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: $\mathcal{S} = -\sin(2\phi_1) = -0.82 \pm 0.85\,(\mathrm{stat}) \pm 0.07\,(\mathrm{syst})$ and $\mathcal{A} = -0.21 \pm 0.28\,(\mathrm{stat}) \pm 0.06\,(\mathrm{syst})$ are obtained.

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 are 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 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.

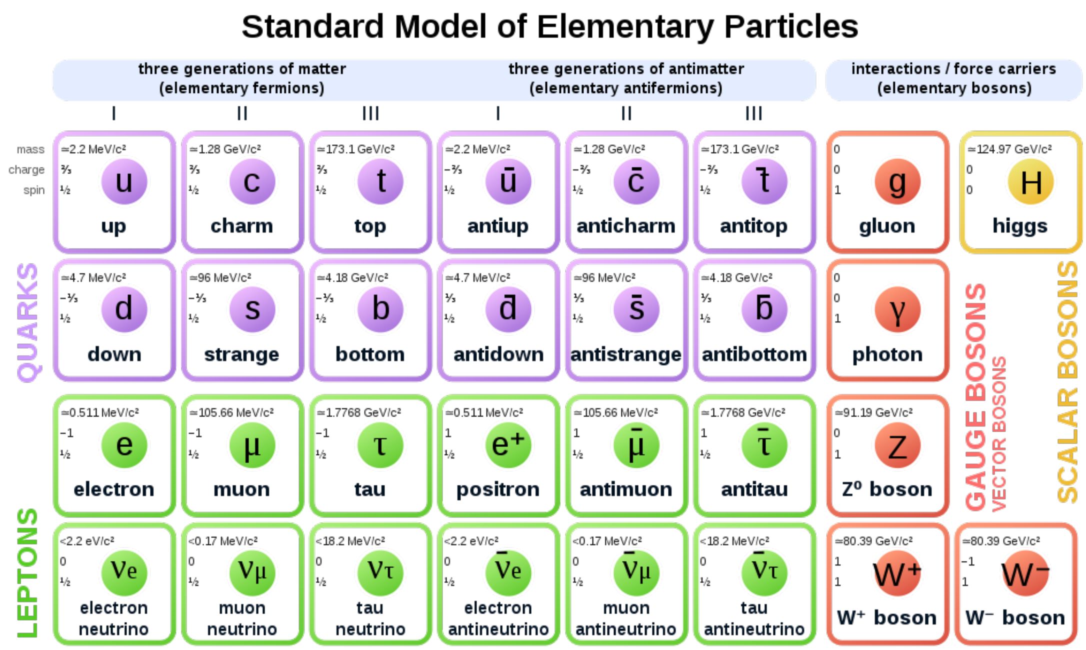


Figure 1-1: Elementary particles in the Standard Model.[2]

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 $\overrightarrow{r} \to -\overrightarrow{r}$. 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 weak interactions. 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_S^0" and "K_L^0" for short and long lifetime particles) with opposite CP eigenvalue. 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_L^0, nearly no K_S^0. But 0.002% of K_L^0 were found to decay into $\pi^+\pi^-$ which is the main decay process of K_S^0.(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_S^0 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 $K_L^0 \to \pi^+\pi^-$ 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 weak interactions 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 $B \to J/\psi K_S^0$. This channel provided a good clearness in theoretical prediction and has relatively large branching fraction, thus it's 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 Phsyics researches.

1.3 CKM mechanism

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

Equation 1.1 is the Higgs potential doublets in the SM, 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 which are appearing when introducing Higgs field ϕ without breaking the gauge symmetry. The Lagrangian for Yukawa interaction of the quark fields[11] can be presented by Equation 1.2.

$$\mathcal{L}^q_{Yuk} = -Q^\dagger Y^d \Phi d'_R - Q^\dagger Y^u \epsilon \Phi^* u'_R + h.c. \tag{1.2}$$

where the primed fields stand for the weak eigenstates of quarks. The ϵ is a $2\times$ 2 matrix and $Q^\dagger$ is the left-handed doublets that stand for weak eigenstates of up and down types quarks, see Equation 1.3 and 1.4.

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{1.3}$$

$$Q = \begin{pmatrix} u' & d' \\ c' & s' \\ t' & b' \end{pmatrix}_L \tag{1.4}$$

Yukawa matrix is an arbitrary 3×3 complex matrix $Y^{u,d}$ which gives the rise of up and down type massive quark field $M^{u,d} = Y^{u,d}\nu$ according to Equation 1.2. The representation of the quark fields using weak eigenstates can be transformed to mass eigenstates by Equation 1.5 and 1.6.

$$S^u_{L,R} \begin{pmatrix} u' \\ c' \\ t' \end{pmatrix}_{L,R} = \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} \tag{1.5}$$

$$S^d_{L,R}\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{L,R} = \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R} \tag{1.6}$$

In Equation 1.5 and 1.6, $S^{u,d}_{L,R}$ are all unitary matrices since they are generated by the normalized eigenstate states of Yukawa matrix. The mass item in the Lagrangian can be presented as Equation 1.7

$$\mathcal{L}_m = -\sum_{q=u,c,t,d,s,b} M_q q^\dagger q \tag{1.7}$$

where the $q = (q_L + q_R)$ is four-component Dirac field, and $q^\dagger_L q_L = q^\dagger_R q_R = 0$. 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 1.8, where $V_{CKM} \equiv S^u_L S^{d\dagger}_L$.

$$\mathcal{L}^q_W = \frac{g}{\sqrt{2}}\left[\begin{pmatrix} \overline{u} & \overline{c} & \overline{t} \end{pmatrix}_L \gamma^\mu W^+_\mu V_{CKM}\begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \begin{pmatrix} \overline{d} & \overline{s} & \overline{b} \end{pmatrix}_L \gamma^\mu W^-_\mu V^\dagger_{CKM}\begin{pmatrix} u \\ c \\ t \end{pmatrix}_L\right] \tag{1.8}$$

The Lagrangian hereby clearly declares the transition of different charged 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 Equation 1.8 transits as Equation 1.9 shows.

$$\begin{pmatrix} \overline{u} & \overline{c} & \overline{t} \end{pmatrix}_L \gamma^\mu W^+_\mu V_{CKM}\begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \Leftrightarrow \begin{pmatrix} u & c & t \end{pmatrix}_L \gamma^\mu W^-_\mu V_{CKM}\begin{pmatrix} \overline{d} \\ \overline{s} \\ \overline{b} \end{pmatrix}_L \tag{1.9}$$

Comparing Equation 1.9 and 1.8, the CP symmetry requires the invariance before and after CP conjugation, meaning that Equation 1.10 is expected.

$$u_L^i V_{ij} \bar{d}_L^j \gamma^\mu W_\mu^- = u_L^n V_{nm}^* \bar{d}_L^m \gamma^\mu W_\mu^- \tag{1.10}$$

The same indices ij and nm are summed over on both side. This is equivalent to Equation 1.11:

$$V_{ij} = V_{ij}^* \tag{1.11}$$

On the one hand, if the CKM matrix is real, CP will be conserved in the weak interaction in the SM due to the natural hold of Equation 1.11. On the other hand, from Equation 1.10, it's still possible to make Lagrangian invariant even if V_{CKM} is not real, which can be achieved by introducing non-physical phases for each quark field $u_L^k e^{(i\phi_{uk})}$ and $d_L^j e^{(i\phi_{dj})}$, the Equation 1.11 becomes Equation 1.12.

$$V_{kj} e^{i(\phi_{dj} - \phi_{uk})} = V_{kj}^* e^{i(\phi_{uk} - \phi_{dj})} \tag{1.12}$$

Assuming the complex phase of the kj-th element in CKM matrix is θ_{kj}, it's obviously required Equation 1.13 to hold.

$$\theta_{kj} = \phi_{uk} - \phi_{dj} \tag{1.13}$$

If the number of generations in quark family is 3 or more, the non-physical phases can not render proper values to ensure the hold of Equation 1.13, and there will always be one irreducible complex phase parameter in the CKM matrix in the existence of three generations of quarks, which means CP symmetry is no longer conserved in the weak interactions.

The 3×3 unitary CKM matrix can be written as Equation 1.14 based on the quark fields it connects using Equation 1.5 and 1.6.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.14}$$

It can be parameterized into the form of Equation 1.15.

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{-i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{-i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{-i\delta} & -c_{12}c_{23} - s_{12}c_{23}s_{13}e^{-i\delta} & c_{23}c_{13} \end{pmatrix} \tag{1.15}$$

where the $c_{jk} = \cos(\theta_{jk})$ and $s_{jk} = \sin(\theta_{jk})$, and δ is the irreducible complex phase. By measuring the relative branching ratio of $b \to c$, $s \to u$ and $b \to u$ in tree level transitions as shown in Equation 1.16.

$$|V_{ub}| \ll |V_{cb}| \ll |V_{us}| \tag{1.16}$$

The relations in Equation 1.17 are often used to simplify CKM matrix presentation.

$$s_{13} = \lambda, s_{23} = A\lambda^2, s_{13}e^{i\delta} = A\lambda^3(\rho - i\eta) \tag{1.17}$$

By using Equation 1.17, CKM matrix is parameterized as Equation 1.18.

$$V_{CKM} = \begin{pmatrix} 1 - 1/2\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - 1/2\lambda^2 & A\lambda^2 \\ A\lambda^3(\rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.18}$$

Using the unitary condition, the Equation1.19 is obtained.

$$1 + \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0 \tag{1.19}$$

Using Equation 1.19, 1.20 and 1.21, the shape of CKM triangle can be defined on the complex plane in Figure 1-2.

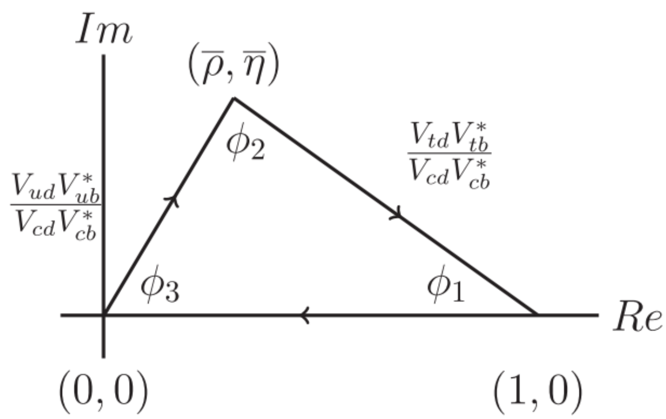


Figure 1-2: The unitary triangles of CKM[12].

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \tag{1.20}$$

$$1 - (\bar{\rho} + i\bar{\eta}) = -\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \tag{1.21}$$

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 been precisely measured due to the small hadronic uncertainties. Flavor-Changing-Neutral-Current (FCNC) processes can rise through the $B_d^0 - \bar{B}_d^0$ mixing in box diagram, and it's believed that potential NP processes might contribute to the difference in 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.

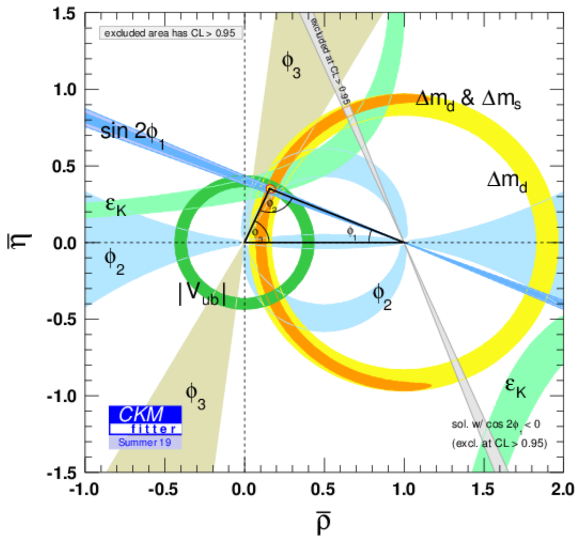


Figure 1-3: The CKM triangle fit in the complex plane of $\bar{\rho} - \bar{\eta}$.[12]

1.4 Time Dependent CP violation

1.4.1 CP violation in neutral B system

The ϕ_1, ϕ_2 and ϕ_3 are essentially measuring the CKM CP violating phase since there's 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 1.22 and 1.23.

$$\phi_1 = Arg(-\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}) \tag{1.22}$$

$$\phi_2 = Arg(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}) \tag{1.23}$$

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 1.24.

$$M_\Gamma = \begin{bmatrix} m - i/2\Gamma & M_{12} - i/2\Gamma_{12} \\ M_{12}^* - i/2\Gamma_{12}^* & m - i/2\Gamma \end{bmatrix} \tag{1.24}$$

Considering the time evolution of mass eigenstates, the time-dependent states can be shown as Equation 1.25 and 1.26 by using the notation of $B_{H,L}$ as physical states at $t = 0$).

$$B_H(t) = e^{-im_H t} e^{-\Gamma_H t/2} B_H \tag{1.25}$$

$$B_L(t) = e^{-im_L t} e^{-\Gamma_L t/2} B_L \tag{1.26}$$

where $M_{H,L}$ and $\Gamma_{H,L}$ are the masses and decay widths of two mass eigenstates. By expanding the mass eigenstates using flavor eigenstates, which are shown in Equation 1.27 and 1.28.

$$B(t) = (1/2p)e^{-im_H t} e^{-\Gamma_H t/2}(pB + q\bar{B}) + (1/2p)e^{-im_L t} e^{-\Gamma_L t/2}(pB - q\bar{B}) \tag{1.27}$$

$$\bar{B}(t) = (1/2q)e^{-im_H t} e^{-\Gamma_H t/2}(pB + q\bar{B}) - (1/2q)e^{-im_L t} e^{-\Gamma_L t/2}(pB - q\bar{B}) \tag{1.28}$$

Replacing $g_\pm(t) = \frac{1}{2}(e^{-im_H t - \Gamma_H/2t} \pm e^{-im_L t - \Gamma_L/2t})$, Equation 1.27 and 1.28 become Equation 1.29 and 1.30.

$$B(t) = g_+(t)B + \frac{q}{p} g_-(t)\bar{B} \tag{1.29}$$

$$\bar{B}(t) = g_+(t)\bar{B} + \frac{p}{q} g_-(t)B \tag{1.30}$$

Considering all the phase-spaces of the decay from flavor eigenstates to final states $f(\bar{f})$ are included in the amplitudes $\mathcal{A}_f(\bar{\mathcal{A}}_f)$, one needs to expand the flavor eigenstates using the final states amplitudes to have the differential decay rate $\Gamma(B \to f, t)$. From $B(t) \propto \mathcal{A}_f \psi_f + h.c$ and ($\bar{B}(t) \propto \bar{\mathcal{A}}_f \psi_{\bar{f}} + h.c$), combined with Equation 1.29 and

1.30, the decay rate can be shown in Equation 1.31 and 1.32.

$$\Gamma(B \to f, t) = |\mathcal{A}_f|(|g_+(t)|^2 + |\lambda_f|^2|g_-(t)|^2 + 2Re(\lambda_f g_+^*(t)g_-(t))) \tag{1.31}$$

$$\Gamma(\bar{B} \to \bar{f}, t) = |\bar{\mathcal{A}}f|(|g_+(t)|^2 + |\bar{\lambda}_{\bar{f}}|^2|g_-(t)|^2 + 2Re(\bar{\lambda}_{\bar{f}} g_+^*(t)g_-(t))) \tag{1.32}$$

where the parameter λ_f and $\bar{\lambda}_{\bar{f}}$ can be defined as Equation 1.33 and 1.34.

$$\lambda_f \equiv (q/p)(\bar{\mathcal{A}}_f/\mathcal{A}_f) \tag{1.33}$$

$$\bar{\lambda}_{\bar{f}} \equiv (q/p)(\mathcal{A}_{\bar{f}}/\bar{\mathcal{A}}_{\bar{f})} \tag{1.34}$$

The q/p is introduced by the coefficient of mass eigenstates from weak eigenstates. Using the Hamiltonian matrix, q/p can be presented using Equation 1.35

$$q/p = \frac{\Delta M - i/2\Delta\Gamma}{2(M_{12} - i/2\Gamma_{12})} \tag{1.35}$$

where the M_{12} and Γ_{12} stands 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. It's 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 1.36.

$$\begin{aligned} A_{CP}(t) &\equiv \frac{\Gamma(B \to f, t) - \Gamma(\bar{B} \to \bar{f}, t)}{\Gamma(B \to f, t) + \Gamma(\bar{B} \to \bar{f}, t)} \\ &= \frac{\mathcal{S} sin(\Delta M t) - \mathcal{A} cos(\Delta M t)}{cosh(\Delta\Gamma t/2) + A^f_{\Delta\Gamma} sinh(\Delta\Gamma t/2)} \end{aligned} \tag{1.36}$$

where

$$\mathcal{S} = \frac{2Im(\lambda_f)}{1 + |\lambda_f|^2} \tag{1.37}$$

$$\mathcal{A} = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \tag{1.38}$$

$$A^f_{\Delta\Gamma} = -\frac{2Re(\lambda_f)}{1 + |\lambda_f|^2} \tag{1.39}$$

From Equation 1.37 and 1.38, the time-dependent CP violation parameters $\mathcal{S}$ and $\mathcal{A}$ are dependent on the parameter λ_f.

1.4.2 ϕ_1 from $B^0 \to J/\psi K_S^0$

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 $B_d^0 - \bar{B}_d^0$ 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 $B^0 \to J/\psi K_S^0$, considering $\Delta\Gamma$ can be treated as zero in the SM in this case[13], Equation 1.36 can be reduced to Equation 1.40.

$$A_{CP}(t) = \mathcal{S}\mathrm{sin}(\Delta M t) - \mathcal{A}\mathrm{cos}(\Delta M t) \tag{1.40}$$

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

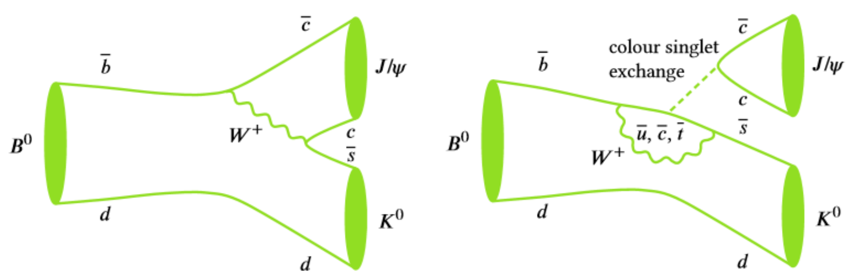


Figure 1-4: The dominated tree-level (left) and the suppressed loop-level (right) of $B \to J/\psi K^0$, in which K^0 particles are detected K_S^0[14].

Using the relation $|V_{ub}| \ll |V_{cb}| \ll |V_{us}| < |V_{cs}|$, it's 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 CP eigenvalue. Given $\eta_f = 1$ and $|\lambda_f| = 1$ in $B^0 \to J/\psi K_S^0$, from 1.37, CP violation parameters can be presented as shown in Equation 1.41.

$$\mathcal{S} = Im(\lambda_f) = -sin(\phi_d)\eta_f = sin(2\phi_1);\ \mathcal{A} = 0 \tag{1.41}$$

From Equation 1.41 , 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[15]. 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 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 1.42

$$\mathcal{A}_f = \lambda_u^s T_f + \lambda_c^s P_f + A_f^{NP} \tag{1.42}$$

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, direct CP violation vanishes and 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 1.43.

$$\Delta S = \mathcal{S}_f - \mathcal{S}_{J/\psi K_s^0} \tag{1.43}$$

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[15]:

- Branching ratios are affected at $\mathcal{O}(|\epsilon_{uc} r_f^T|, |r_f^{NP}|)$
- Direct CP violation 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 $\Delta\mathcal{S} = 2cos2\phi_1 sin\phi_3|\epsilon_{uc}|\mathrm{Re}(r_f^t) + \Delta\mathcal{S}^{NP}$. This

suggests that non-zero ΔS without the NP effect is still allowed in a small scale within the SM. Therefore, in the precised measurement using the future Belle II data, it's important to understand ΔS within the SM correction to properly explain the experiment results. The ΔS should be larger than the SM allowed value by 5σ to be called as the evidence of the NP, 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 reported the time-dependent CP analysis on various $b \to q\bar{q}s$ which experimentally showed that the difference on ϕ_1 has a margin for NP effects[16], 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 $B^0 \to K_S^0 K_S^0 K_S^0$ 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 positive (CP eigenvale $= +1$). There's no up-quark shown in the final states, 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 $B^0 \to K^+K^-K_S^0$[17]. 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 the close CP violation as $J/\psi K_S^0$[17]. To be noted, 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 a calculation on $\Delta\mathcal{S}$ within the SM can be find in Ref.~[18]. 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.~[19]. In general, 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 small compared with the experimental uncertainties. Thus, in the comparison of SM-allowed ΔS and the experimental results, we current only focus on the statistical and systematic uncertainties from the experiment and ignore the QCD model uncertainty. 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.

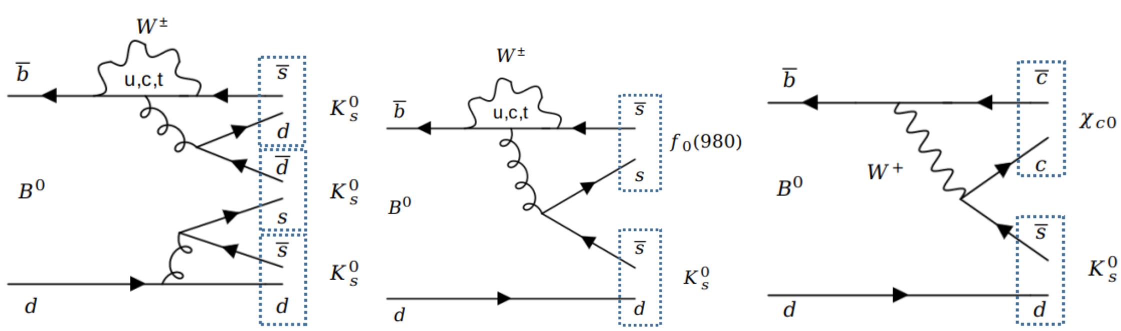


Figure 1-5

The current result of $B^0 \to J/\psi K_S^0$ using the full Belle data is presented as $\mathcal{S}_{J/\psi K_S^0} = +0.670 \pm 0.029(\text{stat}) \pm 0.013(\text{syst})$[15]. In the meantime, the latest result from $B^0 \to K_S^0 K_S^0 K_S^0$ using the full Belle data[20] is presented as: $\mathcal{S}_{3K_S^0} = -0.71 \pm 0.23(\mathrm{stat}) \pm 0.05(\mathrm{syst})$, and the result from BaBar [21] is: $\mathcal{S}_{3K_S^0} = -0.94^{+0.21}_{-0.24}(\mathrm{stat}) \pm 0.06(\mathrm{syst})$. 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 ~ 0.005 at $50\mathrm{ab}^{-1}$ Belle II data[15], 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[22], which requires that the Belle II results should be prepared for $\Delta\mathcal{S} < 0.2$ caused by 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[23]. The 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 $\Delta\mathcal{S}_{3K_S^0} \sim 0.25$, the total uncertainty should be less than 0.04 to claim a 5σ deviation from the SM upper limit 0.05 for the appearance of the NP effects, which is close to the expected total uncertainty in Table 1.1.

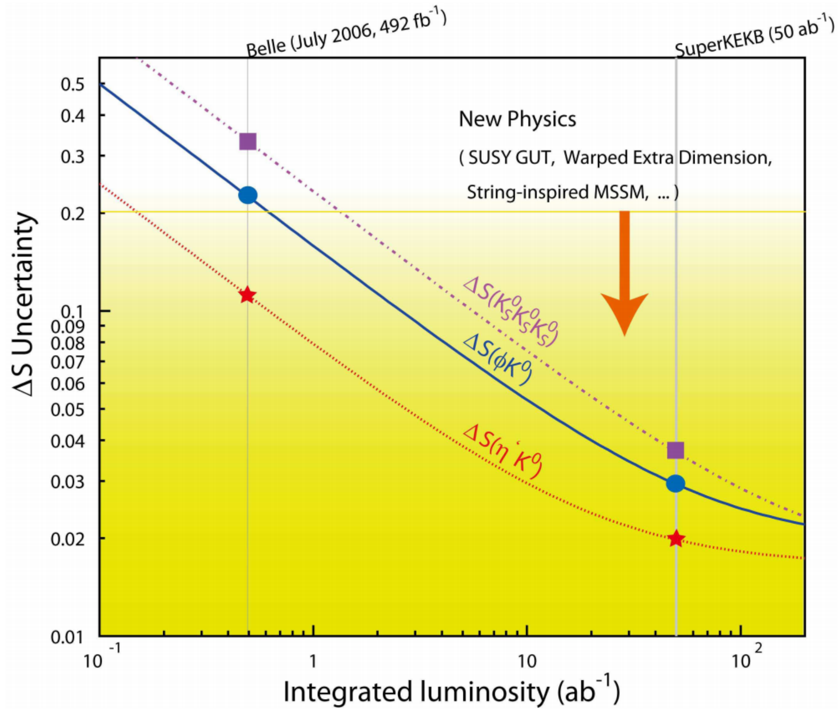


Figure 1-6: Expected sensitivity of $\Delta\mathcal{S}$ with respect to the integrated luminosity of the Belle II future data from the Belle II technical design report[22].

Table 1.1: $\Delta\mathcal{S}$ estimated total uncertainties with respect to the integral luminosities in the Belle II technical design report[22]. Three decay modes receive the same potential NP effects due to the same weak phases involved in the decay processes[17].

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
$\Delta\mathcal{S}_{K_S^0 K_S^0 K_S^0}$	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 $\Upsilon(4S)$ resonances 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 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.

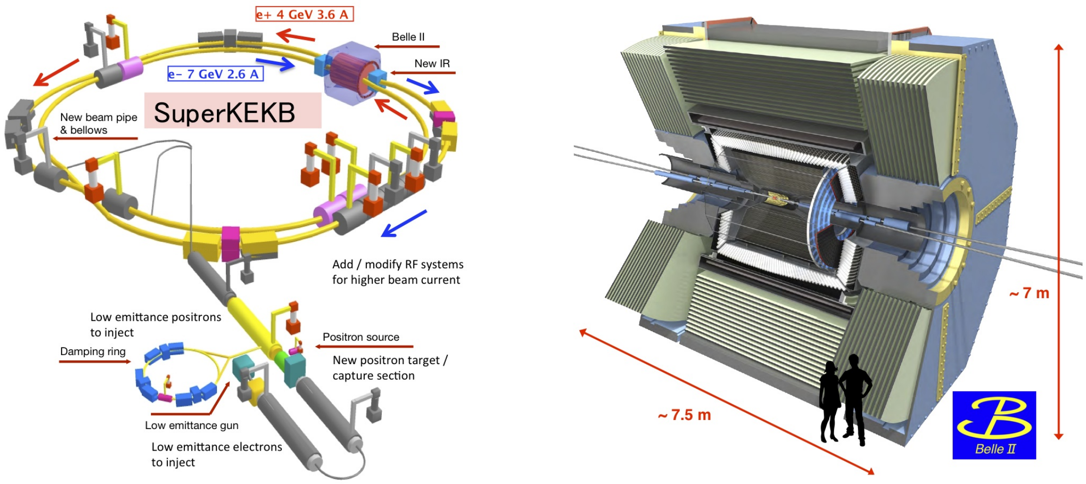


Figure 2-1: The schematic view of SuperKEKB and Belle II detector [22].

Table 2.1: SuperKEKB parameters for low energy (LER) and high energy (HER) rings[15].

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}		cm^{-2}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 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.

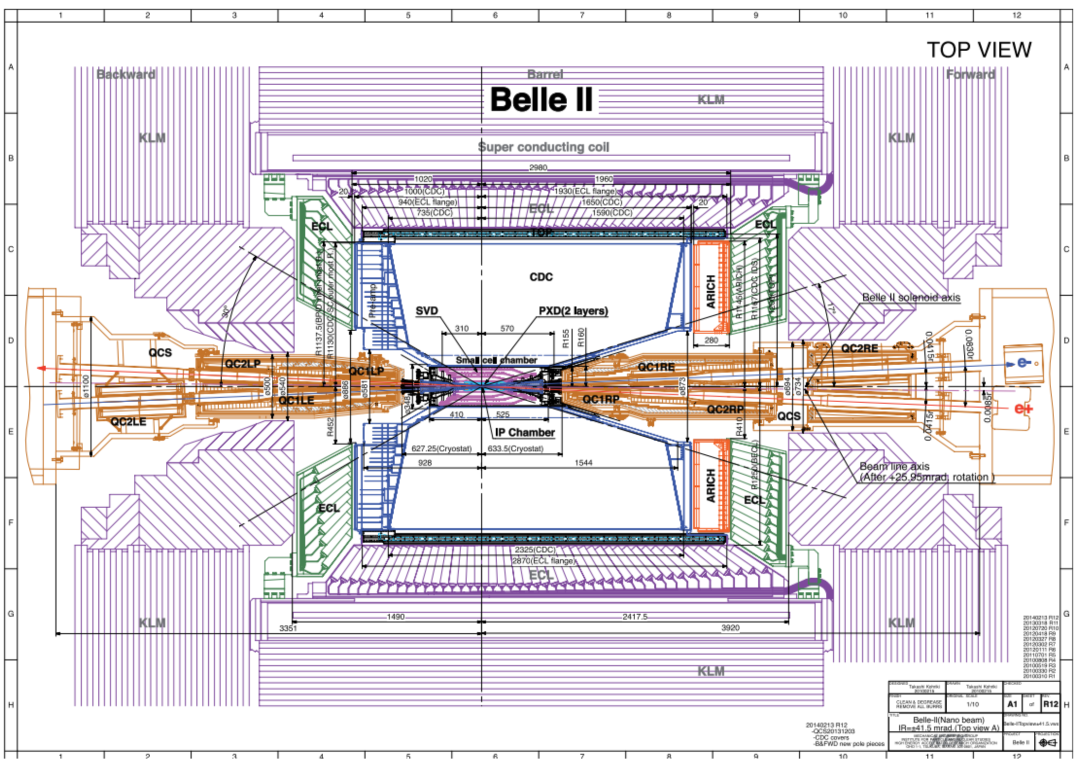


Figure 2-2: The Belle II detector top view [15].

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[22] 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, which is 17 degrees (forward) to 150 degrees (backward). The pixel sensor is a monolithic structure with current-digitizing electronics at the end of the senor which 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

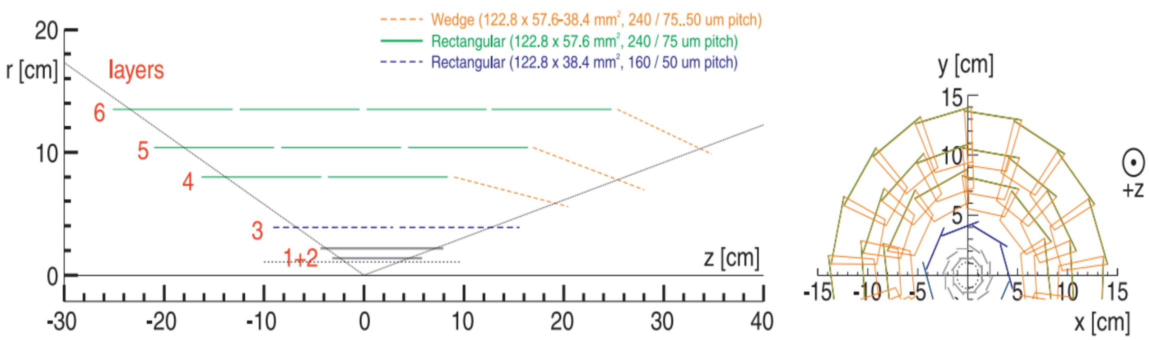


Figure 2-3: A schematic view of of PXD and SVD. Two PXD layers are in grey, SVD layer 3 is in blue and layer 4, 5, 6 are in green[22].

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

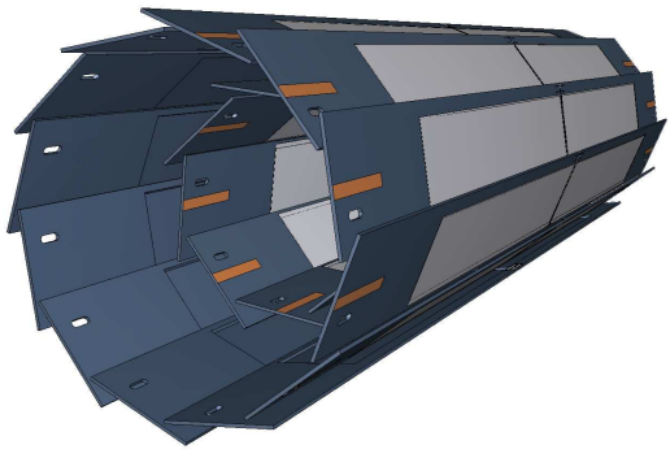


Figure 2-4: The geometry of sensors on PXD where the light grey surfaces are DEPFET sensors with a thinness of 50 microns. The full length including the out modules is 174 mm [22].

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 strips on $r-\phi$ direction (transverse direction) and the latter is for strips on the z direction (beamline direction). To suppress the background hits, a readout chip with a fast shaping time of $\mathcal{O}$(50 ns) is indispensable. The APV25 chip [24] 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 which 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.

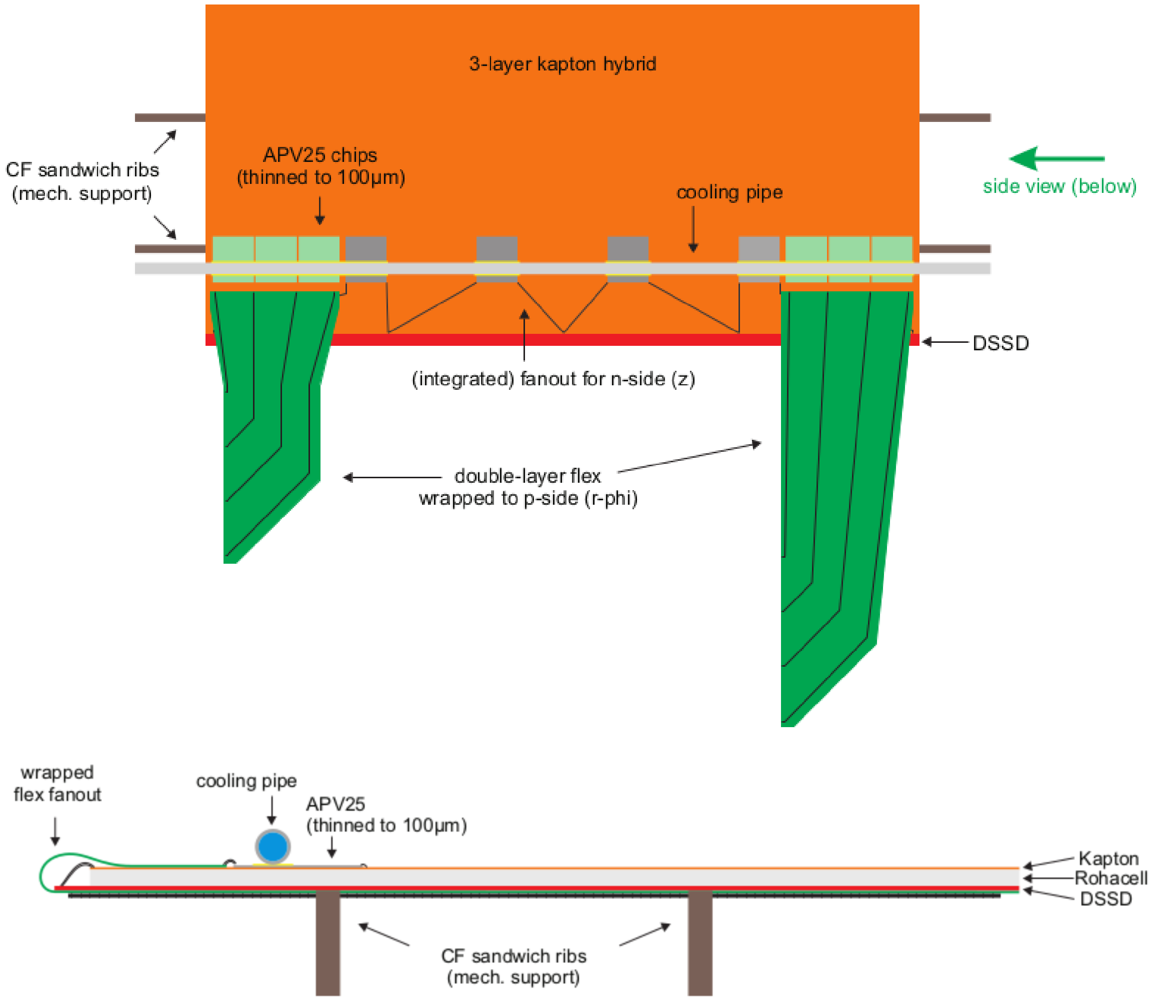


Figure 2-5: The top and side views of Origami chip-on-sensor design for DSSDs of SVD. Top: the APV25 chips in grep read out the same side sensors channel while chips in green read out the sensors on the opposite side using wrapped-around flex pieces. Bottom: side view of the Origami design shows the location of wrapped flex which connects the strips of the bottom sides which are placed at the left edge.[22]

2.3 Central drift chamber (CDC)

The central drift chamber (CDC) is the core component of spectrometer in the Belle II, which 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 out 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[15][22]. Such a design can utilize the information from axial and stereo wires to construct a full 3 dimensional hits which reflects helix tracks in 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. Finally, 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[22] 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 ring (ARICH).

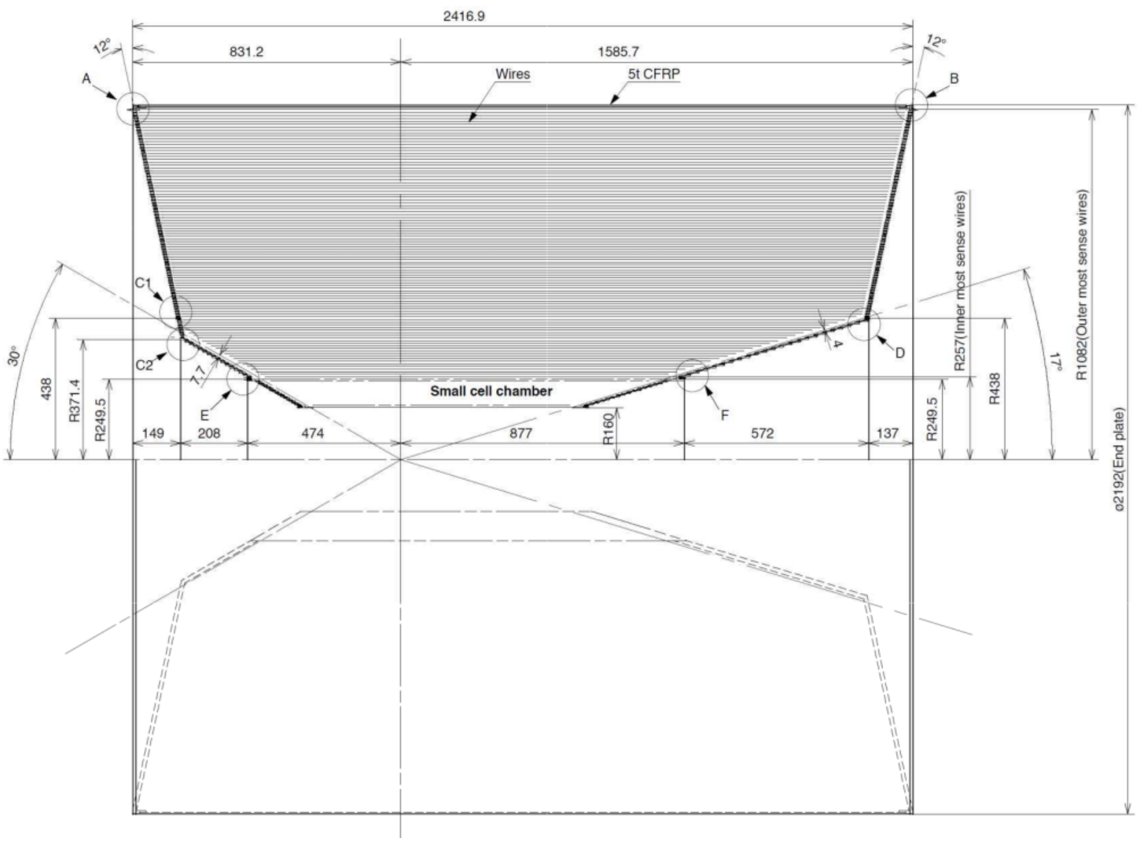


Figure 2-6: CDC structure schematic view [22].

TOP is the specialized detector that can reconstruct Cherenkov radiation's 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 [22]. As the key component of the photon detector, the squared shape MCP PMTs, donated as SL-10 [25], 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 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 image of cosmic muon on the HAPD sensors are shown in Figure 2-10.

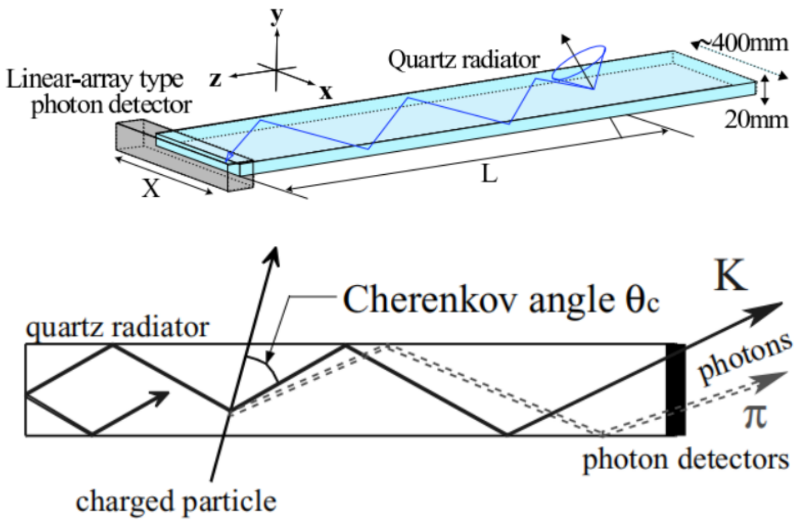


Figure 2-7: Conceptional view of TOP counter (up) and its imaging process of $K^{\pm}$ and $\pi^{\pm}$ (down) [22] for PID purpose.

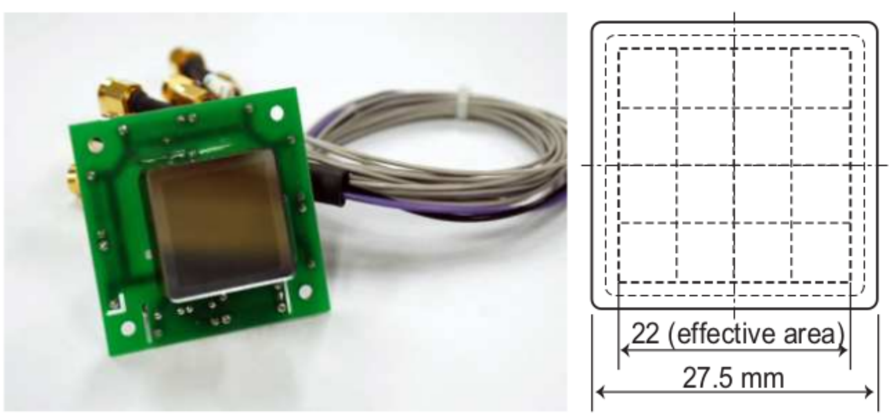


Figure 2-8: SL-10 MCP PMT (left) and the schematic view of 4×4 anode (right) [22]

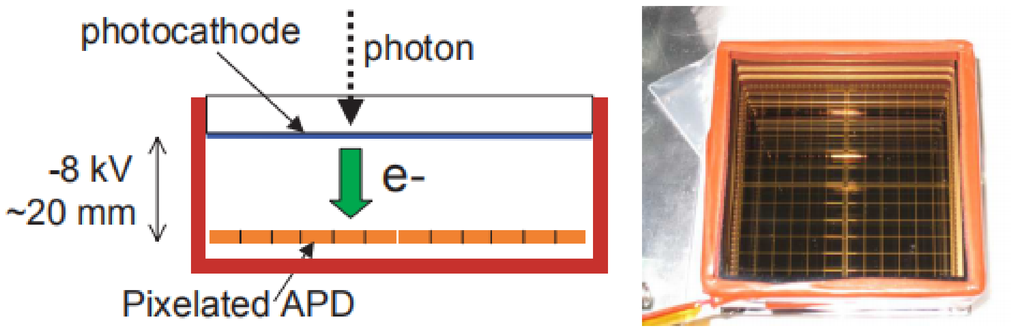


Figure 2-9: Photon-electrons acceleration (left) and pixelated APD (right) at the end [22].

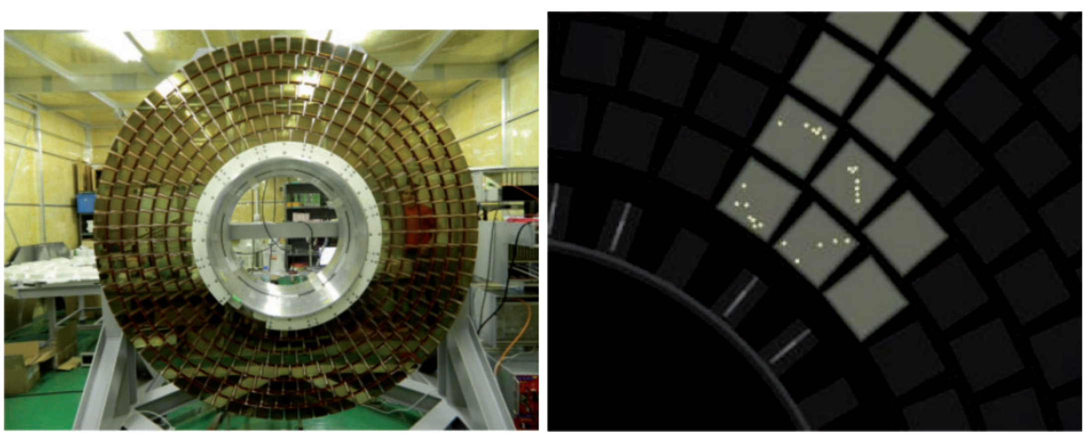


Figure 2-10: ARICH detector (left) and the ring image of cosmic muon on the HAPD sensors[15].

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 [22].

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 [15].

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, as 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

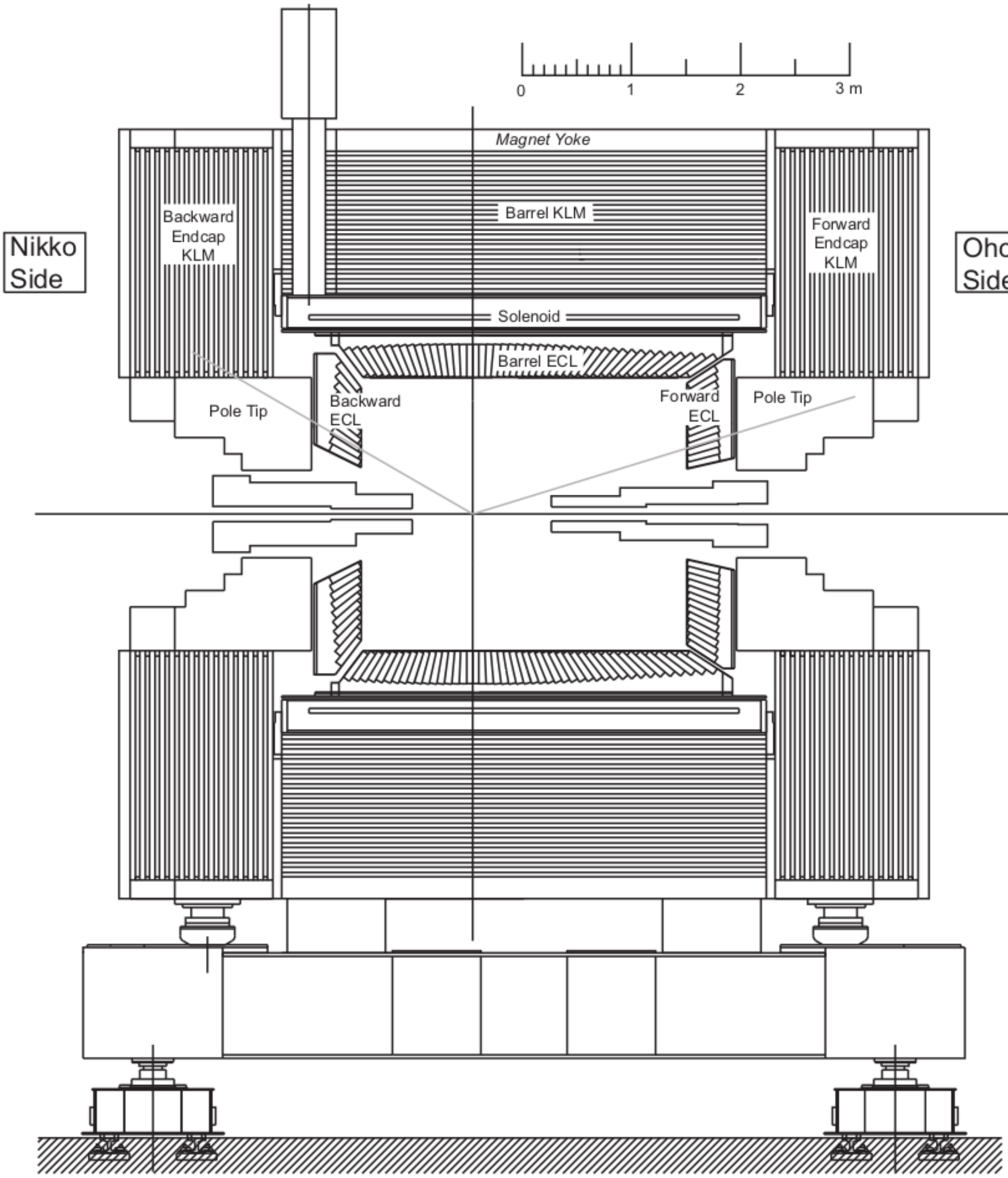


Figure 2-11: The side view of KLM in between the ECL and the solenoid, which the grey lines presents the nominal acceptance angle of the Belle II [22].

event builder and eventually written to the offline storage.

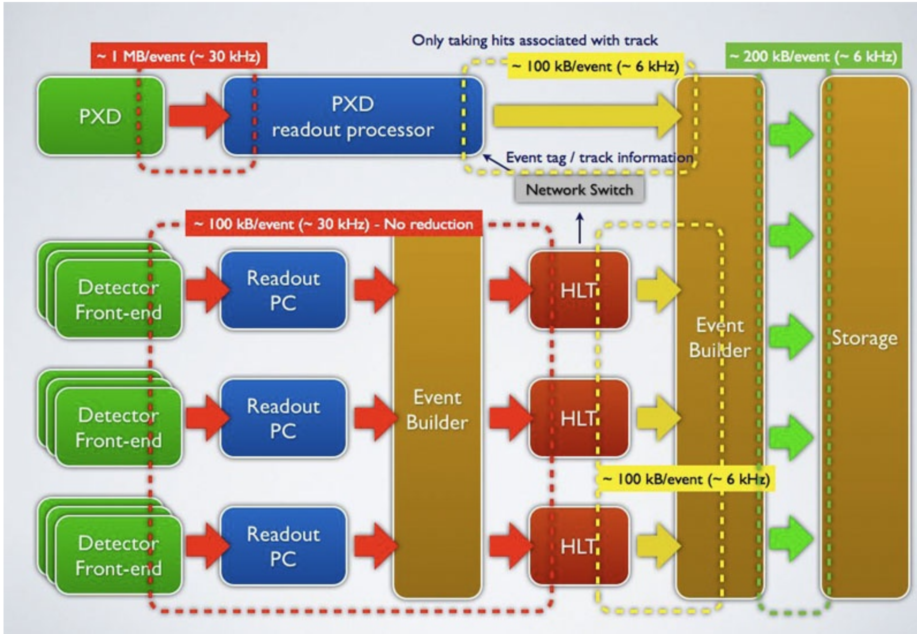


Figure 2-12: The Belle II DAQ workflow with HLT between two event builder to reduce the original 30 kHz event rate down to about 6 kHz for offline storage.

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 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.

Table 2.2: Simulated beam background rate [15]

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 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 languages, but the use of Python is also allowed when it shows clear advantages, such as steering the analysis workflow.

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[26], 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 [27]. The detail information about the core structure of BASF2 can be found in Ref.∼[28].

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.

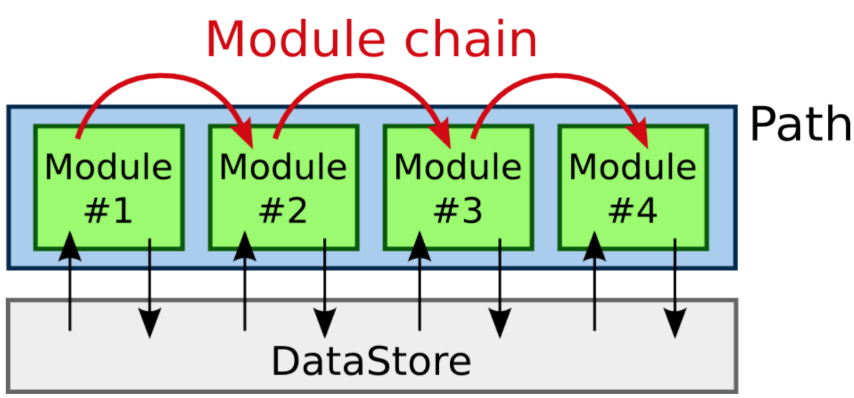


Figure 2-13: The module-based analysis workflow in BASF2.

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’s 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) [29]. 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's 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 [30] and *GEANT4* as the simulator of detectors [31]. 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, 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 $B^0 \to K^0_S K^0_S K^0_S$. The mother particle of the decay chain is Υ(4S), then it decays into a pair of $B^0 - \bar{B}^0$ at branching fraction of 100%, with the model *EvtVSSMix*[30] describing the decay model. Then, one of the B^0 meson is set to decay into three K^0_S 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 $B^0 \to K^0_S K^0_S K^0_S$ 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 Υ(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 Υ(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

Table 2.3: Production cross section for different hadronic flavors from collision at $\sqrt{s} = 10.58$ GeV used in Belle II *generic MC* [15].

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

$B^0 \to K_S^0 K_S^0 K_S^0$ 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) [32]. 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

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}

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 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}[33]. The experiment 14 is not used due to the unfinished processing of the latest experiment data.

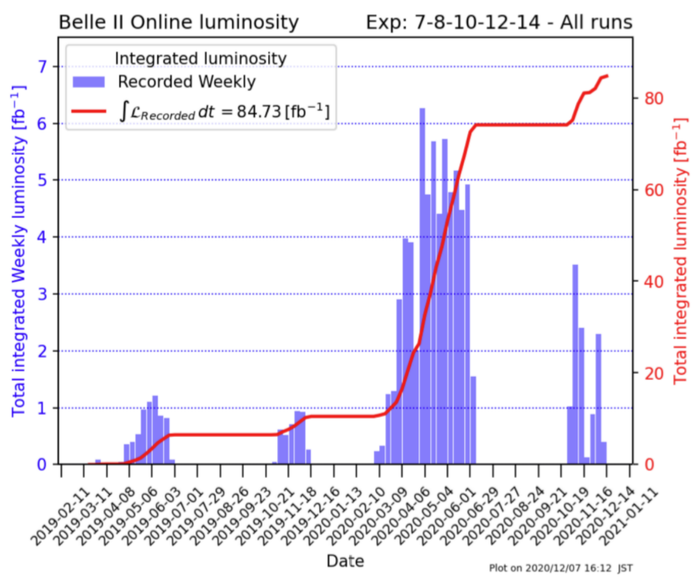


Figure 2-14: Belle II online luminosity from 2019 April to the end of 2020. The experiment 7 and 8 were conducted during 2019 March to June. The experiment 10 was conducted during 2019 October to 2019 November. The experiment 12 was conducted during 2020 February to June. The experiment 14 was conducted 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 [32]. 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 as final states. 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's 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, which leads to the 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, leading to large uncertainties on the vertex positions. Such a degradation of precision of the vertex position eventually results in the large uncertainties in decay time difference, which is 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. Therefore, the flight length of K_S^0 is comparable with the scale of VXD size. In the Belle II energy scale, the flight length of K_S^0 is in a range 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

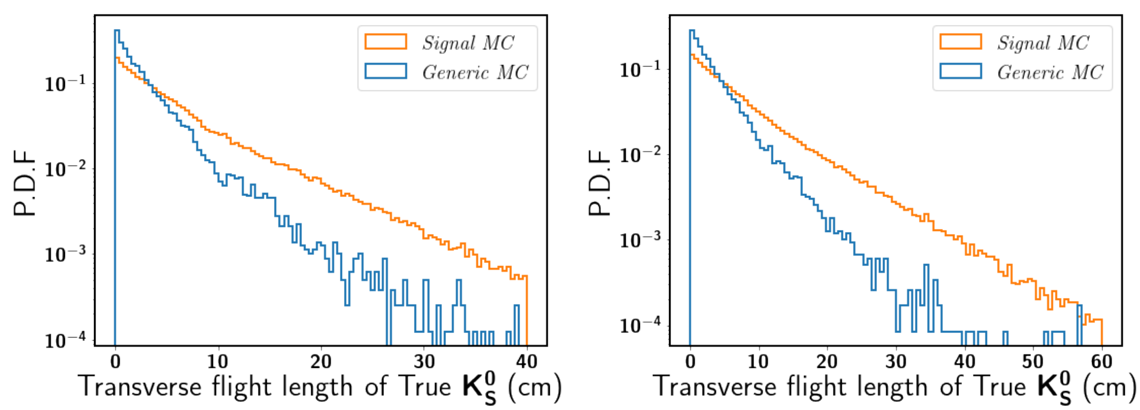


Figure 3-1: The left is the transverse flight length distribution and the right is the total flight length distribution from true K_S^0. The blue is from *generic MC* and the orange is from *signal MC*. Both plots are normalized.

of K_S^0 in *generic MC* is different from the ones from the *signal MC*. 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 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* [34]. After the vertex fit, the fitted invariant mass of K_S^0 is constraint 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's 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 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,

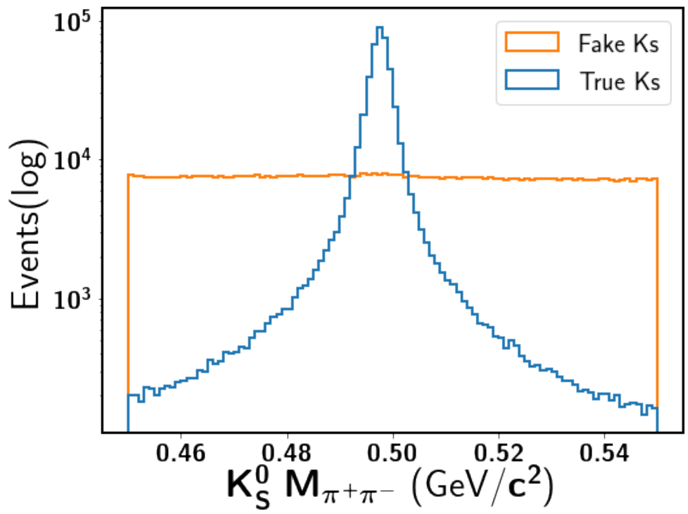


Figure 3-2: M of K_S^0 from cut-based selection in *signal MC*. The blue line is the true K_S^0 and the orange is the fake K_S^0. 200000 candidates are used in total.

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 $\mathrm{M}_{\pi^+\pi^-}$ 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, which significantly reduces 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.

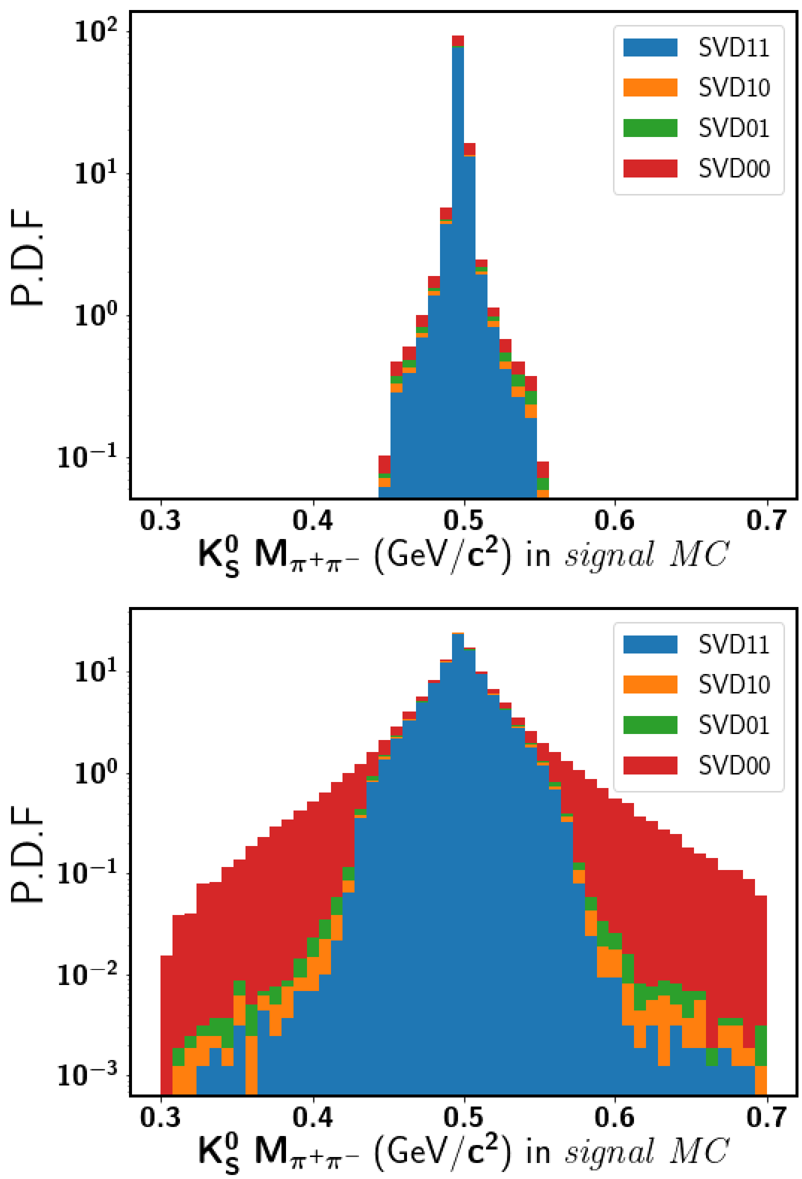


Figure 3-3: The invariant mass before (top) and after (bottom) vertex fit distribution based on SVD types, which shows a clear dispersion in *SVD00* particularly, indicating the inaccurate reconstruction of K_S^0 masses without SVD information.

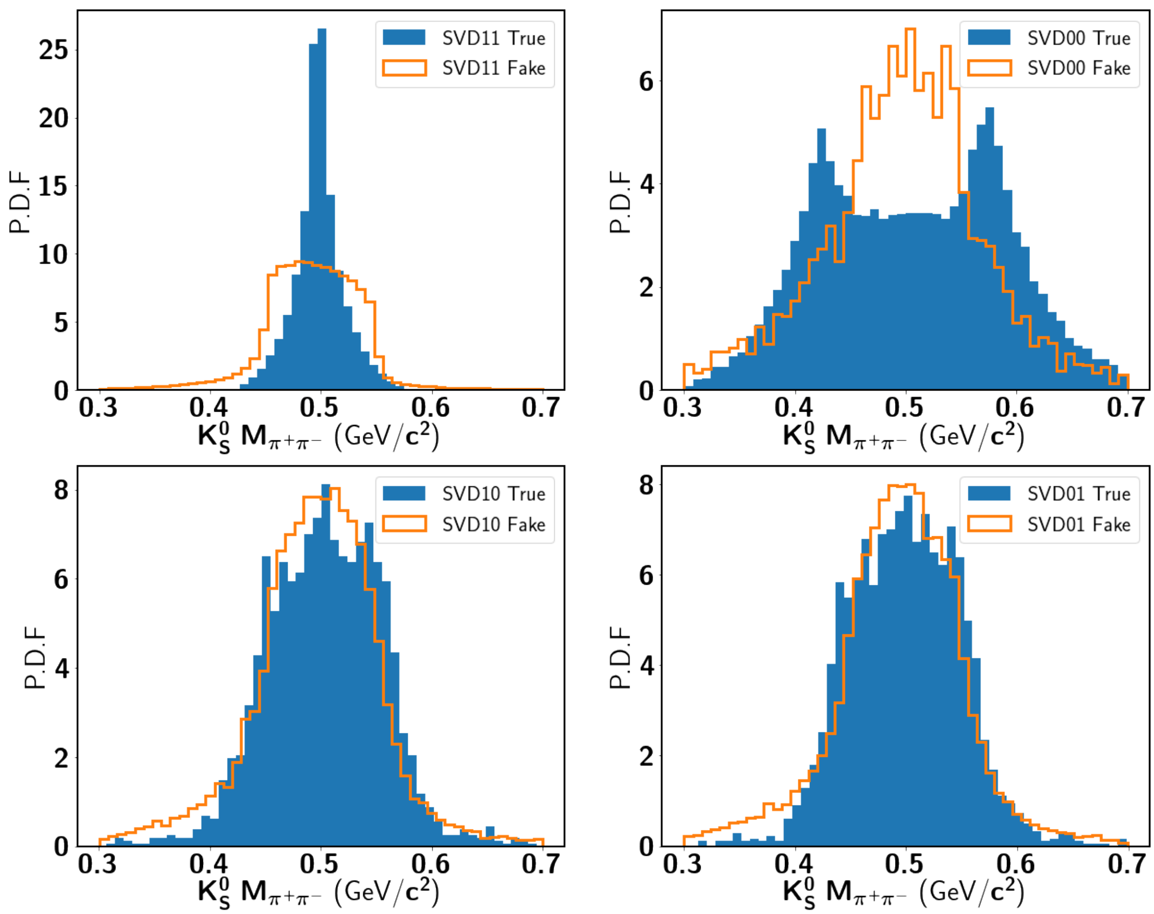


Figure 3-4: K_S^0 invariant mass after vertex fit, where the sideband regions are excluded in these distributions to further reject fake K_S^0

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
$\mathrm{M}_{\pi^+\pi^-}$ 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* [15] which outputs the two likelihood variables based on NeuroBayes algorithm [35]. The Belle tool defines the goodness of K_S^0, called *nb_nolam* and *nb_vlike*. 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 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's 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.

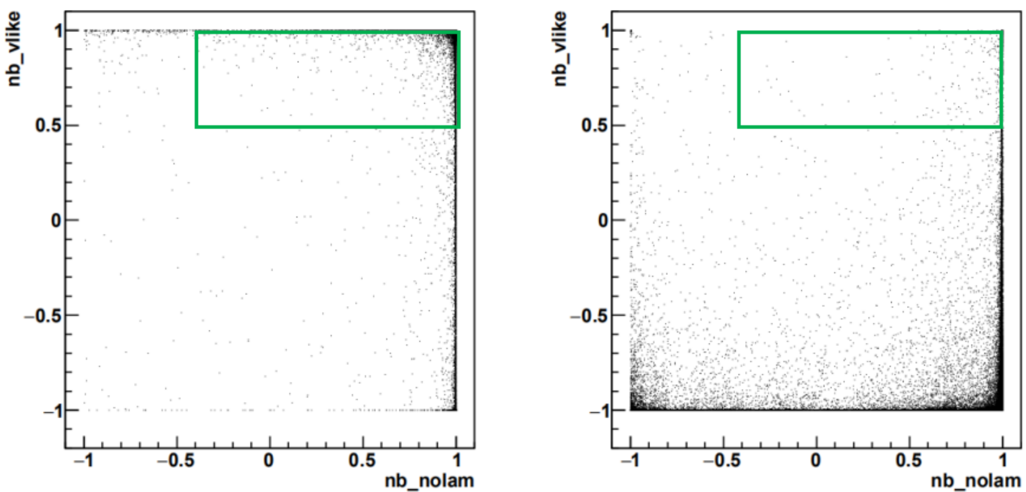


Figure 3-5: The distribution of two variables outputs: *nb_nolam* and *nb_vlike* for K_S^0 candidates from Belle *signal MC*. The left is from true K_S^0 and the right is from the fake K_S^0. In Belle, the standard cuts for K_S^0 is *nb_vlike* > 0.5 and *nb_nolam* > -0.4, which is shown as the green boxes [20].

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 [36]. Compared to other popular classification algorithms such as TMVA [37], scikit-learn [38] and XGBoost [39], FastBDT method is proven to be one order of magnitude faster during the training and applying phases [36]. 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 used 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's 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.

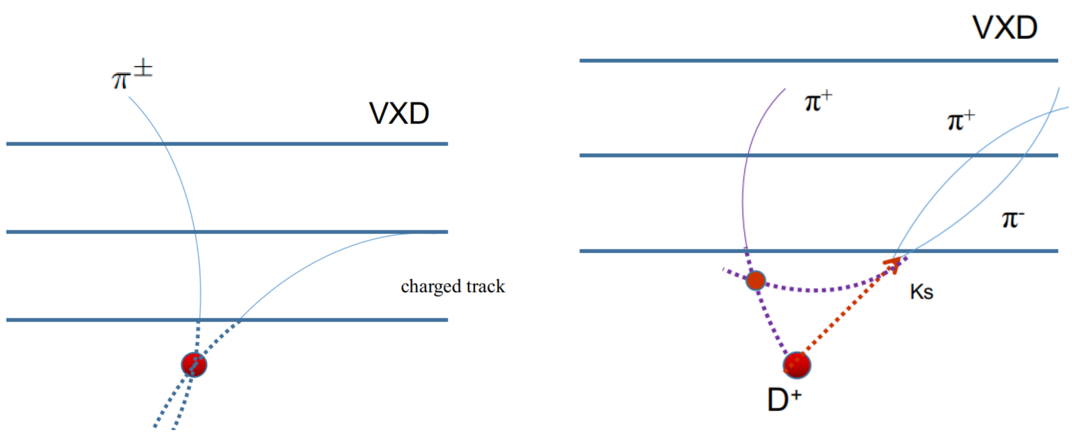


Figure 3-6: The left shows the case when a charged track (not $\pi^{\pm}$) combined with a charged pion to form a fake K_S^0, the right shows the case when two daughters are correctly reconstructed as pions but not from the correct mother particle, which is falsely taken as a K_S^0.

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_S^0, see Figure 3-7 right.

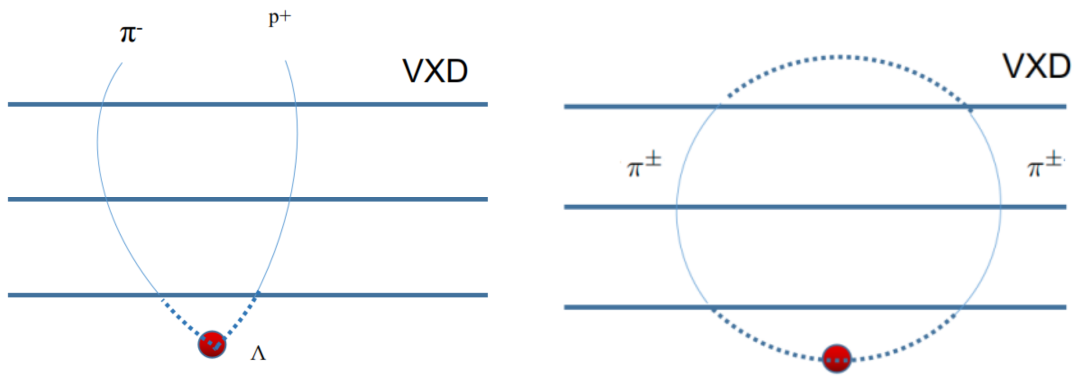


Figure 3-7: The left shows the $\Lambda \to p^+\pi^-$ decay shape that can be treated as K_S^0, the right shows a self-loop formed by a low p_T charged pion reconstructed as two separated tracks with a vertex.

3.2.3 Determination of training variables from K_S^0 decay

Given the characteristics of $K_S^0 \to \pi^+\pi^-$ discussed in the previous section, a set of variables as training features of *KsFinder* can be selected. The set includes variables related to K_S^0 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_S^0 vertex and momentum direction (named *cosVertexMomentum*) is regarding as the most useful variable to separate true and fake K_S^0, which is originally used as an extra cut in the first measurement of $B^0 \to K_S^0 K_S^0 K_S^0$ before MVA based K_S^0 classification tool was developed [40]. For instance, if a falsely reconstructed K_S^0 is made of two tracks, it's likely that the momentum direction of the fake K_S^0 is not aligned with the its vertex direction from IP. So the projection of vertex position of K_S^0 on the reconstructed momentum direction could be negative 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

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)	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

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.

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 and close range of the vertex position. The difference of the true and fake K_S^0 in the distribution of *significanceOfDistance* is not so large compared to that of *cosVertexMomentum*. FastBDT model 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

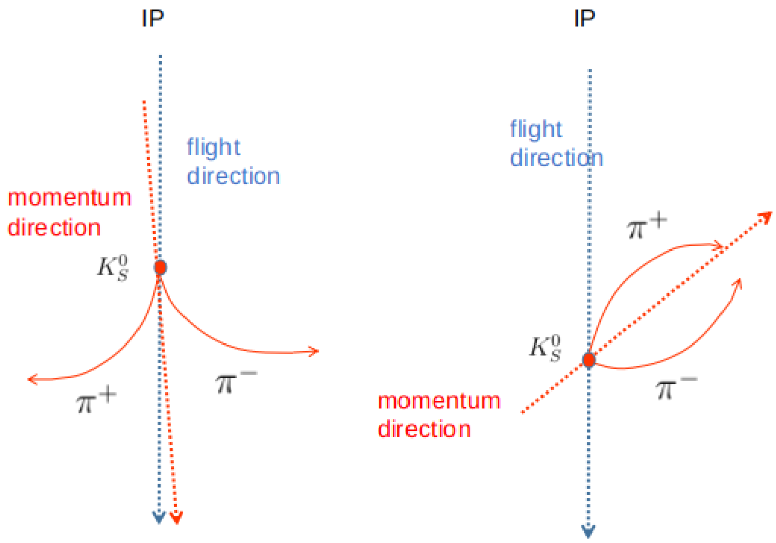


Figure 3-8: The left shows a true K_S^0 decay shape where the cosine angle of K_S^0 vertex position (blue dashed arrow) against reconstructed momentum direction (red dashed arrow) is positive. While the right shows a fake K_S^0 decay shape where the cosine angle of K_S^0 vertex position against the reconstructed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

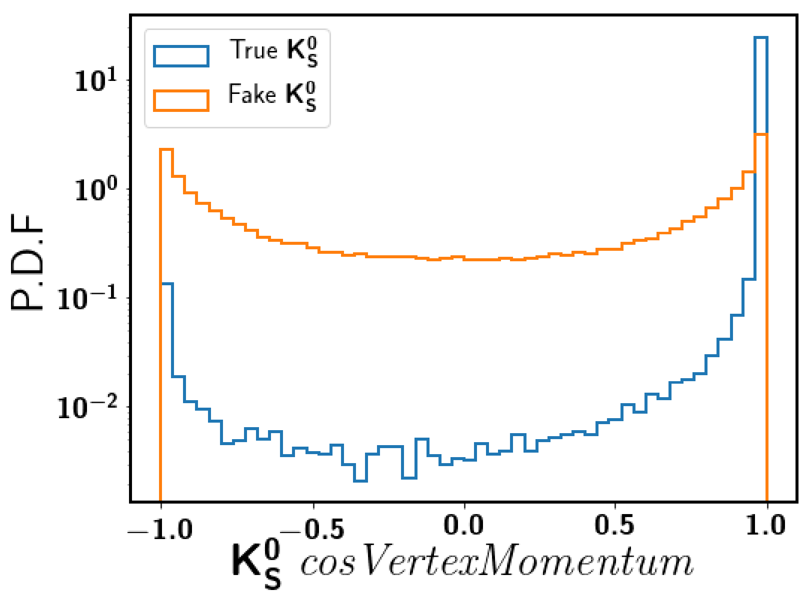


Figure 3-9: The distribution of $cosVertexMomentum$ using *signal MC*. The true and fake K_S^0 ratio is set to be 1:1, where the most true K_S^0 gives $cosVertexMomentum$ > 0.9.

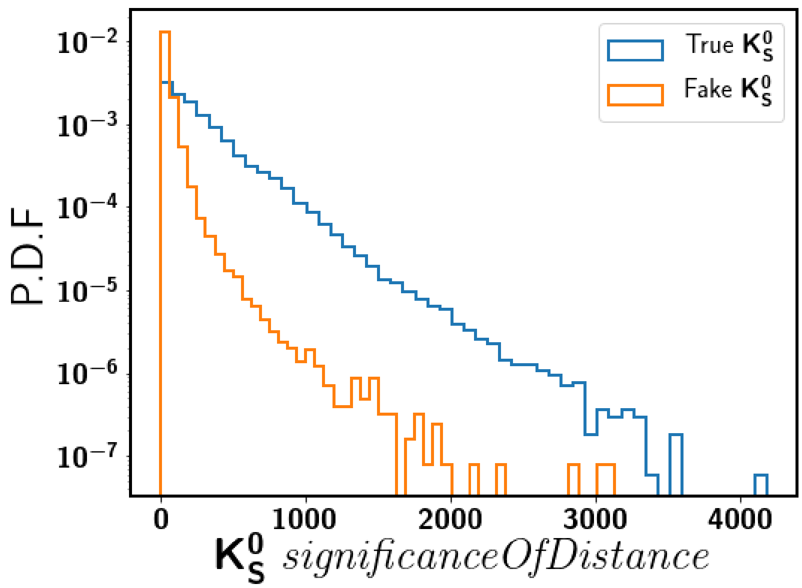


Figure 3-10: The distribution of *significanceOfDistance* using *signal MC*. The true and fake K_S^0 ratio is set to be 1:1, where the most fake K_S^0 are distributed in *significanceOfDistance* < 1000 region.

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 same. Different correlations helps improve the performance of decision trees in tuning of structure. For instance, a true K_S^0 flights longer due to larger momentum in general, so the number of daughter detector hits 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.

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

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

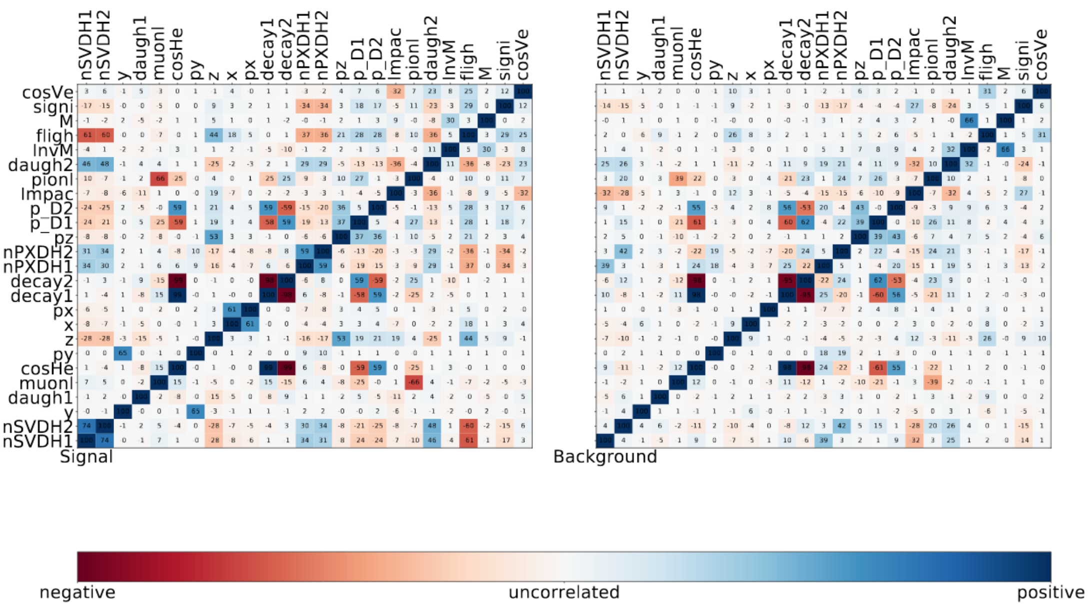


Figure 3-11: The correlation between input variables for *KsFinder*. As the given example, flight length has negative correlation with SVD hits in signal while uncorrelated in background.

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 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 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 and background rejection are calculated by cutting on the different values on *FBDT_Ks*, as defined in Equation 3.1 and 3.2.

$$\text{signal efficency} = \frac{\text{Number of true } K_S^0 \text{ with } \textit{FBDT\_Ks} > \text{cut value}}{\text{Number of all true } K_S^0} \tag{3.1}$$

$$\text{background rejection} = \frac{\text{Number of fake } K_S^0 \text{ with } \textit{FBDT\_Ks} < \text{cut value}}{\text{Number of fake true } K_S^0} \tag{3.2}$$

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

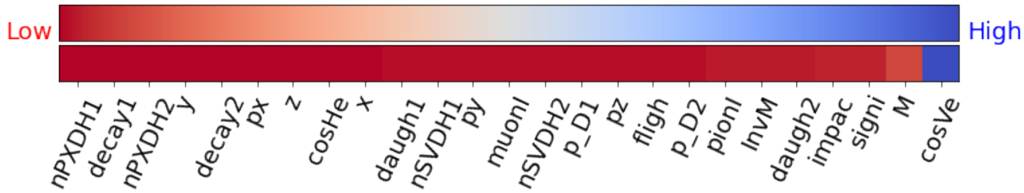
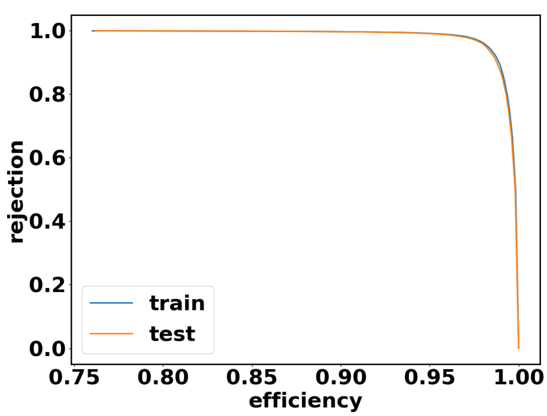


Figure 3-12: The importance rank of the input variables for *KsFinder*. The *cosVertexMomentum* is the most important variable.

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. 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.



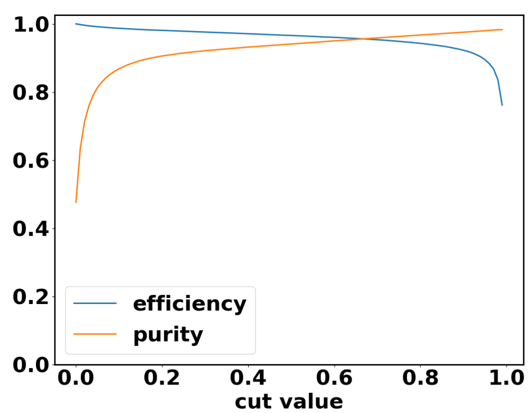


Figure 3-13: The left is ROC curve(blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of *KsFinder* output. The results are obtained by applying the weight file from training sample (a) to testing sample (a).

Because the ROC curves are consistent in the training and testing samples, it

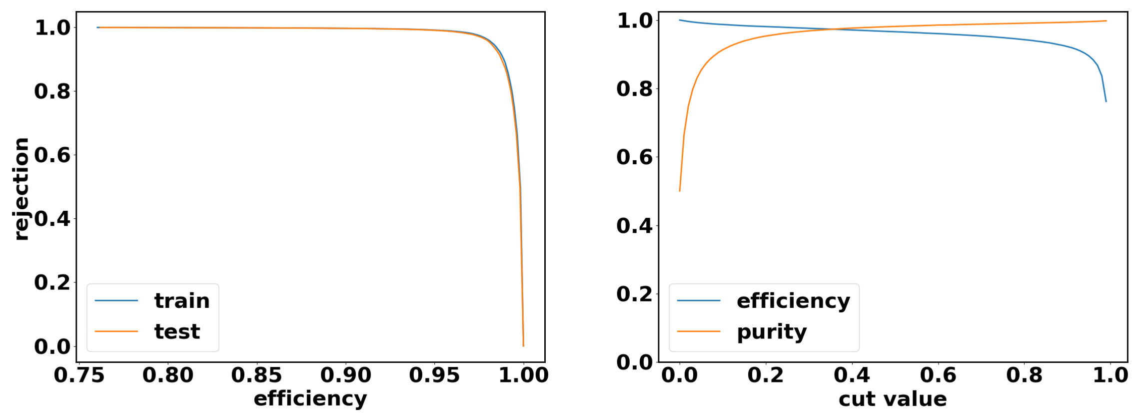
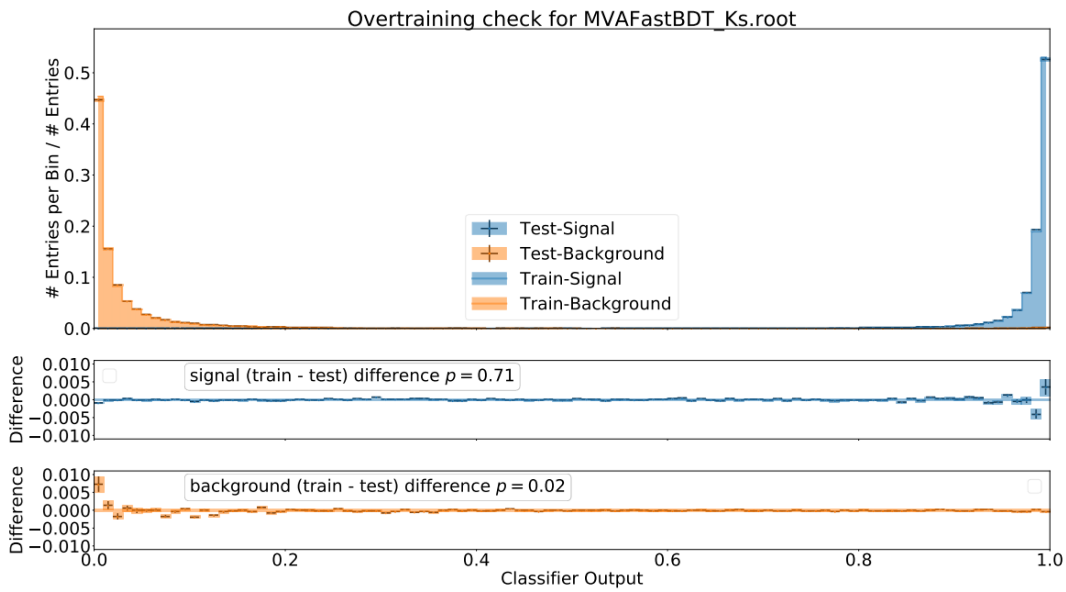
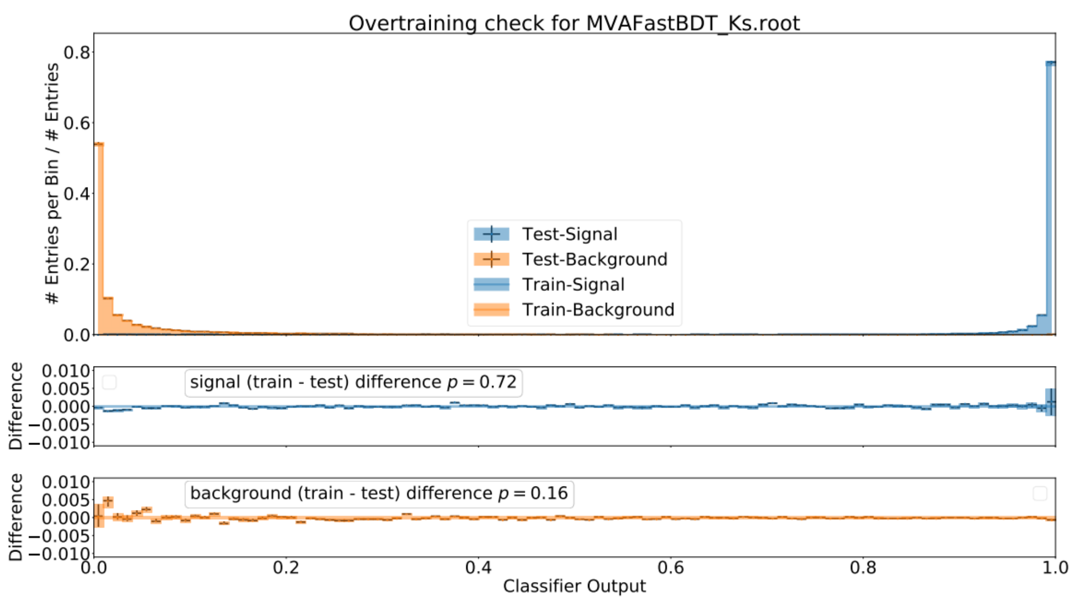


Figure 3-14: The left is ROC curve (blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of *KsFinder* output. The results are obtained by applying weight file from training sample (b) on testing sample (b).

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 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 3.3, where S and B is the number of true and fake K_S^0 after the cut, respectively. 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 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. 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.



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*.

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%

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*.

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.

$$\text{FOM} = \frac{S}{\sqrt{S+B}} \tag{3.3}$$

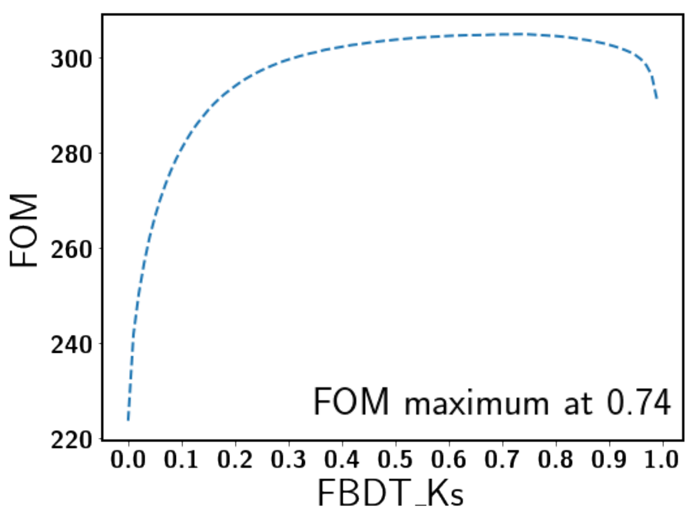


Figure 3-16: FOM of classifier output (*FBDT_Ks*) in *signal MC*, the maximum value is achieved at 0.74, where signal efficiency $\sim 95\%$ and background rejection $\sim 98\%$ are achieved. The FOM curve is almost flat between $0.5 \sim 0.9$, which is insensitive to the cut value in this region.

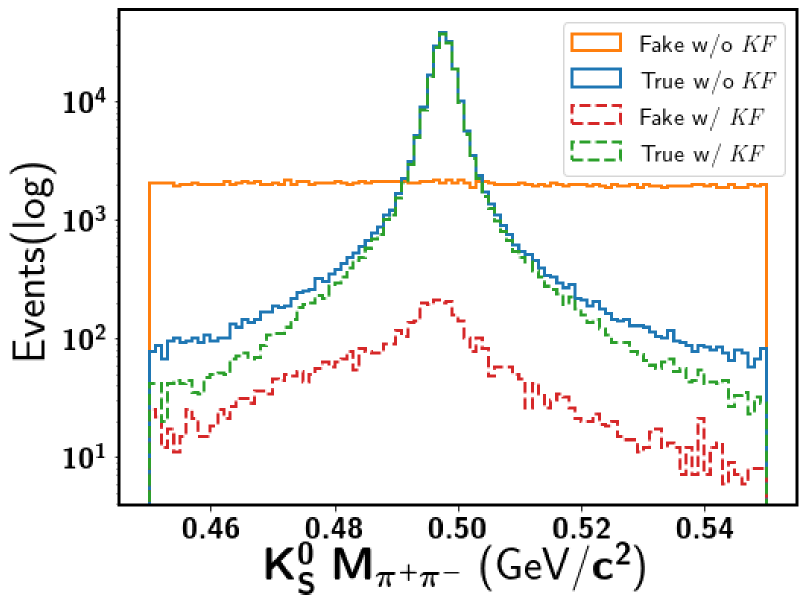


Figure 3-17: K_S^0 purity improvement with $KsFinder(KF)$ cut at 0.74. The blue solid line is true K_S^0 without *KsFinder* and green dashed line is the true K_S^0 with the cut applied. The orange solid line is fake K_S^0 without the cut and the red dashed line is fake K_S^0 with the cut. About 95% of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

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's no MC truth on target variable in real data, the FastBDT method is based on variables in MC samples. If these variables shows close distribution among MC and data, the classification performance is expected to be similar.

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 the validation that approves no clear bias on B^0 for signal extraction is also required. 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 event number distribution in each bin is assumed to follow the *Poisson* Distribution, therefore the standard deviation of each bin is calculated as $\sqrt{N_i}$, where N_i is the event number in the i-th bin. We use three times the standard deviation in each bin as the reference in the drawing error bars. For example, 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 retention ratio before and after applying *KsFinder*, a retention ratio distribution is also produced for each variable below the main comparison plot, where the blue circle dots are the retention ratio for data and red reverse-triangle dots are for MC. In the ratio plot, the reduction fraction by using *KsFinder* is close between the data and MC sample for the majority of the variables. For instance, as the most important variable, the distribution of *cosVertexMomentum* is shown in Figure 3-19. The ratio plot shows a small discrepancy in the background dominated region (*cosVertexMomentum*$<$ 0.0), while remains a good consistence in the signal dominated region (*cosVertexMomentum*$>$ 0.0). This is also true for the invariant mass distribution in Figure 3-18 top left, while other low importance rank variables like *px*, *py* and *pz*, could show a small discrepancies because the *KsFinder* receives much less contributions on classification power from them. Overall, such discrepancies are in an acceptable range, which 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.

Since the FastBDT algorithm relies on the probability density functions to separate signal and backgrounds in each tree node, the similar distribution implies the close classification power in data. However, the correlation in between the variables should also be close, otherwise the FastBDT method may give a bias classification. Since there is no way we could know whether a K_S^0 from data is definitely a true candidate or not, it is impossible to directly compare the correlation among data and MC

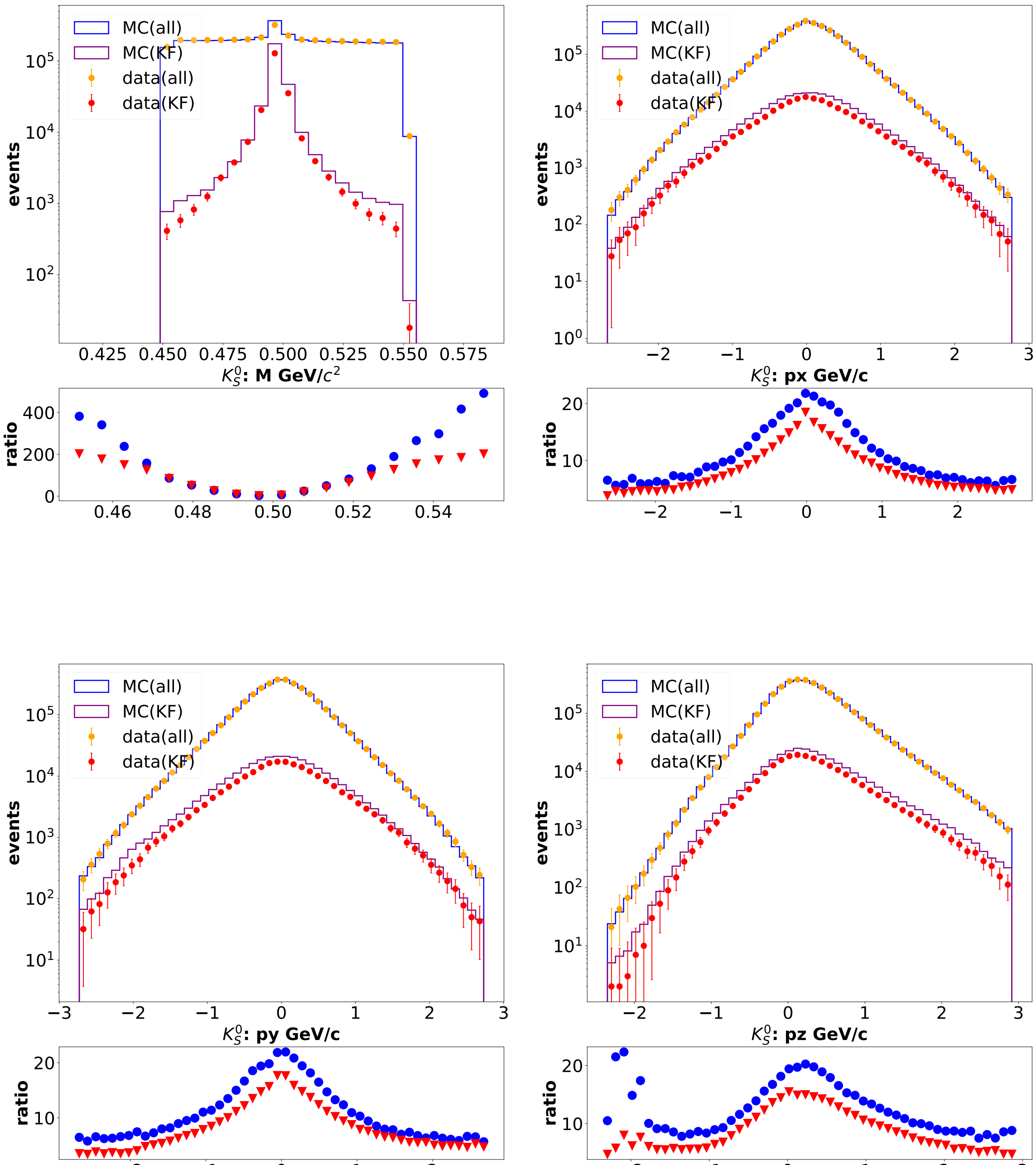


Figure 3-18: The distribution of invariant mass from charged pions and the momentum of K_S^0 in x, y, z directions. The blue line is from all *generic MC* and the purple line is the K_S^0 after *KsFinder* (KF). The yellow dots are data with no *KsFinder* (KF) cut applied and the solid red dots are data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

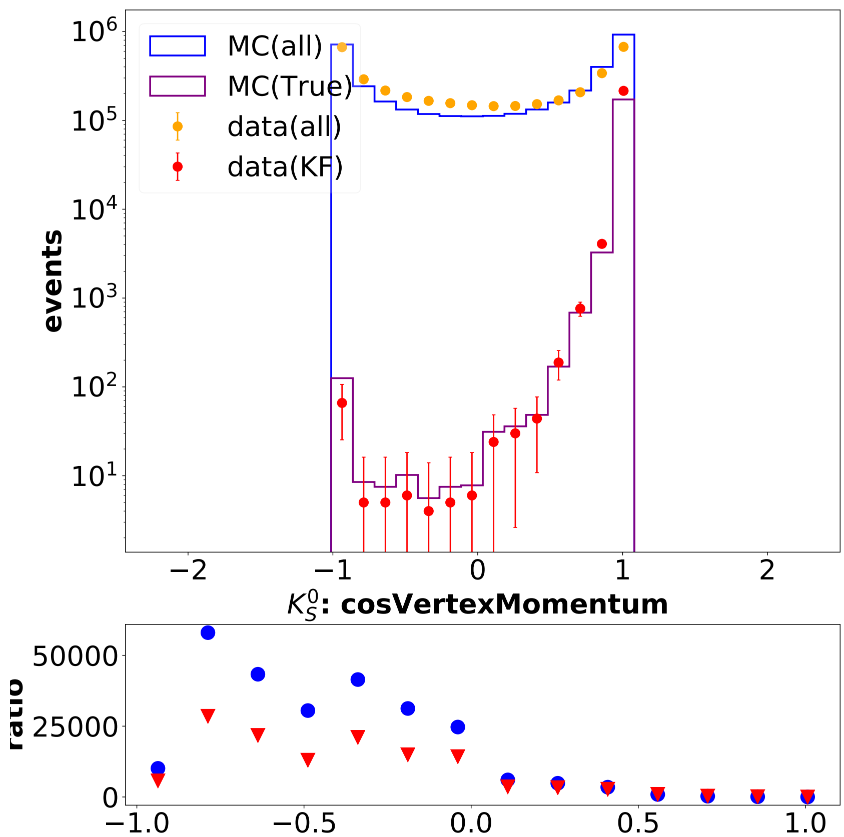


Figure 3-19: The distribution of *cosVertexMomentum* in data and MC with or without *KsFinder* cut applied.

in true and fake K_S^0 separately. Hence, we compare the correlation matrix 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 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 collection.

3.2.7 Data and MC correction by *KsFinder*

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:

$$f_{K_S} = \frac{N_{sig}}{N_{tot}}\,, \tag{3.4}$$

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

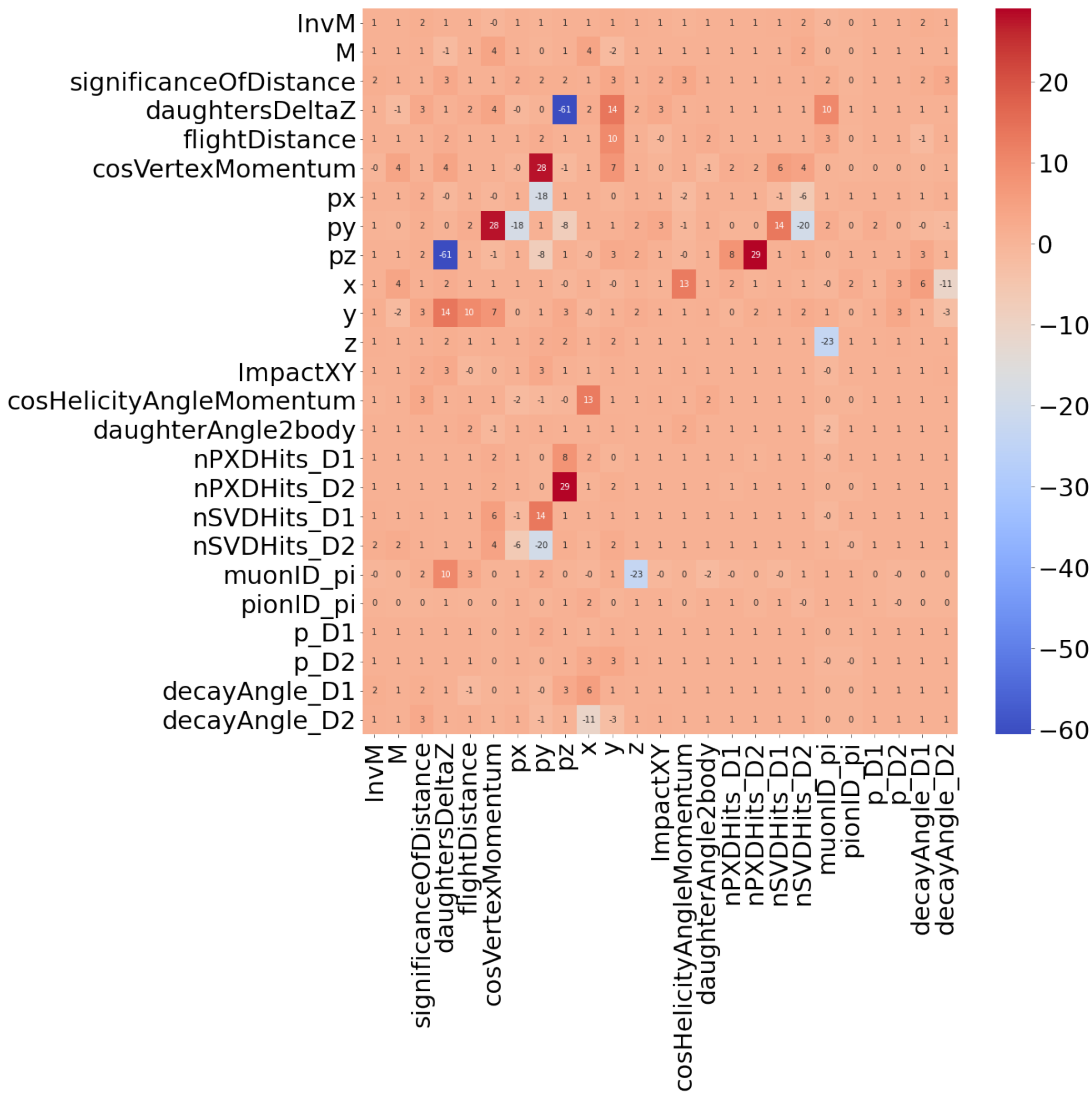


Figure 3-20: The correlation ratio between *KsFinder* input variables, where the value is calculated by using the one from data divided by that from *generic MC*.

shown in Equation 3.5.

$$\mathcal{R}_{K_S} = \frac{f_{K_S}^{MC}}{f_{K_S}^{data}} \tag{3.5}$$

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:

$$\mathcal{R}_{B^0} = \frac{f_{B^0}^{MC}}{f_{B^0}^{data}} \simeq \mathcal{R}_{K_S}^3 , \tag{3.6}$$

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 upper and lower limits of $\mathcal{R}_{B^0}$ by using the center value 1.027 float with ± 0.033, which are 1.060 and 0.994. These two values are applied to the calculation of the signal fraction when performing the systematic uncertainty evaluation in *CP* violation measurement.

a) Data, cut=0.74 b) MC, cut=0.74

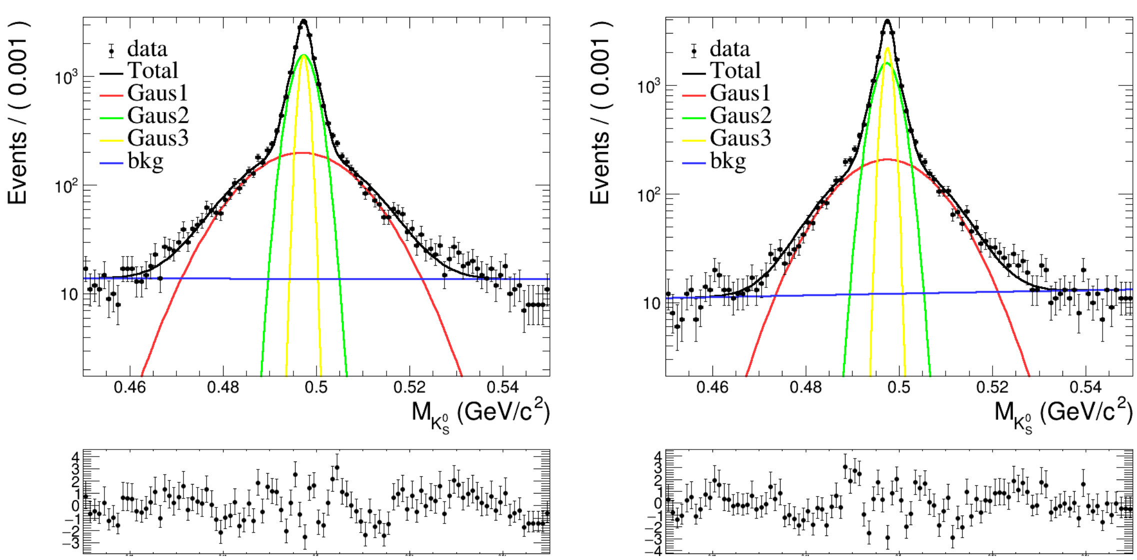


Figure 3-21: The fit on invariant mass $M_{K_S^0}$ where signal component is modeled as a triple-Gaussian and background component is modeled as a Chebyshev polynomial. The bottom plots are the pull of the fitted lines and data points. The signal fraction is slight higher in *generic MC* compared to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

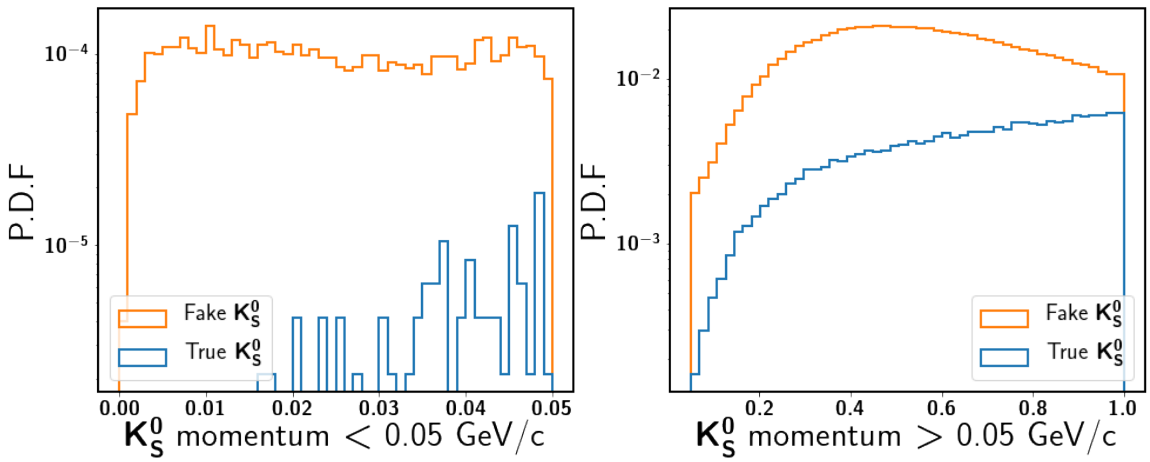


Figure 4-1: The distribution of K_S^0 momentum. Candidates smaller than 0.05GeV/c are rejected.

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. 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 parameters 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* have 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 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.

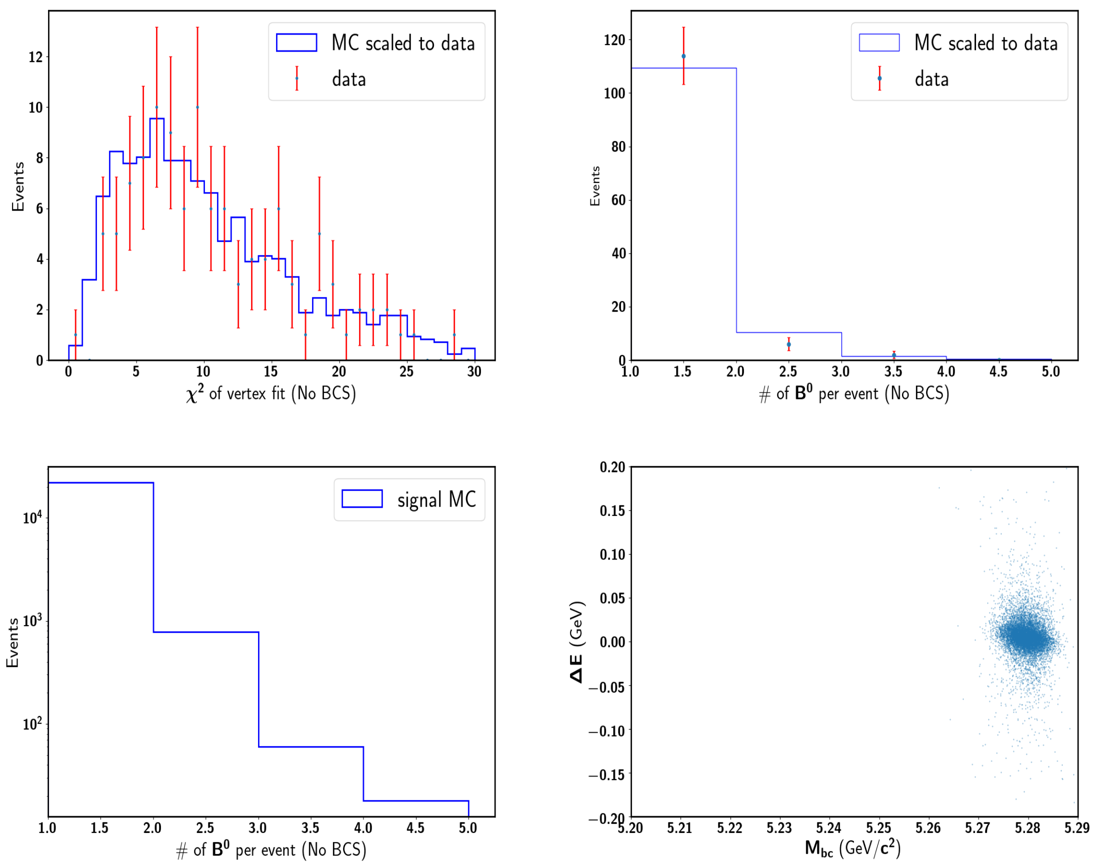


Figure 4-2: Top left is the χ^2 for data and *generic MC* before BCS. Top Right is the B^0 candidates per event in data and *generic MC* before the BCS. Bottom left is the number of B^0 candidates per event from *signal MC*. Bottom right is the 2D M_{bc} and ΔE distribution from *signal MC*.

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 contains 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's 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 [41] developed a set of variables to suppress the continuum background, which has also been implemented into BASF2 framework. The two major sets of variables are the CLEO cone momentum and the modified Super Fox-wolfram momentum.

CLEO cone momentum can be presented as Equation 4.3, where p_i is momentum of i-th particle in the Rest-Of-Event. The particles used in a reconstructed CP-side B^0 are therefore excluded. The θ_i is an angle against momentum thrust of 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.

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

The modified Super Fox-wolfram momentum (KSFW momentum) can be defined as shown in Equation 4.4.

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

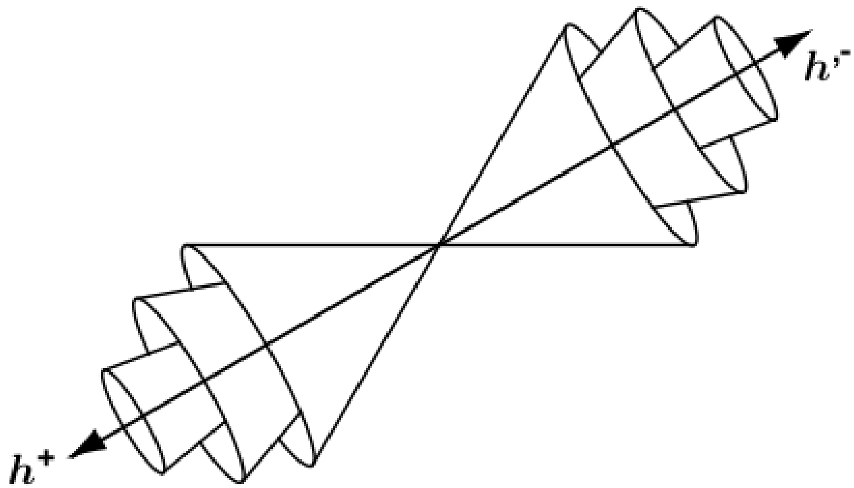


Figure 4-3: A graphical illustration of the CLEO cone. The h^+ and h'^- present the hadronic tracks from a B decay. The first three cones are drawn [41]. The nine CLEO cone momentum can be calculated using the particles in each cone.

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 [42].

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 $cosTBTO$ 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.

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

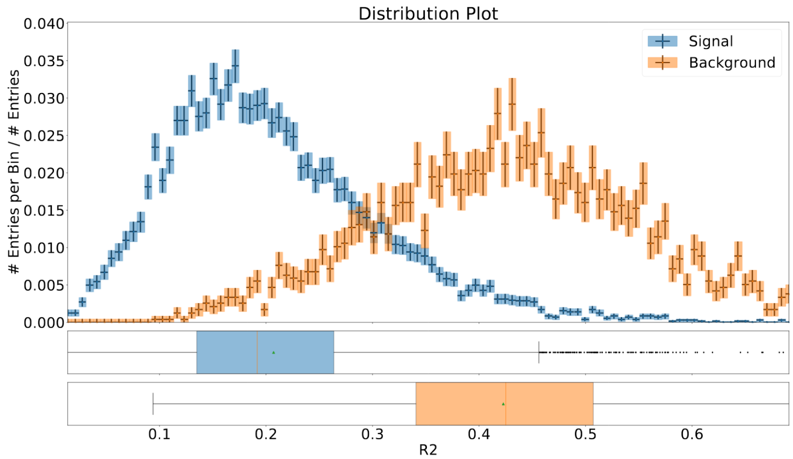


Figure 4-4: R_2 is the ratio of the second to the zeroth KSFW momentum in Equation 4.4 of which the distribution in *signal MC* sample which serves as the highest weight as a variable in discriminating the continuum events, having a quite different distribution between signal and background.

background. Such function is implemented by the BASF2. For the specific decay mode $B^0 \to K_S^0 K_S^0 K_S^0$, the training and testing sample 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 events reconstruction procedures 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 4-5. 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 4-6.

The correlation between these training variables are shown in Figure 4-7 which are varied between signal and continuum background events. The ROC curve and the efficiency/purity with respect to the classifier output are shown in Figure 4-8, yielding a close performance.

The Overtraining check is made by comparing the distribution of signal and background depending on the classifier output in both training and testing samples. The testing samples show about 1% lower in each bin for both signal and background

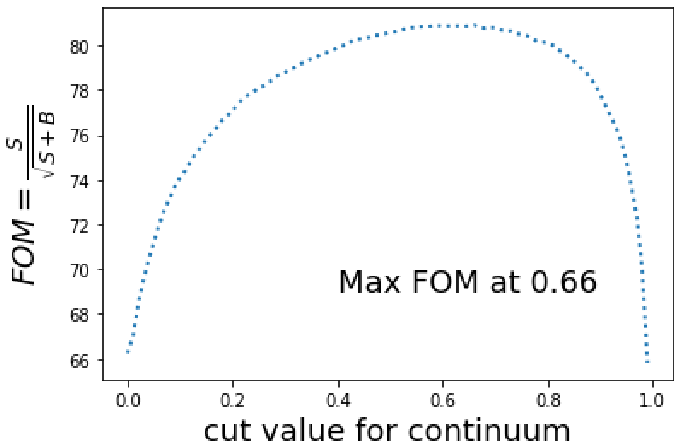
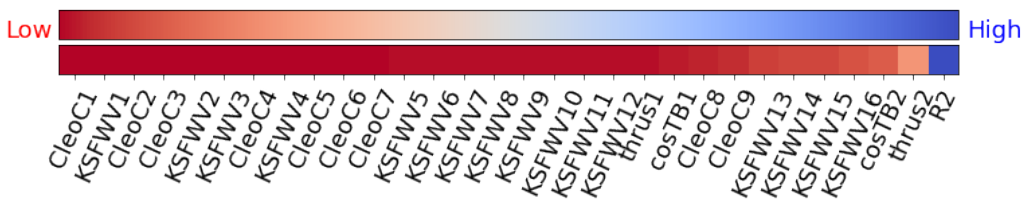


Figure 4-5: FOM depending on the cut value of continuum classifier output, cut value at 0.66 is used for continuum suppression.

Figure 4-6: The importance rank of the input variables for the Belle II CS framework.



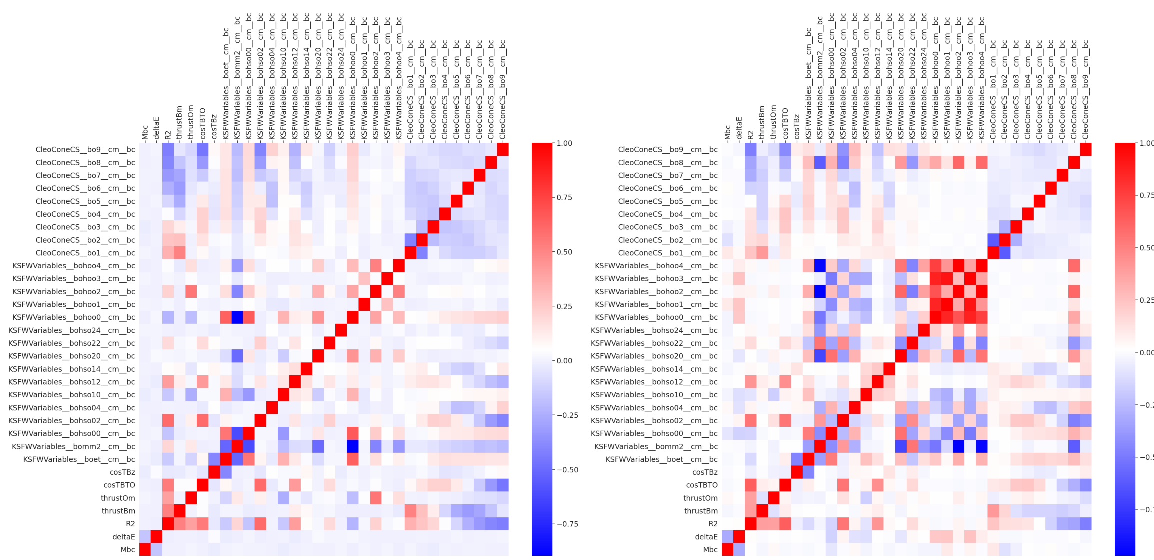


Figure 4-7: The correlation in variables for continuum suppression. The left is for signal and the right is for background.

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

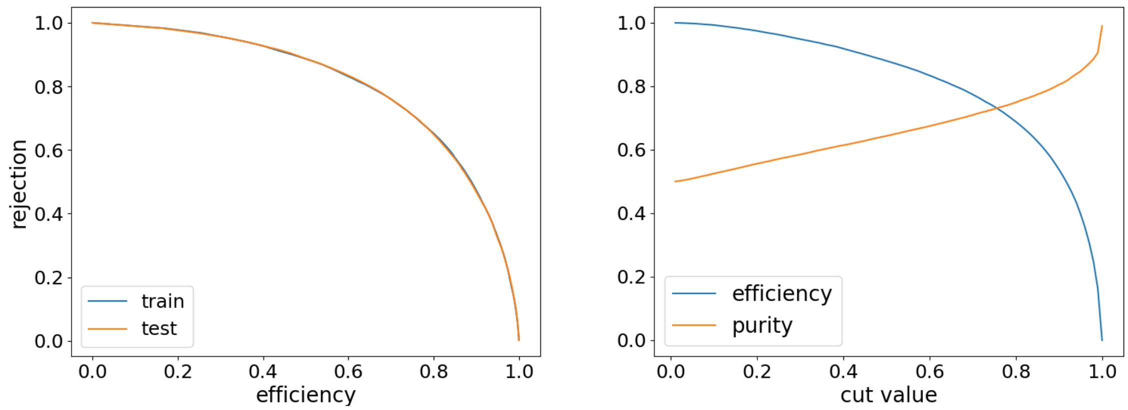


Figure 4-8: The left is the ROC curve (blue for training and orange for testing) and the right is the efficiency(blue) and purity(orange) regarding the classifier output *FBDT_CS*.

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	$M_{bc}(\mathrm{GeV}/c^2)$	$\Delta E(\mathrm{GeV})$	$P(\chi^2)$	*Rank*	*FBDT_CS*	*FBDT_Ks*
Criteria	> 5.20 & < 5.29	$\lvert\Delta E\rvert < 0.2$	> 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.

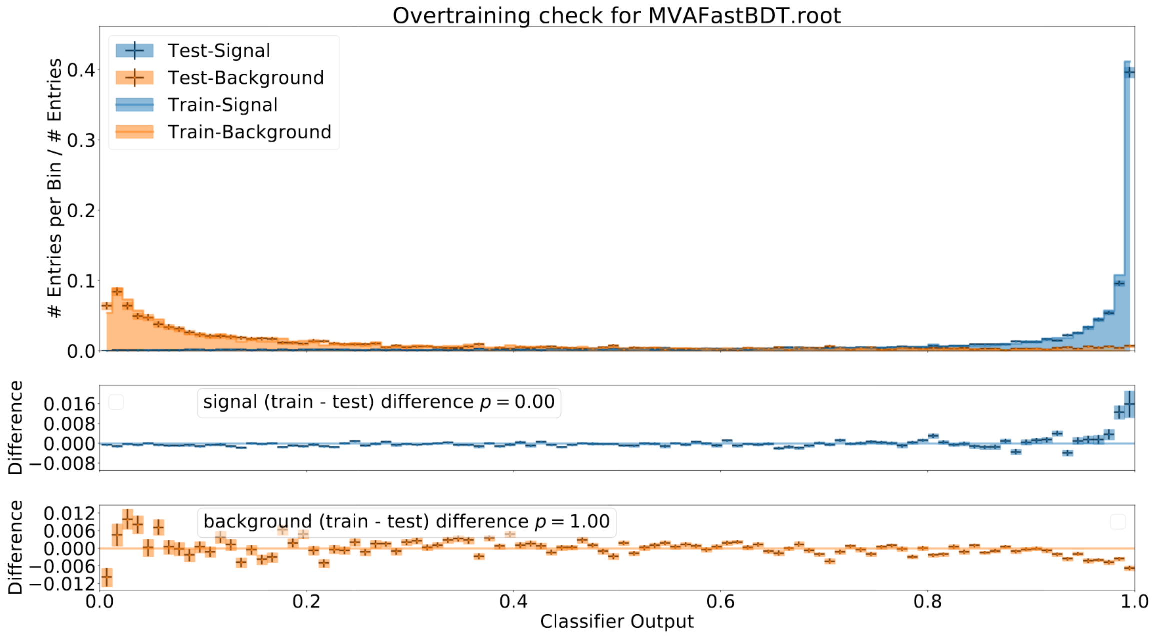


Figure 4-9: Over-training check of continuum classifier, where a very small difference in training and testing (1%) is shown.

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.

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 resonance background, X could be D^0, 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 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 there is 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.

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

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	Br($B \to XK_S$)PDG	Br($X \to 2K_S$)	Br($B \to XK_S$)Dec.	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

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 4-10 which contains 328 signal events, 702 continuum background events and 35 $B\bar{B}$ background events. The distribution of M_{bc} and ΔE from data are also shown in Figure 4-10.

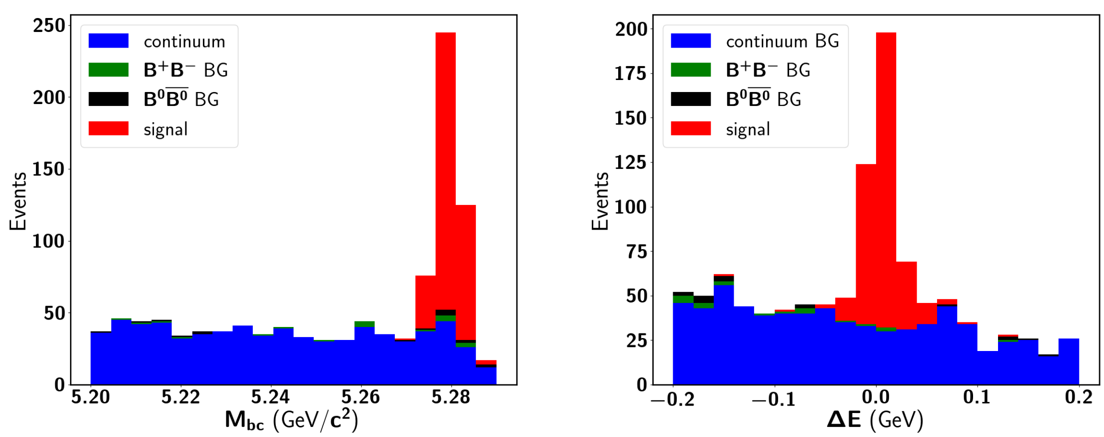


Figure 4-10: The distribution of M_{bc} and ΔE for selected events from *generic MC*, where each background components are stacked with signal. The blue component is the continuum background. The green and the black are the background from charged and neutral $B\bar{B}$ events. The red component is the signal component by checking MC truth matching.

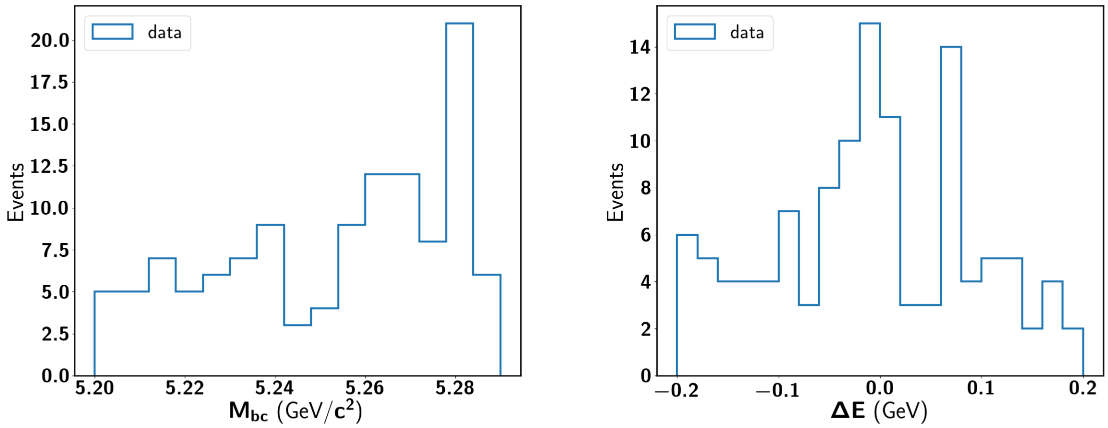


Figure 4-11: The distribution of M_{bc} and ΔE for selected events from data.

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, which is modeled as the Argus distribution [43] 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 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 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 4-12. 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 4-13.

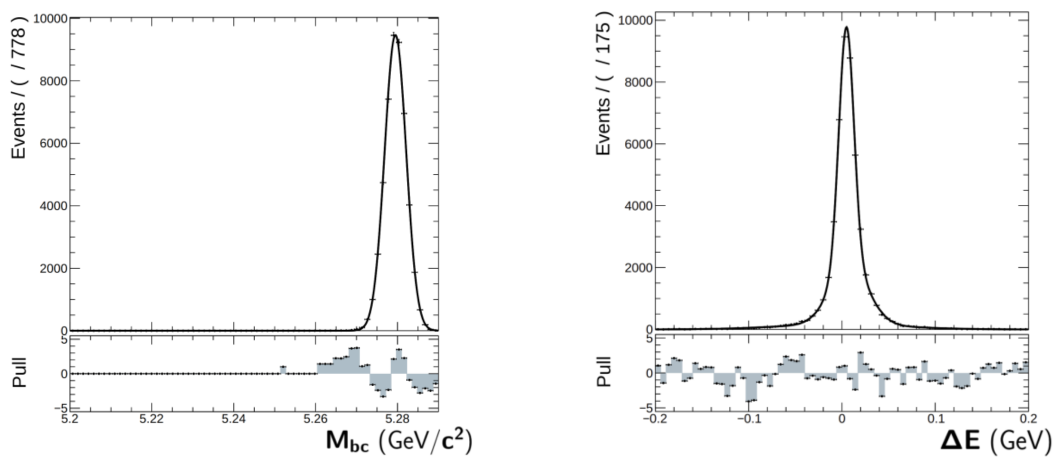


Figure 4-12: The distribution of M_{bc} and ΔE of *signal MC* of $B^0 \to K_S^0 K_S^0 K_S^0$ fitted with single and triple Gaussian functions respectively. The bottom plots are the pull of the data points and the fit result.

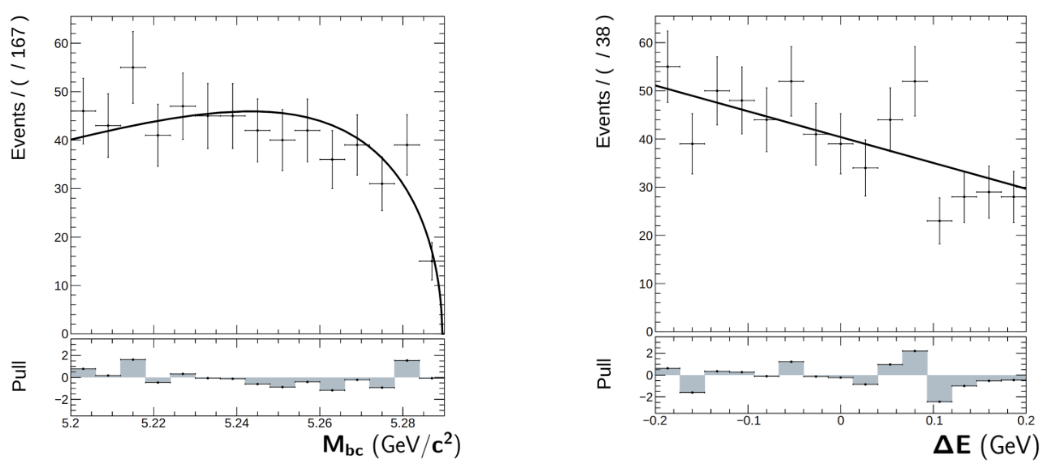


Figure 4-13: The distribution of M_{bc} and ΔE of continuum events in *generic MC* fitted with Argus and Chebyshev polynomial functions, respectively. The bottom plots are the pull of the data points and the fit result.

Then we set the events number for signal and background 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 4-14.

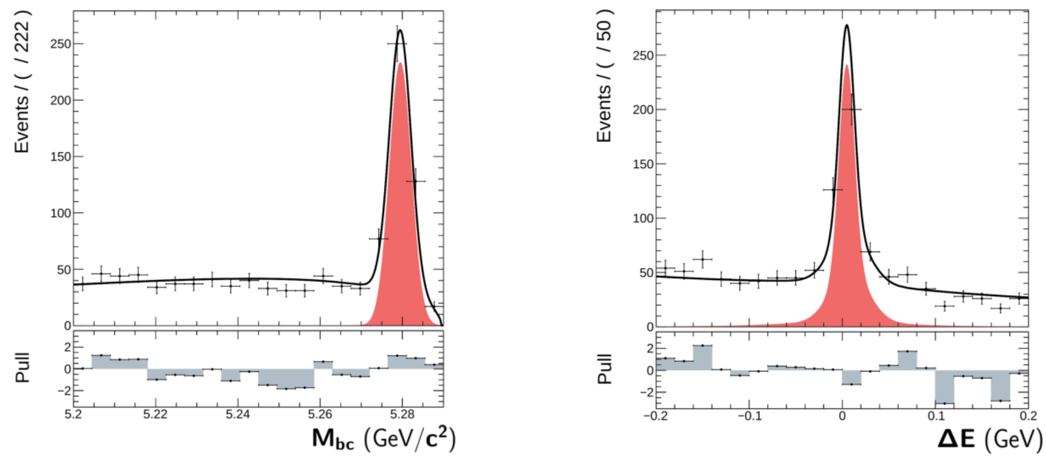


Figure 4-14: Top is the stacked plots for *generic MC* of M_{bc} and ΔE, where each background components are stacked with signal. The bottom is the 2D fit on 1 ab^{-1} *generic MC* projected on M_{bc}(GeV/c^2) and ΔE(GeV), the red is signal component from the fit result in both plots.

Before perform 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 4-15, where the *generic MC* is scaled to the luminosity of experiment data and an agreement within $\sim 1\sigma$ is observed.

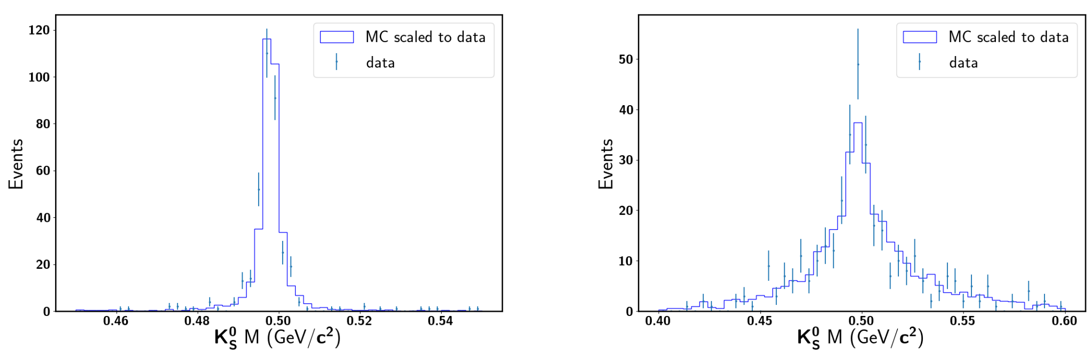


Figure 4-15: Invariant mass before (left) and after (right) B^0 vertex fit from *generic MC* and experiment data.

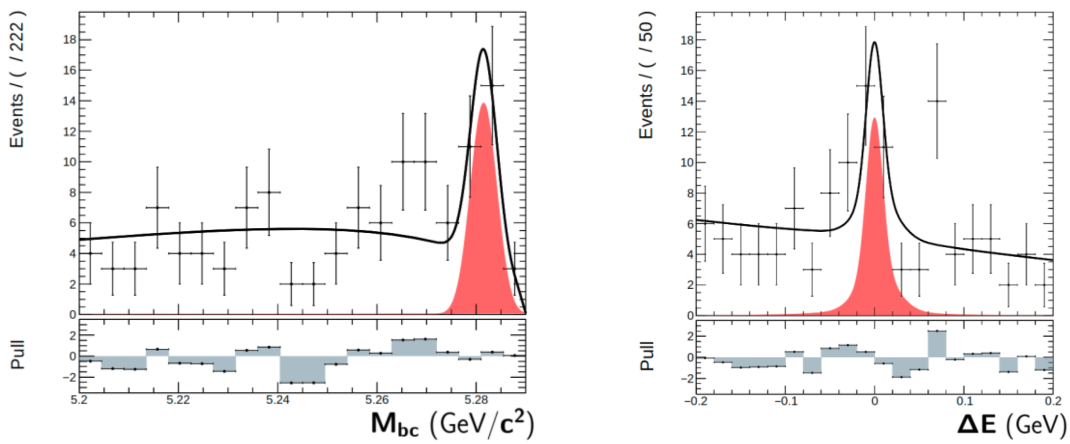


Figure 4-16: M_{bc}(GeV) and ΔE(GeV) 2D fit on 62.8 fb^{-1}data, the red is the signal component.

Same 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 4-16. The number 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 Equation 4.7:

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

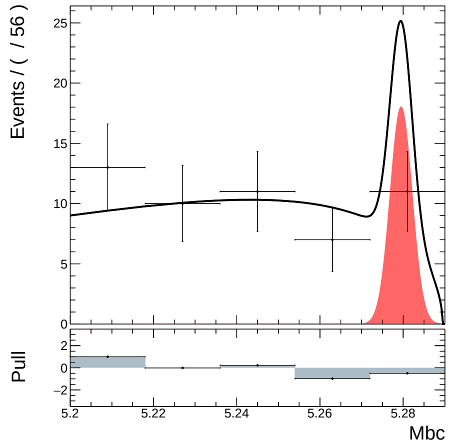
the estimated signal event number can be calculated, where the $N_{B\bar{B}}$ is the number of neutral B meson pairs calculated from the integrated luminosity and ϵ_{rec} is the reconstruction efficiency. The factor 2 accounts for the fact that both B^0 and $\overline{B^0}$ can decay to three K_S^0 in the final states.

Using 62.8 fb^{-1} data from the Belle II, 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. The number of signal and background events from the same amount of *generic MC* is calculated to be ~ 20.6 and ~ 4.1. These values are summarized in Table 4.5. If 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 latter.

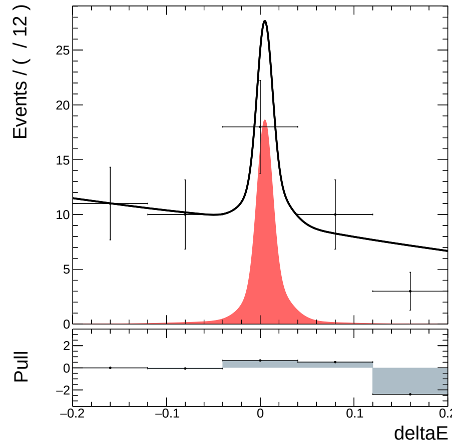
Table 4.5: Reconstructed signal and background events using M_{bc} and ΔE 2D fit, compared with expected numbers from *generic MC*.

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

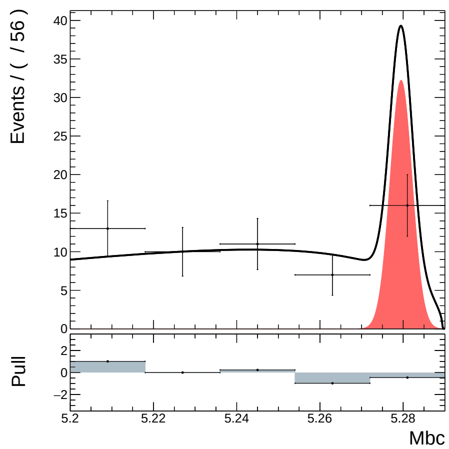
To check reliability of the number of events fitted from the M_{bc} and ΔE in this low statistics case, we test the 2D fit result, by merging the continuum events and the different number of signal events, to check the linearity of the input and output. From the 2D fit on *generic MC*, the expected number of continuum events in the full range of M_{bc} and ΔE is ~ 46 at 62.8 fb^{-1}, which is set to be the constant number of continuum events used in the linearity test. Then the number of signal events from 5 to 30 with 5 events per step are injected into the continuum 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 in each injection test are shown in Figure 4-17. The output signal and background events depending on the injected number of the signal events are presented in Figure 4-18, where the dependence on both signal and background events are fitted with linear function $y = ax + b$. The error bar on each data point is taken from the statistical uncertainty of the number of signal or continuum 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. The signal events yield from the current luminosity is considered as a reliable result.



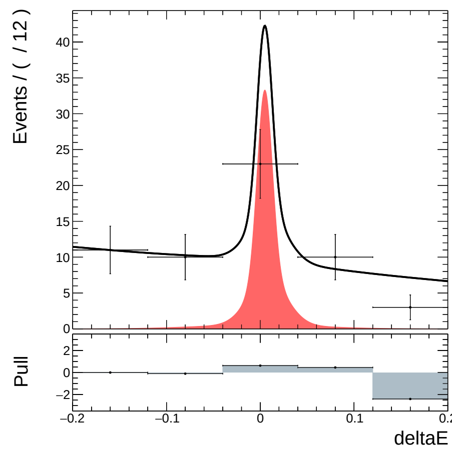
a) signal injected: 5



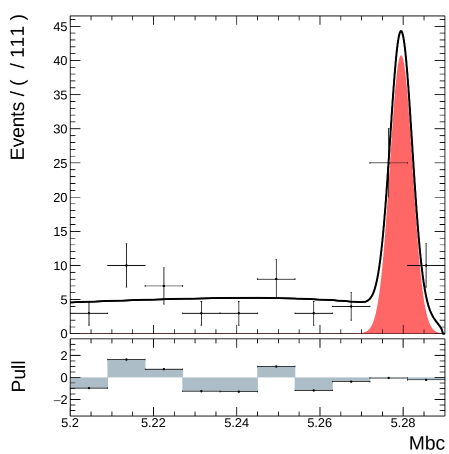
b) signal injected: 5



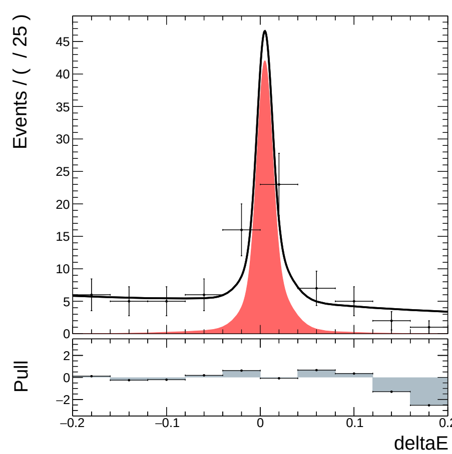
c) signal injected: 10



d) signal injected: 10



e) signal injected: 30



f) signal injected: 30

Figure 4-17: 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.

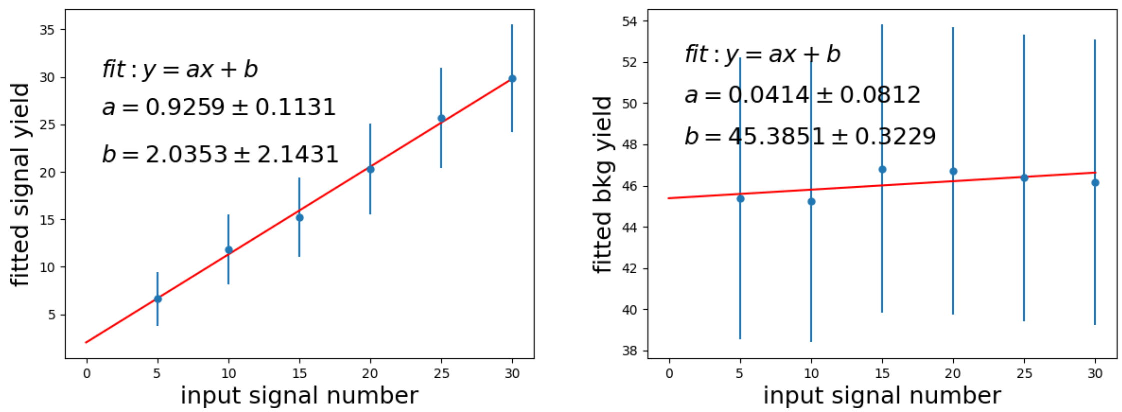


Figure 4-18: Injection test for signal extraction. The linearity is clear between input and output signal events number. The error bar on each point is taken from the statistical uncertainty of the number of signal or continuum events from 2D fit results.

Chapter 5

CP parameters measurement

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

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

The Equation 5.1 describes the physics distribution of signal events only. To perform the unbinned maximum likelihood fit on data, a complete model for i-th event that includes the overlay of background components and outlier bands can be defined as Equation 5.2.

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

where f_{sig} and f_{ol} are the fraction of signal and outlier components, respectively. The $\mathcal{P}_{bkg}$ and $\mathcal{P}_{ol}$ are defined by Equation 5.3 and 5.4.

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

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

where $\delta(\Delta t_i - \mu^{\delta}_{bkg})$ is Dirac δ function and G is 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.

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

The R_{sig} stands for the resolution function for signal events, which receives smearing effect from CP and tag side separately, namely R_{cp} and R_{tag}. The treatment of CP side and tag side is different because of vertexing strategies. For CP side, vertex of B^0 is reconstructed by fully fitting all the daughter particles. Instead, in tag side, there's no full reconstruction of B^0 so vertex fit is applied for the selected charged tracks in the rest-of-event. The background events have its own resolution model which is independent from CP violation parameters. The outlier is used to smooth fit for large Δt events. In the low statistics case as the current luminosity is, the outlier is not included in the fit to have a more realistic model for data.

For signal events, the resolution functions are studied for CP-side and tag-side based on each possible degradation such as detector resolutions, effect of tracks from non-primary B vertex and so on. Such a method is used in Belle analysis and named as artificial model. Details are summarized in [44]. Considered the vertex position difference Δz for signal events as shown in Equation 5.6.

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

where the primed ones stands for physics truth of the position and the non-primed is the measured value, the resolution function receives contribution from both CP and tag-side effects. In the meantime, the resolution functions on both sides also depend on the applied constraint. Considering that the fine structure of IP profile is not yet fully understood and small discrepancies have been observed between data and simulation[45], there's no IP constraint applied for both sides in vertex fit, which avoids potential bias from IP profile under this low statistical situation. The combined contributions can be presented as Equation 5.7.

$$R_{sig} = R_{cp} \otimes R_{tag} \quad (5.7)$$

5.1.1 CP-side resolution function

CP-side vertex is fitted with all tracks from a reconstructed B^0, thus the resolution models only depend on detectors' effect. For each event, the resolution effect 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.

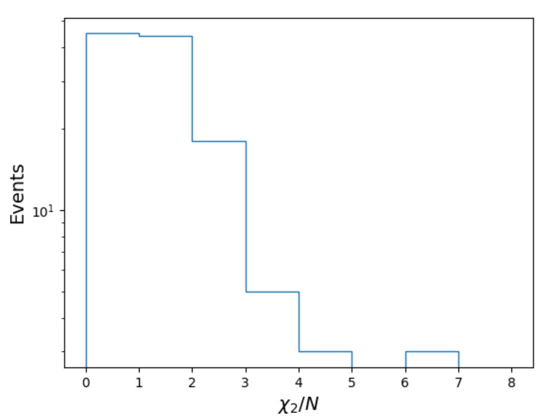


Figure 5-1: χ^2/N of selected events from data.

Therefore, we model the resolution functions on CP-side by using a double Gaussian function, where the mean is fixed to zero and the standard deviation is scaled by χ^2/N and the error of reconstructed vertex $\sigma_{z_{cp}}$, as shown in Equation 5.8.

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

where s_{cp}^{main} and s_{cp}^{tail} are defined in Equation 5.9.

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

The dependence of resolution models on χ^2/N is shown in Figure 5-2. Restrictively speaking, the CP-side resolution for $B^0 \to K_S^0 K_S^0 K_S^0$ is slight different from $B^0 \to J/\psi K_S^0$, due to the absence of the direct charged tracks 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 fitted which are listed in Table 5.1.

Table 5.1: Parameters in R_{cp}.

f_{cp}^{tail}	0.07424 ± 0.0008
s_0^{main}	0.9151 ± 0.0077
s_1^{main}	0.2142 ± 0.0064
s_0^{tail}	2.0477 ± 0.0779
s_1^{tail}	1.3470 ± 0.0720

5.1.2 Tag-side resolution function

For the tag-side, the vertexing is done by using *KFit* and no IP constraint used. Due to the charged tracks from non-primary B vertex, the resolution functions on tag-side not only receives contribution from detectors' effect R_{det}^{tag} but also the resolution degradation from secondary vertex, called R_{np}^{tag}. To the contrary, if all tracks that

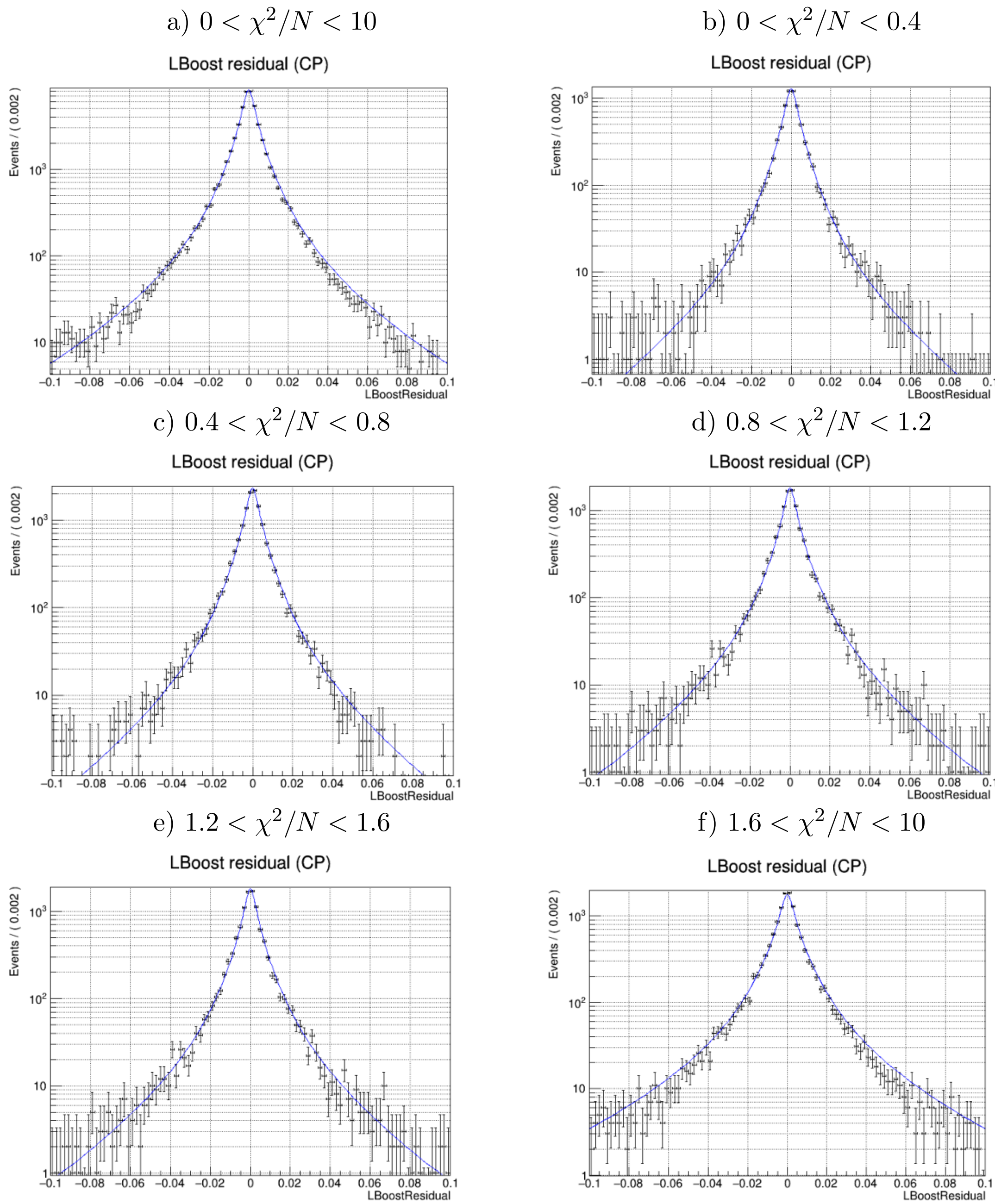


Figure 5-2: The z position of B^0 vertices on CP-side, which is dependent on the χ^2/N. The first plot is the fit in the full range and the rest are the fit in each slices of χ^2/N.

are used for tag-side vertexing are primary tracks, the resolution will only be affected by the detectors' effect. The vertex position difference is defined as Equation 5.10. Therefore, the effects from both detectors and non-primary tracks contributes to the total resolution on tag-side as Equation 5.11 shows.

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

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

Similarly to CP-side resolution function, detectors' effect is presented in Equation 5.12

$$R_{det}^{tag}(\delta z_{tag}^{det}) = (1 - f_{tag}^{tail})G(0, s_{tag}^{main} \cdot \sigma_{z_{tag}}) + f_{tag}^{tail}G(0, s_{tag}^{tail} \cdot \sigma_{z_{tag}}) \tag{5.12}$$

where main and tail Gaussian functions have the same central value at zero, but the standard deviation is scaled by χ^2_{tag}/N on the tag-side as shown in Equation 5.13.

$$s_{tag}^{main/tail} = s_0^{main/tail} + s_1^{main/tail} \cdot \chi^2_{tag}/N \tag{5.13}$$

Technically R_{det}^{tag} can be fitted with MC samples of which tag-side tracks are all from primary vertex. After obtaining the fitted parameters of R_{det}^{tag} , R^{tag} will only be dependent on R_{np}^{tag}. The fit model of R_{np}^{tag} is shown in Equation 5.14. It consists of three functions, including one Dirac δ function and two single-side exponential functions E_p and E_n. The $E_p(x, \tau_p) = (1/\tau_p)e^{-x/\tau_p}$ when $x > 0$ and the $E_n(x, \tau_n) = (1/\tau_n)e^{x/\tau_n}$ when $x < 0$. The exponential factors in both positive and negative components are scaled by the tag-side vertex uncertainty $\sigma_{z_{tag}}$.

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

Also, since tag-side has no dependence on how CP-side is reconstructed, the resolution functions on tag-sdie are almost mode-independent. Thus these parameters are obtained by fitting to the control sample . The control sample consists of multiple exclusive $D^{(*)}$ hadronic decays, of which the details are summarized in Appendix B. The The fit plots for tag-side resolution functions are shown in Figure 5-3 and 5-4. The parameters obtained from the fit are listed in Table 5.2 and Table 5.3.

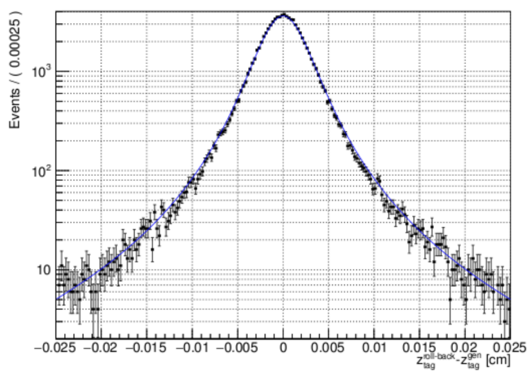


Figure 5-3: R_{det}^{tag} fit

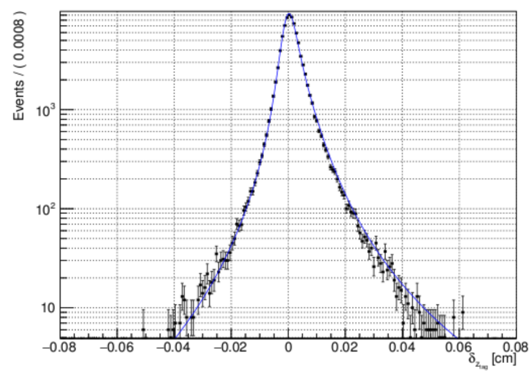


Figure 5-4: R_{np}^{tag} fit

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}	3.4480 ± 0.0897
s_1^{tail}	0.2666 ± 0.0276

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

f_δ	0.6256 ± 0.0049
f_p	0.8316 ± 0.0051
τ_n	2.9141 ± 0.0758
τ_p	2.4846 ± 0.0269

The boost direction of each event is not constant event-by-event, so the position of vertex may not be optimized by calculating $\Delta t_i = \Delta z/\beta\gamma c$. This effect can be reduced by replacing vertex position difference on z-axis with the relative distance along the boosting direction, or introducing another resolution function called R_k[46]. The R_k has not been implemented in Belle II resolution model. Therefore, Δz projection on the boosted direction of each event is used for reducing this kinematics effect on resolution function.

5.1.3 Background events Δt distribution

The R_{bkg} is uncorrelated to vertex reconstruction method approximately. Because the background mainly comes from continuum events passing the selection, it's reasonable to model its resolution by a Gaussian-like function. A double-Gaussian with its standard deviation scaled by the measured uncertainties from both sides is used as Equation 5.15. To be noted, unlike resolution functions on CP or tag-side, the standard deviations of the double Gaussian are scaled by both the vertex position uncertainties $\sigma_{z_{cp}}$ and $\sigma_{z_{tag}}$.

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

The background events Δt shapes $\mathcal{P}_{bkg} \otimes R_{bkg}$ can be determined by fitting to side-band data. There are totally seven floating parameters which are listed in Table 5.4 with fitted values using 60 sideband events at $M_{bc} < 5.26$ GeV, shown in Figure 5-5

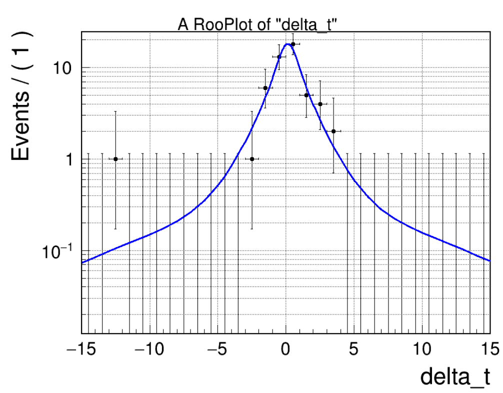


Figure 5-5: $\mathcal{P}_{bkg} \otimes R_{bkg}$ fit using 60 sideband events at $M_{bc} < 5.26$ GeV.

μ_δ^{bkg}	0.1310 ± 0.1902
μ_l^{bkg}	0.1638 ± 0.5030
τ_{bkg}	1.0541 ± 0.4370
f_δ^{bkg}	0.5861 ± 0.2570
f_{tail}^{bkg}	0.0417 ± 0.0408
σ_{main}^{bkg}	1.4348 ± 0.3940
σ_{tail}^{bkg}	28.0930 ± 8.8221

Table 5.4: Parameters in Background Δt distribution.

5.2 Flavor Tagging

In order to determine the flavor of tag side B^0, flavor tagging algorithm is being developed. The flavor tagging uses information from $\mu^\pm$, $\pi^\pm$,$K^\pm$ and Λ which are categorized into 13 different types as illustrated in Figure 5-6.

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	

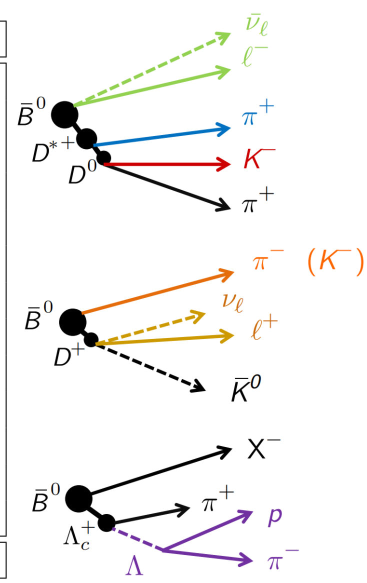


Figure 5-6: Particles and their categories used in flavor tagging algorithm[47].

For each particle that has been used from above categories, PID and kinematics information are extracted and feed to the combiner as training variables, to obtained a classifier response corresponding to each category. Then for all responses from these categories, a total classifier is trained to present the likelihood of flavor q. This algorithm is called category-based method and used in this thesis. After the reconstruction on the CP-side B^0 is done, the rest-of-events tracks used to form the particle lists are selected5-6. The FastBDT as the back-end algorithm is chosen for performing training on the classifier of flavor tagging. Targeted variable is true q of tag-side neutral B in MC. 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 are completely invisible.

Considering the limited power of flavor tagging accuracy, there is a certain fraction of events that are wrongly tagged among all that can be flavor tagged using the final state particles in the rest-of-event. Thus, the flavor tagging efficiency ϵ and wrong tag fraction w are defined, respectively. Taking into account of the performance of flavor tagging, the observed distribution of Equation 5.2 becomes Equation 5.16.

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

Compared to the original, the term with $\mathcal{S}$ and $\mathcal{A}$ is scaled by factor $r \equiv |1-2w|$, defined as 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 which makes w an important source of systematic uncertainty, too. The validation of flavor tagger using flavor specific decay modes in 2019 Belle II data is summarized here[48]. 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 with $q\cdot r$ opposite to its MC flavor. The distribution of $q\cdot r$ is shown in Figure 5-7 using *signal MC* of $B^0 \to K_S^0 K_S^0 K_S^0$.

Besides, w can be different between B^0 and $\overline{B^0}$, where $\bar{w} = (w_{B^0} + w_{\overline{B^0}})/2$ and

$\Delta w = w_{B^0} - w_{\overline{B^0}}$. Due to the small value of Δw, the contribution from Δw is treated as zero in Equation 5.16 in this analysis. Similarly, for ϵ, the values calculated based on each r-bin are summed and used in Equation 5.16, where the total efficiency is $(99.72 \pm 0.02)\%$, treated as 1. The difference $\mu = \epsilon_{B^0} - \epsilon_{\overline{B^0}}$ is about 1 % to 2 % in each r-bin, thus treated as zero. The distributions of w, Δw, ϵ, and μ in each r-bin are shown in Figure 5-8. The values obtained from *signal MC* are agreed with those from the control sample study[48]. Thus the values from control samples are used for CP fit.

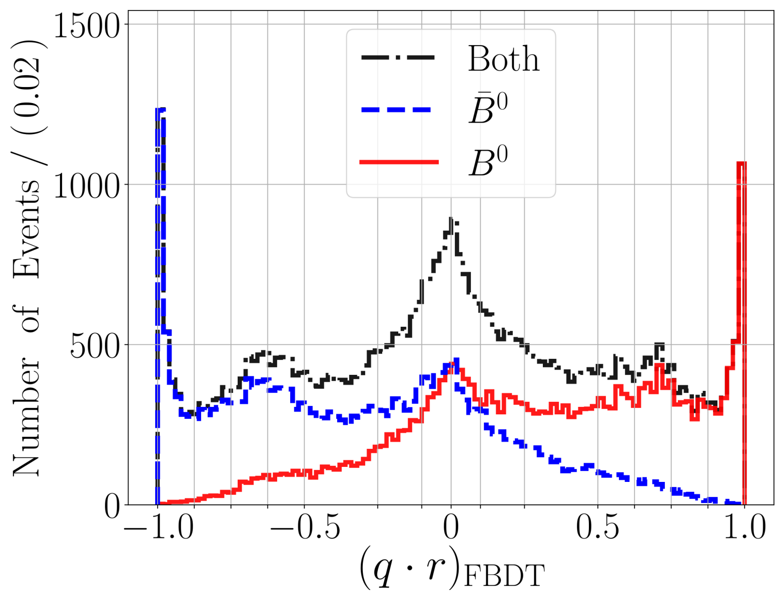


Figure 5-7: The distribution of flavor tagger output $(q \cdot r)$ for both tag-side of B^0 and $\overline{B^0}$

5.3 CP Fitter

The parameters that are needed for measuring $\mathcal{S}$ and $\mathcal{A}$ are studied and obtainable. Using observed Δt distribution from selected events, Equation 5.5 can be fitted using unbinned maximum likelihood fit which takes Δt, signal fraction f_{sig}, the flavor charge q as observables. 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 is developed based on Python and RooFit, which is naturally easy

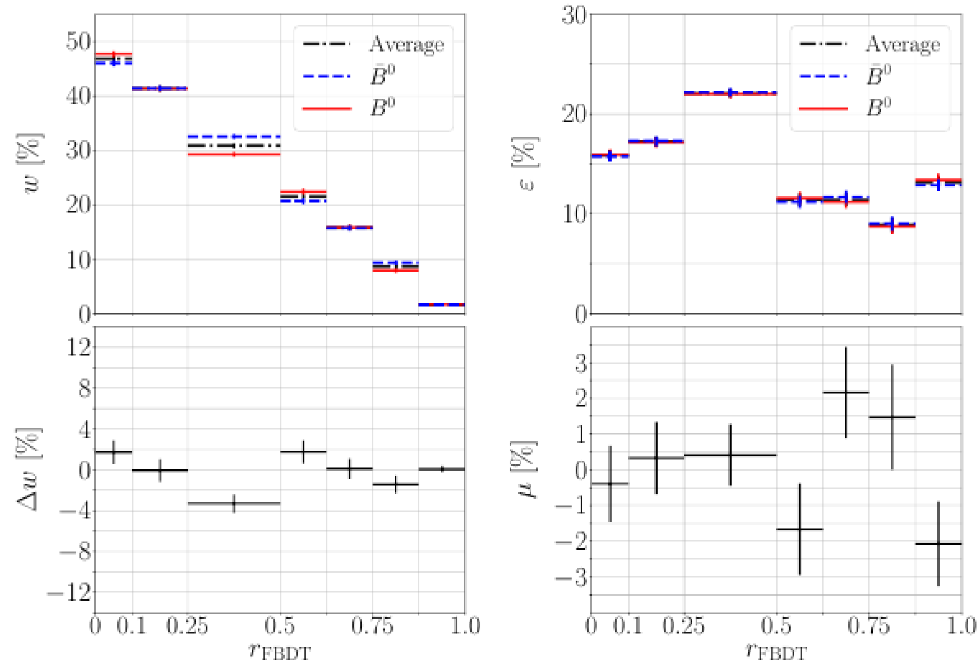


Figure 5-8: The flavor tagging efficiency, wrong tagging fraction, and their difference between different flavors sorted in each r-bin.

to use and maintained with BASF2. The fitter requires a configuration files which contains all the parameters' definitions including their ranges, initial values, floating states and uncertainties.

5.4 Blind analysis and fit

As a required procedure to make sure the CP parameters are measured without 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 number of events used. To check the reliablity of fit result from CP fitter, a linearity test and toy MC study are also performed.

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

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

Table 5.6: The CP fit results using *signal MC* and *generic MC* with only statistical uncertainties.

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)	$\sin(2\phi_1) = 0.20 \pm 0.85$	$\mathcal{A} = -0.06 \pm 0.30$

5.4.1 *CP* fit on MC samples

Using CP fitter, we first perform the CP fit on events in *signal MC* and *generic MC*. The *signal MC* and *generic MC* are generated with phase-space model which contains zero CP violation ($\mathcal{S} = \mathcal{A} = 0$). The events that pass the selections in Table 5.5 are used for CP parameters fit. We have 10000 (8873 passing selections) events from signal sample and 415 (373 passing selections) events from 1 ab^{-1} *generic MC* to fit CP parameters. To mimic the events number expected in data sample, 30 events randomly taken from *generic MC* are used to perform the fit as well. The plots are shown in Figure 5-9, 5-10 and 5-11. The fit results of $\mathcal{S}$ and $\mathcal{A}$ are summarized in Table 5.6.

The fit results are consistent with expectation in non-CP violation from MC input, and the statistical uncertainties has the tendency $\delta \propto 1/\sqrt{N}$ as poission distribution, where N is events number used for CP fit. To test fit on non-zero CP violating MC, the fit on $B^0 \to J/\psi K_S^0$ *signal MC* is also done, the details of events selection as well as fit model determination can be found[45]. The fit result over 10000 events is shown in Figure 5-12, which results in $\sin(2\phi_1) = 0.70 \pm 0.05$ and $\mathcal{A} = -0.01 \pm 0.02$. The results agree with the input.

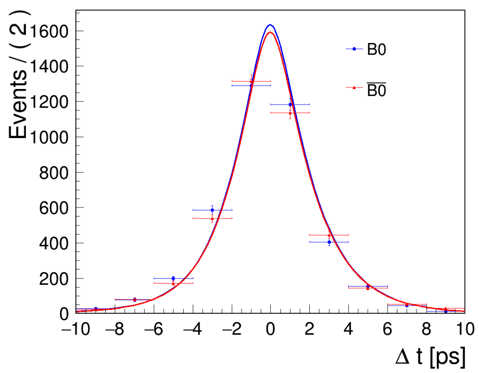


Figure 5-9: *CP* fit on 8873 *signal MC*.

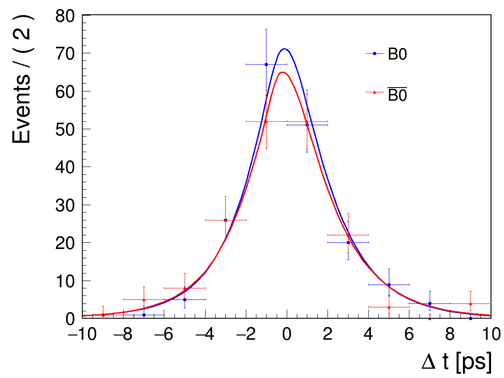


Figure 5-10: *CP* fit on 373 *generic MC*.

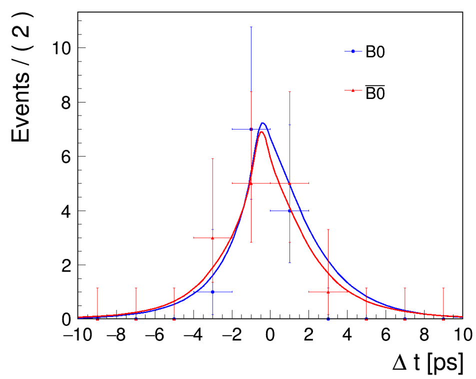


Figure 5-11: *CP* fit on 30 *generic MC*.

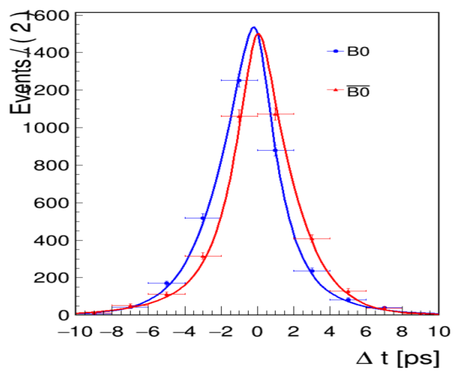


Figure 5-12: *CP* fit over 10000 $B^0 \to J/\psi K^0_S$ *signal MC*.

5.4.2 Linearity Test

To validate the *CP* fit linearity, a series of toy MC samples is generated, which the χ^2 from vertex fit, events number N and vertex errors on *CP* and tag-side are sampled from the distribution of *signal MC*. The resolution functions parameters are kept as 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. Each dataset contains 10000 events. The dependence between input and output are shown in Figure 5-13. The linearity fit shows a good agreement between input and output.

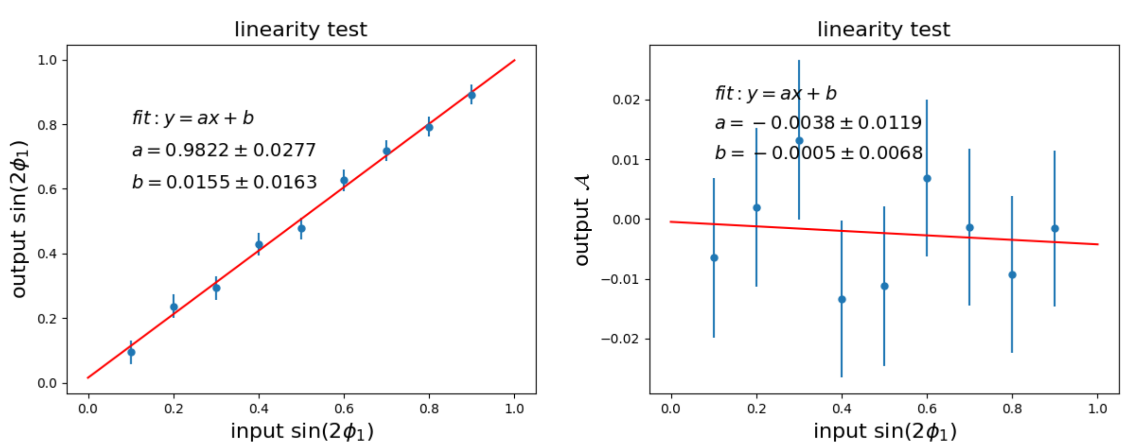


Figure 5-13: Linearity test of *CP* fit.

Also, we fix $\sin(2\phi_1)$ at zero while floating $\mathcal{A}$ from 0.1 to 0.9, the dependence between input and output are as Figure 5-14 shows. The linearity fit shows a good agreement as well.

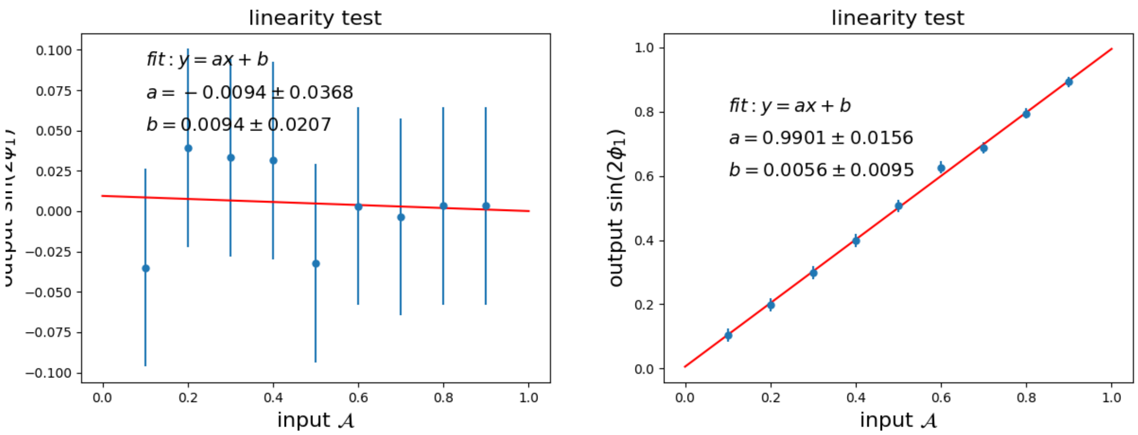


Figure 5-14: Linearity test of CP fit.

5.4.3 Toy MC Fit Pull

In order to check the fit bias with input-output method, a series of 1000 dataset of toy MC has been created containing about 26 events in each. The event number is set based on the expected number from signal region in data after the selection. The χ^2 from vertext fit, events number N and vertex errors on CP and tag-side are sampled from the distribution of data. The fit to dataset is performed with zero input $\sin(2\phi_1)$ and $\mathcal{A}$ as floating parameters. We expect to use the normal distribution to fit the pull of $\sin(2\phi_1)$ and $\mathcal{A}$.

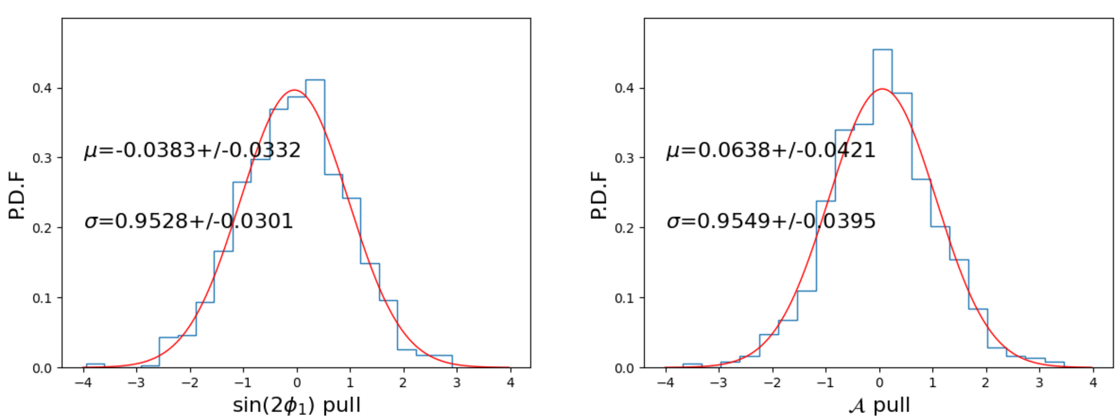


Figure 5-15: Pull of $\sin(2\phi_1)$ and $\mathcal{A}$ fitted with the standard normal distribution.

The fit results shows a good recovery of input $\sin(2\phi_1)$ and $\mathcal{A}$ with no clear bias is spotted.

5.4.4 Lifetime and Δm_d Fit

Before looking at *CP* parameters in data, we need to check if the physics parameters are consistent when setting the *CP* fitter to fit them in float. 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 $B^0 \to K^0_S K^0_S K^0_S$ is 1.431 ± 0.382 ps. The result is consistent with PDG value. The distribution of Δt in lifetime fit is shown as Figure 5-16. The B^0 and B^+ lifetime fit using control sample is also performed and summarized in here[45]. The results are consistent with PDG values as input in MC generator.

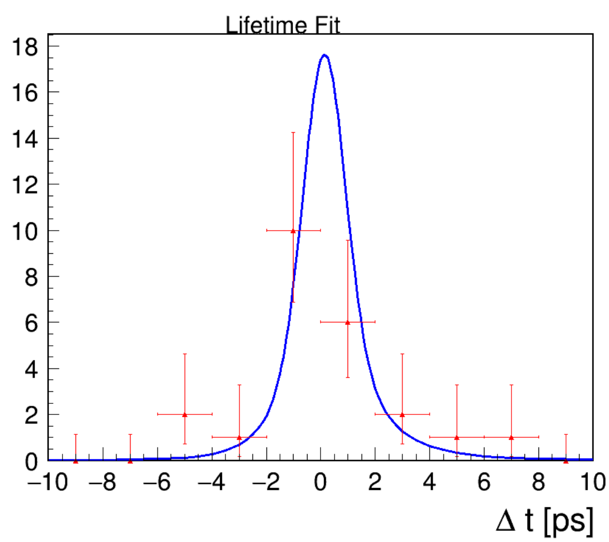


Figure 5-16: Lifetime fit on data

To test the fit on physics parameter Δm_d, we generate 200 toy MC sets of $B^0 \to K_S^0 K_S^0 K_S^0$ with input $\Delta m_d = 0.507$ GeV/c^2 where each set contains 26 events as same as data. The fit result is close to normal distribution and the pull of Δm_d is shown in Figure 5-17.

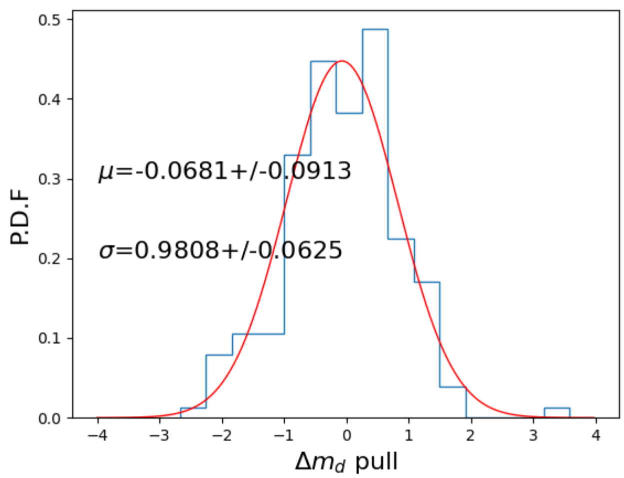


Figure 5-17: Pull of Δm_d

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 5-18.

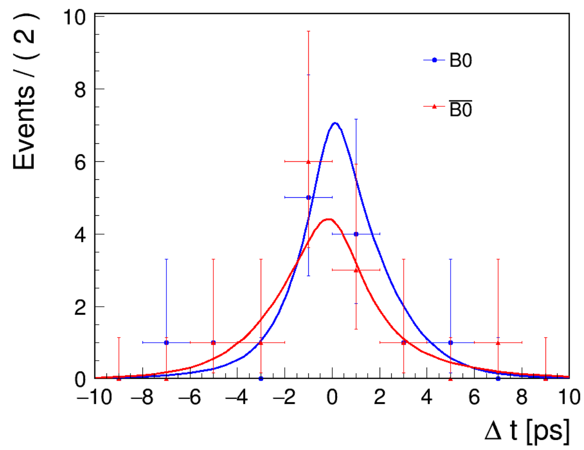


Figure 5-18: The CP fit from data.

The results of CP parameters are:

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

5.6 Systematic Uncertainty

The systematic uncertainty that affects the fit results may come from many aspects of the measurement setup. Currently there is no estimation on systematic uncertainty from the tag-side interference in the Belle II, which is caused by the interference between CKM-favored or CKM-suppressed tree-level decays. The Belle results suggested the contribution is ~ 0.001[49] which is trivial at current systematic uncertainty. As shown in Table 5.7 to 5.11, the total contribution from each source is calculated by adding-in-quadrature using each systematic uncertainty caused by the fit model parameters belonging to that category. To be specific, in each category 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 impact on the CP fit results from every parameter in each category 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. On the other hand, from Table 5.12 to 5.14, the contributions are not directly from the parameters of the fit model, but events number in fit bias test, *KsFinder* correction factor and vertex reconstruction options, we modified their values and repeat the fit to obtain the systematic uncertainties. After obtaining the $\delta\mathcal{S}$ and $\delta\mathcal{A}$ from each category, the total systematic uncertainty is calculated by add-in-quadrature again from all the sources, as shown in Table 5.16.

The signal resolution functions' parameters are determined from MC study for signal component. The impact on fit results is summarized as follows Table 5.7. The background Δt shapes' parameters are determined from data sideband $M_{bc} <$

Table 5.7: 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}	0.005443	0.001299	-0.005675	-0.001404
s_{1CP}^{main}	0.019934	-0.000903	-0.020204	0.000633
s_{0CP}^{tail}	-0.003233	-0.001623	0.003270	0.001596
s_{1CP}^{tail}	-0.002044	-0.000063	0.002009	0.000048
f_{tag}^{tail}	0.003140	-0.001257	-0.003117	0.001266
s_{0tag}^{main}	0.002011	-0.001395	-0.001956	0.001398
s_{1tag}^{main}	0.005059	-0.000840	-0.004969	0.000825
s_{0tag}^{tail}	-0.000135	-0.000393	0.000101	0.000435
s_{1tag}^{tail}	0.000101	0.000027	-0.000472	0.000129
f_{δ}	-0.007248	-0.000552	0.007231	0.000591
f_p	0.003037	0.004347	-0.003069	-0.004314
τ_n	-0.001010	-0.002841	0.000937	0.002940
τ_p	0.004497	0.002502	-0.004648	-0.002478
Total	$\delta\mathcal{S} = 0.033793$		$\delta\mathcal{A} = 0.009283$	

5.26 GeV. The impact on fit results is summarized in Table 5.8. The flavor tagging parameters wrong tagging fraction w in each rbin is determined by applying flavor tagging on control sample. The impact in each rbin on fit results is summarized in Table 5.9: The physics parameters Δm_d and τ_{B^0} uncertainties are included using the PDG average value. The impact on fit results is summarized in Table 5.10. The signal fraction is determined using 2D fit results of M_{bc} and ΔE from data. The impact on fit results is summarized in Table 5.11.

The fit bias uncertainties are determined by taking the larger ones among the fit error of 300000 *signal MC* events with zero *CP* violation and the difference between

Table 5.8: Systematic uncertainty from background Δt shapes

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
μ_{δ}^{bkg}	-0.014294	-0.016581	0.006758	0.006537
μ_{l}^{bkg}	-0.002798	-0.012567	0.003789	0.012783
τ_{bkg}	0.001377	0.001689	-0.004159	0.000085
f_{δ}^{bkg}	-0.011315	0.001365	0.011187	-0.001395
f_{tail}^{bkg}	-0.002661	0.001530	0.002480	-0.001368
σ_{main}^{bkg}	0.020702	0.022041	-0.023618	-0.015690
σ_{tail}^{bkg}	-0.000275	-0.000159	0.000179	0.000141
Total	$\delta\mathcal{S} = 0.039430$		$\delta\mathcal{A} = 0.037269$	

Table 5.9: Systematic uncertainty from wrong tagging fraction

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
w_1	-0.001892	0.001911	0.001855	-0.002004
w_2	-0.001645	0.001104	0.001609	-0.001155
w_3	-0.000490	0.001344	0.000473	-0.001341
w_4	0.000656	0.000264	-0.000654	-0.000255
w_5	-0.000123	0.000204	0.000123	-0.000195
w_6	0.000095	0.000054	0.000096	-0.000045
w_7	0.000191	-0.000396	-0.000191	0.000402
Total	$\delta\mathcal{S} = 0.003709$		$\delta\mathcal{A} = 0.003790$	

Table 5.10: Systematic uncertainty from physics parameters

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
Δm_d	-0.001767	-0.000687	0.001778	0.000696
τ_{B^0}	-0.004561	-0.000546	0.004565	0.000555
Total	$\delta\mathcal{S} = 0.006923$		$\delta\mathcal{A} = 0.001250$	

Table 5.11: Systematic uncertainty from signal fraction

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
mu1_mbc	0.000822	-0.003888	-0.000797	0.003849
sigma1_mbc	0.000476	0.008442	-0.000628	-0.008733
m0_argus	-0.000707	0.004140	0.001448	-0.005781
c_argus	-0.005544	0.001449	0.000922	-0.000078
f1_de	0.027826	0.020589	-0.019237	-0.008409
f2_de	0.020809	0.017649	-0.016129	-0.007005
mu1_de	-0.000443	-0.000153	0.000496	0.000088
mu2_de	-0.000563	0.001446	0.000591	-0.001446
mu3_de	-0.003164	-0.000834	0.003354	0.000981
sigma1_de	-0.000172	-0.000966	0.000206	0.000906
sigma2_de	-0.003150	0.002958	0.002635	-0.002475
sigma3_de	-0.001926	-0.002550	0.002470	0.002985
a0_cheb	0.000952	0.000057	-0.000893	-0.000102
N_sig_f	-0.004640	0.003987	0.004922	-0.003504
Total	$\delta\mathcal{S} = 0.044387$		$\delta\mathcal{A} = 0.033932$	

input and output of the center value. The fit result is $\mathcal{S} = 0.000127 \pm 0.009817$ and $\mathcal{A} = 0.000265 \pm 0.005702$. So the values of fit errors are used as listed in Table 5.12.

Table 5.12: Systematic uncertainty from fit bias

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
fit bias	0.009817	0.005702

Applying *KsFinder* cut at 0.74 based on MC study may introduce small impact on data due to the different response on the classifier between data and MC. Therefore the contribution of systematic uncertainty from KsFinder is considered. At cut value 0.74, the $\mathcal{R}_{\mathcal{B}'}$ presenting MC and data signal yield ratio is $\mathcal{R}_{B^0} = 1.027 \pm 0.033$, where the upper and lower limit is 1.060 and 0.994, respectively. These two ratios are applied on the signal fraction obtained by data to repeat the fit, and the difference of fit results compared to the original values are used as systematic uncertainty, see Table 5.13.

For the contributions from vertex reconstruction, the impacts from the selections in Table 5.5 are considered. Given the fact that cut values in Table 5.5 are very loose and the statistics from data is very limited, the changing of the these values doesn't affect events collected from data so that systematic uncertainty can not be reflected

Table 5.13: Systematic uncertainty from *KsFinder*.

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
$\mathcal{R}_{B^0} = 1.06$	0.004826	-0.000606
$\mathcal{R}_{B^0} = 0.994$	-0.000508	0.000007
Total	0.004852	0.000606

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 impact from the IP constraint options as well as the potential bias are not considered. The summarized systematic uncertainties are listed in Table 5.14, where the zero values appear due to the unchanged events input under the modified ranges.

Table 5.14: Systematic uncertainty from vertex reconstruction

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
$\sigma_{z_{tag}} < 0.05$ cm	0.004369	-0.003599
$\sigma_{z_{tag}} < 0.15$ cm	0.000000	0.000000
$\chi^2/N(CP) < 3$	0.018197	-0.020242
$\chi^2/N(CP) < 13$	0.000000	0.000000
$\chi^2/N(tag) < 40$	0.000000	0.000000
$\chi^2/N(tag) < 60$	0.000000	0.000000
$\lvert\Delta t\rvert < 50$ ps	0.003325	-0.000396
$\lvert\Delta t\rvert < 90$ ps	0.000000	0.000000
IP constraint	0.000000	0.000000
Total	0.019007	0.020563

The detailed study of the tag-side interference is not available from the current Belle II thus the Belle result $\delta\mathcal{S} \sim 0.001$ and $\delta\mathcal{A} \sim 0.008$ are referenced[49], as shown in Table 5.15.

The contributions at the curent stage of Belle II are summarized based on the all the above sources, as shown Table 5.16.

Table 5.15: Systematic uncertainty of the tag-side interference .

Sources	$\delta\mathcal{S}$	$\delta\mathcal{A}$
tag-side interference	0.001000	0.008000

Table 5.16: The contributions of each source and the total of systematic uncertainty.

Sources	$\delta\mathcal{S}$	$\delta\mathcal{A}$
signal fraction	0.044387	0.033932
background Δt shapes	0.039430	0.037269
signal Δt shapes	0.033793	0.009283
wrong tag fraction	0.003709	0.003790
fit bias	0.009817	0.005702
physics parameters	0.006923	0.001250
KsFinder impact on data	0.004852	0.000606
vertex reconstruction	0.019007	0.020563
tag-side interference	0.001000	0.008000
Total	0.072186	0.056233

Chapter 6

Conclusion, discussion and prospect

The CP parameters measurement is performed based on the validation of analysis strategies by blind analysis. In the MC study, the fit is performed which shows a consistent result for CP parameters compared to the simulation input. The linearity and pull of the CP fit are checked to validate the reliability of the fit procedures. The fit result on B^0 lifetime using experiment data is also agreed with the current value in PDG with a relatively large statistical uncertainty due to the low statistics from data.

After the CP fit procedures are validated, the CP parameters $\mathcal{S}$ and $\mathcal{A}$ using the Belle II early data with 62.8 ab^{-1}in 2019 and 2020 spring and summer is performed. The result is shown in Equation 6.1.

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

The result agrees with the prediction of the Standard Model and the previous results from Belle[20] and BaBar[50]. The systematics study is performed considering the main contributing sources at the moment. The measurement precision of CP

parameters in this study is majorly limited due to the large statistical uncertainties from low statistics, which leads to no clear evidence or hint on the NP effects.

In this thesis, the analysis strategies that aim to maximize the efficiency and purity are developed and a conservative measurement approach is taken for the Belle II early data. The newly developed *KsFinder* contributes much in improving the signal significance by effectively rejecting fake K_S^0, which is also useful in background rejection for other channels with K_S^0 in the final state. The model of the resolution of vertex positions has been studied using MC sample and sideband data, which is benefiting from 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.

6.1 Improvements on statistical uncertainty

This study has shown a good potential of performing *CP* measurement in Belle II for the incoming years with more data recorded. The precision on $\mathcal{S}$ requires the large luminosity as shown in Figure 1-6 and the statistical uncertainty in this thesis fits in the scale. With 50 ab^{-1} luminosity from the full Belle II data sample in future, the statistical uncertainty is expected to be 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, due to the flavor tagging accuracy limitation, the expected statistics of tagged events could be increased as well, however, at this moment no clear sign of large boosting on flavor tagging efficiency or reduction of wrong tagging fraction is foreseen. Therefore, the expected reduction of statistical uncertainty is mostly due to the increased data sample, with current reconstruction efficiency assumed, which is shown in Figure 6-2. As the current luminosity is quite far away from 50 ab^{-1}, it is hard to precisely predict the uncertainty in full Belle II luminosity. Given that the expected sensitivity of ΔS is in a precision level of three decimal places for both $\eta' K_S^0$ and ϕK_S^0[15], therefore such a precision is taken to estimate the future uncertainties based on the current results. The estimated statistical uncertainty is ~ 0.030 as shown in Figure 6-2.

From Table 4.3, the current B^0 reconstruction efficiency is about 35% which could be further optimized mainly by improving K_S^0 efficiency. As discussed in chapter 3, the K_S^0 reconstruction efficiency and quality becomes worse for long-flight ones. 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 considered from a same track so that they can be used together during tracking. If a K_S^0 decay outside of layer 5 at 10.4 cm, even though the daughter tracks pass the SVD layer 6, unless they hit the overlapping region of the SVD edges which will create two hits instead of one, otherwise only SVD hits information is missing and such a track become a CDC-only track. This effect is shown in Figure 6-1, where *SVD00* type K_S^0 start to appear at SVD layer 5. Such a algorithm is to suppress the beam background and SVD noise strips that create a large fraction of random single hits. Thus, the actual sensitivity volume of SVD is reduced and the K_S^0 efficiency is negatively affected. In future, the improvement of tracking algorithm is expected to remove this requirement while still be able to effectively reduce single hits background.

In general, the expected B^0 signal yield can be improved in future and help to further reduce the statistical uncertainty. From Figure 6-2, the extrapolated statistical uncertainty at 0.711 ab^{-1} using current value 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 ~ 0.030 at 50 ab^{-1} as a conservative value for estimating total uncertainty later.

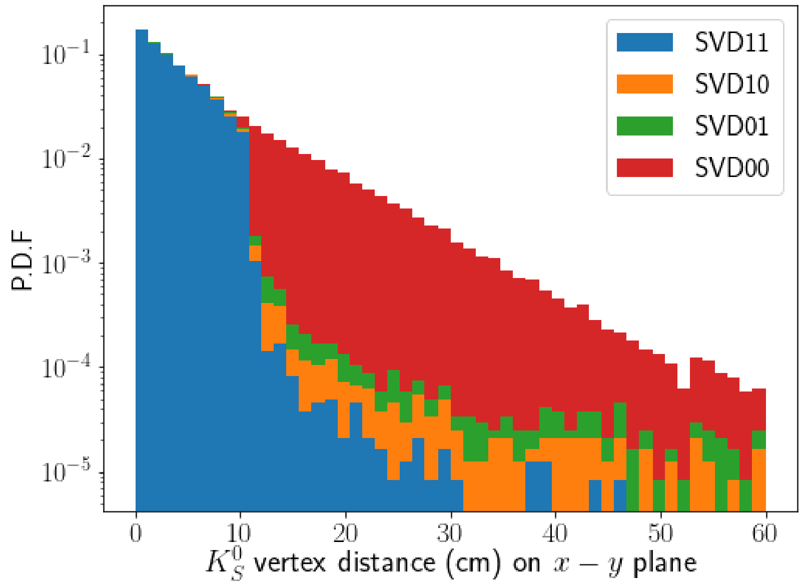


Figure 6-1: K_S^0 flight length on $x-y$ plane for each category of K_S^0, where *SVD00* type K_S^0 start to appear at about SVD layer 5.

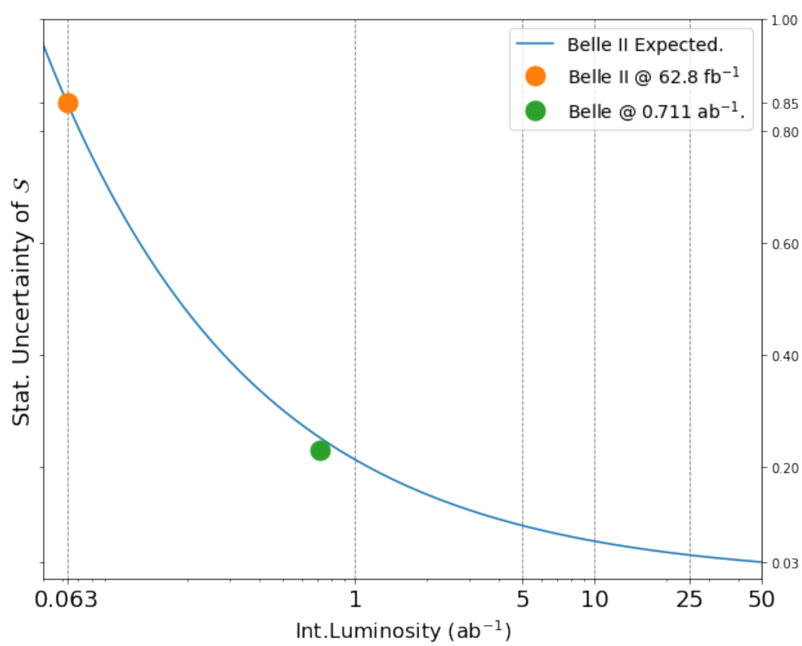


Figure 6-2: Statistical uncertainty of $\mathcal{S}$ extrapolation based on the current result of $B^0 \to K_S^0 K_S^0 K_S^0$ in Belle II, where the orange is the current value and the green is the Belle result at 0.711 ab^{-1} Υ(4S) data.

6.2 Improvements on systematic uncertainty

As discussed in the introduction chapter, the NP effects that potentially contributes to the $B^0 \to \phi K_S^0$ could also affect $B^0 \to K_S^0 K_S^0 K_S^0$. In the current SM correction, the QCD factorization (QCDF) scan approach suggests the expected upper-limit for $\Delta\mathcal{S}$ is about 0.05[15], and the "add-in-quadrature" QCDF predicts the $\Delta S = 0.01 \sim 0.02$[15]. In this scale, it's 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. The irreducible sources mainly refer to the terms that are not scaled or improved with increased luminosity, such as the irreducible vertexing related terms and tag-side interference[15]. The reducible sources are mainly the signal Δt shape , the signal fraction, the background Δt shape, the flavor tagging, and the fit bias. As an initial study in early operation with very low statistics, it's summed that the reducible terms are achieved from the increased future luminosity where the reduction of parameters' uncertainties are scaled by the fraction of squared-root of increased events statistically. If no improvement on systematic uncertainty is expected from the current evaluation, with the full Belle II data it is still challenging to validate the evidence of the NP effects against the small theoretical predictions. In the Belle II original prospect, the new PXD detector would be contributive to the reduction of vertexing related systematic uncertainties by 50% on average. However, the PXD is not fully installed at the current Belle II and the vertex reconstruction in this analysis doesn't include IP constraint options. Hence the conservative scenario that systematic uncertainties do not change due to the largely improved vertexing quality is assumed, which makes the increased luminosity the major factor for precision improvements. To show the estimated systematic uncertainties in future, the data collection as well as MC simulation at 50 ab^{-1} is taken as reference. Similar to the discussion in statistical uncertainties, the precision level of the estimation is set to three decimal places[15].

For signal Δt shape, we assumed that the CP-side resolution functions parameters can optimized by increased *signal MC* and better vertex reconstruction. The current contribution from CP-side resolution parameters is 0.029946. The overall 50% reduction in the uncertainties of CP side resolution parameters is assumed, which can be calculated to be 0.015 by using the CP-side related terms in the Table 5.7. The current tag-side parameters contribution is 0.015658. The tag-side parameters are determined from 2 ab^{-1} control sample MC which could be further reduced at the Belle II 50 ab^{-1} luminosity so the reduced uncertainties are expected to be about 20% of the current value, which is ~ 0.003. In Table 6.1, the improvement where both CP and tag side uncertainties are reduced is calculated to be 0.015 while a conservative case that CP side resolution remain as the present ones 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 term in the systematic uncertainties currently, it mainly suffers from the very low statistics in data that causes the inaccurate description of signal shape against background. The signal fraction is determined event-by-event using M_{bc} and ΔE 2D fit shape in Figure 4-16. Hence, the uncertainties of signal fraction parameters are expected to be reduced quickly with the increased data collection in future. Using 1 ab^{-1} *generic MC*, the signal fraction contribution is estimated for comparison, where the combined contribution on $\mathcal{S}$ uncertainty by floating $\pm 1\sigma$ for each parameter is ~ 0.013. This estimation is close to the observed systematic uncertainty from signal fraction in the Belle data which is ~ 0.015[20] and the data scaled value ~ 0.011, both with $\mathcal{O}(0.01)$ level difference. Therefore, the current contribution is taken as the baseline to extrapolate the reduced systematic uncertainty in 50 ab^{-1} Belle II data, reduced by luminosity ratio to ~ 0.002, listed in Table 6.2.

For the background Δt shape, it can be reduced by the increased luminosity since

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

Luminosity (50 ab^{-1})	Improved uncertainty
Signal fraction	~ 0.002

the current parameters are determined by sideband data with very low statistics. Therefore, the systematic uncertainty from background Δt shape is reduced by factor ~ 28 calculated from $\sqrt{50/0.063}$ to be ~ 0.001 as listed in Table 6.3. This estimation is checked by comparing the background Δt shape uncertainties in between the Belle result and the one in this analysis, where the background Δt shape contribution in the Belle result is 0.017[20] with a tighter sideband region used, which is consistent with scaled luminosity of the Belle data.

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	~ 0.001

For the contributions of wrong tag fraction, the flavor tagging related information is expected to be determined by experimental data using flavor-specific decay modes with increased the Belle II data in future. To estimate the expected uncertainties from the Belle II future data, the results on w in each r-bin by using about 8.7 fb^{-1} Belle II in 2019 are compared with the Belle results[48]. The flavor tagging performance studied by Belle II early data presented a close efficiency and wrong tag fraction values compared to the Belle results. In each r-bin, the uncertainties of w is reduced averagely by factor of ~ 11, which is consistent with the squared root of the luminosity ratio. Hence, the expected uncertainties at 50 ab^{-1} Belle II data is assumed to be ~ 7 times smaller than those from the Belle result[20], which is about ~ 0.002 as listed in Table 6.4.

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	~ 0.002

For the fit bias contribution as systematic uncertainty, currently the values are taken by the statistical fit error using 300000 *signal MC*. In future, the fit bias contribution is expected to be estimated by taking the larger one among the fit statistical error using more MC sample and the center value difference between the input and output. Thus, if MC sample used in fit could be at least 100 times more than one million in future, then the fit error is possible to be smaller than input-output difference. From the current MC production plan of 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 0.009817 from Tabel 5.12 as an estimation similar to the one used in *CP*-side resolution functions, 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 necessary with much larger data sample. In addition, the impact of using *KsFinder* receives contribution from the data MC mismatch on signal purity, which is expected to be improved by better data MC consistency in future. For physics parameters Δm_d and τ_{B^0}, the uncertainties could be further reduced by better physics input. Finally, vertex reconstruction options are not contributing much in this analysis mostly because of the very loose cuts and no IP constraint used. However, this does not mean their contributions are missing since the resolution function parameters obtained under such loose IP conditions receives more inaccuracies than the ones from using tighter cuts and IP constraint. The signal Δt shape contribution partially absorbs the expected uncertainties from vertex reconstruction options. In future, by using proper IP constraint and tighter vertex cuts, the signal Δt shape will contribute less and vertex reconstruction will contribute more to systematic uncertainties. The current systematic uncertainty from the tag-side interference is referenced as ~ 0.001 as a negligible term from the Belle study of $B^0 \to (c\bar{c}K^0)$[49]. In the full Belle II data, such a value is still not considered as an large contribution. We keep it as 0.001 considering it's an irreducible source. The current estimations from *KsFinder*(0.004852), physics parameters(0.006923), the vertex reconstruction(0.019007) and tag-side interference(~ 0.001) are unchanged in the 50 ab^{-1} Belle II luminosity to have a conservative expectation on $\mathcal{S}$ systematic uncertainty, shown in Table 6.7.

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	$\sim$ 0.002
fit bias	$\sim$ 0.005
Total	$\sim$0.016(0.031)

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	0.004852
physics parameters	0.006923
vertex reconstruction	0.019007
tag-side interference	$\sim$ 0.001
Total	0.020826

6.3 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 terms from Table 6.6 and the unchanged ones 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

~ 0.026. Both are shown in Table 6.8 where the major contribution is from signal Δt shapes and the vertex reconstruction options which are both related to the vertexing quality.

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($\sim$0.063)	CP/tag(50)	only-tag(50)
Syst.Uncert.($\mathcal{S}$)	~ 0.072	$\sim$0.026	~ 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. With no CP vertex resolution improvement, total uncertainty of ~ 0.048 is expected. On the other hand, the total uncertainty will be at ~ 0.040 if 50% CP-side resolution improvement is assumed, which is a conservative estimation considering that the current CP-side vertexing performance is not optimized and some unchanged systematic uncertainties could possibly be improved. Given the total uncertainty from $B^0 \to J/\psi K_S^0$ at that time is expected to be ~ 0.005[15], ΔS sensitivity is 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 ~ 0.05 without taking into account any uncertainty, which is close to the theoretical predicted upper-limit. 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($\sim$0.063)	CP/tag(50)	only-tag(50)
Tot.Ucert.($\mathcal{S}$)	~ 0.853	~ 0.040	~ 0.048

6.4 *KsFinder* importance

While monitoring the uncertainties of the *CP* parameters is crucial in searching the hidden NP effects, avoiding bias 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 biased and 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 bias introduced by the large fraction of background events that yield random *CP* asymmetry due to the statistical fluctuation. The larger background events without using *KsFinder* cut in Table 4.2 can produce the wrongly estimated signal fraction (f_{sig}) in Equation 5.5 so that the *CP* parameters are biased using the biased fit model. Especially when the luminosity is increased in future, the signal fraction uncertainty is expected to be largely reduced, such a biased signal fraction will be treated as a large contribution to the systematic uncertainty compared to the correct ones. To demonstrate the effect, 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 significance in M_{bc} and ΔE 2D fit is considerately lower than the ones 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. Clearly the signal fraction defined by the 2D fit in the latter case is biased from the MC truth, which shows a positively biased signal fraction on average, as summarized in Table 6.10.

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

Table 6.10: The signal and background events using different K_S^0 selection cuts compared with the *generic MC* counts, showing that signal events reconstructed using *KsFinder* is more precise 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

is 39.3% and the fit result is $(43.7 \pm 1.4)\%$, which shows over 3σ deviation as a strong bias. If such a bias is taken into account as the systematic uncertainty, signal fraction difference is used as a floating value to check the impact on the CP fit results, which leads to an extra systematic uncertainty from the biased f_{sig} at level of ~ 0.006, already larger than any other reducible source except for the signal Δt shapes as listed in Table 6.6. Therefore, the development of *KsFinder* is particularly important in the precised CP measurement for $B^0 \to K_S^0 K_S^0 K_S^0$. The current performance of *KsFinder* is presenting a purity about 95% in K_S^0 reconstruction which means there is still small room for improvements, as well as the background rejection power. 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.

6.5 Prospect

Even though the current result on CP parameters are dominated by the large uncertainty, mostly from the statistical one, the previous discussions about the future uncertainties 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, where the statistical and reducible systematic

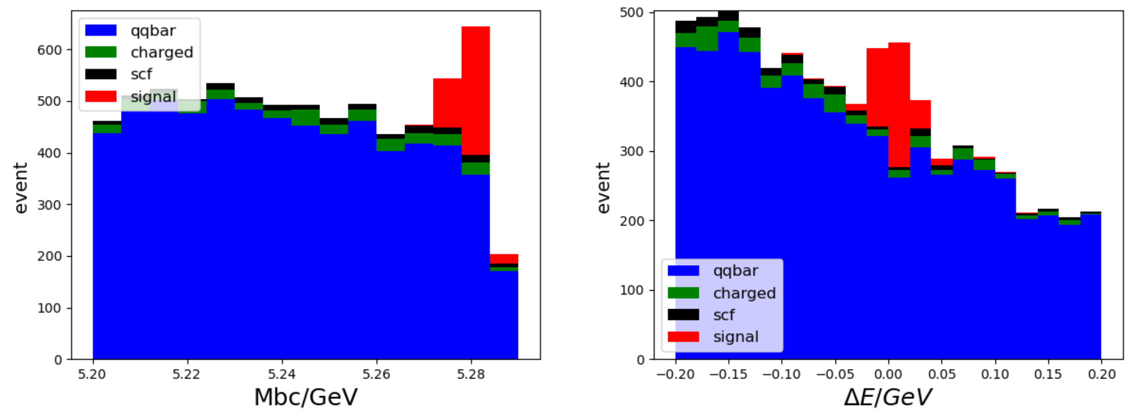


Figure 6-3: M_{bc} and ΔE stacked histogram of 1 ab^{-1} *generic MC* sample replacing $FBDT\_Ks > 0.74$ by $cosVertexMomentum > 0.9$ in Table 4.2 as a selection criteria, showing a much worse signal significance.

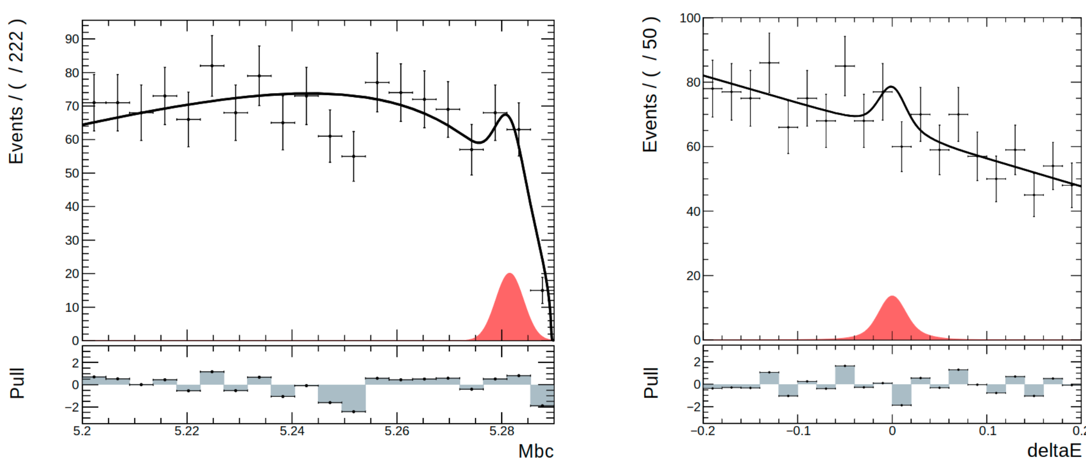


Figure 6-4: M_{bc} and ΔE 2D fit on 1 ab^{-1} *generic MC* sample, replacing $FBDT\_Ks > 0.74$ by $cosVertexMomentum > 0.9$ in Table 4.2. The red is signal component.

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. In conclusion, the progress that has been made so far in this thesis paves a well-constructed and solid path for searching the NP effects in time dependent CP violation study of $B^0 \to K_S^0 K_S^0 K_S^0$.

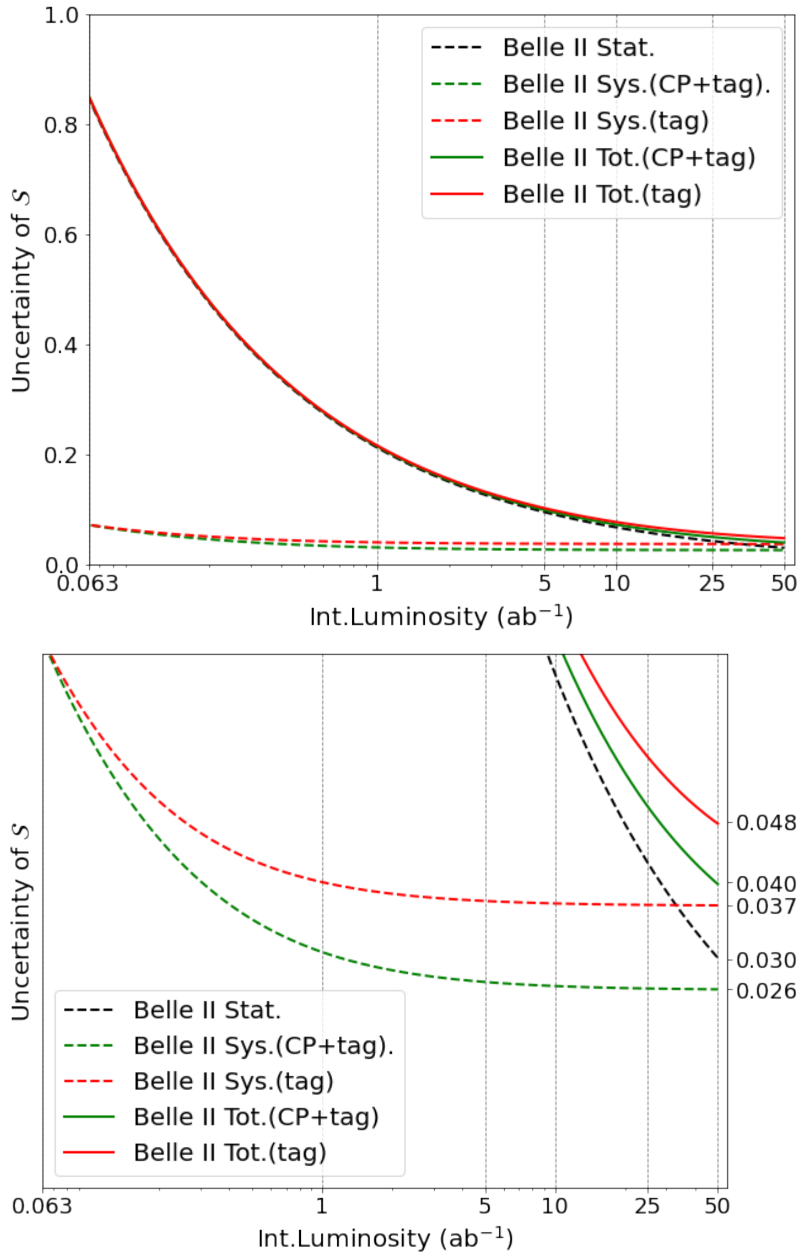
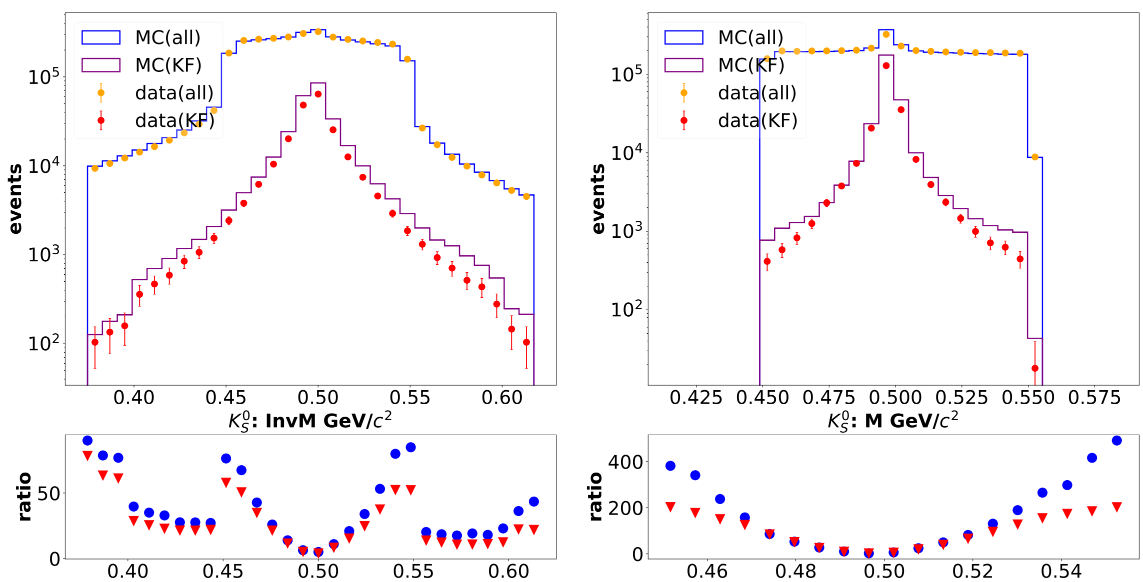


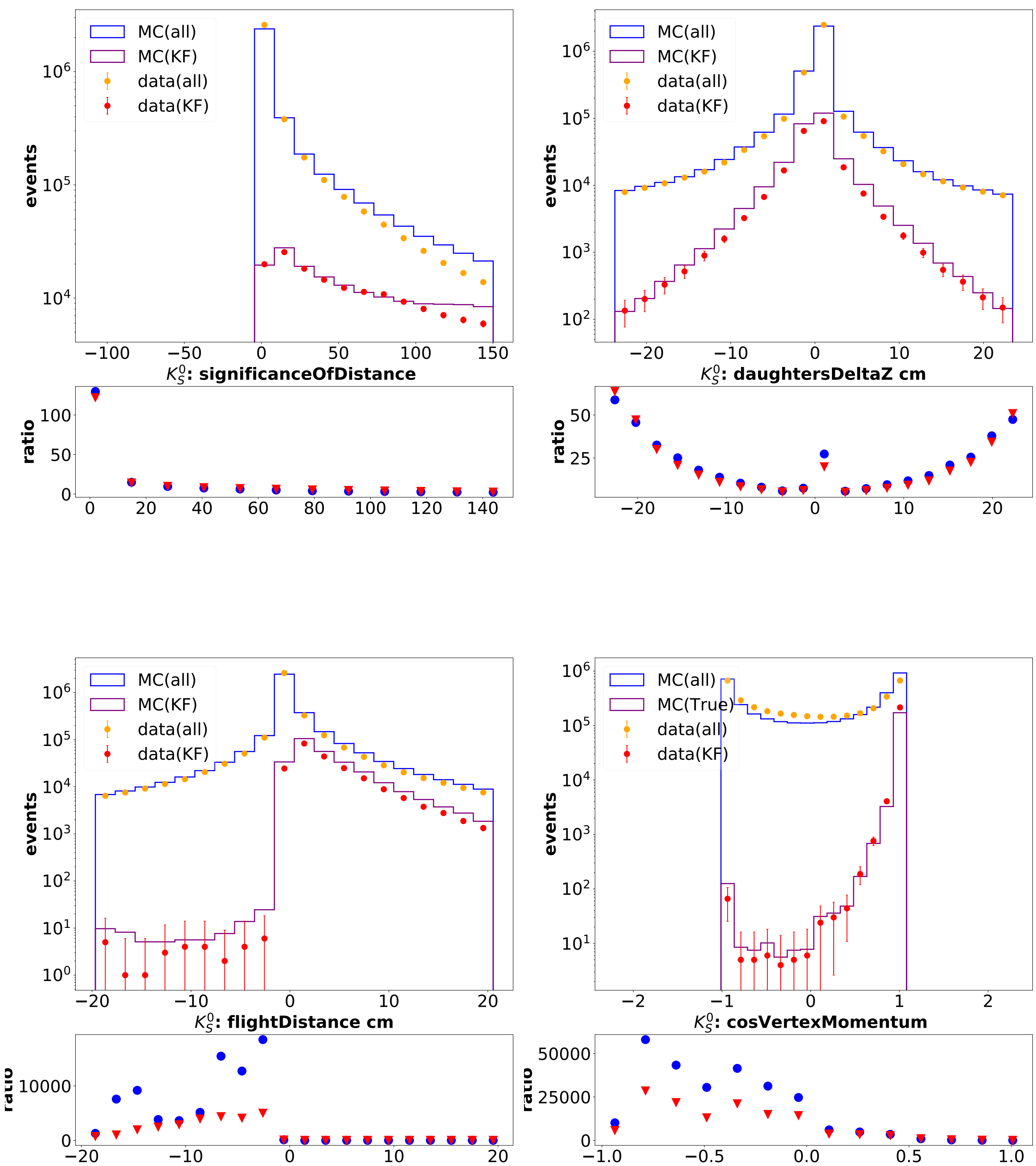
Figure 6-5: The expected total uncertainty of $\mathcal{S}$ in $B^0 \to K_S^0 K_S^0 K_S^0$, where the dashed lines are the statistical(black), CP/tag-side improved systematic(green) and only tag-side improved systematic(red) uncertainties, with the corresponding solid lines as the total uncertainties.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

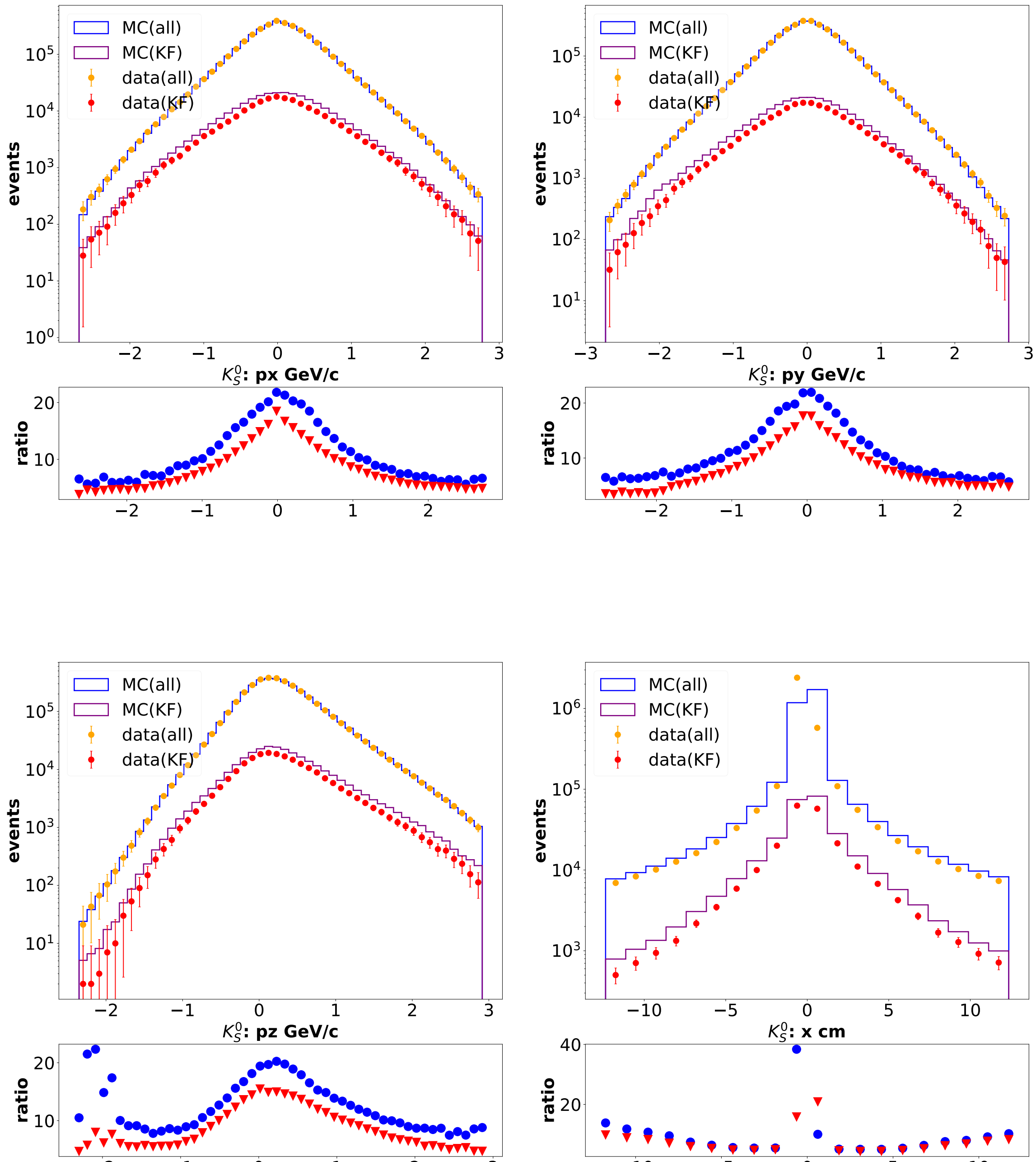
Appendix A

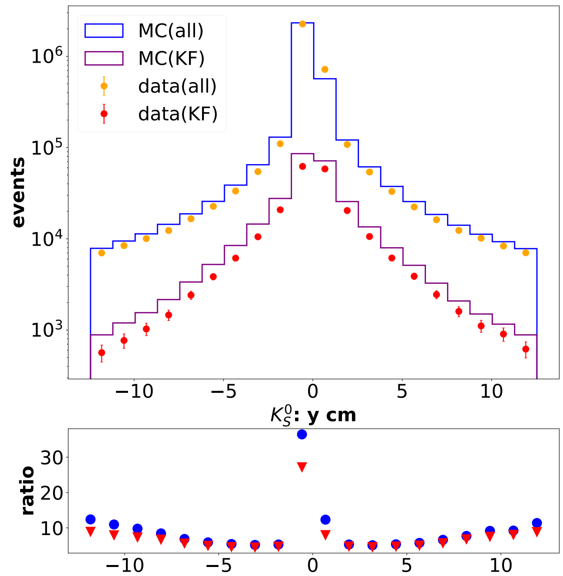
Data Validation Plots for K_S^0

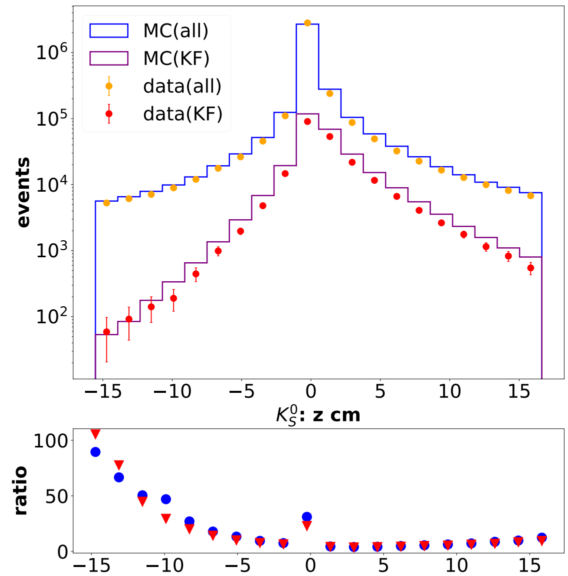
Figure A-1: The distribution of the training variables in KsFinder. The blue and purple solid lines are the total and true K_S^0 distributions from generic MC, respectively. The yellow and red dots are the data distribution before and after applying KsFinder cut. The uncertainties in data are taken as three times the Poisson standard deviation.

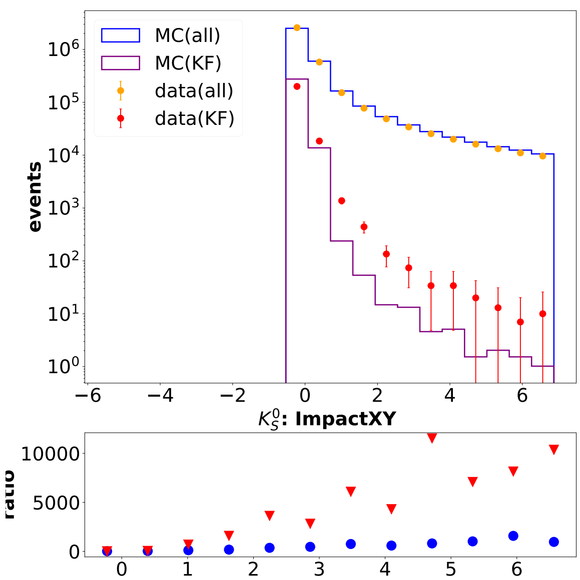


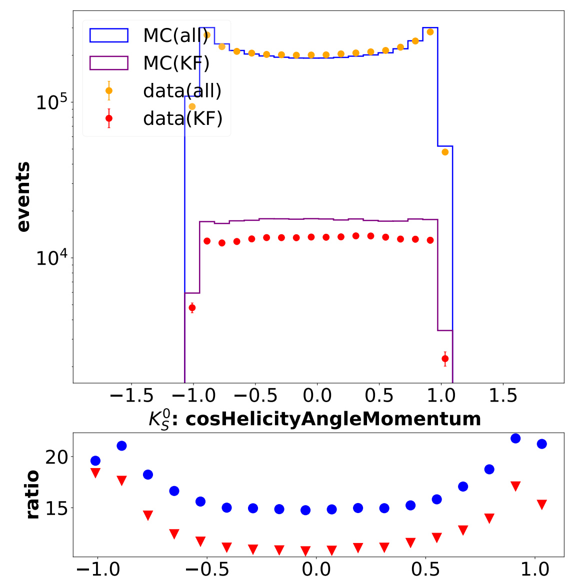

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: significanceOfDistance
ratio
K_S^0: daughtersDeltaZ cm
K_S^0: flightDistance cm
MC(True)
K_S^0: cosVertexMomentum

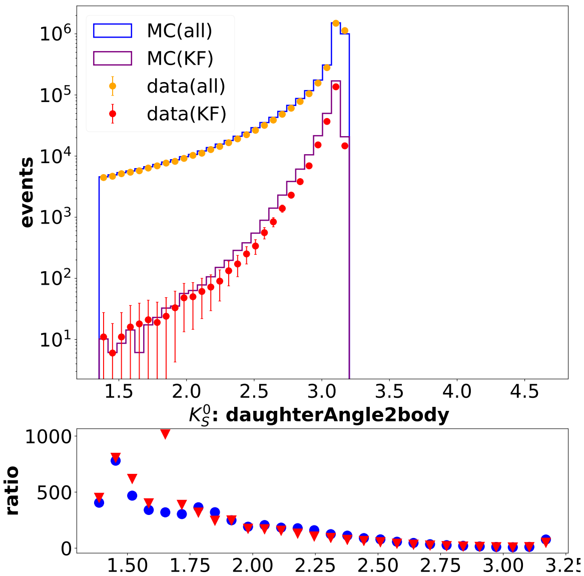

MC(all)
MC(KF)
data(all)
data(KF)
events
ratio
K_S^0: px GeV/c
K_S^0: py GeV/c
K_S^0: pz GeV/c
K_S^0: x cm

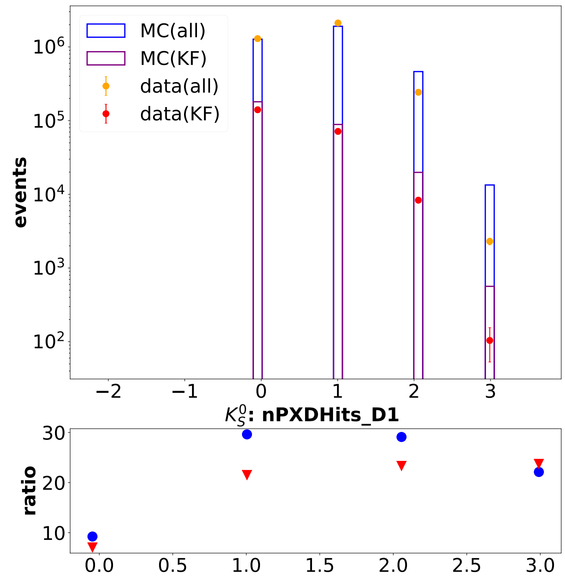

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: y cm
ratio

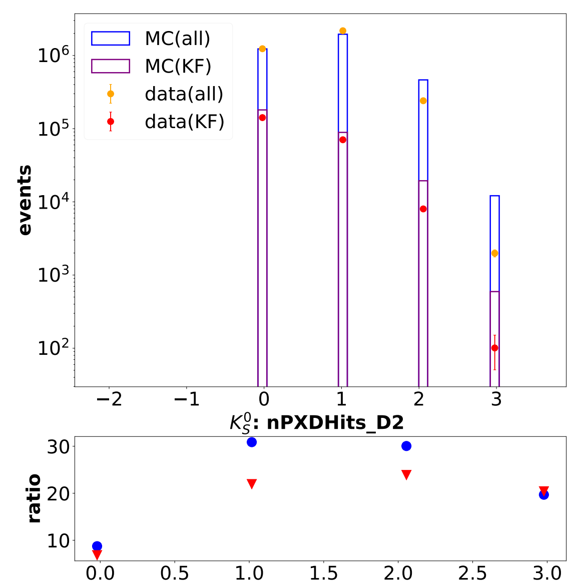

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: z cm
ratio

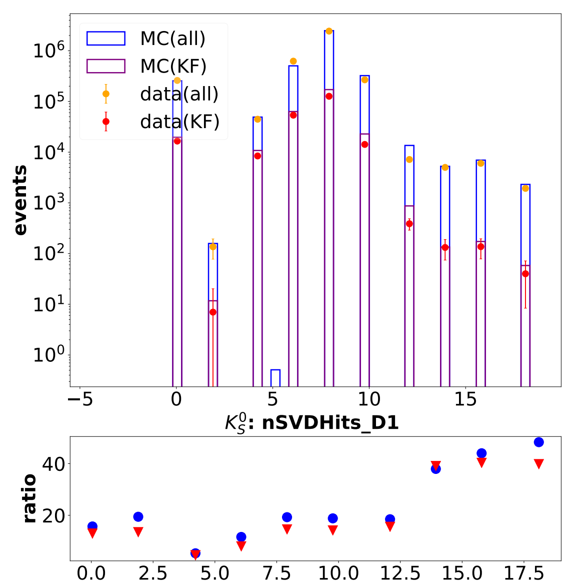

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: ImpactXY
ratio

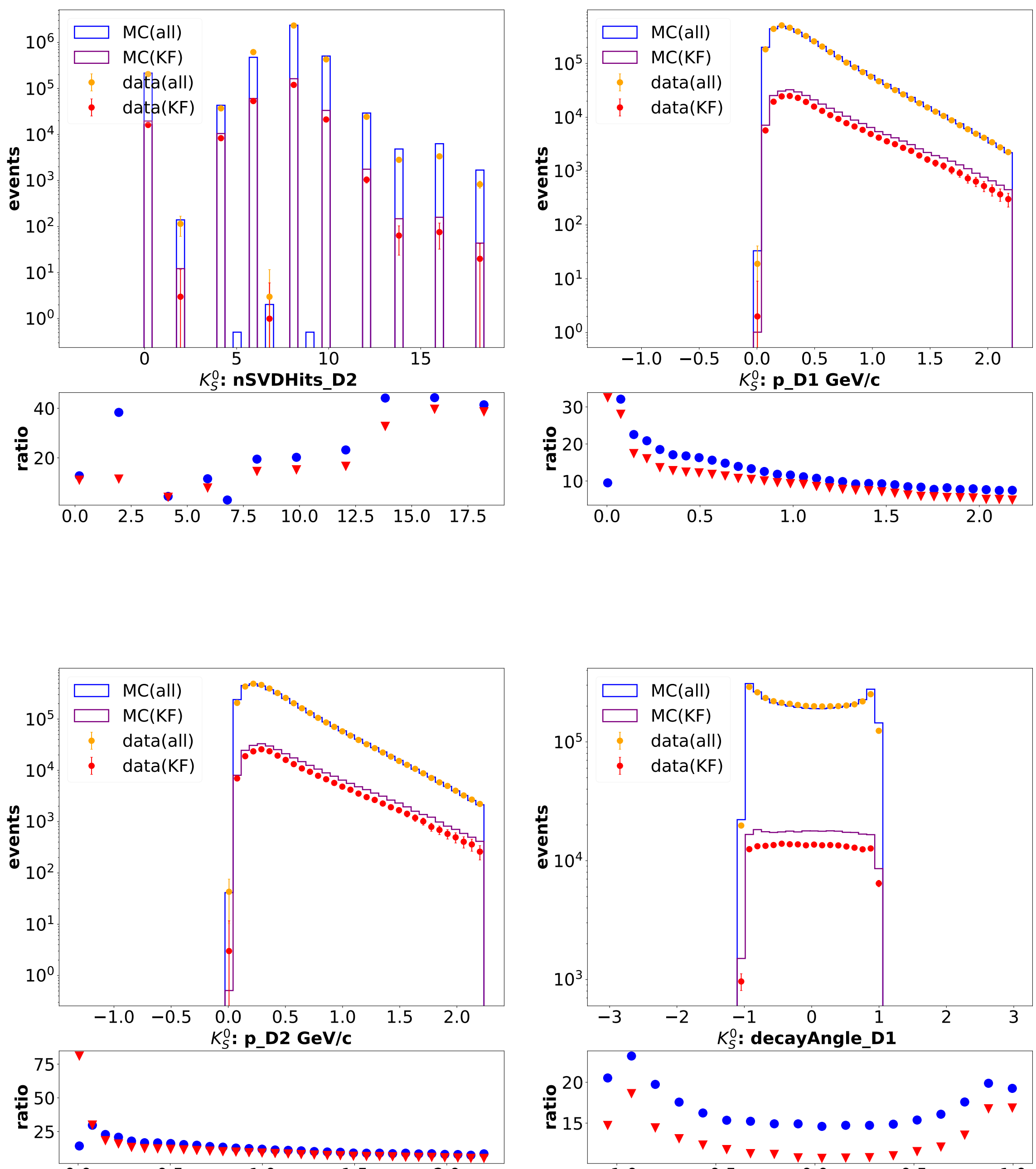

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: cosHelicityAngleMomentum
ratio

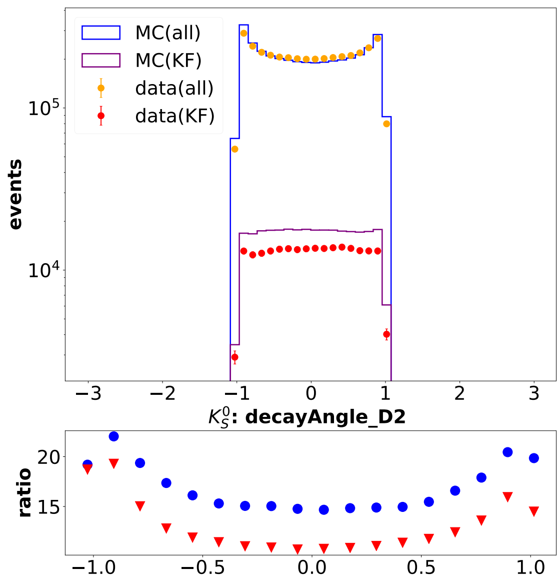

MC(all)
MC(KF)
data(all)
data(KF)
events
K_S^0: daughterAngle2body
ratio


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
ratio
K_S^0: nSVDHits_D2
K_S^0: p_D1 GeV/c
K_S^0: p_D2 GeV/c
K_S^0: decayAngle_D1



MC(all)
MC(KF)
data(all)
data(KF)
10^5
10^4
events
−3
−2
−1
0
1
2
3
K_S^0: decayAngle_D2
20
15
ratio
−1.0
−0.5
0.0
0.5
1.0

Appendix B

Control Samples

Table B.1: Hadronic control sample reconstruction criteria. 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 $\rho^+ \to \pi^+\pi^0$[45].

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 where A,B and C are in the increasing order of momentum

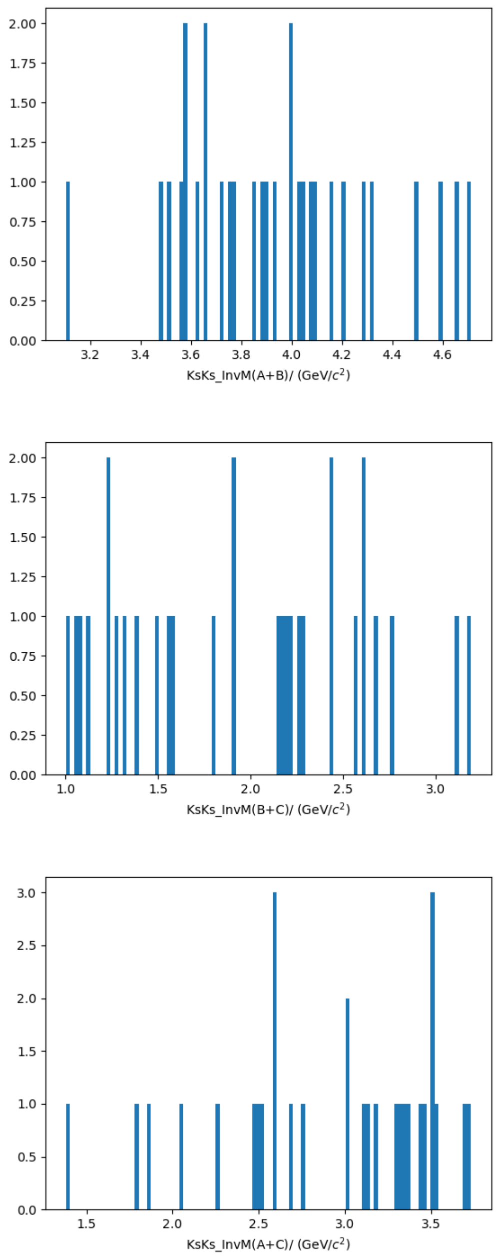


Figure C-1: Data in signal region

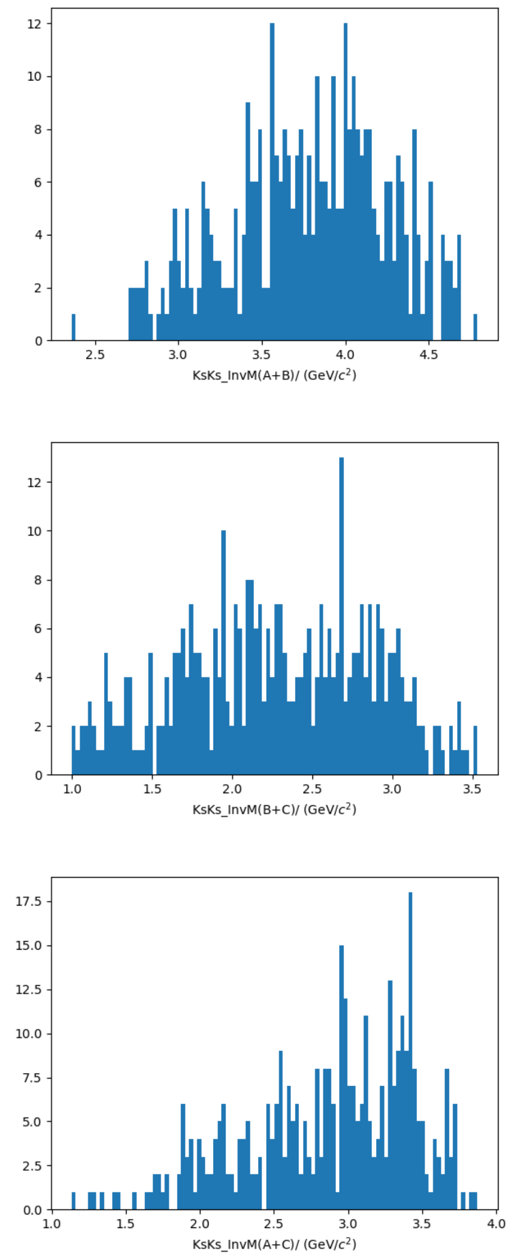
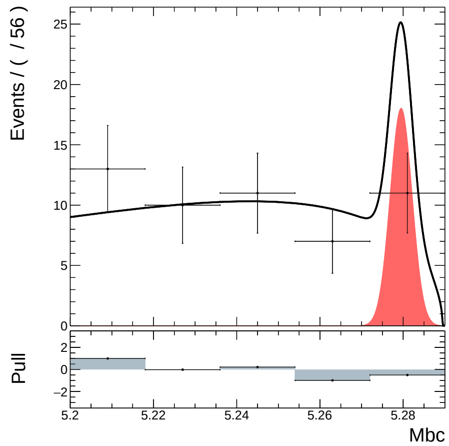


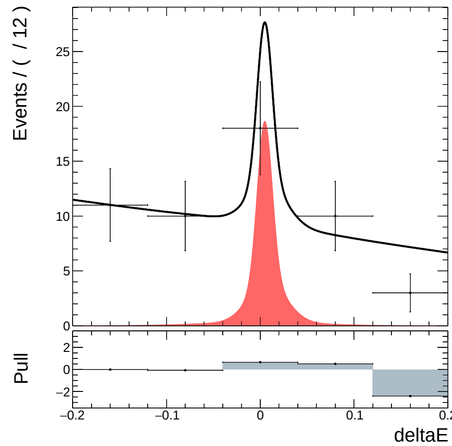
Figure C-2: Generic MC in signal region.

Appendix D

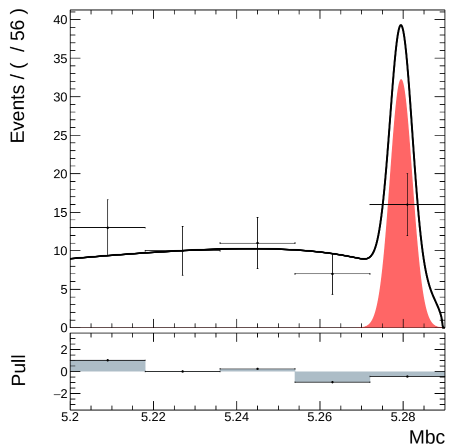
Injection test for B^0 signal yield



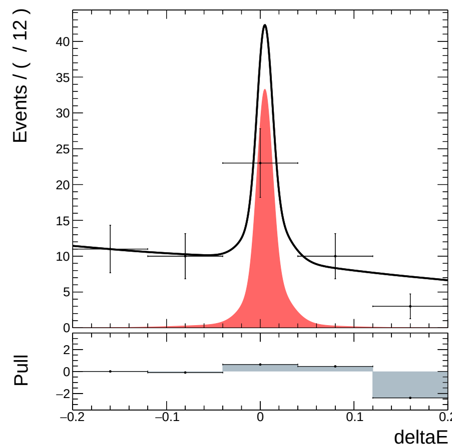
a) signal injected: 5



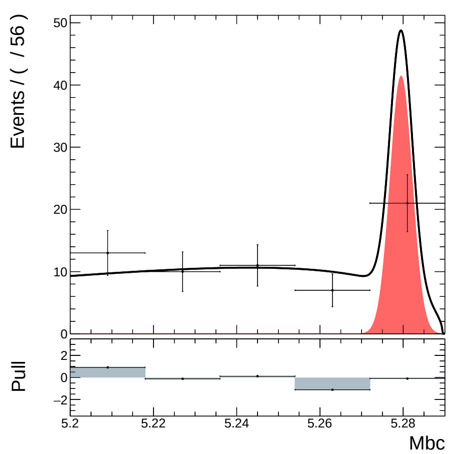
b) signal injected: 5



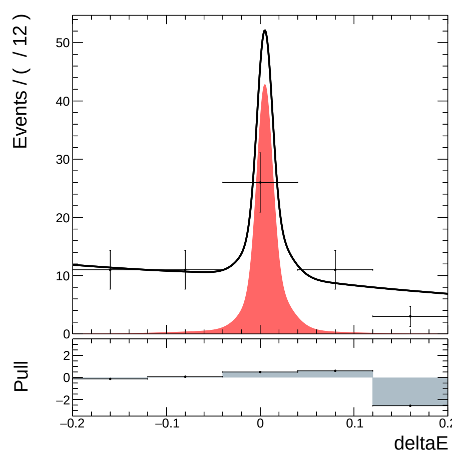
c) signal injected: 10



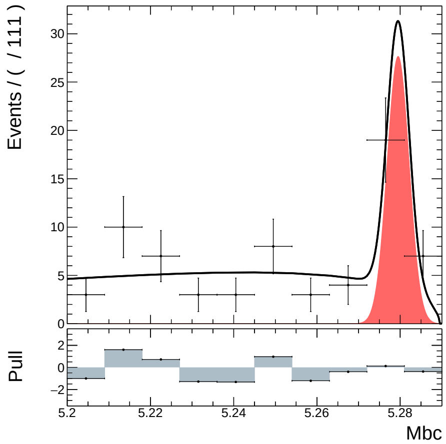
d) signal injected: 10



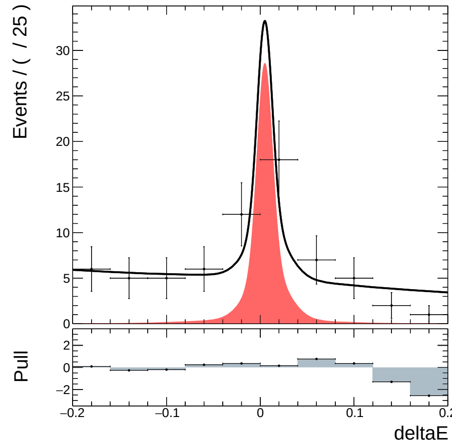
e) signal injected: 15



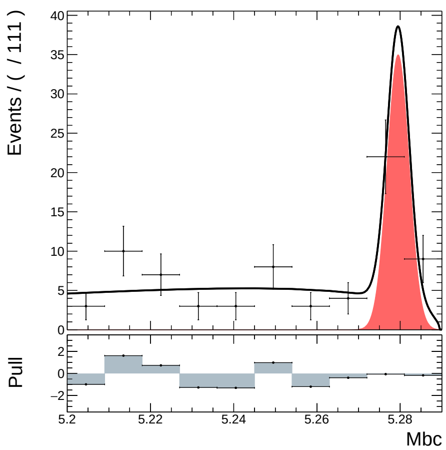
f) signal injected: 15



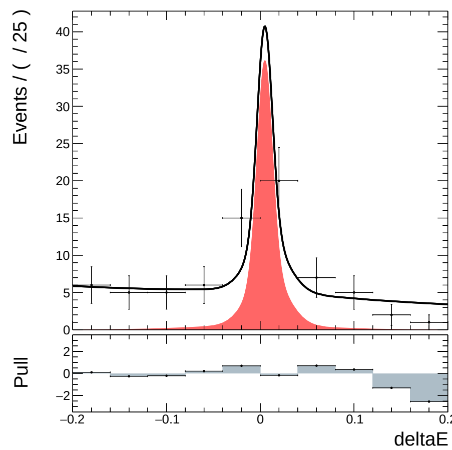
g) signal injected: 20



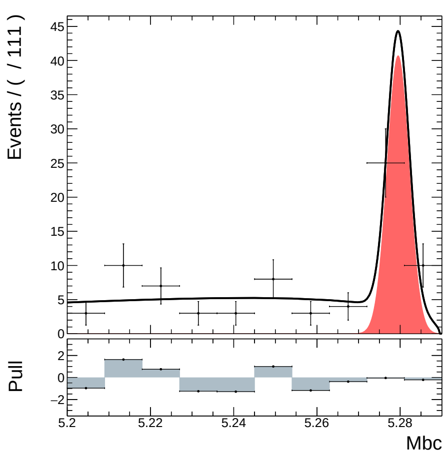
h) signal injected: 20



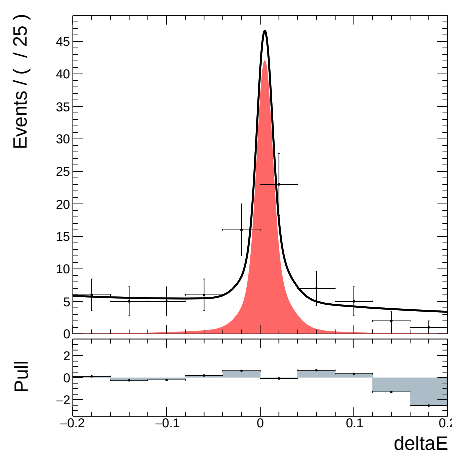
i) signal injected: 25



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.

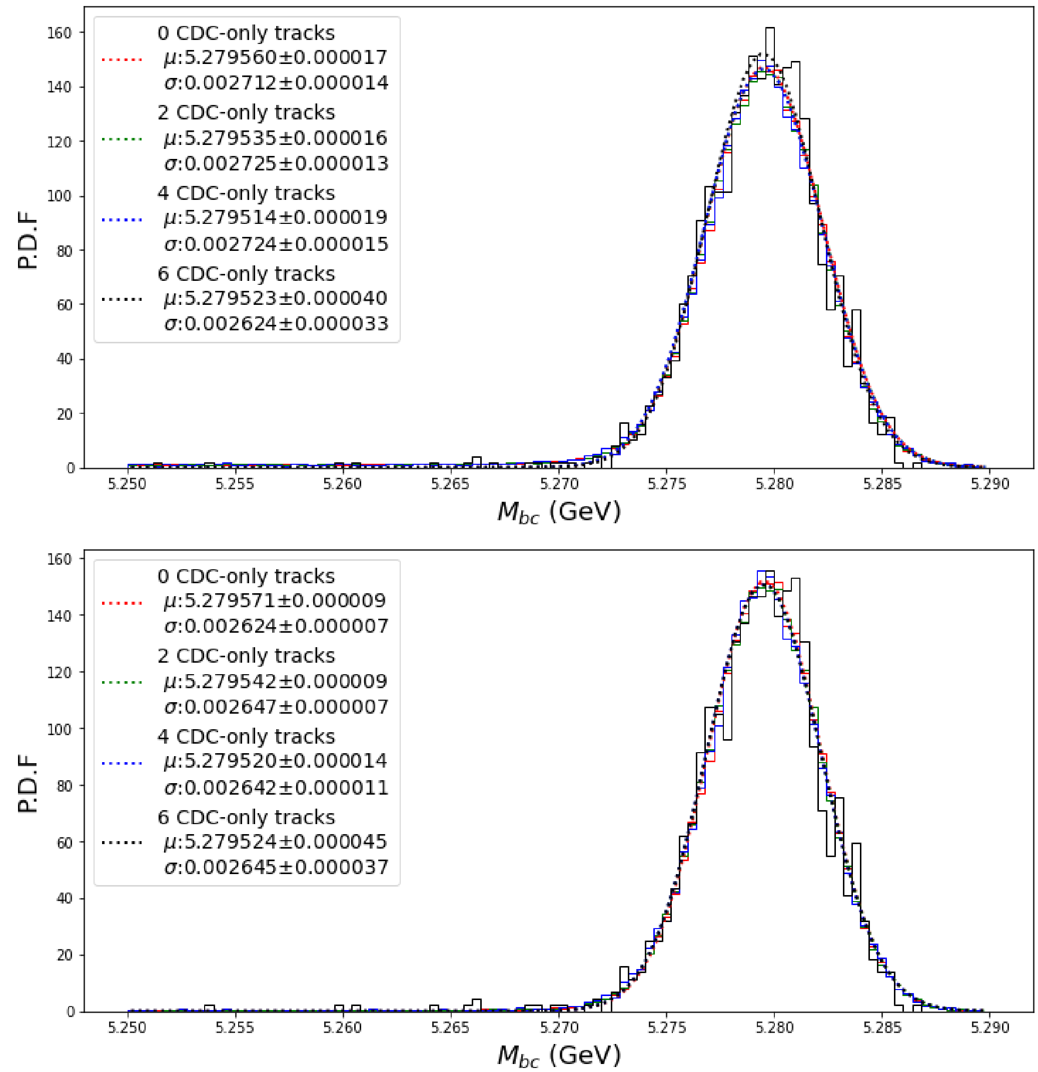


Figure E-1: M_{bc} distribution based on the number of CDC-only tracks in final states. Top: no *KsFinder* used; Bottom: *KsFinder* used. The μ and σ are the mean and standard deviation of the Gaussian function.

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

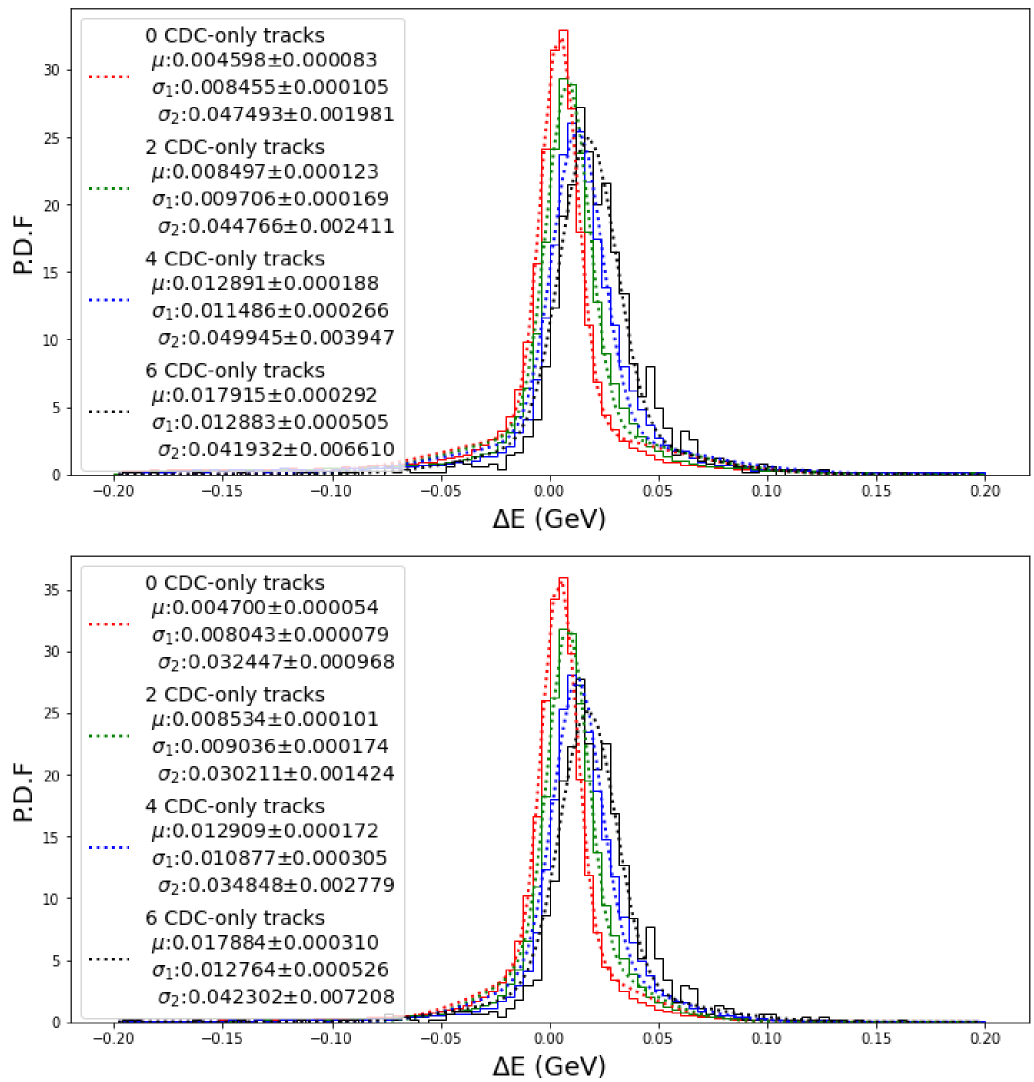


Figure E-2: ΔE distribution based on the number of CDC-only tracks in final states. Top: no *KsFinder*; Bottom: *KsFinder* used. The μ is the common mean for double Gaussian. The σ_1 and σ_2 are the standard deviations of the Gaussian function.

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

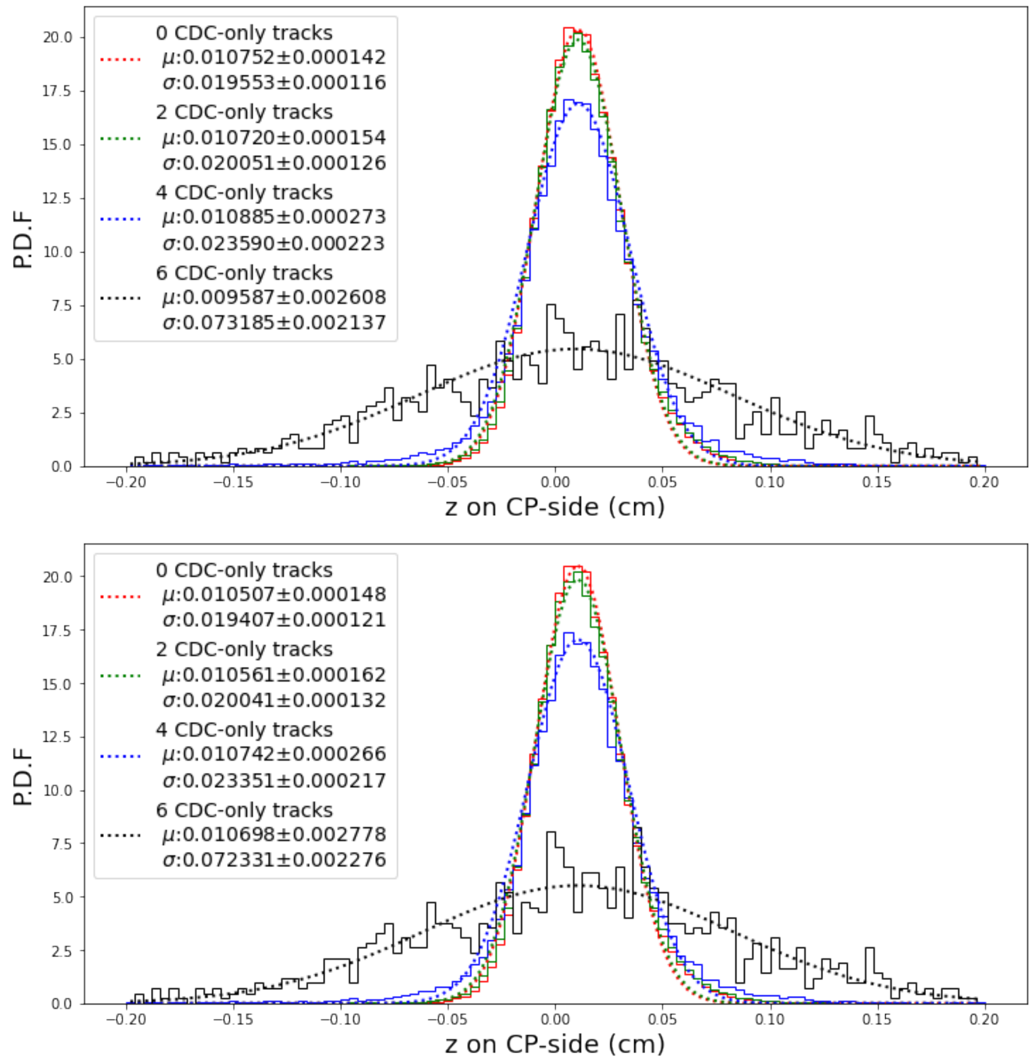


Figure E-3: Δz distribution based on the number of CDC-only tracks in final states. Top: no *KsFinder*; Bottom: *KsFinder* used. The μ and σ are the mean and standard deviation of the Gaussian function.

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.

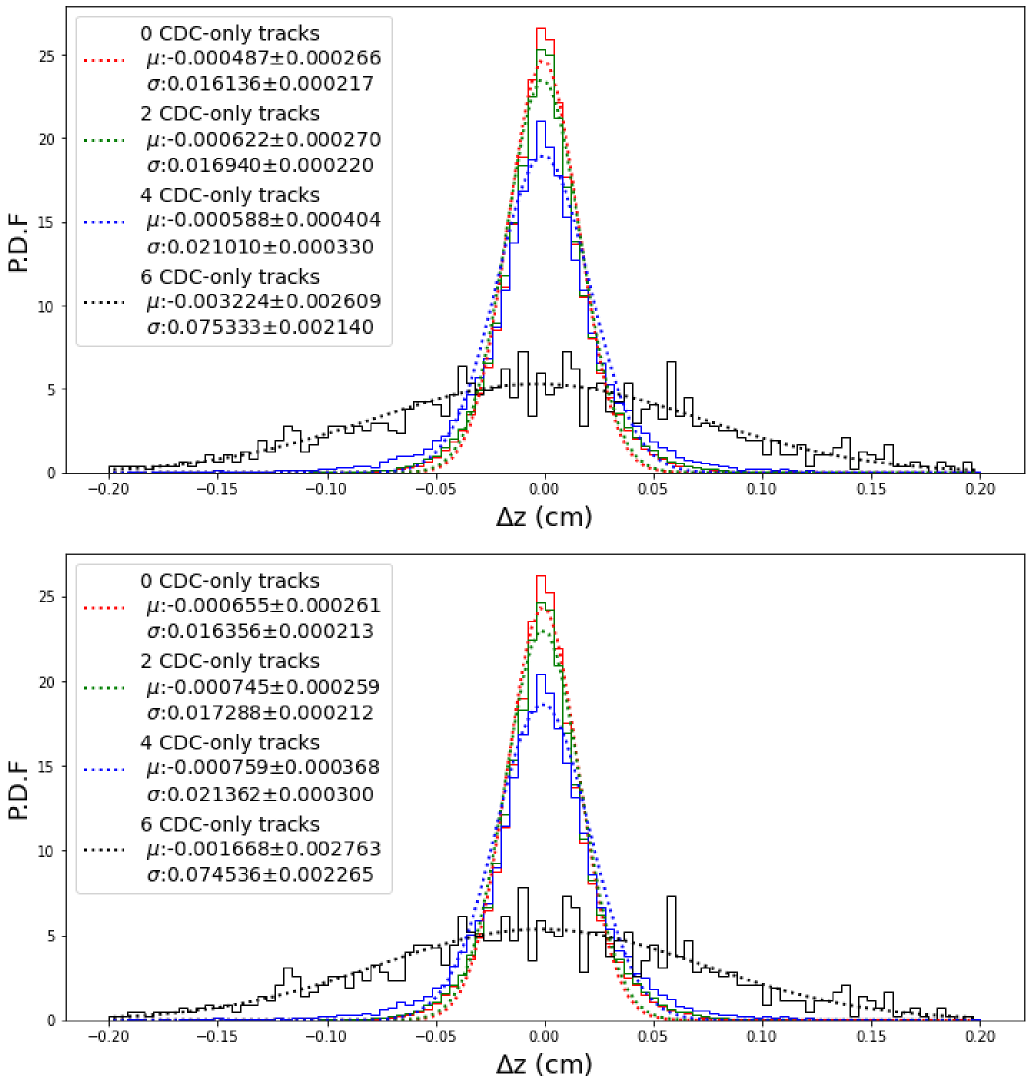


Figure E-4: Δz distribution based on th number of CDC-only tracks in final states. Top: no *KsFinder*; Bottom: *KsFinder* used. The μ and σ are the mean and standard deviation of each Gaussian function.

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. 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] URL https://commons.wikimedia.org/wiki/File:Parity_violation_principle_Wu_experiment_(English).jpg.

[5] James H Christenson, Jeremiah W Cronin, Val L Fitch, and René Turlay. Evidence for the 2 π decay of the k 2 0 meson. *Physical Review Letters*, 13(4):138, 1964.

[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] Yoshiki Teramoto. Cp violation in b physics at belle. 2002.

[8] David Lange. Study of cp violation at babar.

[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] Marc Sher. Electroweak higgs potential and vacuum stability. *Physics Reports*, 179(5-6):273–418, 1989.

[11] Augusto Ceccucci. The ckm quark-mixing matrix. 2008.

[12] ckmfitter. 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 $B^0 \to J/\psi K^0_S$ Decays with the LHCb Experiment.* PhD thesis, Dortmund U., 2014.

[15] 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.

[16] 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 c p violation in b0 $\to \eta^{'}$ k0 decays and improved measurements of c p asymmetries in b0 $\to \varphi$ k0, ks ks ks and b0$\to$ j/ψ k0 decays. *Physical review letters*, 98(3): 031802, 2007.

[17] Tim Gershon and Masashi Hazumi. Time-dependent cp violation in b0$\to$ p0p0x0 decays. *Physics Letters B*, 596(3-4):163–172, 2004.

[18] 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.

[19] 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.

[20] 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 $b^0 \to k_s k_s k_s$ decays at belle. *arXiv preprint arXiv:2011.00793*, 2020.

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

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

[23] 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.

[24] 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.

[25] 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.

[26] URL `https://root.cern/`.

[27] URL `https://dirac.readthedocs.io/en/latest/DeveloperGuide/Overview/`.

[28] 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.

[29] 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`.

[30] URL `https://evtgen.hepforge.org/doc/models.html`.

[31] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G 2 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.

[32] URL `https://pdglive.lbl.gov/Particle.action?init=0&node=S042&home=MXXX045`.

[33] URL `https://confluence.desy.de/display/BI/Phase+3+luminosity+monitoring`.

[34] 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.

[35] 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.

[36] 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`.

[37] 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.

[38] 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.

[39] 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.

[40] 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 inb0→ks0ks0ks0decay. *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`.

[41] 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`.

[42] Adrian Bevan, Bostjan Golob, Thomas Mannel, Soeren Prell, and Bruce Yabsley. *The physics of the B factories*. Springer Nature, 2017.

[43] 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→ u decays. *Physics Letters B*, 241(2):278–282, 1990.

[44] *BELLE-II-NOTE-TE-2019-010 (2020).* URL `https://docs.belle2.org/record/1384/files/belle2-note-resol_belle.pdf`.

[45] URL https://docs.belle2.org/record/1941/files/BELLE2-NOTE-PH-2020-027.pdf.

[46] 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.

[47] URL https://docs.belle2.org/record/1905/files/BELLE2-NOTE-PH-2020-013_v3.2.pdf?version=1.

[48] 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.

[49] Y Yosuke and K Miyabayashi. Measurement of time-dependent cp violation in b0→(cc) k 0 decays with 772× 106bb. Technical report, Belle Note, 2011.

[50] JP Lees, V Poireau, V Tisserand, J Garra Tico, E Grauges, M Martinelli, DA Milanes, A Palano, M Pappagallo, G Eigen, et al. Amplitude analysis and measurement of the time-dependent c p asymmetry of b 0→ k s 0 k s 0 k s 0 decays. *Physical Review D*, 85(5):054023, 2012.