analyzing petabytes of data, scientifically. URL `https://root.cern/`.

[26] The DIRAC project. URL `https://dirac.readthedocs.io/en/latest/DeveloperGuide/Overview/`.

[27] Thomas Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II core software. *Computing and Software for Big Science*, 3(1):1–12, 2019.

[28] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The belle ii core software. *Computing and Software for Big Science*, 3(1), Nov 2018. ISSN 2510-2044. doi:10.1007/s41781-018-0017-9. URL `http://dx.doi.org/10.1007/s41781-018-0017-9`.

[29] The EvtGen project. URL `https://evtgen.hepforge.org/doc/models.html`.

[30] Sea Agostinelli, John Allison, K. al Amako, John Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al. GEANT4 – a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[31] BOTTOM MESONS (PDG). URL `https://pdglive.lbl.gov/Particle.action?init=0&node=S042&home=MXXX045`.

[32] The Belle II online luminosity monitor (Internal Link). URL https://confluence.desy.de/display/BI/Phase+3+luminosity+monitoring.

[33] J-F. Krohn, F. Tenchini, P. Urquijo, F. Abudinén, S. Cunliffe, T. Ferber, M. Gelb, J. Gemmler, P. Goldenzweig, T. Keck, et al. Global decay chain vertex fitting at Belle II. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 976: 164269, 2020.

[34] M. Feindt and U. Kerzel. The NeuroBayes Neural Network Package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):190–194, 2006.

[35] Thomas Keck. FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification. *arXiv preprint arXiv:1609.06119*, 2016. URL https://arxiv.org/pdf/1609.06119.

[36] Jan Therhaag and TMVA Core Developer Team. TMVA-Toolkit for multivariate data analysis. In *AIP Conference Proceedings*, volume 1504, pages 1013–1016. American Institute of Physics, 2012.

[37] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[38] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[39] K. Sumisawa, Y. Ushiroda, M. Hazumi, K. Abe, K. Abe, I. Adachi, H. Aihara, Y. Asano, V. Aulchenko, T. Aushev, and et al. Measurement of Time-DependentCP-Violating Asymmetries in $B^0 \to K^0_S K^0_S K^0_S$ Decay. *Physical Review Letters*, 95(6), Aug 2005. ISSN 1079-7114. doi:10.1103/physrevlett.95.061801. URL http://dx.doi.org/10.1103/PhysRevLett.95.061801.

[40] A. J. Bevan, B. Golob, Th. Mannel, S. Prell, B. D. Yabsley, H. Aihara, F. Anulli, N. Arnaud, T. Aushev, M. Beneke, and et al. The Physics of the B Factories. *The European Physical Journal C*, 74(11), Nov 2014. ISSN 1434-6052. doi:10.1140/epjc/s10052-014-3026-9. URL http://dx.doi.org/10.1140/epjc/s10052-014-3026-9.

[41] H. Albrecht, R. Gläser, G. Harder, A. Krüger, AW. Nilsson, A. Nippe, T. Oest, M. Reidenbach, M. Schäfer, W. Schmidt-Parzefall, et al. Search for Hadronic $b \to u$ Decays. *Physics Letters B*, 241(2):278–282, 1990.

[42] Yosuke Yusa, Δt resolution function based on artificial parametrization, BELLE-II-NOTE-TE-2019-010 (Belle II internal note). URL https://docs.belle2.org/record/1384/files/belle2-note-resol_belle.pdf.

[43] TDCPV Working Group, Belle II, Measurement of the CP Violation Parameter sin2ϕ_1 in $B^0 \to J/\psi K^0$ Decay, BELLE2-NOTE-PH-2020-027 (Belle II internal note). URL `https://docs.belle2.org/record/1941/files/BELLE2-NOTE-PH-2020-027.pdf`.

[44] H. Tajima, H. Aihara, T. Higuchi, H. Kawai, T. Nakadaira, J. Tanaka, T. Tomura, M. Yokoyama, M. Hazumi, Y. Sakai, et al. Proper-time resolution function for measurement of time evolution of B mesons at the KEK B-Factory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 533(3):370–386, 2004.

[45] F. Abudinén, I. Adachi, R. Adak, K. Adamczyk, P. Ahlburg, JK. Ahn, H. Aihara, N. Akopov, A. Aloisio, F. Ameli, et al. First flavor tagging calibration using 2019 Belle II data. *arXiv preprint arXiv:2008.02707*, 2020.

[46] Y. Yosuke and K. Miyabayashi. Measurement of Time-dependent CP Violation in $B^0 \to (c\bar{c})K^0$ Decays with 772×10^6 BB. Technical report, Belle Note, 2011.