

博士論文

Doctoral Dissertation

Time Dependent Charge-Parity Violation in  $B^0 \rightarrow K_s^0 K_s^0 K_s^0$  in Belle

II early operation

(Belle II 初期データを使った  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  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 琏

Wan Kun



**Time Dependent Charge-Parity Violation in  $B^0 \rightarrow K_s^0 K_s^0 K_s^0$  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**

Belle II experiment is a next-generation super B-factory experiment. The targeted instantaneous luminosity is  $8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$  and the expected integrated luminosity is  $50 \text{ ab}^{-1}$  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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  decay to precisely measure the  $CP$  parameters  $\mathcal{S}$  and  $\mathcal{A}$  in penguin-dominated  $b \rightarrow s$  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  $62.8 \text{ fb}^{-1}$  recorded experiment data from Belle II detector 2019 and 2020 (spring and summer) operation. The measurement results:  $\mathcal{S} = -\sin(2\phi_1) = -0.82 \pm 0.85 \text{ (stat)} \pm 0.07 \text{ (syst)}$  and  $\mathcal{A} = -0.21 \pm 0.28 \text{ (stat)} \pm 0.06 \text{ (syst)}$  are obtained.

Thesis Supervisor: Hiroaki Aihara

Title: Professor



# Contents

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	The Standard Model . . . . .	9
1.2	Symmetry Violation . . . . .	10
1.3	CKM mechanism . . . . .	12
1.4	Time Dependent $CP$ violation . . . . .	17
1.4.1	$CP$ violation in neutral B system . . . . .	17
1.4.2	$\phi_1$ from $B^0 \rightarrow J/\psi K_S^0$ . . . . .	20
1.4.3	$\phi_1$ from penguin-dominated mode $b \rightarrow q\bar{q}s$ . . . . .	21
1.4.4	$\phi_1$ from $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ . . . . .	22
<b>2</b>	<b>Belle II experiment</b>	<b>25</b>
2.1	Belle II and SuperKEKB overview . . . . .	25
2.2	Vertex detector (VXD) . . . . .	28
2.3	Central drift chamber (CDC) . . . . .	31
2.4	TOP and ARICH detectors . . . . .	32
2.5	Electromagnetic calorimeter (ECL) . . . . .	34
2.6	$K_L^0$ muon detector (KLM) . . . . .	36
2.7	Trigger and DAQ system . . . . .	36
2.8	Analysis software framework . . . . .	40
2.8.1	BASF2 Core Structure . . . . .	40
2.8.2	Event processing workflow . . . . .	41
2.8.3	mDST structure . . . . .	43
2.8.4	Conditional Database . . . . .	44

2.9	Belle II simulation . . . . .	44
2.10	Belle II data taking . . . . .	47
<b>3</b>	<b><math>K_S^0</math> reconstruction study</b>	<b>49</b>
3.1	Cut-based $K_S^0$ Reconstruction . . . . .	50
3.2	MVA-based $K_S^0$ selection . . . . .	53
3.2.1	Belle II $K_S^0$ classification . . . . .	53
3.2.2	Decay Topology of $K_S^0 \rightarrow \pi^+\pi^-$ . . . . .	55
3.2.3	Determination of training observables from $K_S^0$ decay . . . . .	57
3.2.4	Training, Applying and Testing of KsFinder . . . . .	59
3.2.5	The Performance and Over-fitting check . . . . .	61
3.2.6	Data Validation for Classifier . . . . .	64
3.2.7	KsFinder Effects on Kinematics Evaluation . . . . .	66
<b>4</b>	<b>Analysis strategies</b>	<b>69</b>
4.1	Data Sample and Event Selection . . . . .	69
4.1.1	$K_S^0$ Selection . . . . .	69
4.1.2	$B^0$ Reconstruction . . . . .	72
4.1.3	Continuum Suppression . . . . .	73
4.1.4	Resonance Background . . . . .	80
4.1.5	$B\bar{B}$ background and self-cross feed . . . . .	81
4.1.6	Signal Extraction . . . . .	81
4.1.7	Kinematics and Vertexing Dependence on KsFinder . . . . .	86
<b>5</b>	<b><math>CP</math> parameters measurement</b>	<b>95</b>
5.1	Vertex Resolution Model . . . . .	96
5.1.1	$CP$ -side resolution function . . . . .	97
5.1.2	Tag-side resolution function . . . . .	98
5.1.3	Background events $\Delta t$ distribution . . . . .	102
5.2	Flavor Tagging . . . . .	103
5.3	$CP$ Fitter . . . . .	105

5.4	Blind analysis and fit . . . . .	106
5.4.1	$CP$ fit on MC samples . . . . .	106
5.4.2	Linearity Test . . . . .	108
5.4.3	Toy MC Fit Pull . . . . .	109
5.4.4	Lifetime and $\Delta m_d$ Fit . . . . .	110
5.5	$CP$ fit on data . . . . .	112
5.6	Systematic Uncertainty . . . . .	112
<b>6</b>	<b>Conclusion and Prospect</b>	<b>117</b>
<b>A</b>	<b>Training Observables for <math>K_S^0</math> Classifier</b>	<b>121</b>
<b>B</b>	<b>Data Validation Plots for <math>K_S^0</math></b>	<b>125</b>
<b>C</b>	<b><math>K_S^0</math> mass fit with varied cut value</b>	<b>129</b>
<b>D</b>	<b><math>2K_S^0</math> invariant mass distribution</b>	<b>135</b>



# 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.

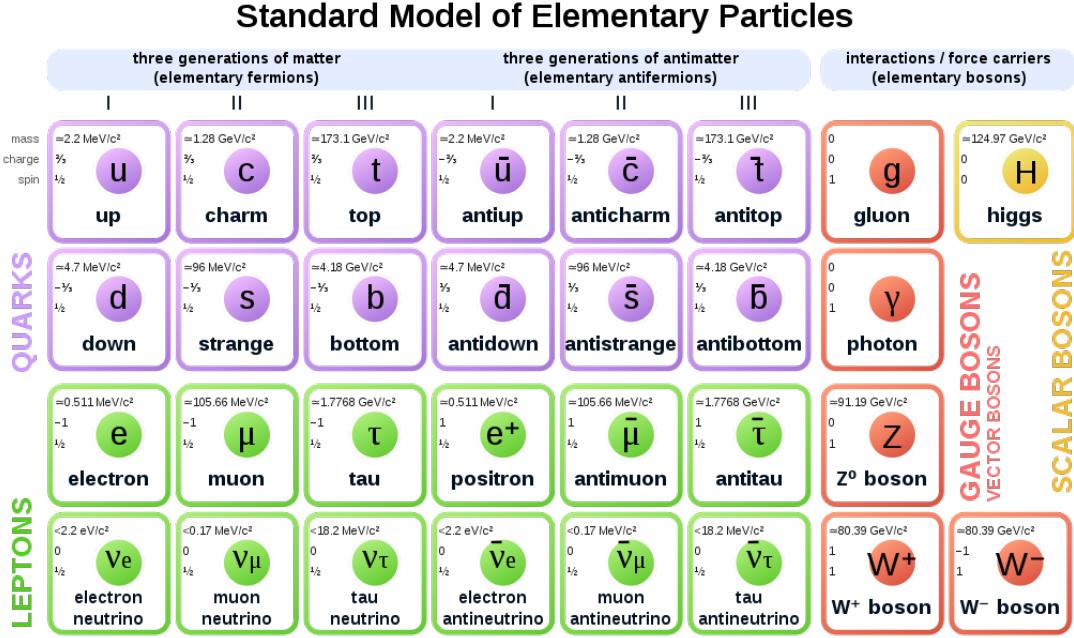


Figure 1-1: Elementary particles in the Standard Model.[2]

## 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  $\vec{r} \rightarrow -\vec{r}$ . 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  $\beta$  decay of  $^{60}\text{Co}$  was presented that the electrons emitted from  $^{60}\text{Co}$  decay prefer 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  $CP$ . 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  $K_S^0$ . ( $CP = 1$  in  $\pi^+\pi^-$  final states, while  $K_L$  has  $CP = -1$ ). 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  $K_L^0 \rightarrow \pi^+\pi^-$  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  $B \rightarrow J/\psi K_S^0$ . 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 Physics researches.

### 1.3 CKM mechanism

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

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  $\phi$  and  $\chi$  are the psuedo-Goldstone fields which are appearing when introducing Higgs field  $\phi$  without breaking the gauge symmetry. The Lagrangian for Yukawa interaction of the quark fields[11] can be presented by Equation 1.2.

$$\mathcal{L}_{Yuk}^q = -Q^\dagger Y^d \Phi d'_R - Q^\dagger Y^u \epsilon \Phi^* u'_R + h.c. \quad (1.2)$$

where the primed fields stand for the weak eigenstates of quarks. The  $\epsilon$  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.

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (1.3)$$

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

Yukawa matrix is an arbitrary  $3 \times 3$  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.

$$S_{L,R}^u \begin{pmatrix} u' \\ c' \\ t' \end{pmatrix}_{L,R} = \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} \quad (1.5)$$

$$S_{L,R}^d \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{L,R} = \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R} \quad (1.6)$$

In Equation 1.5 and 1.6,  $S_{L,R}^{u,d}$  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

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

where the  $q = (q_L + q_R)$  is four-component Dirac field, and  $q_L^\dagger q_L = q_R^\dagger q_R = 0$ . As a result of diagonalizing  $Y^{u,d}$ , the charged current  $W^\pm$  interactions couple to the physical quarks and the Lagrangian is written as Equation 1.8, where  $V_{CKM} \equiv S_L^u S_L^{d\dagger}$ .

$$\mathcal{L}_W^q = \frac{g}{\sqrt{2}} \left[ \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L \gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \begin{pmatrix} \bar{d} & \bar{s} & \bar{b} \end{pmatrix}_L \gamma^\mu W_\mu^- V_{CKM}^\dagger \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L \right] \quad (1.8)$$

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.

$$\begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L \gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \Leftrightarrow \begin{pmatrix} u & c & t \end{pmatrix}_L \gamma^\mu W_\mu^- V_{CKM} \begin{pmatrix} \bar{d} \\ \bar{s} \\ \bar{b} \end{pmatrix}_L \quad (1.9)$$

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.

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

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

$$V_{ij} = V_{ij}^* \quad (1.11)$$

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.

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

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

$$\theta_{kj} = \phi_{uk} - \phi_{dj} \quad (1.13)$$

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 3 generations of quarks, which means  $CP$  symmetry is no longer conserved in the weak interactions.

The  $3 \times 3$  unitary CKM matrix can be written as Equation 1.14 based on the

quark fields it connects using Equation 1.5 and 1.6.

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

It can be parameterized into the form of Equation 1.15.

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

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

$$|V_{ub}| \ll |V_{cb}| \ll |V_{us}| \quad (1.16)$$

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

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

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

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

Using the unitary condition, the Equation 1.19 is obtained.

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

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.

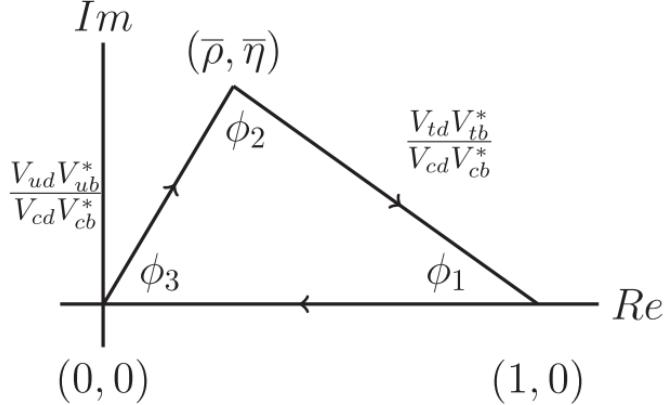


Figure 1-2: The unitary triangles of CKM[12].

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

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

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  $\phi_1$  and  $\phi_2$  are mainly obtained from the time-dependent  $CP$  violations (TDCPV) measurement. The  $\phi_1$  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  $\phi_1$  value in tree-dominated processes and penguin-dominated processes, where both involve  $b \rightarrow s$  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.

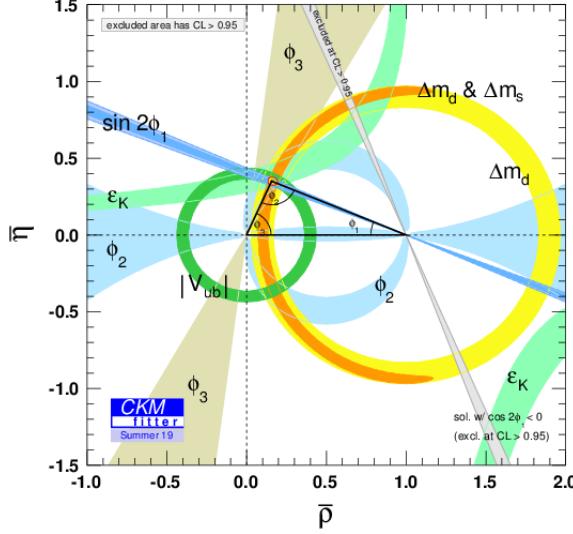


Figure 1-3: The CKM triangle fit in the complex plane of  $\bar{\rho} - \bar{\eta}$ .[12]

## 1.4 Time Dependent $CP$ violation

### 1.4.1 $CP$ violation in neutral B system

The  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  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  $\phi_1$ , TDCPV measurements provide a good experimental environment. From Figure 1-2, one can obtain  $\phi_1$  and  $\phi_2$  by Equation 1.22 and 1.23.

$$\phi_1 = \text{Arg}\left(-\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right) \quad (1.22)$$

$$\phi_2 = \text{Arg}\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right) \quad (1.23)$$

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:  $|B\rangle_{H,L} = p|B\rangle \pm q|\bar{B}\rangle$ , where  $H$  and  $L$  stand for the heavier and lighter mass eigenvalues. The  $|B\rangle$  and  $|\bar{B}\rangle$  present the flavor eigenstates of neutral  $B$  mesons. The Hamiltonian

matrix can be written using flavor eigenstates as shown in Equation 1.24.

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

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  $t = 0$ .

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

where the  $M_{12}$  and  $\Gamma_{12}$  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  $|\mathcal{A}_f| \neq |\bar{\mathcal{A}}_{\bar{f}}|$ , direct  $CP$  violation will occur. The time-dependent decay rate difference is defined as Equation 1.36.

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

, where

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

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

$$A_{\Delta\Gamma}^f = -\frac{2\text{Re}(\lambda_f)}{1 + |\lambda_f|^2} \quad (1.39)$$

From Equation 1.37 and ??, the time-dependent CP violation parameters  $\mathcal{S}$  and  $\mathcal{A}$  are dependent on the parameter  $\lambda_f$ , which are determined by Equation 1.33.

### 1.4.2 $\phi_1$ from $B^0 \rightarrow J/\psi K_S^0$

If final states are  $CP$  eigenstates, the amplitude is 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 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  $\phi_d = \text{Arg}(V_{td}^* V_{tb}) / (V_{tb}^* V_{td}) \approx 2\phi_1$  based on negligible correction to the SM. In mode  $B^0 \rightarrow J/\psi K_S^0$ , 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.

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

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

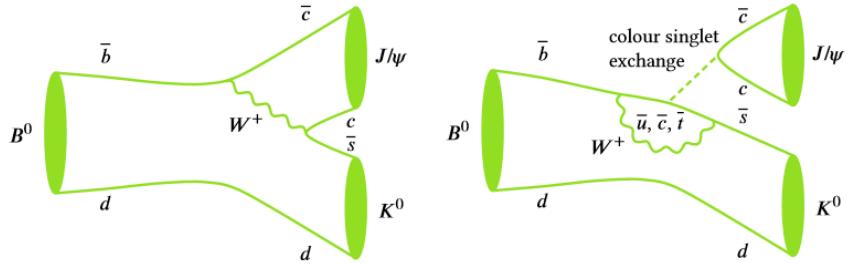


Figure 1-4: The dominated tree-level (left) and the suppressed loop-level (right) of  $B \rightarrow J/\psi K^0$ , in which  $K^0$  particles are detected  $K_S^0$ . [14]

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. Defining  $\eta_f$  as the  $CP$  eigenvalue,

Given  $\eta_f = 1$  and  $|\lambda_f| = 1$  in  $B^0 \rightarrow J/\psi K_S^0$ , from 1.37,  $CP$  violation parameters

can be presented as shown in Equation 1.41.

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

From Equation 1.41,  $\phi_1$  can be obtained precisely in the measurement of time-dependent  $CP$  violation in  $B^0 \rightarrow J/\psi K_S^0$ .

### 1.4.3 $\phi_1$ from penguin-dominated mode $b \rightarrow q\bar{q}s$

Compared to  $B^0 \rightarrow J/\psi K_S^0$  channel, the measurement of  $\mathcal{S}$  and  $\mathcal{A}$  from penguin-dominated channels through  $b \rightarrow q\bar{q}s$  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  $B^0 - \bar{B}^0$  through different short-distance vertices than the tree-level dominated decays. Second, the tree-level decay amplitude is suppressed and penguin-level amplitude is dominated, where 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  $\phi_1$  or NP contributions to the weak Hamiltonian.

Considering possible New Physics contribution as  $A_f^{NP}$ , the decay amplitude can be rendered as Equation 1.42

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

where  $T_f$  and  $P_f$  are tree-level and penguin-level amplitudes. The coefficients  $\lambda_u^s$  and  $\lambda_c^s$  are determined from CKM matrix elements by  $\lambda_i^q \equiv V_{ib}^* V_{iq}$ . Note that compared to the  $B^0 \rightarrow J/\psi K_S^0$ , the tree level amplitude  $T_f$  is suppressed and penguin amplitude  $P_f$  is dominated in  $b \rightarrow q\bar{q}s$ . 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  $\epsilon$ ,  $b \rightarrow q\bar{q}s$  modes are pure penguin with the same weak phase

as  $B^0 \rightarrow J/\psi K_S^0$  has. Thus, direct  $CP$  violation vanishes and time-dependent  $CP$  violation reflects  $\mathcal{S}$  in the same way as  $B^0 \rightarrow J/\psi K_S^0$  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  $\phi_1$ . 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 CPV in the SM are of  $\mathcal{O}(\epsilon_{uc} \text{Im}(r_f^T))$
- $-n_f^{CP} \mathcal{S} = \sin(2\phi_1) + \Delta\mathcal{S}$ , where  $\Delta\mathcal{S} = 2\cos 2\phi_1 \sin \phi_3 |\epsilon_{uc}| \text{Re}(r_f^t) + \Delta\mathcal{S}^{NP}$

#### 1.4.4 $\phi_1$ from $B^0 \rightarrow K_S^0 K_S^0 K_S^0$

Since Belle reported the time-dependent  $CP$  analysis on various  $b \rightarrow q\bar{q}s$  which experimentally showed that the difference on  $\phi_1$  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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is one of the most promising modes for this purpose. The  $CP$  eigenvalue of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is positive ( $CP = +1$ ). Since there's no up-quark shown in the final states, the potential contribution of  $b \rightarrow u\bar{u}s$  rescattered into  $b \rightarrow s\bar{s}s$  is almost of absence. This makes  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  a cleaner channel than  $B^0 \rightarrow K^+ K^- K_S^0$  which has a different weak phase contribution [17]. Any NP effects expected in the  $B^0 \rightarrow \phi K_S^0$  should also affect  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  and the absence of NP effects will lead the same  $CP$  violation just as  $J/\psi K_S^0$ [17]. Currently there's no specific theoretical calculation on the  $\Delta\mathcal{S}$  for three-body  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ . However, due to the same weak phase of this decay as  $\eta' K_S^0$  and  $\phi K_S^0$ , the theoretical prediction on  $\Delta\mathcal{S}$  is usually applied to  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  as well[17]. The expected range on  $\Delta\mathcal{S}$  is typically at level of  $\sim 0.05$ [18], which requires the expected precision improvement for both statistical and systematic uncertainty in future data. The result of  $\phi_1$  from  $B^0 \rightarrow J/\psi K_S^0$  using 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 of  $\phi_1$  from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  using full Belle data[19] is presented as:  $\mathcal{S}_{3K_S^0} = -0.71 \pm 0.23(\text{stat}) \pm$

0.05(syst), and the result from BaBar [20] is:  $\mathcal{S}_{3K_S^0} = -0.94^{+0.21}_{-0.24}(\text{stat}) \pm 0.06(\text{syst})$ . Both results have shown a small deviation from the result in  $B^0 \rightarrow J/\psi K_S^0$  for NP effects while the statistical uncertainties are dominated which prevents the claim whether NP effects are spotted. In  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ , the experimental sensitivity of  $\Delta\mathcal{S}$  will be dominated by  $\mathcal{S}_{3K_S^0}$  because the total uncertainty from  $J/\psi K_S^0$  will be reduced to about 0.005 at  $50 \text{ ab}^{-1}$  Belle II data[15]. The Figure 1-5 shows the scaled  $\Delta\mathcal{S}$  uncertainty regarding the luminosity in Belle II prospective[21], which only includes the statistical uncertainty of Table 1.1 . If the conservative estimation of systematic uncertainty from Belle is considered, the red arrow shows the approximate luminosity where the experimental sensitivity becomes comparable with theoretical prediction at  $\sim 0.05$ . If the future result is different from that in  $J/\psi K_S^0$  with 5 times of the uncertainty at 0.05, then it could be an evidence for NP effects. Of course, smaller the total uncertainty is, easier identifying the existence of NP effects will be.

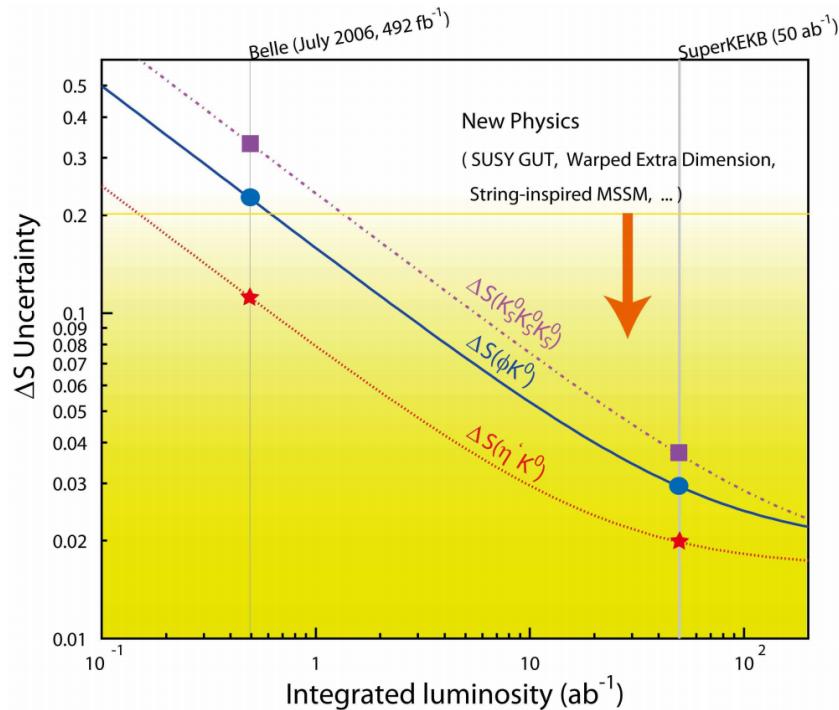


Figure 1-5: Expected sensitivity of  $\Delta\mathcal{S}$  regarding the integrated luminosity of Belle II prospective.[21]

Table 1.1:  $\Delta\mathcal{S}$  scaled uncertainty (statistical) with integral luminosity [21].

Observable	Belle ( $0.5 \text{ ab}^{-1}$ )	Belle II ( $5 \text{ ab}^{-1}$ )	Belle II ( $50 \text{ ab}^{-1}$ )
$\Delta\mathcal{S}_{\phi K_S^0}$	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 about 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  $8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$ , a factor of 40 times higher than its predecessor, the KEKB. Some key parameters of the SuperKEKB are listed in Table 2.1. The overview of SuperKEKB and Belle II are shown in Figure 2-1.

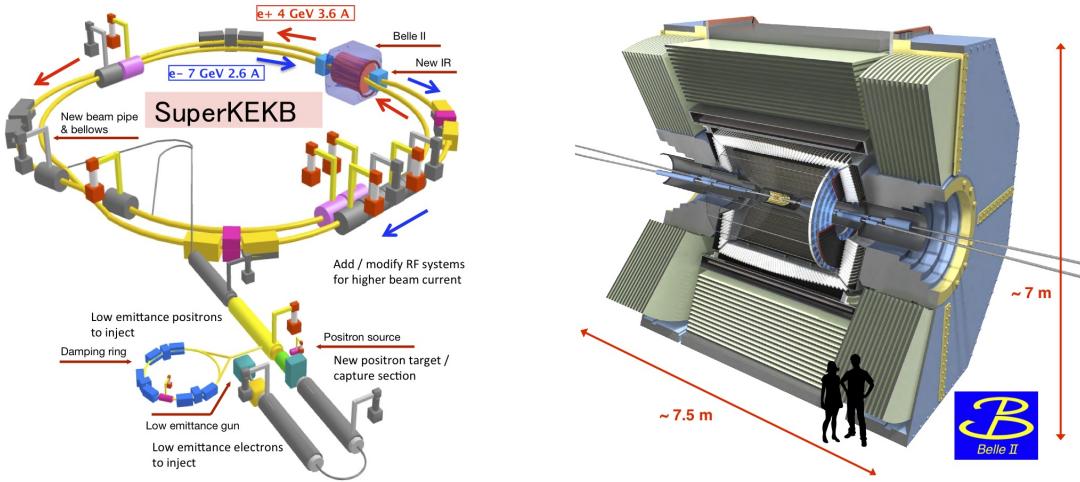


Figure 2-1: Overview of SuperKEKB and Belle II detector [21].

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	$8 \times 10^{35}$		$\text{cm}^{-2}\text{s}^{-1}$
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 events 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 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. Belle II detector in the top view is shown in Figure 2-2.

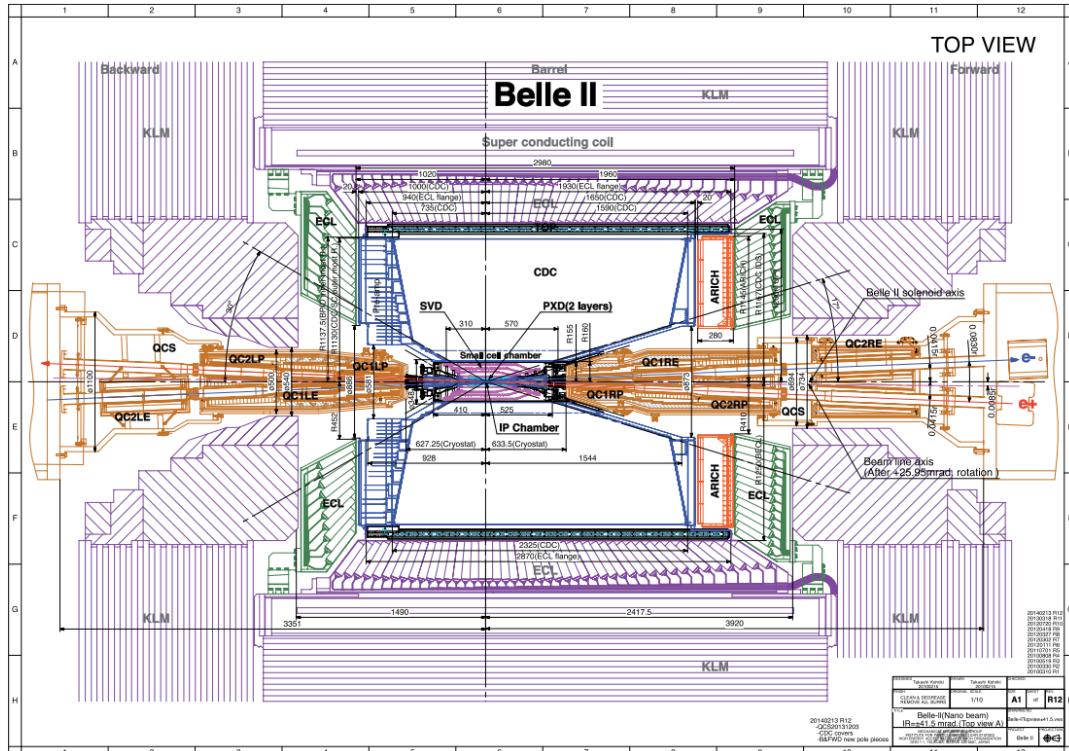


Figure 2-2: Belle II detector top view [15].

The success of Belle II detector depends on the complex of sub-detectors which 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 components, 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 ???. The PXD is placed at a radii of  $r = 14$  mm and  $r = 22$  mm with DEPFET[21] type pixel sensors, which is designed to provide two dimensional hit position information. The inner layer leaves a sufficient space for possible variations of the beampipe layout. The size of two layers are determined by the required acceptance angle, 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 sensor which makes the a very thin a 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

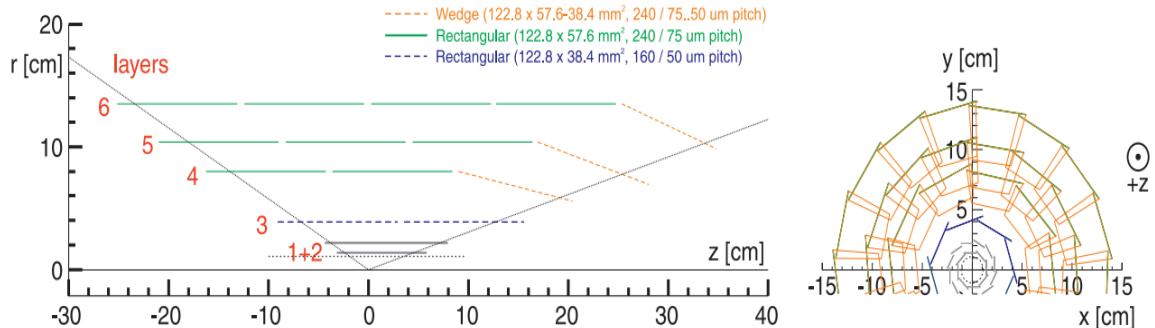


Figure 2-3: A schematic view of PXD (2 layers in gray) and SVD (4 layers in green and orange). [21].

are exposed to 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 vertex detector and central drift chamber to

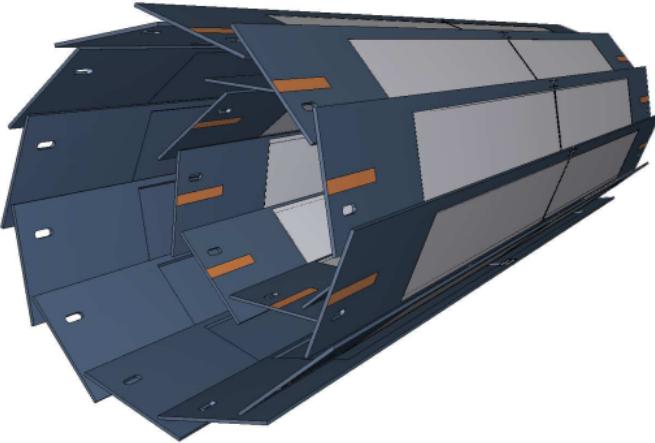


Figure 2-4: The geometry of sensors on PXD where the light grey surfaces are DEPFET sensors with a thinness of 50 microns. The full length including the out modules is 174 mm[21].

extrapolate the fitted 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 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 and the latter is for strips on the  $z$  direction. To suppress the background hits, a readout chip with a fast shaping time of  $\mathcal{O}(50\text{ ns})$  is indispensable. The APV25 chip[22] 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. 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.

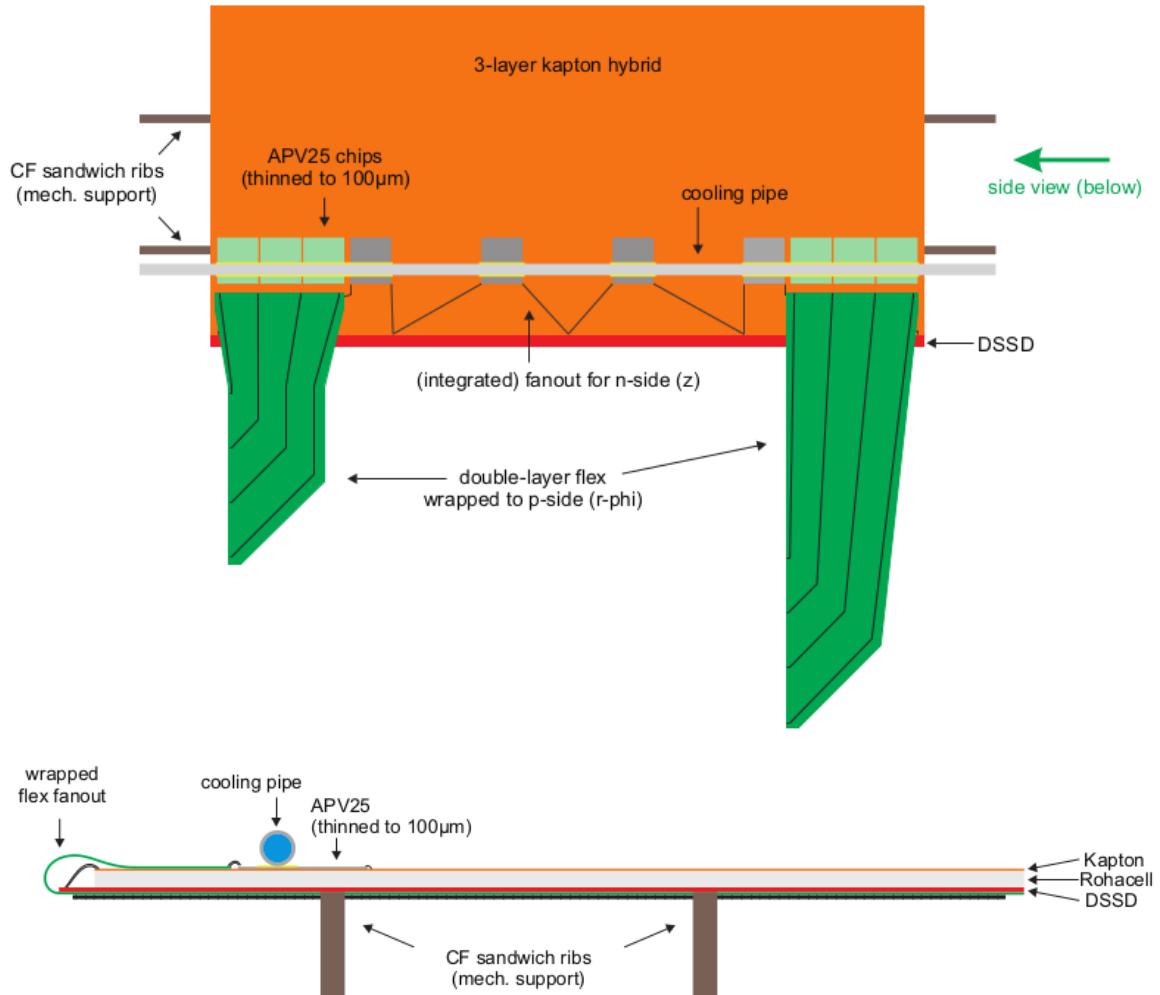


Figure 2-5: The top and side views of Origami Chip-on-Sensor design for DSSDs of SVD. Top: the APV25 chips in grey read out the same side sensors channel while chips in green read out the sensors on the opposite side using wrapped-around flex pieces. Bottom: side view of the Origami design shows the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

## 2.3 Central drift chamber (CDC)

The Central drift chamber (CDC) is the core component of spectrometer in 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<sub>2</sub>H<sub>6</sub> 50%:50% mixture with an average drift velocity of 3.3 cm  $\mu\text{s}^{-1}$  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 thanks to the much thinner layers in barrel region. The whole CDC contains 14336 sense wires in 56 layers, placed in the axial direction and the stereo direction[21][15]. 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.

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 Belle II, also to create more space for SVD volume, the inner radius of CDC in Belle II is 160 mm and the outer radius is 1130 mm. Within this volume, CDC can create three dimensional trigger information. A dedicated trigger type called  $z$  trigger[21] based on the 3D tracking achieved by an FPGA using axial and stereo wires is implemented.

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.

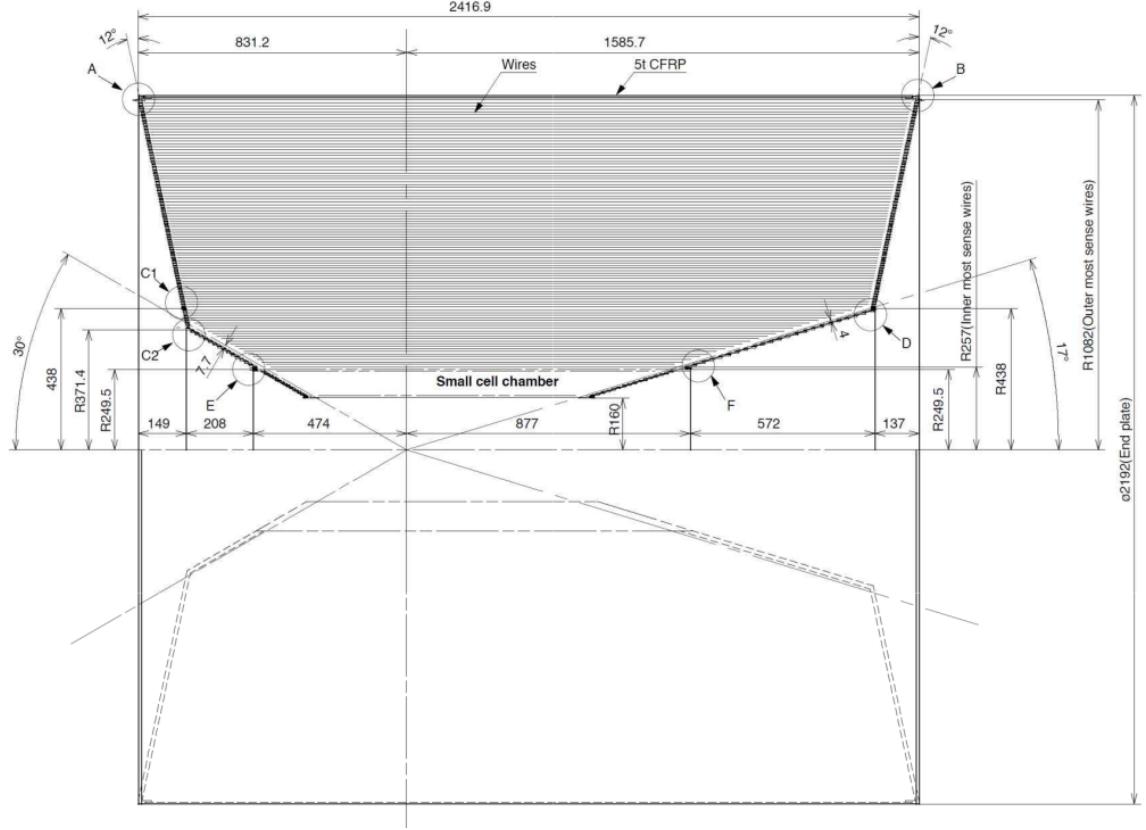


Figure 2-6: CDC structure schematic view[21].

## 2.4 TOP and ARICH detectors

The particle identification system of Belle II mainly consists of two parts, time-of-propagation counter (TOP) and aerogel based Cherenkov radiation imaging ring (ARICH).

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 internally reflected inside a quartz radiator is measured. The quartz radiator is composed of three components: a long bar for radiating Cherenkov light and propa-

gating this light via total internal reflection to the bar end, where the MCP-PMTs are mounted; a spherical mirror mounted on the forward end of the bar for focusing the light; and, for the 1-bar design, a prism that attaches to the backward end of the bar and allows the Cherenkov ring image to expand before the photons are recorded by the PMTs. A 3-dimentional information with  $x - y$  position and a timing information are obtained by micro-channel plate (MCP) PMTs at the end surfaces of the quartz bar. The resolution of starting time is achieved about 50 ps[21]. As the key component of the photon detector, the squared shape MCP PMTs, donated as SL-10[23], have been developed with a  $4 \times 4$  anode array, a multi-alkali photocathode, two MCP plates with  $10 \mu\text{m}$  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.

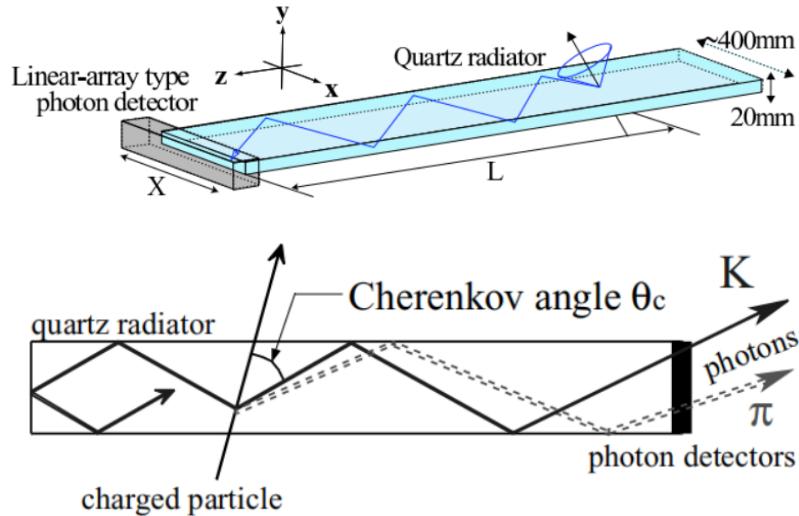


Figure 2-7: Conceptional view of TOP counter (up) and its imaging process of  $K^\pm$  and  $\pi^\pm$  (down)[21] for PID purpose.

Aerogel Ring-Imaging Cherenkov detector (ARICH) is located at the forward end-cap in Figure 2-2 to separate charged particles in a momentum range from 0.5 GeV to 4 GeV, which requires single-photon-sensitive high-granularity sensor to reconstruct the Cherenkov angle with small photon yield. ARICH Japan has developed a hybrid avalanche photon detector (HAPD) to meet the requirements. Each sensor is

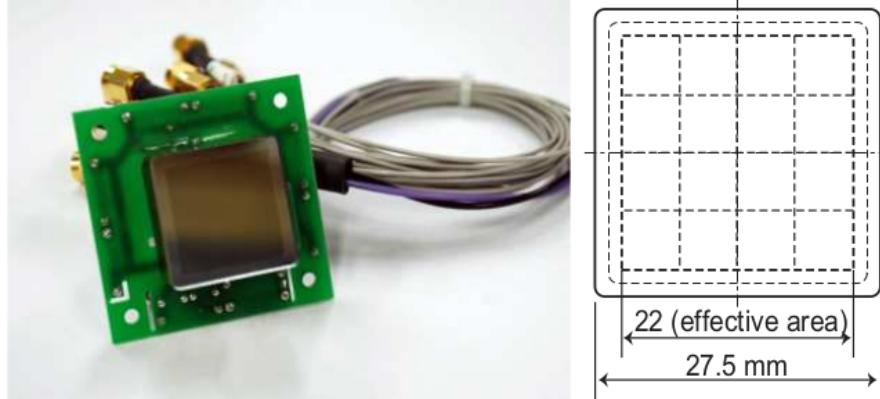


Figure 2-8: SL-10 MCP PMT (left) and the schematic view of  $4 \times 4$  anode (right)[21]

$73 \times 73 \text{ mm}^2$  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.

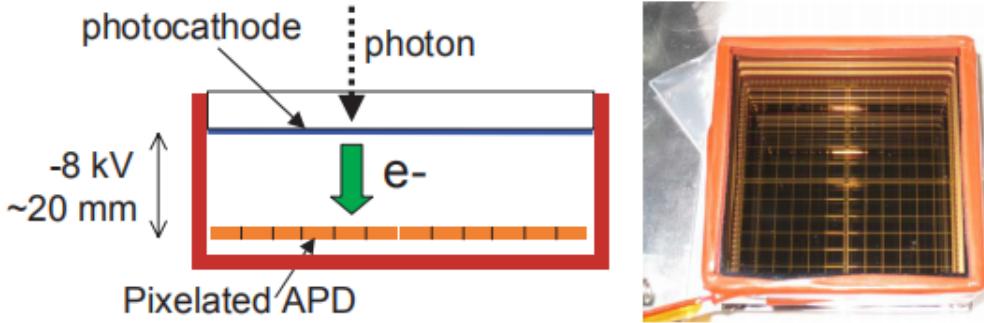


Figure 2-9: Photon-electrons acceleration (left) and pixelated APD (right) at the end[21].

## 2.5 Electromagnetic calorimeter (ECL)

The electromagnetic calorimeter (ECL) in Belle II is mainly responsible for the detection of  $\gamma$  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

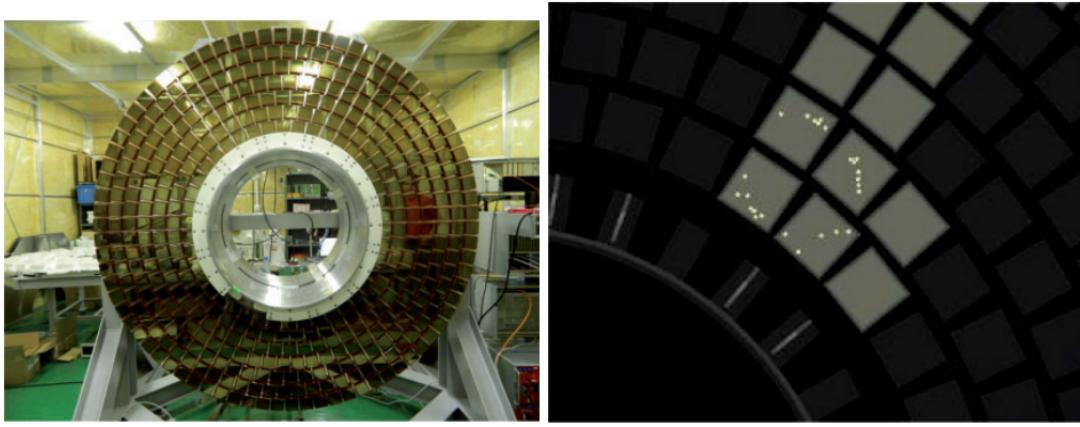


Figure 2-10: ARICH detector (left) and the ring image of cosmic muon on the HAPD sensors[15]

at  $z = 1.96$  m (forward) and  $z = -1.02$  m (backward) from the IP. The barrel section contains 6624 CsI(Tl) crystals of 29 distinct shapes and each crystal is a pyramid shape with about  $6 \times 6$  in cross section and 30 cm in length. The endcaps section contains 2112 CsI crystals of 69 shapes and the total number of crystals is 8736, with a total mass of about 43 tons[21].

As the basic component of ECL, the thallium doped caesium iodide CsI(Tl) 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(Tl). 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). Especially in the forward direction of the electron beamline, where the level of beam background is much higher, the effect of pile-up noise becomes even worse and the performance of ECL will be of trouble if no special measure taken. Therefore, the pure CsI crystal is considered to be chosen as the material of detector to achieve a fast wave-shaping time and higher radiation tolerance compared to the dosed CsI(Tl), which is an backup option for the future upgrade. ECL is the most important detector for providing trigger information for low multiplicity events, since the main feature of these events

is one or two energetic photon(s) emitted from IP region while the charged tracks are missing.

## 2.6 $K_L^0$ muon detector (KLM)

The  $K_L^0$  and muon detector (KLM) system of Belle II consists of a sandwich stacked iron plates at outside of the superconducting solenoid. The iron plates serve as the interaction materials with  $> 3.9$  times the interacting length of material 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 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  $\sim 5\mu\text{s}$  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

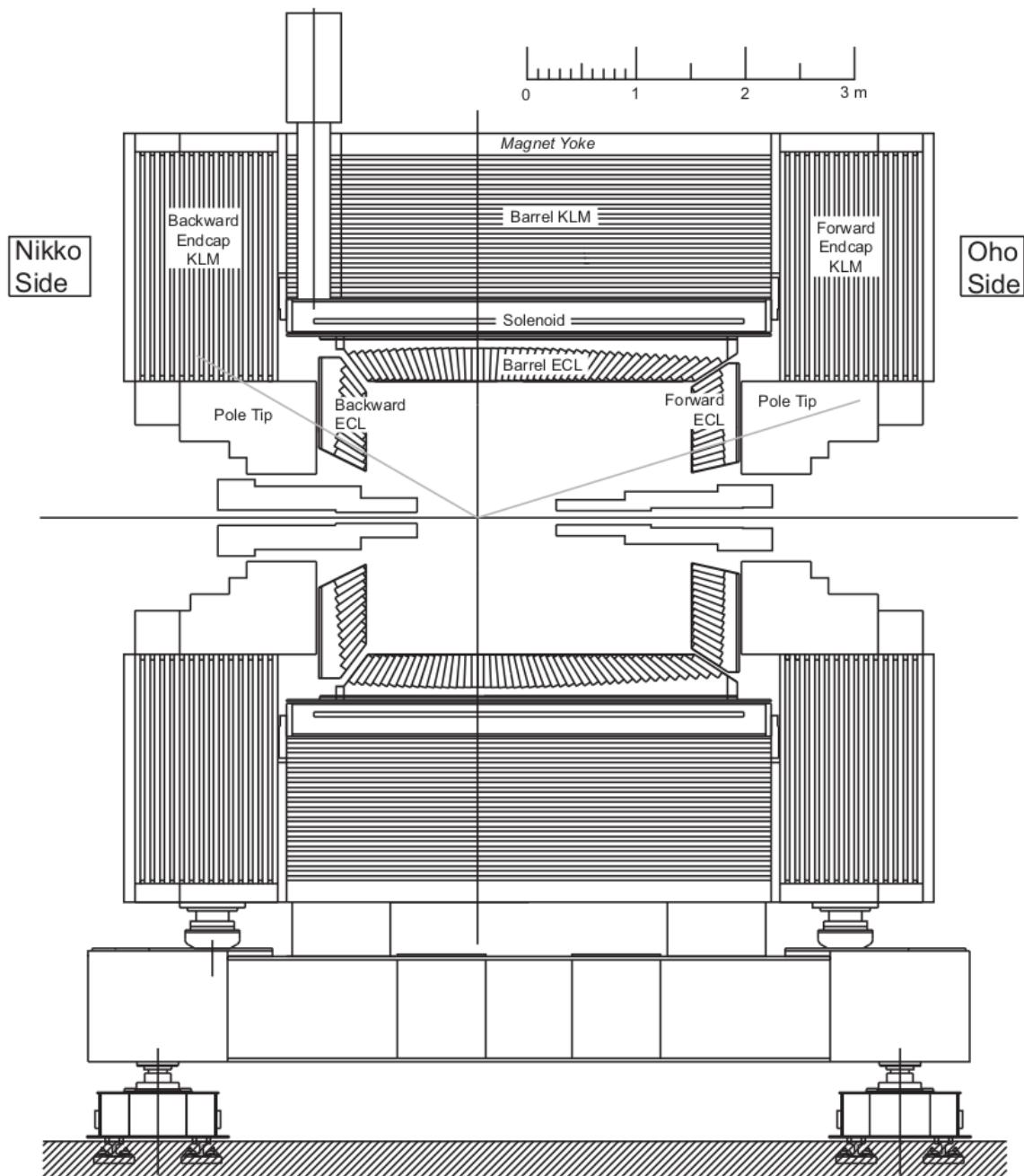


Figure 2-11: The side view of KLM in between the ECL and the solenoid, which the grey lines presents the nominal acceptance angle of the Belle II[21].

have been implemented for L1 trigger. The key improvements of L1 come from the firmware-based reconstruction algorithm and trigger logic.

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 the data size. The event rate reduction relies on the tracking extrapolation on PXD plane, which is the key function of HLT. HLT can suppress the event rate down to 15 kHz using the information from the CDC tracking and ECL reconstruction. The information from the detectors other than PXD are fed into the first event builder for full event reconstruction. The event rate is further reduced to below 10kHz by using 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 HLT on PXD and other detector read-out systems are combined into the second event builder and eventually written to the offline storage.

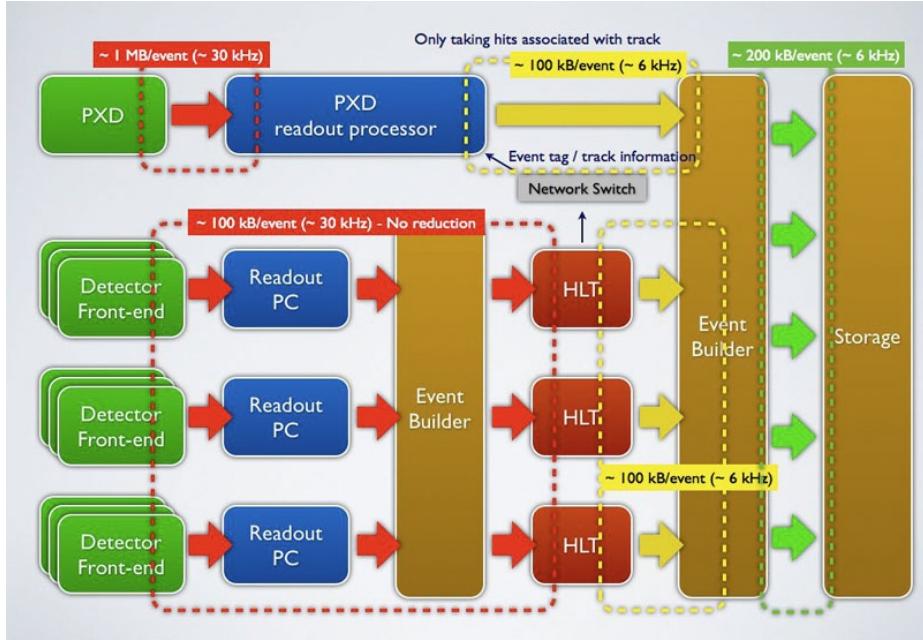


Figure 2-12: The Belle II DAQ workflow with HLT between two event builder to reduce the original 30 kHz event rate down to about 6 kHz for offline storage.

Since the primary goal of 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 offline reconstructing the simulated events and studying the efficiency, close to 100%  $B$  decays are recorded by Belle II trigger system. However, the extensive capabilities of studying a large range of physics not only in  $b$  sector brings a challenge to Belle II trigger and DAQ system. The control of beam background becomes essential. The main sources of beam background are 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 is 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

The improvements on both L1, HLT and the DAQ system are dedicated to serve the operation of the Belle II detector with much higher event rate in future. The further upgrades are also considered such as the replacement of the current DAQ module (COPPER board[21] on the read-out PCs) with the latest PCIe-40 platform[24] to enlarge the bandwidth between detector front-end electronics and the event builder in Figure 2-12.

## 2.8 Analysis software framework

The data acquired by the Belle II experiment or simulation can be processed by 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. 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 such as ROOT, and the tools for configuring and installing BASF2 which are mostly Python and shell scripts. Physics 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 analysis 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 one 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 management service provided by DIRAC system[25]. The detail information about the core structure of BASF2 can be found here[26].

## 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 a certain detectors' condition are packed together, marked with the conditions' database index that is 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 a event to properly set up constants needed for this run.
- beginRun: called at the start of calling this module, including setting up database conditions used in this run (run-dependent configurations) or event (event-based configurations).
- event: called for each event. 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. For example, the “Path” object declared in a steering file stores the sequence of modules that will be executed, to which allow other modules such as “mdstInput” or “reconstructDecay” to be added. Users can use “boolean” type variable set in “event” function to create a conditional branch of a “Path” in case that one event needs to be processed with different modules at the same time. For instance, in the decay reconstruction package, if a decay chain is not fulfilled by missing one particle in the “event” functions, other back-up decay chains can be checked to see if a successful reconstruction is possible.

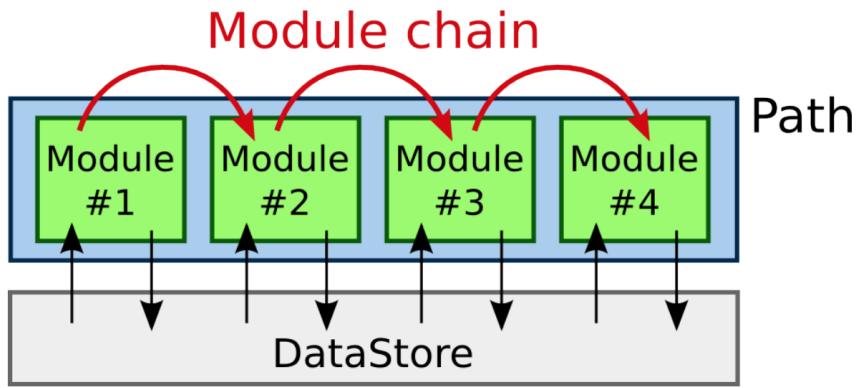


Figure 2-13: The module-based analysis workflow in BASF2.

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[27] format files, or simply prints out processing statistics to the standard output. In practice, BASF2 starts running when it checks there is at least one module specifying the number of events to be processed in a “path” from the “steering file”, then it reads in the information from Datastore in the input ROOT file, execute all the requested modules in the “steering file” and return the time and number of events as information printed in 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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  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  $K_S^0$ ,  $\Lambda$  and photon converted to electron pairs. V0 also stores their relation to the charged daughter tracks and track fit results for further selections.
- PIDLikelihood: it presents for the possibility of a charged track to be an electron, muon, charged kaon and pion, proton and deuteron provided by particle identification system.
- MCParticle: simulated particles and particle-detectors relations are created if simulated particles are correctly reconstructed as tracks or clusters.

#### 2.8.4 Conditional Database

In addition to the physics data, analysis relies on various conditional data that are different calibration of detector, weight files for multi-variate analysis usage like PID and so on. This data is stored in a central database server called central Conditional Database (CDB)[28].

Conditions are made of payloads and each payload has its own “Intervals of Validity” (IoV). It defines in which runs the payload is valid. A set of payloads and IoVs are called a global tag (GT). Considered the GT that is required by the different analysis purposes may change even though the experiment condition is still same, GT is subjected to be updated once new calibrations of detectors or weight files for analysis tools are available.

Users instantiate one of the creation classes, add objects to them and commit them 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 CDB 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 for the focus of this analysis is in  $b$  sector which is mainly from  $\Upsilon(4S)$  decay, the discussed simulation is based on the collisions with center-of-mass (CMS) energy at  $\sqrt{s} = 10.58$  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[29] and *GEANT4* as the simulator of detectors[30]. For the simulation and the reconstruction used in this analysis, the latest release of BASF2 was used. Thanks to 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, 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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ . The mother particle of the decay chain is  $\Upsilon(4S)$ , then it decays into a pair of  $B^0 - \bar{B}^0$  at branching fraction of 100%, with the model *EvtVSSMix*[29] describing the decay model. Then, one of the  $B$  meson is set to decay into three  $K_S^0$  based on phase-space model (“PHSP”) at 100% branching fraction. The default configuration of *evtgen* can not handle multi-bodies charmless  $B$  decay with TDCPV. A modified decay model profile is under-development and not fully validated yet. Thus, MC sample of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  yields zero *CP* parameters by default. As for the other  $B$  meson, it decays into all possible final states that are described by Belle II generic decay file.

As for generic MC, all hadronic processes in a  $\sqrt{s} = 10.58$  GeV collision are simulated. The total production cross section receives contributions from not only  $\Upsilon(4S)$  ( $b$ -flavor decay dominated), but also  $u, d, s, c$ . Their relative branching fractions are taken from cross sections at  $\sqrt{s} = 10.58$  GeV as shown in Table 2.3. Generic MC sample contains 6 types of MC samples due to this production arrangement, where  $\Upsilon(4S)$  produces *mixed* (neutral) and *charged*  $B$  meson pairs and the rest are *uubar*, *ddbar*, *ccbar* and *ssbar*, respectively. In this thesis, the latter 4 types of MC samples are combined and called *qqbar* for simplicity. In the mixed MC sample, the branching fraction of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is set at  $6 \times 10^{-6}$  and the branching fraction of  $K_S^0 \rightarrow \pi^+ \pi^-$

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]	$1.110 \pm 0.008$	1.61	0.40	0.38	1.30

is set at 0.692. Both values are taken from Particle Data Group (PDG)[31]. 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, MC 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  $\Delta t$  in which the background is mixed shown in Equation 2.1:

$$\bar{N} = sR_{BG}\Delta t \quad (2.1)$$

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 precise 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 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: M

Events number	BG0	BG1
<i>signal MC</i>	$10^6$	$10^6$
<i>generic MC</i>	$1 \text{ ab}^{-1}$	-

million) is produced without beam background for cross-checking the reconstruction performance. For generic MC,  $1 \text{ ab}^{-1}$  sample including mixed, charged and  $q\bar{q}$  events are produced with beam background at  $\sqrt{s} = 10.58 \text{ GeV}$ . 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 Phase 3 operation that marked the beginning of official physics runs has started. By the end of 2020, Belle II has been operating in Phase III for 4 total run seasons. The integrated luminosity collected during this period of time is about  $84.73 \text{ fb}^{-1}$ , 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 number 7, 8, 10 and 12 is used. The data of experiment 7, 8 and 10 were collected during 2019 and the data of experiment 12 and 14 were collected during 2020, as shown in Figure 2-14. Correspondingly, the integrated luminosity for offline reconstruction that is used for this thesis is about  $62.8 \text{ fb}^{-1}$ [32] which is from experiment 7, 8, 10 and 12. The experiment 14 is not used due to the unfinished processing of the latest experiment data.

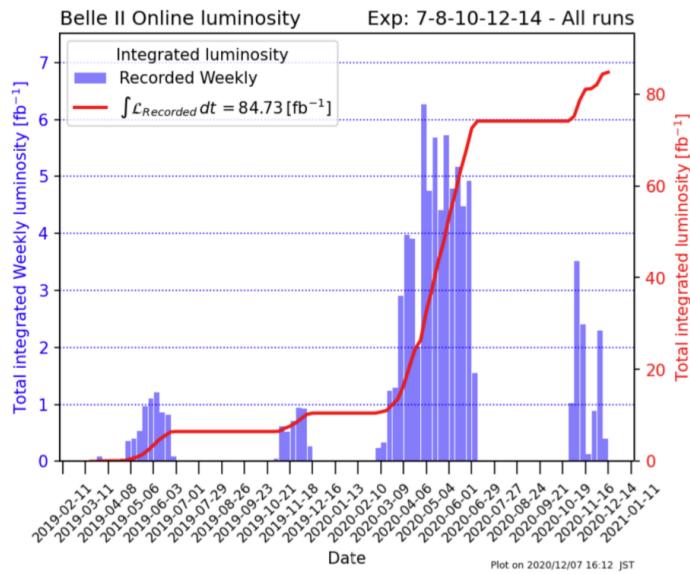


Figure 2-14: Belle II online luminosity from 2019 April to the end of 2020[32].

# Chapter 3

## $K_S^0$ reconstruction study

The final states of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  only depends on the decay of  $K_S^0$ . The main decay channels of  $K_S^0$  is to either  $\pi^+ \pi^-$  at branching fraction of about 0.692, or to  $\pi^0 \pi^0$  at branching fraction of 0.307, referenced from PDG[31]. The characteristics of these two decays are much different in terms of the response from the Belle II detector. The charged decay that yields  $\pi^+ \pi^-$  leaves two tracks originating from VXD or CDC volumes with opposite charges. On the other hand, the  $\pi^0$  main decay channel is  $\pi^0 \rightarrow \gamma\gamma$  which typically results in the photon clusters on the ECL. There are mainly two reasons for not selecting  $\pi^0$  as final states. First,  $\pi^0 \rightarrow \gamma\gamma$  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 quality in this case. The photons could be originating from many other resources, such as beam background and charged track radiation. Besides, the main viable selection is the 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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  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  $B^0 \rightarrow J/\psi K_S^0$ , which has two direct charged tracks of  $e^+ e^-$  or  $\mu^+ \mu^-$  from  $J/\psi$ . 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  $B^0$ , which eventually leads to a large uncertainties in defining the decay time of signal  $B^0$  and the decay time difference as the key observables of TDCPV measurement. Therefore, only  $K_S^0$  reconstructed using charged pions is considered to reconstructed  $B^0$  in this analysis.

### 3.1 Cut-based $K_S^0$ Reconstruction

The  $K_S^0$  has average life time at  $(8.954 \pm 0.004) \times 10^{-11}$  s in 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  $\mu\text{m}$  away from  $B$  vertex to more than 13.5 cm that is further than the outmost layer of SVD ladders, see Figure 3-1.

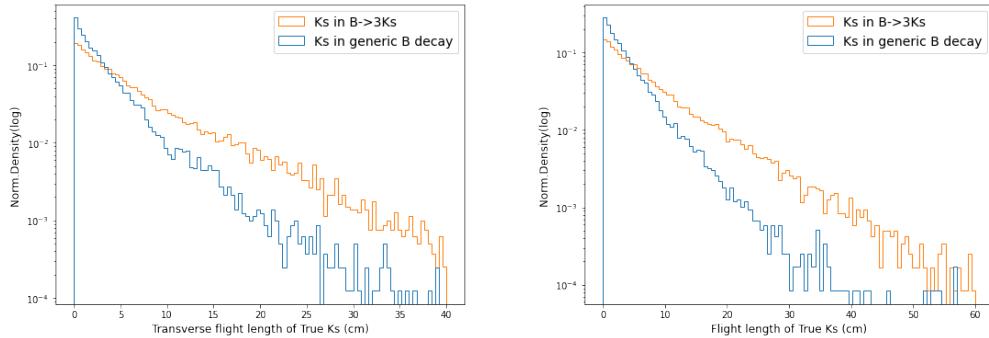


Figure 3-1: The left is the transverse flight length distribution and the right is the total flight length distribution from true  $K_S^0$ . The blue is from generic MC and the orange is from signal MC. Both plots are normalized.

Due to the different topology of  $B^0$  decay, the average momentum of  $K_S^0$  in generic MC is different from the ones from the signal MC. In general, 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 tracks with mass hypothesis of

$\pi^\pm$  is used, which the tracks and PID of charged pions are pre-selected by the criteria in Table 3.1. 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. The V0 based  $K_S^0$  and offline reconstructed  $K_S^0$  are merged and the vertex fit is performed using “TreeFit”[33]. The duplication of  $K_S^0$  between two  $K_S^0$  collections are possible. Therefore, the objects’ index of two charged pions’ 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 from signal MC as shown in Figure 3-2.

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

Selection Criteria	$\theta$	CDC Hits Number	PID
CDC acceptance		$> 20$	pionID $> 0.1$

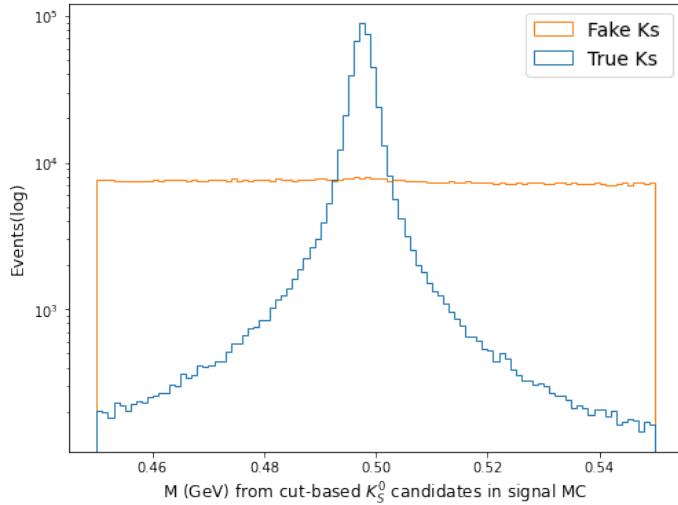


Figure 3-2: “M” of  $K_S^0$  from cut-based selection in signal MC. The blue line is the true  $K_S^0$  and the orange is the fake  $K_S^0$ . 200000 candidates are used in total.

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 from its charged

daughters, which is critical in performing a proper tracking. Belle II tracking efficiency gets poor due to higher beam background when performing track finding for tracks without inner detector hits such as SVD. For certain fraction of  $K_S^0$  decaying outside of layer 5 of SVD, it's much likely that there is no SVD hits assigned to their daughters' tracks. This is due to the feature of SVD track finding, where a track candidate needs either at least 3 SVD hits to form a good SVD track, or 2 hits to form a hit double-lets to be used as a complete track. Single hit on layer 5 or layer 6 is filtered out to suppress the large fraction of beam background induced by random single hits. This effect is shown in Figure 3-3.  $K_S^0$  are categorized based on how many SVD hits their daughters have, in which  $SVD10$  and  $SVD01$  stands for  $K_S^0$  that only  $\pi^+$  and  $\pi^-$  has non-zero SVD hits number,  $SVD11$  and  $SVD00$  stands for  $K_S^0$  that both or neither charged pions have SVD hit non-zero SVD hits number. This is related to the track quality of  $K_S^0$  where  $SVD11$   $K_S^0$  has the best quality and  $SVD00$  has the worst. Thus, the efficiency and purity of  $K_S^0$  with long flight length is reduced. It's clear that  $SVD00$   $K_S^0$  show up at about 11 cm where SVD layer 5 is placed. Most of  $SVD10(SVD01)$   $K_S^0$  start to show up at the similar range. The geometric structure of PXD and SVD is shown in Figure 2-3 and the fraction for each types of  $K_S^0$  in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is listed in Table 3.2.

$K_S^0$ type	SVD11	SVD00	SVD10	SVD01
% in Belle II	52%	39%	5%	5%

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

Fake  $K_S^0$  candidates costs a large extra processing time and the number of combinatorial backgrounds in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  becomes high, which largely 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.

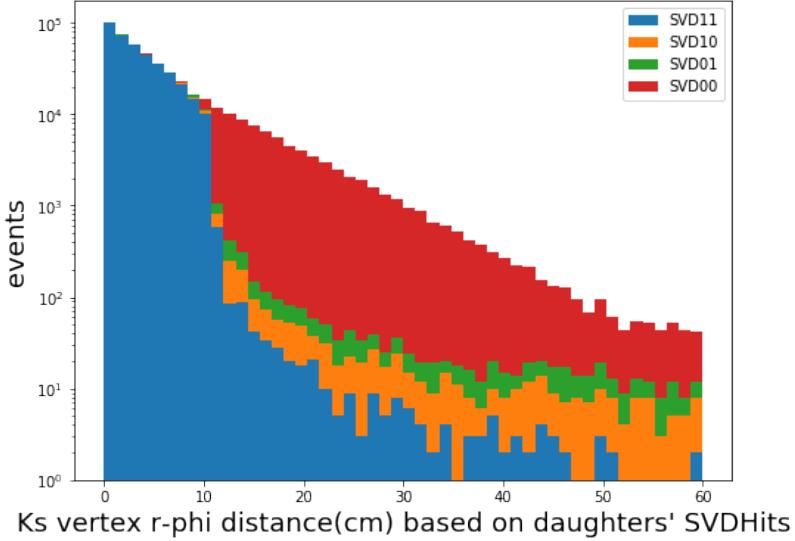


Figure 3-3:  $K_S^0$  transverse flight length based on SVD hits of pions.  $SVD11$ : both pions have SVD hits,  $SVD10(SVD01)$ , positive(negative) pions have SVD hits, and  $SVD00$ : no SVD hits from pions. The result is from signal MC of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ .

## 3.2 MVA-based $K_S^0$ selection

### 3.2.1 Belle II $K_S^0$ classification

As mentioned in the last section, in order to improve the reconstruction performance of  $K_S^0$  from cut-based selection, a MVA-based package called “KsFinder” has been developed. 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  $K_S^0 \rightarrow \pi^+ \pi^-$  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)  $K_S^0$ . It aims to improve the limitations in the Belle  $K_S^0$  MVA classification tool.

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[34]. The Belle tool defines the goodness of  $K_S^0$ , called *nb\_nolam* and *nb\_vlike*, respectively. As their names suggest, *nb\_nolam* is the likelihood of not being a  $\Lambda$  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  $\Lambda$  particle and a high likelihood of being a V0-like particle, assuming the major backgrounds for  $K_S^0$  is the mis-identified  $\Lambda$  among V0-like particles. By putting cuts on these two variables, a purification of  $K_S^0$  can be made, see Figure 3-4. It can effectively reduce fake  $K_S^0$  from cut-based selected candidates, however, there are a few 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 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. Last but not least, the classification of  $K_S^0$  is not the directly targeted output of the neuro-network. Instead, it classifies the V0-like particle and “ $\Lambda$ ”. Besides, the computation speed of NeuroBayes is not optimized.

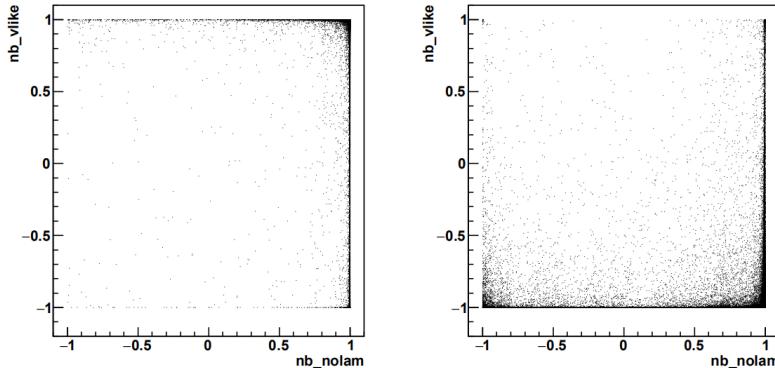


Figure 3-4: The distribution of two variables outputs:  $nb\_nolam$  and  $nb\_vlike$  for  $K_S^0$  candidates from Belle signal MC. The left is from true  $K_S^0$  and the right is from the fake  $K_S^0$ . In Belle, the standard cuts for  $K_S^0$  is  $nb\_vlike > 0.5$  and  $nb\_nolam > -0.4$ [19].

Such a dedicated  $K_S^0$  classification tool is not implemented yet in BASF2 framework until 2019. Considered the limitation of NeuroBayes, the development of  $K_S^0$  classifier demands another algorithm and structure. The *Boosted Decision Trees* (BDT) is widely employed for multivariate classification and regression tasks in high energy physics field. Particularly, a speed-optimized and cache-friendly implementation of such a method called FastBDT (FBDT) is popularly used[35]. Compared

to other popular classification algorithms such as TMVA[36], scikit-learn[37] and XGBoost[38], FastBDT method is proven to be one order of magnitude faster during the training and applying phases[35]. By using FastBDT algorithm, KsFinder in Belle II is expected to give a single output which directly presents the goodness of a candidate of being a true  $K_S^0$ . Since the FastBDT algorithm depends on the variables that are different 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 section 3.1. KsFinderTeacher is responsible for extracting variables to perform training of the FastBDT model and generate a weight file containing all the nodes information in ROOT format, which also provides a function to communicate with BASF2 CDB so that users can share or download others' 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, purity. By providing MC sample(s) from a certain decay mode(s), users can easily generate their own weight file(s) of  $K_S^0$  classification that suits different decay modes despite  $K_S^0$  variables distribution may be varied. 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 $K_S^0 \rightarrow \pi^+ \pi^-$

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  $K_S^0 \rightarrow \pi^+ \pi^-$  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  $\pi$ , 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  $D^+ \rightarrow K_S^0(\rightarrow \pi^+ \pi^-) \pi^+$ . The decay shape resembled the above cases are illustrated in Figure 3-5.

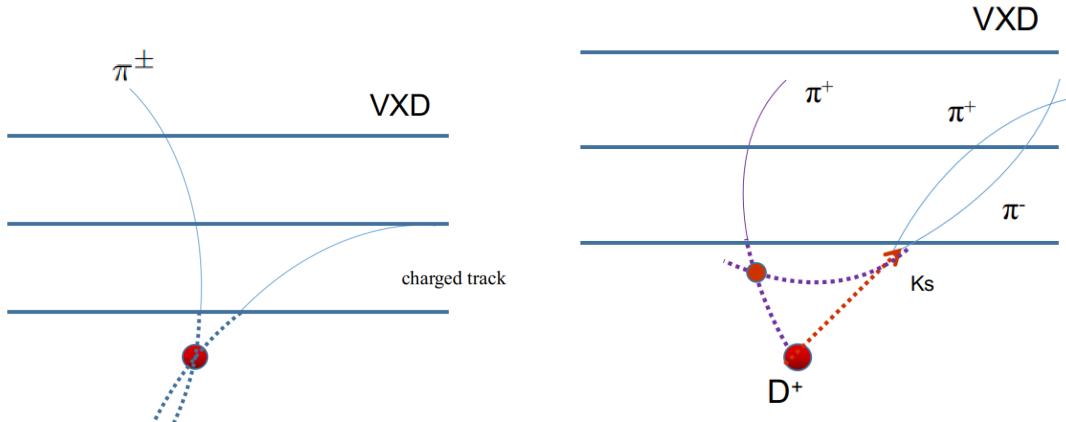


Figure 3-5: The left shows the case when a charged track (not  $\pi^\pm$ ) combined with a charged pion to form a fake  $K_S^0$ , the right shows the case when two daughters are correctly reconstructed as pions but not from the correct mother particle, which is falsely taken as a  $K_S^0$ .

The V0-like particles mainly refer to  $K_S^0$ ,  $\Lambda$  and  $\gamma$ .  $\gamma \rightarrow e^+ e^-$  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  $\Lambda \rightarrow p^+ \pi^-$ , it happens when the positive charged tracks (proton track) is wrongly identified as  $\pi^+$ , see Figure 3-6 left. The key observable to distinguish this background is the invariant mass of mother particle, which is 1.115  $\Lambda$  GeV, much larger than the  $K_S^0$ . The number of left-over  $\Lambda$  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  $\text{PID}(\pi^\pm)$  smaller than  $\text{PID}(p)$ .

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 3-6 right.

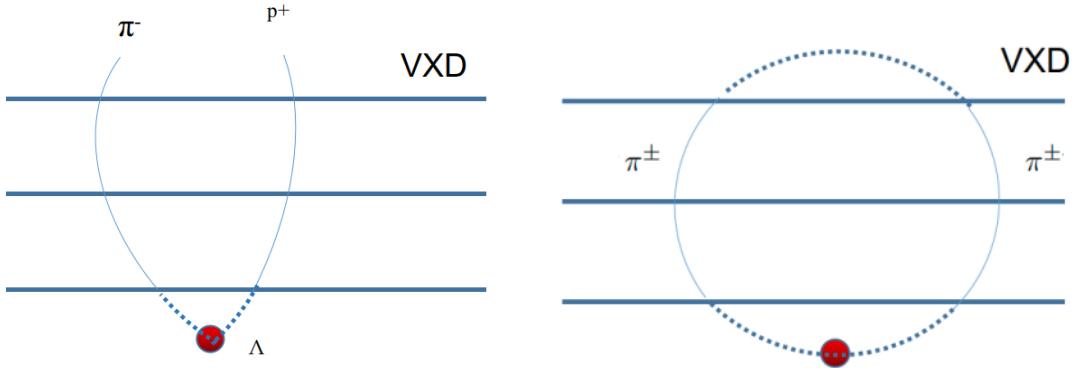


Figure 3-6: The left shows the  $\Lambda \rightarrow p^+\pi^-$  decay shape that can be treated as  $K_S^0$ , the right shows a self-loop formed by a low  $p_T$  charged pion reconstructed as two separated tracks with a vertex.

### 3.2.3 Determination of training observables from $K_S^0$ decay

Given the characteristics of  $K_S^0 \rightarrow \pi^+\pi^-$  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 3.3.

The cosine between  $K_S^0$  vertex and momentum direction (named “cosVertexMomentum”) is of the most importance because it demonstrates the best separation between a true and a fake  $K_S^0$ . 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

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

Table 3.3: 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.

3-7. This often happens when the two tracks taken as  $\pi^\pm$  are accidentally crossed, or due to the misidentified track(s). The abbreviations and importance rank of input variables from KsFinderTest function is shown in Table 3.4.

As a 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 training data to the algorithm or applying the classification. 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 DTs whose structures are not same. Different correlations helps improve the performance of DTs in tuning of structure. For instance, a true  $K_S^0$  flights longer due to larger momentum in general, so its daughters’ detector hits number becomes fewer. Then these two observables have negative correlations in true  $K_S^0$ . In case 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 3-8 . At last, one should also avoid using many observables with too strong correlations, since in this

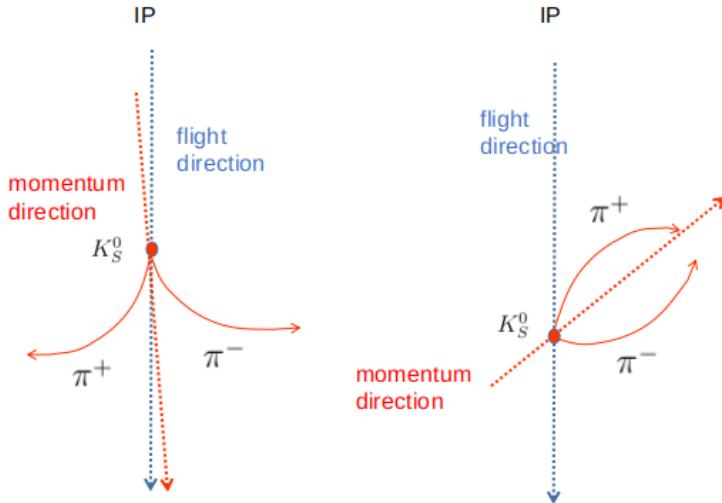


Figure 3-7: The left shows a true  $K_S^0$  decay shape where the cosine angle of  $K_S^0$  vertex position (blue dashed arrow) against reconstructed momentum direction (red dashed arrow) is positive. While the right shows a fake  $K_S^0$  decay shape where cosine angle of  $K_S^0$  vertex position against reconstructed momentum direction can be negative.

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 3-8.

### 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. KsFinder configures that the depth of each DT is 3 and boosting steps is 200. In both MC samples, the ratio of true and fake  $K_S^0$  is set to 1:1 and each component contains 200000 data points. The distribution of input variables in signal MC is shown in Appendix A.

To train the KsFinder, KsFinderTeacher function is called for the training samples

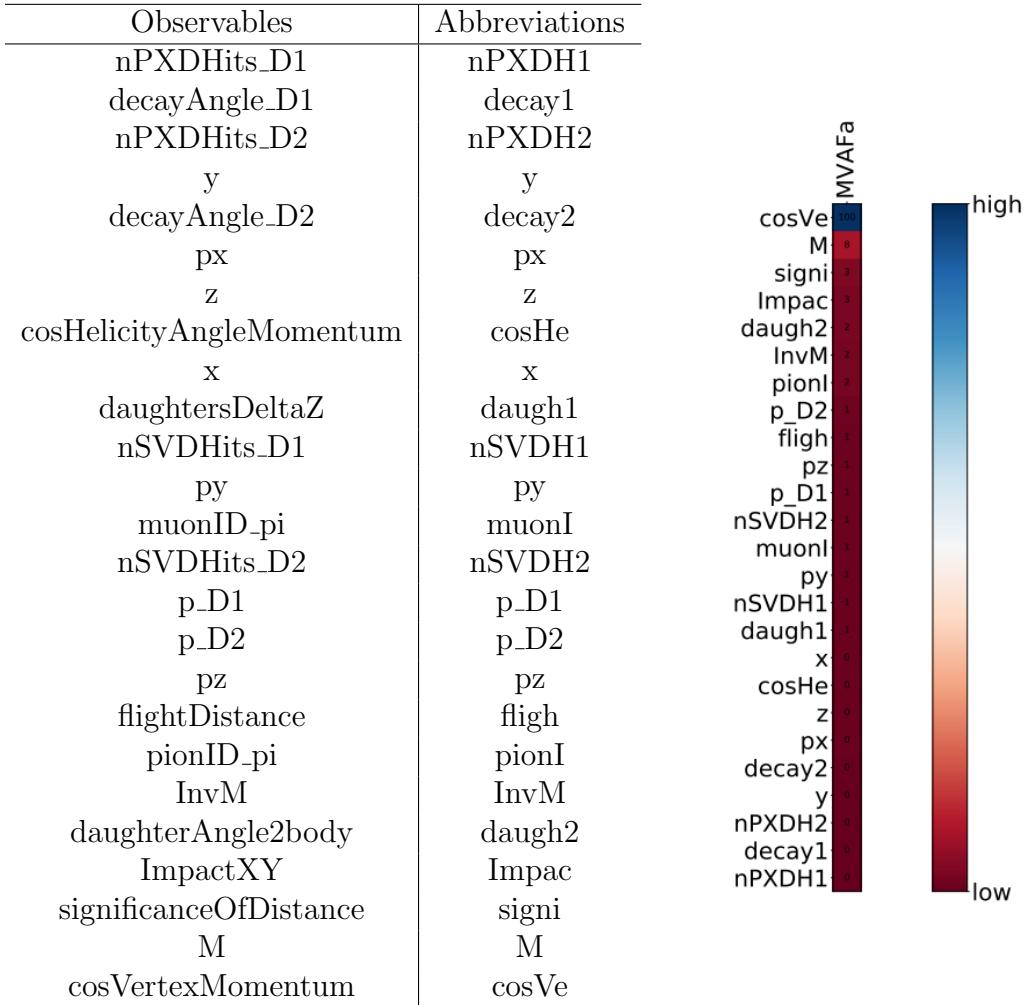


Table 3.4: The abbreviations (left) and importance rank (right) of input variables from KsFinderTest, where the most important variable is “cosVertexMomentum”.

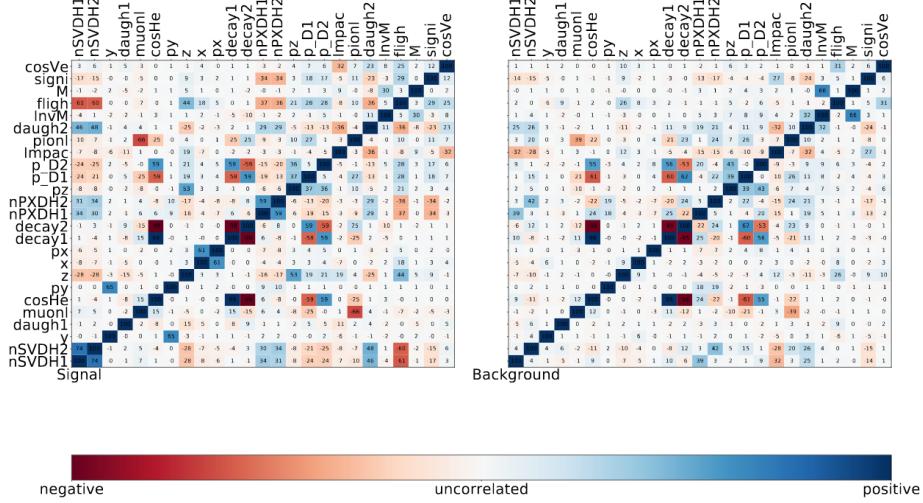


Figure 3-8: The correlation between input variables for KsFinder. As the given example, flight length has negative correlation with SVD hits in signal while uncorrelated in background.

from signal/generic MC and weight files are saved. To apply the classification of  $K_S^0$ , KsFinderApplier reads in the testing samples of signal/generic MC and calculate 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. After the applying of KsFinder on the testing samples, KsFinderTest is called to check the performance and over-training of KsFinder on the testing samples, which will be discussed in the next section.

### 3.2.5 The Performance and Over-fitting check

To evaluate the performance of KsFinder on both signal and generic MC 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.

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

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

The ROC (receiver operating characteristics) curve is usually taken as an indicator of the performance where the curve shows the dependence of rejection power with respect to the signal purity. The larger area under a ROC curve means that the better performance is achieved since background rejection drops slower when increasing the cut. The ROC curves and efficiency/background rejections are shown in Figure 3-9 and 3-10, where the former is for signal MC and the latter is for generic MC.

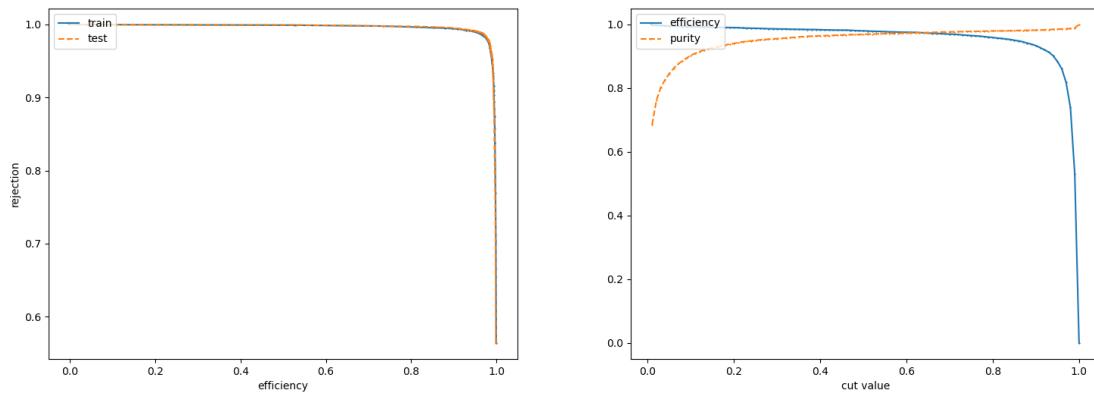


Figure 3-9: The left is ROC curve(blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of KsFinder output. Results are from signal MC sample.

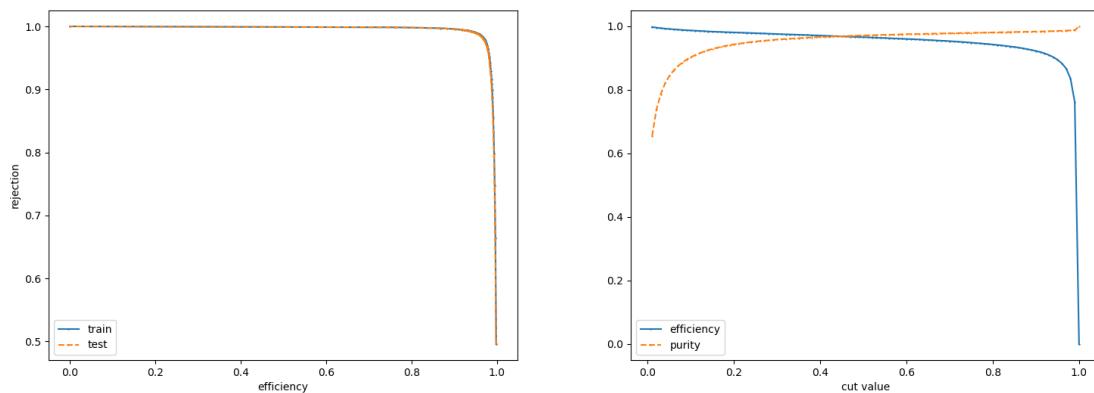
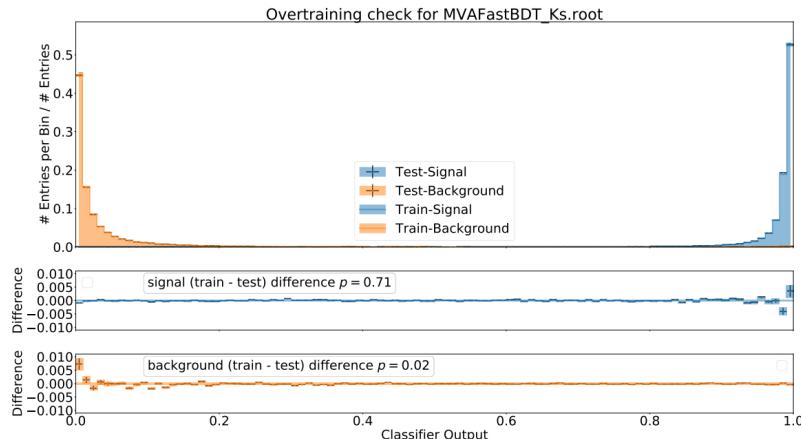
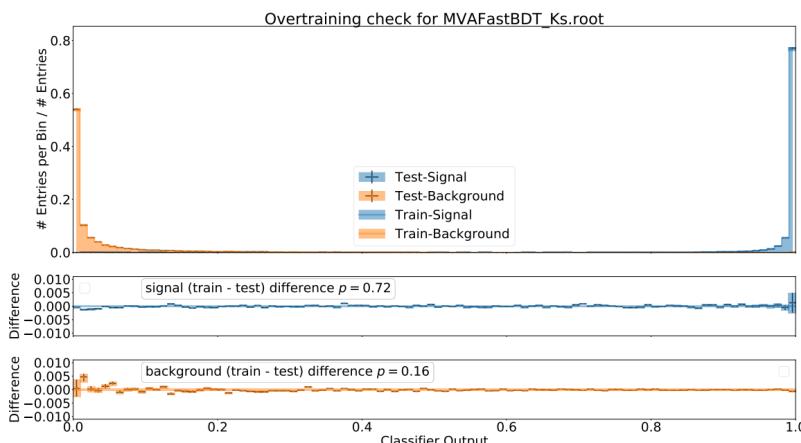


Figure 3-10: The left is ROC curve (blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of KsFinder output. Results are from generic MC sample.

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. To be noted, the curves are consistent in training and testing samples. While the ROC curve has shown the absence of noticeable overfitting in classification, the detailed check can be made by comparing the distributions of classifier output on true and fake  $K_S^0$  in training and testing samples. Therefore, the distribution of signal and background in training and testing sample with respect to the KsFinder output is plotted, where a distinctive separation for both signal MC and generic MC is shown and no over-training is found, as shown in Figure 3-11.



a) Over-fitting check for signal MC sample.



b) Over-fitting check for generic MC sample.

Figure 3-11: The over-training check based on the comparison between training/testing data points in both signal and generic MC.

The 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 3-12. The maximum FOM is achieved at  $\text{FBDT\_Ks} = 0.74$  in signal MC, which is going to be used as the cut value to further reject fake  $K_S^0$ .

$$\text{FOM} = \frac{S}{\sqrt{S+B}} \quad (3.3)$$

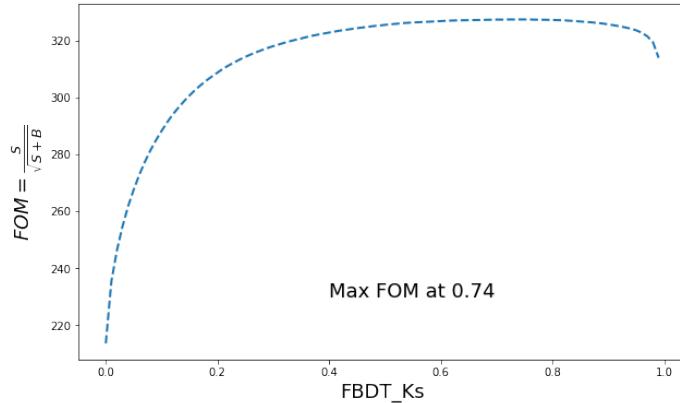


Figure 3-12: FOM of classifier output (FBDT\_Ks) in signal MC, the maximum value is achieved at 0.74 and the curve is almost flat between  $0.5 \sim 0.9$ .

In the signal MC sample, the true  $K_S^0$  fraction before applying KsFinder cut is 39%, and 95.3% of them are kept after the cut is applied. In the meantime, the fake  $K_S^0$  fraction before applying the cut is 61%, and 97.6% of them are rejected after the cut is applied. The purity of the  $K_S^0$  candidates is improved largely as shown in Figure 3-13.

### 3.2.6 Data Validation for Classifier

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.

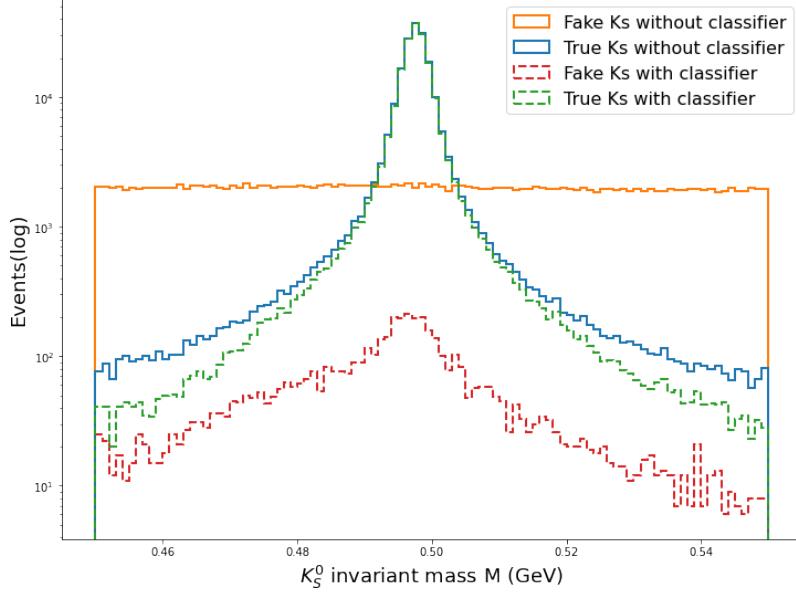


Figure 3-13:  $K_S^0$  purity improvement with cut value of FBDT\_Ks at 0.74 applied. The blue solid line is true  $K_S^0$  without KsFinder and green dashed line is the true  $K_S^0$  with the cut applied. The orange solid line is fake  $K_S^0$  without the cut and the red dashed line is fake  $K_S^0$  with the cut. About 95.3% of true  $K_S^0$  are kept while 97.6% of the fake ones are rejected by applying the cut.

In addition, due to the fact that  $K_S^0$  candidates are used for 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  $B^0$ 's variables that are used 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 about  $5.17 \text{ fb}^{-1}$ . The MC sample is extracted from generic MC with equivalent events number which includes two types of MC ( $MC12b$  is the one produced in early 2019, and  $MC13$ ).

The Figure 3-14 shows the invariant mass and momentum distributions from data and MC samples, and full comparison of all training variables is included in Appendix B. Most of the distribution shows a good consistence before and after using KsFinder. It shows that kinematics of  $K_S^0$  in data and MC yield fairly close distributions and no clear bias is seen by applying the KsFinder cut.

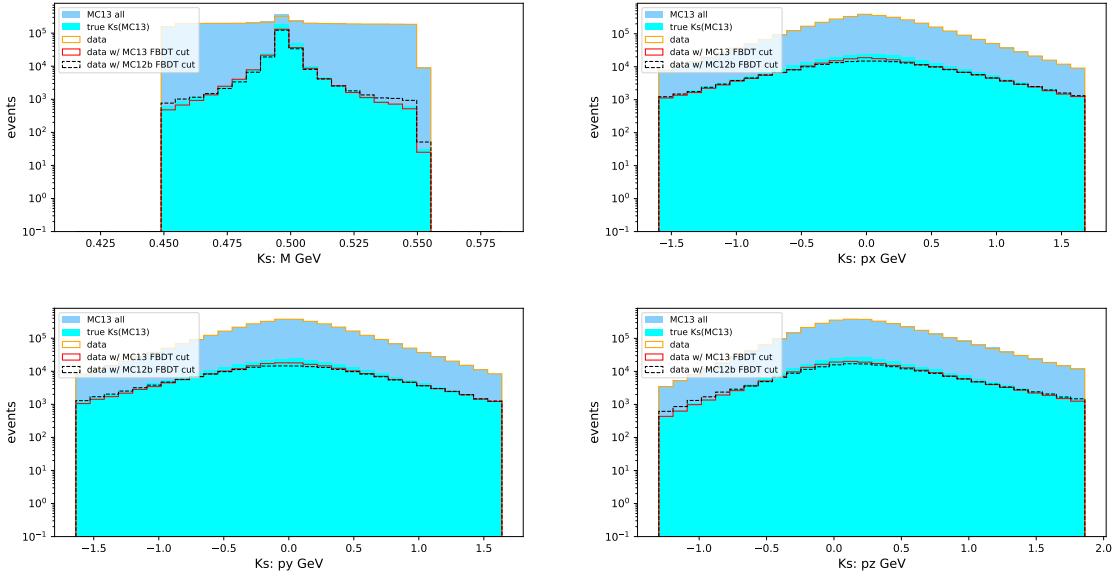


Figure 3-14: The distribution of invariant mass from daughters and the momentum of x,y,z direction. Blue bar is from all MC13, cyan bar is the true  $K_S^0$  in it. Yellow step histogram is data with no cut, solid red data with MC13 trained cut, and the dashed is with MC12b cut. Experimental data has a good agreement with MC before and after applying the KsFinder.

### 3.2.7 KsFinder Effects on Kinematics Evaluation

Implementing KsFinder for  $K_S^0$  may induce extra bias on the event numbers for  $K_S^0$ . It's not easy to directly evaluate the impact of each variables in training towards the final signal yield because the output is non-linear dependence on those variables. However, we can directly use the output of KsFinder and introduce the scale factor when checking the data and MC signal yields.

A fit on invariant mass  $M$  of  $K_S^0$  candidates after varied cut on KsFinder output is done by modeling signal shape as double-Gaussian and background as Chebyshev polynomial. Significance is defined as  $S_{data/MC} = N_{signal}/N_{total}$  in data and MC from fitting using RooFit. A list of intervals of cut value on KsFinder output is made and the significance is calculated within each interval. Fitting results are shown in Fig 3.17 using loose and tight cut respectively. The fit plots in all cut intervals are included in Appendix C. Data/MC correction is defined as:

$$R = \frac{S_{MC}}{S_{data}} \quad (3.4)$$

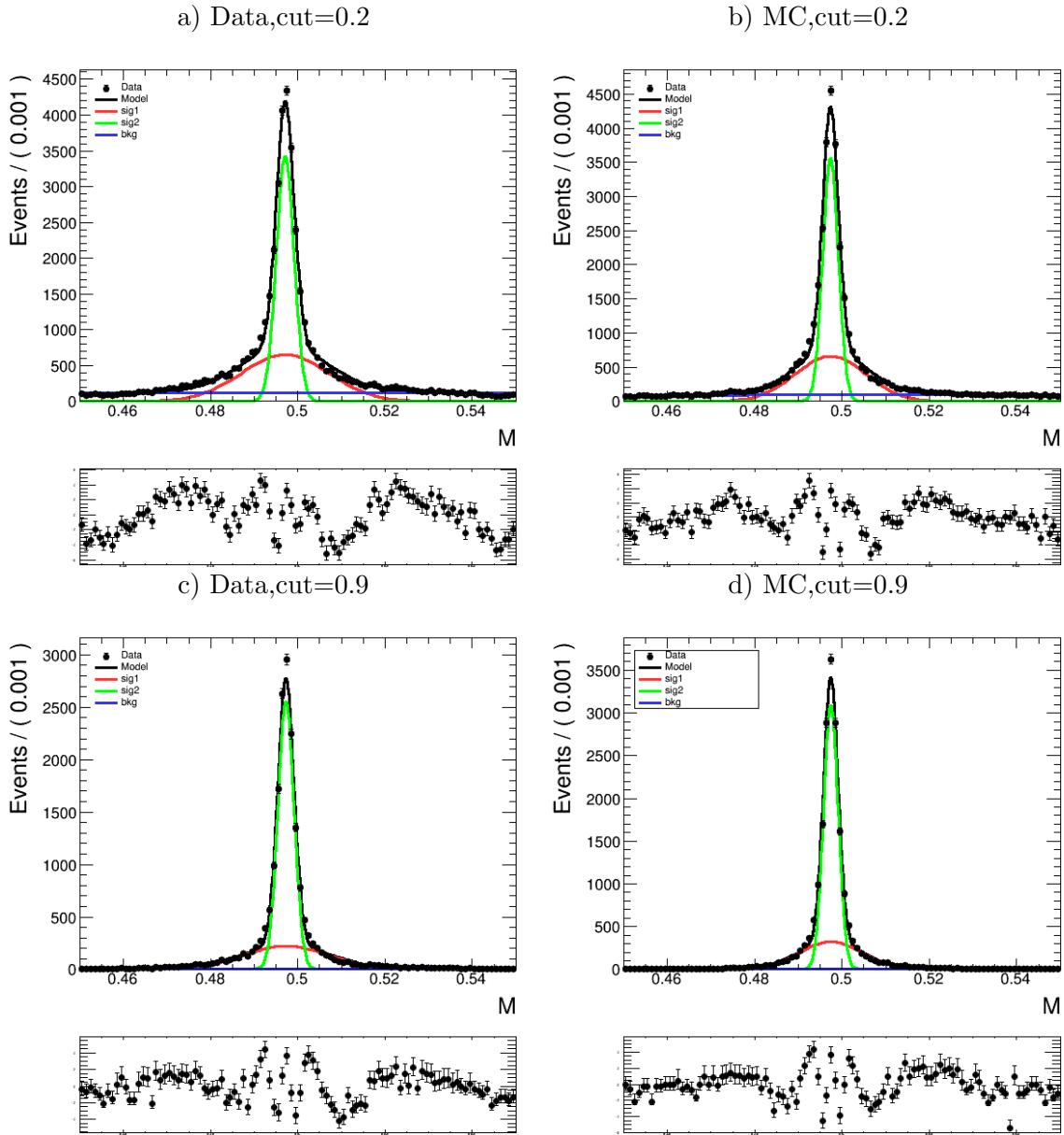


Figure 3-15: Invariant mass fit of  $K_S^0$  using cut at 0.2(loose) and 0.9(tight) to calculate  $S_{data/MC}$ .

The correction is defined by taking the R value within the chosen interval. Uncertainty of R is defined by the difference of maximum and minimum of R in all intervals. R is distributed as:

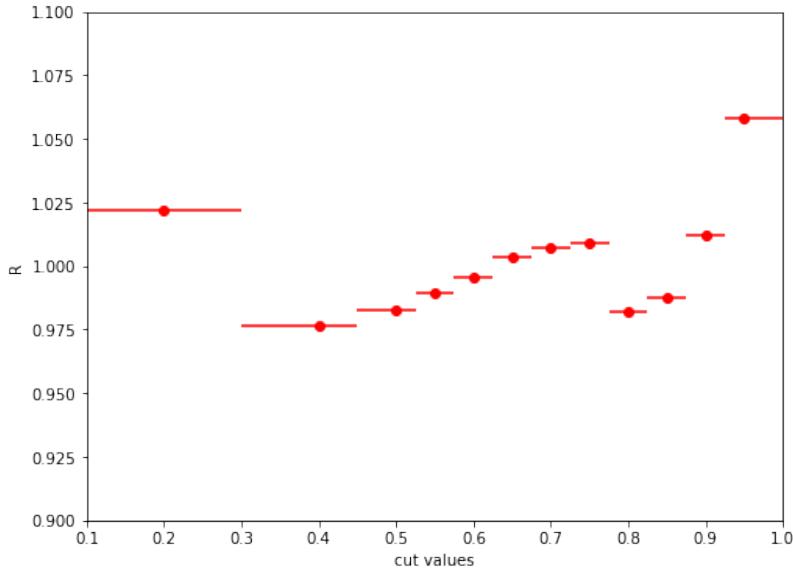


Figure 3-16: Data MC correction induced by  $K_S^0$  classifier.

The R, for example in cut value 0.74 of maximum FOM, is  $R = 1.009 \pm 0.081$ . In  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ , the correction of  $B^0$  events should be proportional to R to the three. Correction that is implemented for  $B^0$  is  $1.027^{+0.33}_{-0.18}$ . In most of the cut intervals, the R is within 2.5% so the bias on  $K_S^0$  numbers is very small.

# Chapter 4

## Analysis strategies

### 4.1 Data Sample and Event Selection

As introduced in section 2.9, the branching fraction of  $\mathcal{B}(B^0 \rightarrow K_S^0 K_S^0 K_S^0) = 6.0 \times 10^{-6}$ . The simulation takes the  $\Upsilon(4S)$  as the mother particle and generate its decay process to two scalar  $B^0$  mesons with mixing.  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is simulated based on only the possible phase-space of kinematics that final states could have, which no  $CP$  violation is assigned for  $\mathcal{S}(\sin 2\phi_1)$  and  $\mathcal{A}$  in both signal and generic MC.

#### 4.1.1 $K_S^0$ Selection

$K_S^0$  reconstruction is first reconstructed by the cut-based reconstruction which contains a large fraction of fake candidates, as discussed in chapter 3. In addition, a momentum cut on  $K_S^0$  is used considering distribution in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ . Only the  $K_S^0$  candidates with momentum larger than 0.05 GeV are selected, as shown in Figure 4-1. Next, the vertex fit of  $K_S^0$  is performed using *TreeFit* package[42]. Using fitted  $K_S^0$  momentum and energy, fitted invariant mass is different from the one obtained directly using daughters' 4-vector. This quantity often receives impact of the measurement uncertainties. If we check out the distribution of fitted invariant mass based on daughters' SVD hits as section 3.1 discussed, *SVD00*  $K_S^0$  shows a large dispersion from the central region of the distribution of invariant mass after vertex

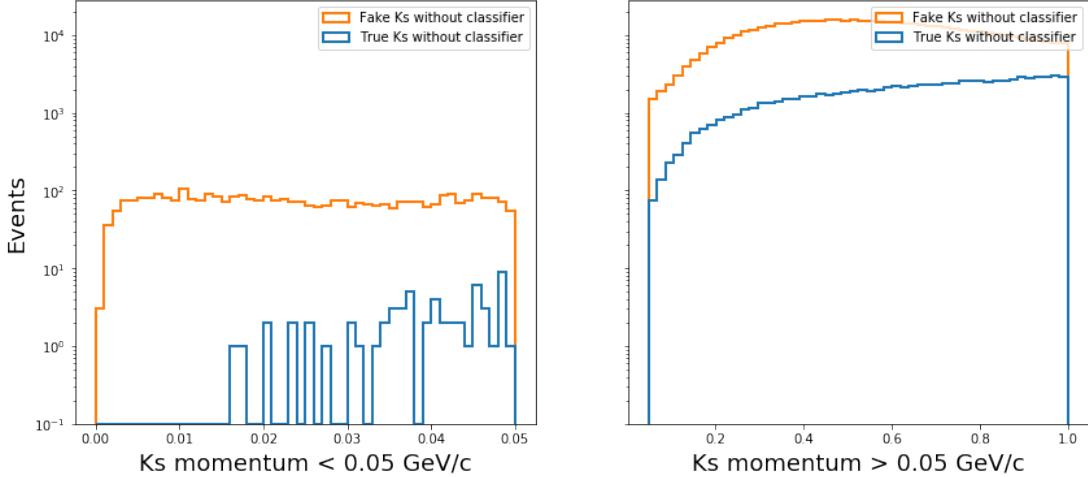


Figure 4-1: The distribution of  $K_S^0$  momentum. Candidates smaller than 0.05 GeV/c are rejected. (Both plots share the same y-axis scale on the left side.)

fit, while *SVD11*  $K_S^0$  shows a small dispersion as shown in Figure 4-2. Therefore, considering the  $K_S^0$  candidates with different SVD hits, see Figure 3-3 and Table 3.2, reflecting the different tracking quality of daughter  $\pi^\pm$ , the different cut on invariant mass  $M_{\pi^+\pi^-}$  after vertex fit is applied depending on the SVD hits number of pion tracks. As shown in Figure 4-3, 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 4.1. This improves the reconstruction purity.

In summary,  $K_S^0$  are first reconstructed using BASF2 standard library with all converged vertex fit candidates are kept using *TreeFit*. Then we reject  $K_S^0$  candidates with momentum smaller than 0.05 GeV/c. After this, tuned cuts on invariant mass after vertex fit for each  $K_S^0$  based on their SVD hits of pions are applied according to Table 4.1. Eventually, only  $K_S^0$  with “FBDT\_Ks” larger than 0.74 are kept based on Figure 3-12.

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

Table 4.1: Invariant mass cut window after vertex fit for  $K_S^0$  based on Figure 4-3.

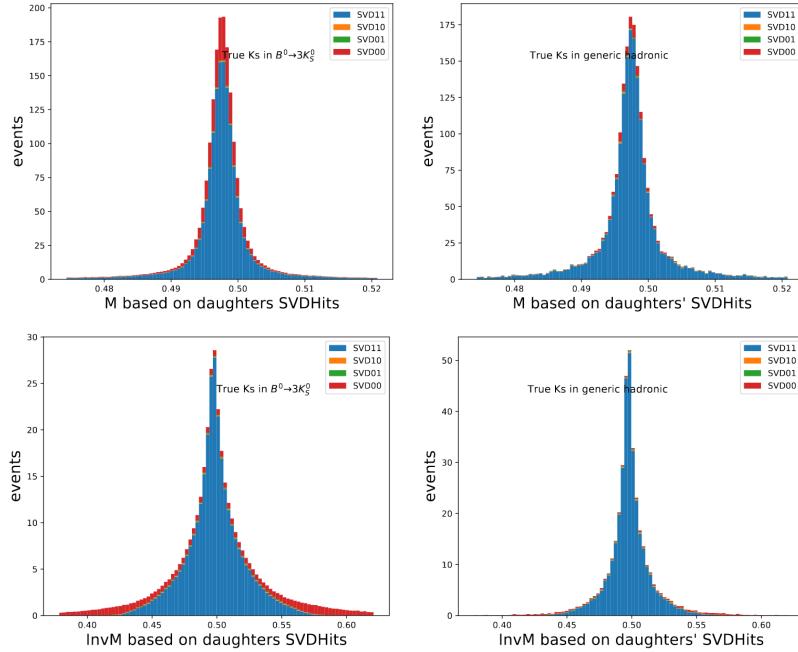


Figure 4-2: (Top two plots)The invariant mass from daughters' 4-vector, the left is  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  signal MC and the right is generic MC; (bottom two plots)The invariant mass after vertex fit, the left is  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  signal MC and the right is generic MC; In both cases, the red shows the clear dispersion on  $SVD00$  type  $K_S^0$ .

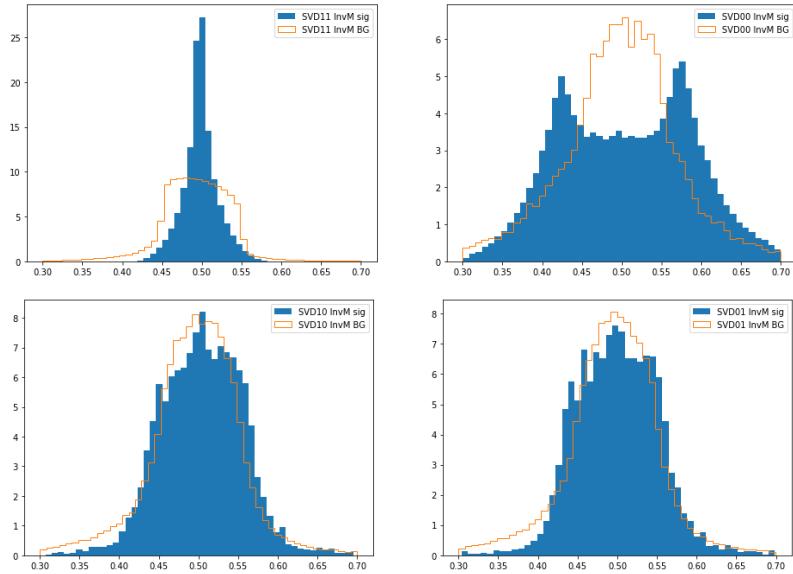


Figure 4-3: Invariant mass after vertex fit of  $K_S^0$ , the sideband is excluded in each distribution to further reject fake  $K_S^0$

### 4.1.2 $B^0$ Reconstruction

By combining three  $K_S^0$  particles from selected dataset, we can reconstruct  $B^0$ . The beam-constraint mass  $M_{bc}$  and energy difference  $\Delta E$  are used to extract signal, as defined in Equation 4.1 and 4.2, respectively. For  $M_{bc}$ ,  $s$  is defined as the invariant mass and  $p_B^*$  is the reconstructed  $B$  momentum, both in the center-of-mass frame of  $e^+e^-$ . For  $\Delta E$ ,  $E_B^*$  is the reconstructed energy in the center-of-mass frame of  $e^+e^-$ . 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 requested to be reconstructed. The  $B^0$  vertex fit using *TreeFit* is performed on each  $B^0$  candidate and only  $B^0$  with converged vertex fit result is kept (by a very loose cut of "chiProb > 0.001"), which is essential to obtain the decay vertex positions of  $B$  mesons for further  $CP$  violation analysis. When multiple  $B^0$  candidates are obtained in a single event, the best candidates selection (BCS) is performed by ranking their vertex fit quality. Since the BCS is based on the vertexing quality that might introduce bias in the vertex positions for  $CP$  fit, we check the distribution of the vertex  $\chi_2$ , as shown in Figure 4-4 top right where the data and generic MC present a good consistence within  $1\sigma$  on average. The distribution of candidates number per event without BCS is shown in top left of Figure 4-4 as well, showing an agreement between data and generic MC within  $1\sigma$ . The distribution of candidates per event is also in an agreement with that from signal MC (bottom left of Figure 4-4). The 2D distribution of  $M_{bc}$  and  $\Delta E$  from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  signal MC is shown in Figure 4-4 bottom right, where the correlation factor is about 15% between two observables.

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

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

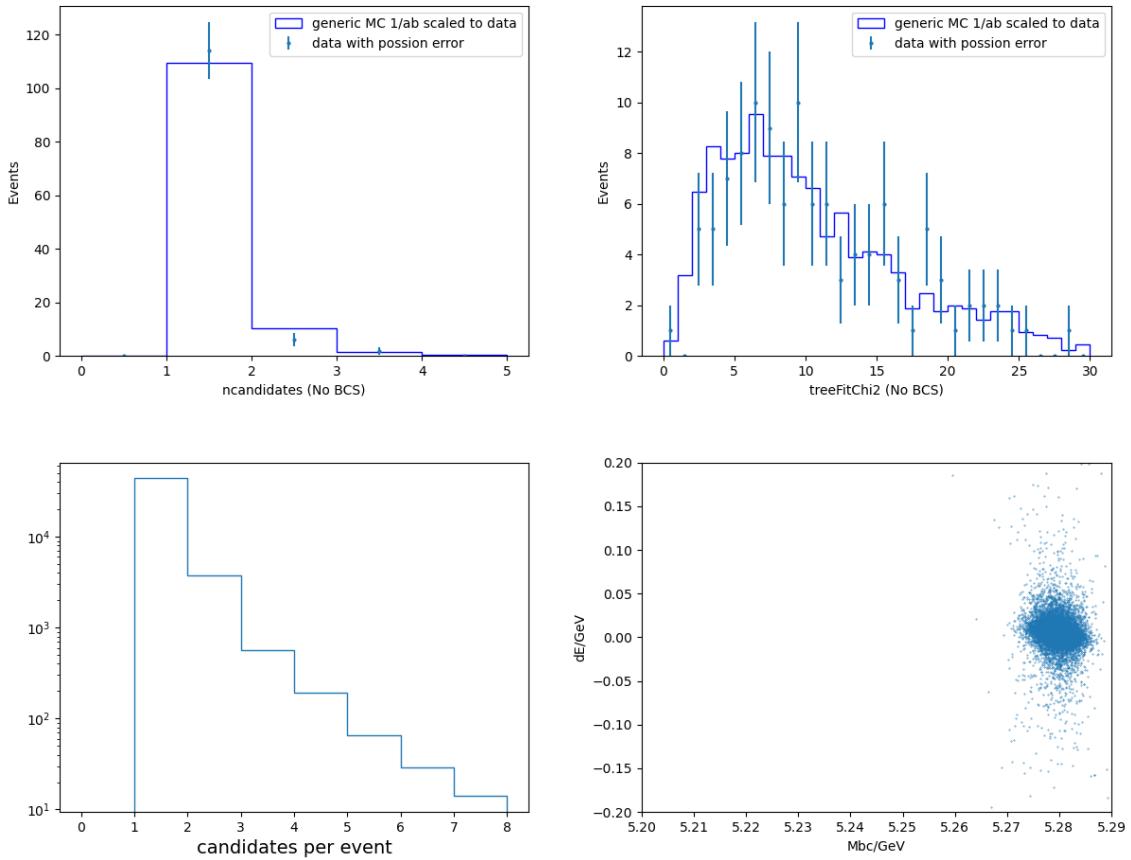


Figure 4-4: Top left is the candidates per event in data and generic MC before the BCS. Top Right is the  $\chi_2$  for data and generic MC before BCS. Bottom left is the number of candidates  $B^0$  per event from signal MC. Bottom right is the 2D  $M_{bc}$  and  $\Delta E$  distribution from signal MC.

### 4.1.3 Continuum Suppression

The production cross-section of  $B\bar{B}$  from  $\Upsilon(4S)$  receives a sizable contribution from other flavor of quarks other than b quark. This calls a demand to distinguish a specific  $B\bar{B}$  decay events from combinatorial background from  $e^+e^- \rightarrow q\bar{q}$ , so called continuum suppression (CS). The rejection is essential because it's the dominated background. In the case of  $b \rightarrow s$  charmless decay like  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ , the number of continuum background can exceed the signals by a few orders of magnitudes without suppression. 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 beam energy. As a result

, decay products are emitted more isotopically, compared to continuum background. The ARGUS and CLEO collaboration[43] developed a set of variables to suppress the continuum, which has also been implemented into BASF2 framework.

CLEO cones momentum can be presented as Equation 4.3, where  $p_i$  is momentum of i-th particle in Rest-Of-Event (ROE) particles in a event except for the ones used to reconstructed  $CP$ -side  $B^0$ ,  $\theta_i$  is angle against momentum thrust of reconstructed  $CP$ -side of  $B$  meson.

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

Besides, the modified Super Fox-wolfram momentum named KSFW momentum, are calculated in each event. The KSFW momentum are defined as shown in Equation 4.4.

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

where the first term is shown in Equation 4.5.

$$R_l^{so} = \frac{\alpha_{cl} H_{cl}^{so} + \alpha_{nl} H_{nl}^{so} + \alpha_{ml} H_{ml}^{so}}{E_{beam}^* - \Delta E} \quad (4.5)$$

when l is odd in Equation 4.5:

$$H_{nl}^{so} = H_{ml}^{so} = 0 \quad (4.6)$$

and  $H_{cl}^{SO}$  is defined as shown in Equation 4.7:

$$H_{cl}^{so} = \sum_i \sum_{jx} Q_i Q_{jx} |p_{jx}| P_l(\cos\theta_{i,jx}) \quad (4.7)$$

$i$  runs over  $B$  daughter particles and  $jx$  for other particles in ROE.  $Q$  is charge and  $p_{jx}$  is momentum for each particle.  $P_l(\cos\theta_{i,jx})$  is the  $i$ -th order Legendre polynomial of cosine of  $i$  and  $jx$ -th particles. On the other hand, for  $l$  is even,  $H_{xl}^{SO}$  can be written

in Equation 4.8.

$$H_{xl}^{so} = \sum_i \sum_{jx} |p_{jx}| P_l(\cos\theta_{i,jx}) \quad (4.8)$$

The second term in Equation 4.4, when  $l$  is odd, can be defined as Equation 4.9.

$$R_l^{oo} = \sum_j \sum_k \beta_l Q_j Q_k |p_j| |p_k| P_l(\cos\theta_j, k) \quad (4.9)$$

$j$  and  $k$  runs over ROE particles and others are same as Equation 4.7. For an even  $l$ :

$$R_l^{oo} = \sum_j \sum_k \beta_l |p_j| |p_k| P_l(\cos\theta_j, k) \quad (4.10)$$

$\beta$  is Fisher coefficients to be determined. Using above definitions, we can form the possibility density functions for KSFW, cosine angle against  $B$  meson thrust  $\cos\theta_B$  and  $\Delta Z$  of two side vertices. Then based on each event's variables' value, we can calculate a ratio  $\mathcal{R}$  as Equation 4.11, where the likelihood  $L$  of signal( $L_S$ ) and background( $L_B$ ) are obtained from the possibility density functions defined in Equation 4.12. The  $\mathcal{R}$  is:

$$\mathcal{R} = \frac{L_S}{L_S + L_B} \quad (4.11)$$

$$L_{S/B} = P(KSFW)_{S/B} \times P(\cos\theta_B)_{S/B} \times P(\Delta Z)_{S/B} \quad (4.12)$$

where  $P$  is probability density function for signal and continuum, depending on the discriminating variables in the parentheses. For example, the distribution of a variable called  $R_2$  shown in Figure 4-5 where the possibility density function is different for signal and continuum events in generic MC.

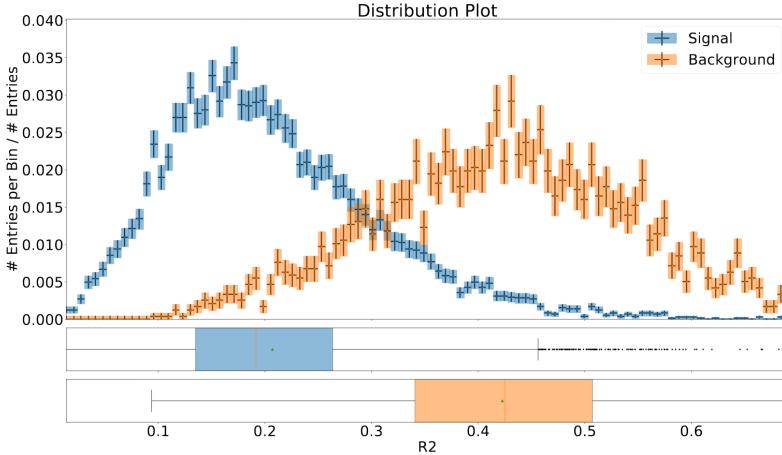


Figure 4-5:  $R_2$  is the ratio of the second to the zeroth KSFW momentum in Equation 4.4. Its distribution in generic MC sample which serves as the highest weight as a variable in discriminating the continuum events, having a quite different distribution between signal and background.

In order to maximize suppression power in this analysis, these variables (KSFW, CLEO cone momentum and angular distributions) are combined as an input for FastBDT classifier. The targeted variable is the continuum event truth. The MC samples using signal  $B\bar{B}$  events from signal MC and continuum events from generic MC ( $q\bar{q}$  component) are prepared in a ratio of their cross-section at  $\Upsilon(4S)$  energy. The same events reconstruction procedures for  $B^0$  is applied for both MC samples. Events passing the reconstruction for  $B^0$  using  $M_{bc}$  and  $\Delta E$  are used for training the continuum suppression classifier. The fraction of signal and background is set to 1:1 during the training. The output of continuum suppression classifier is renamed as “FBDT\_CS”. Then we determine the cut value at 0.66 based on the maximum of *FOM* curve, as shown in Figure 4-9. The variables used in training are listed in Table 4.2 with their abbreviations and the rank of important variables is in Table 4.3.

The correlation between these training variables are shown in Figure 4-6. The correlation among the variables are varied in signal and background. The ROC curve and the efficiency/purity with respect to the classifier output are shown in Figure 4-7.

Overtraining check is made by comparing the distribution of signal and background depending on the classifier output in both training and testing samples. The

Table 4.2: Variables and the abbreviations for CS.

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

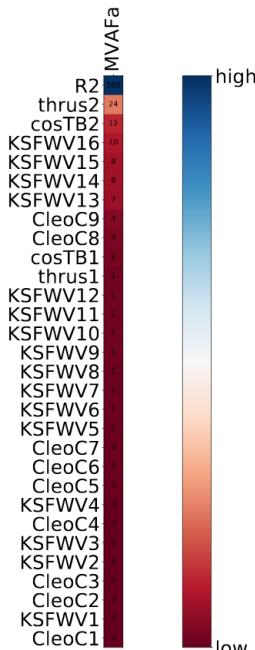


Table 4.3: The rank of important variables for CS.

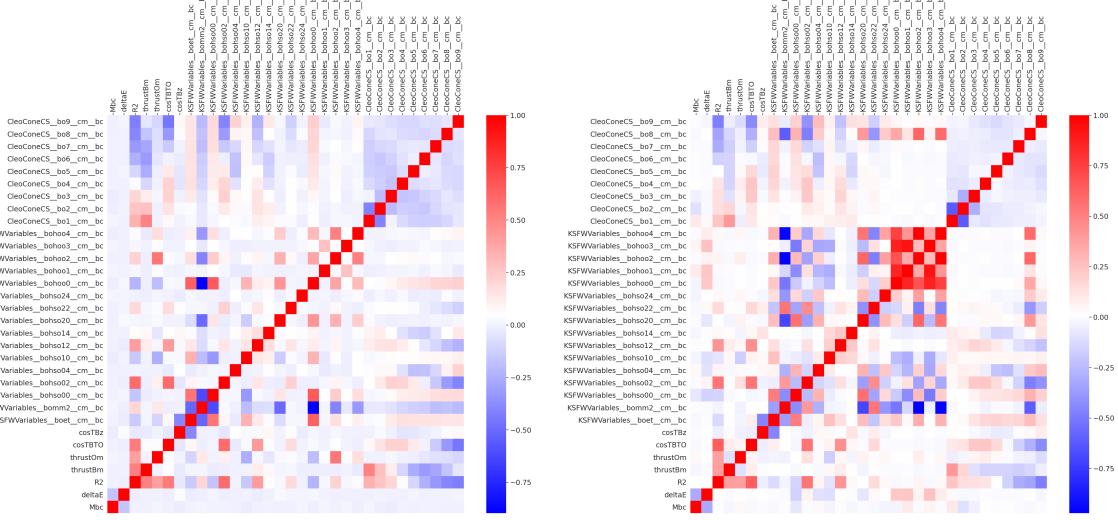


Figure 4-6: The correlation in variables for continuum suppression. The left is for signal and the right is for background.

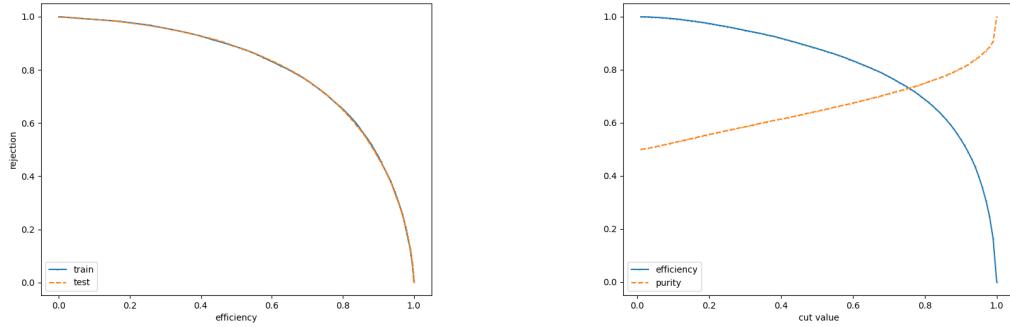


Figure 4-7: The left is the ROC curve (blue for training and orange for testing) and the right is the efficiency(blue) and purity(orange) regarding the classifier output “FBDT\_CS”

testing samples show about 1% lower in each bin for both signal and background events, which is within the acceptable range.

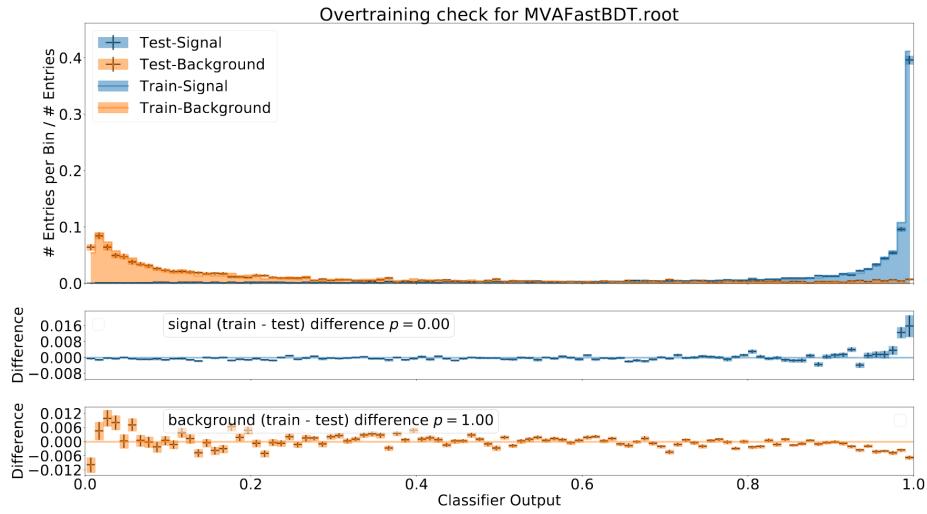


Figure 4-8: Over-training check of continuum classifier, where a very small difference in training and testing (1%) is shown.

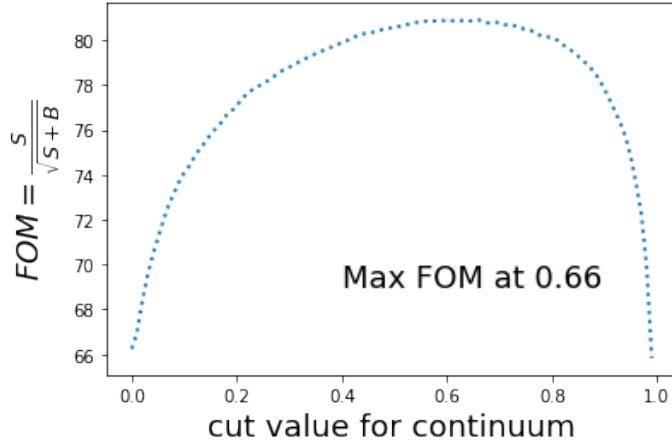


Figure 4-9:  $FOM$  depending on the cut value of continuum classifier output, cut value at 0.66 is used for continuum suppression.

The summary of  $B^0$  selections is listed in Table 4.4, including the application of KsFinder (by “FBDT\_Ks”) and continuum suppression (by “FBDT\_CS”).

$B^0$	$M_{bc}/\text{GeV}$	$\Delta E/\text{GeV}$	chiProb	Rank	FBDT_CS	FBDT_Ks
Selection	$> 5.20 \& < 5.29$	$ \Delta E  < 0.2$	$> 0.001$	= 1	$> 0.66$	$> 0.74$

Table 4.4:  $B^0$  selection criteria, “chiProb” is from  $B^0$  vertex fit and “Rank” is from best candidate selection.

Explicitly, the reconstruction performance of  $B^0$  is summarized in Table 4.5, the efficiency, purity, fraction of multiplicity events and best candidates fraction of  $B^0$  is slight improved in Belle II compared to Belle.

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 4.5: 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  $\Delta E$  in Table 4.4.

#### 4.1.4 Resonance Background

Besides the major contribution from continuum background, charmonium resonance that mediates through  $b \rightarrow c$  transition brings odd  $CP$  eigenvalue in the final states as same as  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ . Monitoring their contribution is also important. Basically, one needs to check the resonance states formed by two  $K_S^0$  with corresponding invariant mass. In  $B^0 \rightarrow X(K_S^0 K_S^0)K_S^0$ , there are two types of resonant events that give out same final states, one is resonant signal and the other is resonant background. For  $b \rightarrow s$  transitions as resonance signal because of the  $CP$ -even final states,  $X$  could be  $f_2(1270)$ ,  $f_0(1500)$ ,  $f'_2(1525)$ ,  $f_0(980)$ ,  $f_0(1710)$  and  $f_2(2010)$ . For  $b \rightarrow c$  transition as resonance background because of  $CP$ -odd final states,  $X$  could be  $D^0$ ,  $J/\psi$ ,  $\psi(2S)$ ,  $\chi_{c0}$ ,  $\chi_{c1}$ , and  $\chi_{c2}$ .

The number of these background in signal reconstruction could be further reduced by implementing veto on invariant mass of  $2K_S^0$ . However, such veto should be carefully validated with data. The distribution of invariant mass of  $X$  should agree well in MC and data, which is hard to check in the low luminosity. The distribution of  $2K_S^0$  invariant mass in generic MC and data are shown in the Appendix

Some of these resonance have not been implemented inside generic MC production in the current Belle II simulation. Given the very limited statistics of data accumulation we used in this analysis, we only present the expected number of these resonances in  $400\text{fb}^{-1}$  luminosity (about  $2.14 \times 10^8$  events) from generic  $\Upsilon(4S)$  events. These numbers should be re-checked in the future when data accumulation increases, and veto must be based on the structure of  $2K_S^0$  invariant mass from data as well. Details about the expected yields can be found in Table 4.6. Currently there is no veto applied for rejecting these resonant background considering the estimated background number is about 1 event in the current luminosity.

#### 4.1.5 $B\bar{B}$ background and self-cross feed

Another possible contribution of backgrounds are from  $B\bar{B}$  events including the charged and the neutral particles. The estimated contributions of these types can be checked with charged  $B\bar{B}$  samples and the mixed samples. For this channel, the number of the events is very limited. Self-cross feed backgrounds stands for the events from the signal-like events but the tag-side particle(s) is associated as a fake signal. The combined contributions from  $B\bar{B}$  background and self-cross feed is about 3% in the channel and therefore we don't perform special treatment on them.

#### 4.1.6 Signal Extraction

The event selections defined in Table 4.4 is applied to signal MC, generic MC and experiment data for signal extraction. As introduced in chapter 2, the integral luminosity in generic MC is  $1\text{ ab}^{-1}$  and experiment data used in this analysis is about  $62.8\text{ fb}^{-1}$  from the latest official processing.

The unbinned maximum likelihood fit using RooFit is performed to extract the signal. The 2D fit using both  $M_{bc}$  and  $\Delta E$  are done by taking the probability density function as shown in Equation 4.13.

$$\mathcal{P}(M_{bc}, \Delta E) = f_{sig} \times \mathcal{P}_{sig}^{M_{bc}} \times \mathcal{P}_{sig}^{\Delta E} + (1 - f_{sig}) \mathcal{P}_{bkg}^{M_{bc}} \times \mathcal{P}_{bkg}^{\Delta E} \quad (4.13)$$

Table 4.6: Expected yield for signal and background resonances  $2.14 \times 10^8 B\bar{B}$  in generic MC. The branching fraction of  $B \rightarrow X K_S$  and  $X \rightarrow 2K_S$  are listed for both PDG value and value in Belle II generic decay profile (see section 2.9). The events from  $CP$ -odd contamination is expected to be very low at current luminosity ( $62.8 \text{ fb}^{-1}$ ).

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

where  $\mathcal{P}_{sig}^{M_{bc}}$  and  $\mathcal{P}_{sig}^{\Delta E}$  are the single Gaussian and triple Gaussian functions. The  $f_{sig}$  is fraction of signal events based on  $M_{bc}$  and  $\Delta E$ . The  $\mathcal{P}_{bkg}^{M_{bc}}$  is primarily continuum events, and presented as Argus distribution as Equation 4.14 shows, with a preset mass threshold at  $c = 5.29$  GeV.

$$f(x; \chi, c) = \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} \cdot \frac{x}{c^2} \sqrt{1 - \frac{x^2}{c^2}} \cdot \exp\left\{-\frac{1}{2}\chi^2(1 - \frac{x^2}{c^2})\right\} \quad (4.14)$$

$x$  is defined in  $0 < x < c$ .  $\chi$  and  $c$  are parameters of the distribution,  $\Psi(\chi) = \Phi(\chi) - \chi\phi(\chi) - \frac{1}{2}$  where  $\Phi(\chi)$  and  $\phi(\chi)$  cumulative distribution and probability density functions of the standard normal distribution, respectively.

The  $\mathcal{P}_{bkg}^{\Delta E}$  is modeled by the first order Chebyshev polynomials. The shape parameters of signal events are determined by fitting to signal MC, and then fixed as constants in fitting of Equation 4.13 on  $M_{bc}$  and  $\Delta E$  for generic MC and experiment data. Fitting results on signal MC are shown in Figure 4-10. The continuum back-

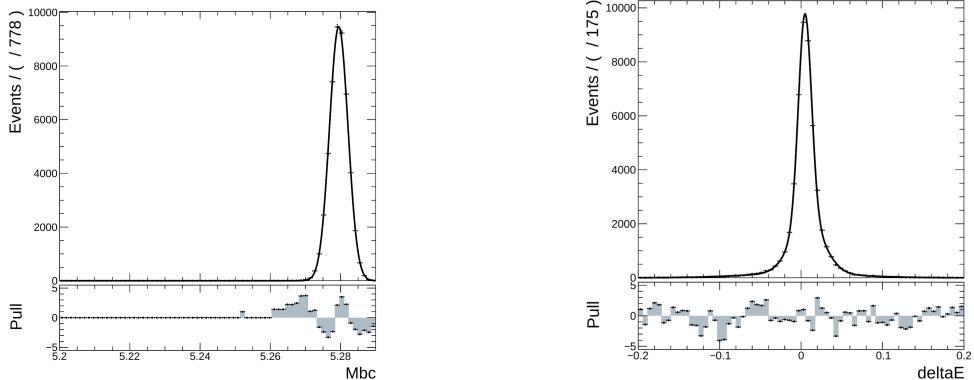


Figure 4-10: The distribution of  $M_{bc}$  and  $\Delta E$  of signal MC of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  fitted with single and triple Gaussian functions respectively.

ground is fitted by using off-resonance generic MC to determine the shapes then fix them as constants for 2D fit as shown in Figure 4-11.

Then we set the events number for signal and background as floating parameters and use Equation 4.13 as 2D fit model on  $1 \text{ ab}^{-1}$  generic MC and experiment data, which is also done by using unbinned maximum likelihood fit. For  $B^0$  in generic MC, the stacked histogram of each contribution and the 2D fit result projected on  $M_{bc}$  and  $\Delta E$  is shown in Figure 4-12.

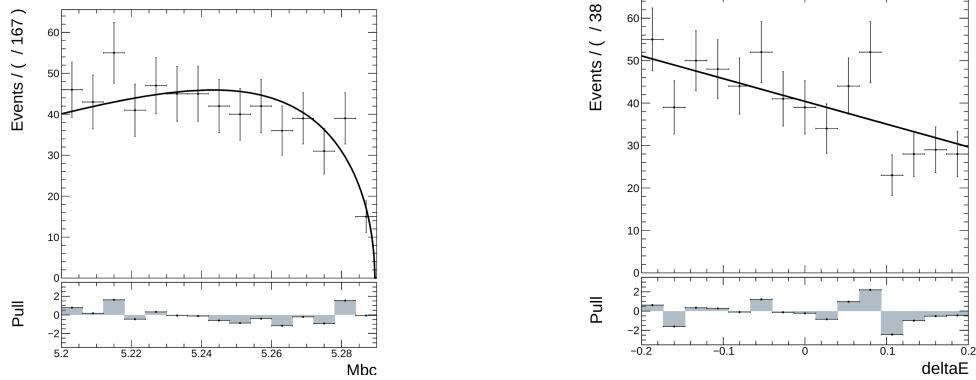


Figure 4-11: The distribution of  $M_{bc}$  and  $\Delta E$  of continuum events in generic MC fitted with Argus and Chebyshev polynomial, respectively.

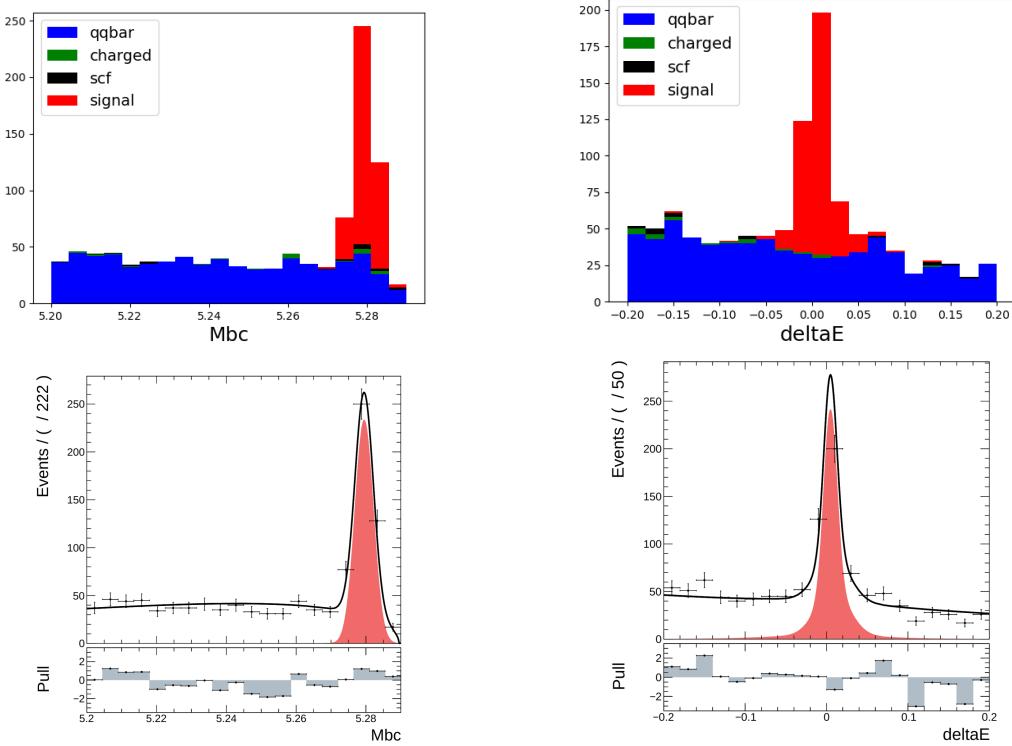


Figure 4-12: Top is the stacked plots for generic MC of  $M_{bc}$  and  $\Delta E$ , where each background components are stacked with signal. The bottom is the 2D fit on  $1 \text{ ab}^{-1}$  generic MC projected on  $M_{bc}$  and  $\Delta E$ , the red is signal component from the fit result in both plots.

Before perform fitting 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 distributions are shown in Figure 4-13, where the generic MC is scaled

to the luminosity of experiment data and an agreement within  $\sim 1\sigma$  is observed on average.

The 2D fit of experiment data projected on  $M_{bc}$  and  $\Delta E$  is in Figure 4-14.

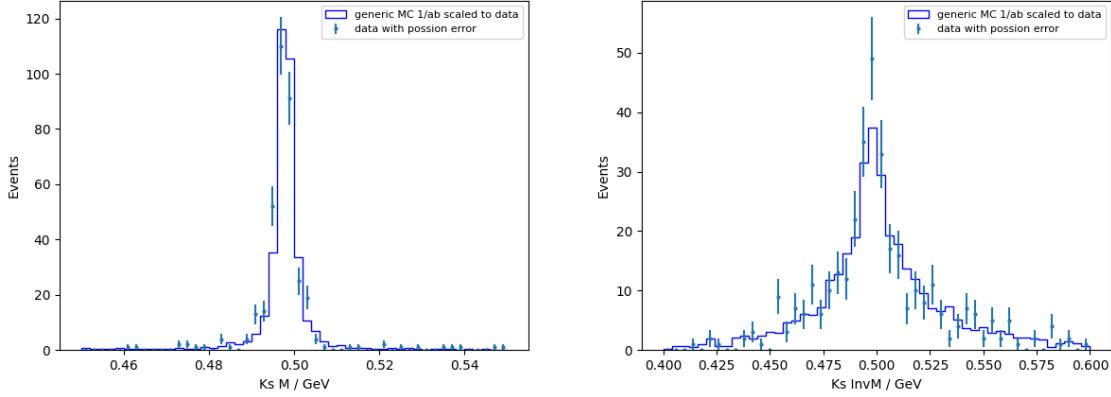


Figure 4-13: Invariant mass before(left) and after vertex fit(right) from generic MC and experiment data.

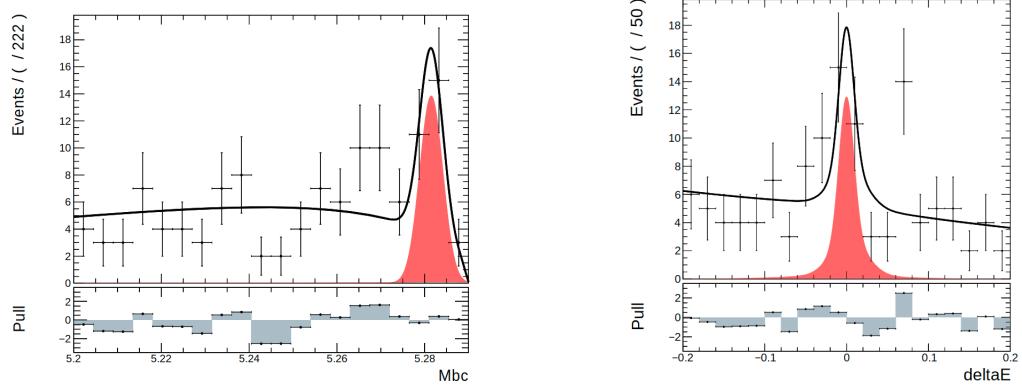


Figure 4-14:  $M_{bc}$  and  $\Delta E$  2D fit on  $62.8 \text{ fb}^{-1}$  data, the red is the signal component.

The number of signal events is extracted by the integral of fit model over the signal region which is defined as  $5.27 < M_{bc} < 5.29 \text{ GeV}$  and  $-0.1 < \Delta E < 0.1 \text{ GeV}$ . The expected signal events with  $\sim 35\%$  efficiency is calculated as Equation 4.15.

$$\mathcal{B}(B^0 \rightarrow K_S^0 K_S^0 K_S^0) = \frac{N_{sig}}{\mathcal{B}(K_S^0 \rightarrow \pi^+ \pi^-)^3 \times \epsilon_{rec} \times N_{B\bar{B}}} \quad (4.15)$$

In  $1 \text{ ab}^{-1}$  generic MC, the expected signal number is  $7.7 \times 10^8 \times 6 \times 10^{-6} \times 21\% \times 35\% \simeq 339$ . The 2D fit result from  $M_{bc}$  and  $\Delta E$  yields  $341 \pm 20$  events which agrees with

expected number within  $1\sigma$ . The event number in sideband defined as  $M_{bc} < 5.26$  GeV in generic MC is 507. Compared to Belle result with  $772 \times 10^6$  ( $\sim 1 \text{ ab}^{-1}$ )  $B\bar{B}$  pairs used, signal from data yields  $327 \pm 19$ . In  $62.8 \text{ fb}^{-1}$  data fit in Belle II, we extract  $N_{sig} = 17.4 \pm 4.2$  in signal region. The sideband region  $M_{bc} < 5.26$  GeV contains 60 events in data.

To check linearity of the event number fitted from the  $M_{bc}$  and  $\Delta E$  in this low statistics case, we extract the fraction of continuum backgrounds from generic MC sample rescaled to the experimental data luminosity, which includes about 46 continuum events. 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  $\Delta E$  fit to check the output signal events number. The  $M_{bc}$  and  $\Delta E$  distributions and fit in each injection test are shown in Figure 4-15. The fitted signal and background events depending on the injected numbers (linearity test) are presented in Figure 4-16, where the dependence on both signal and background events number are fitted with linear functions. The fit results show a good linearity on the input and output of signal numbers while the background numbers remain constant close to the input number.

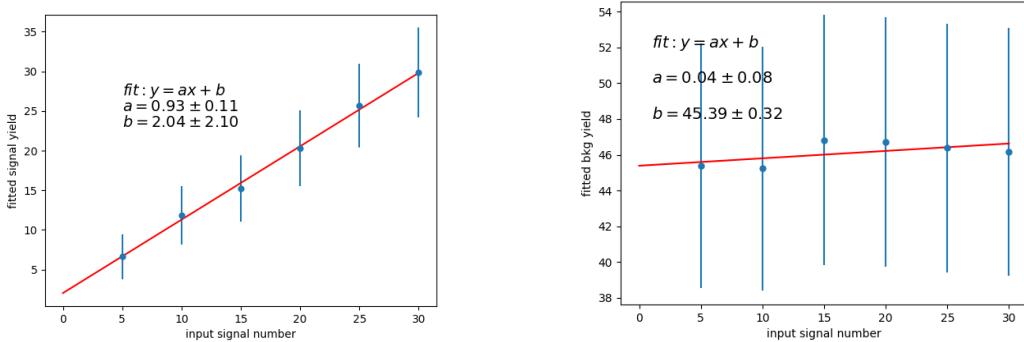
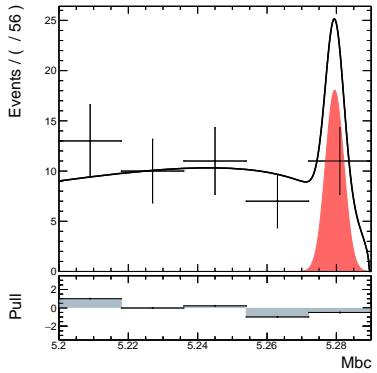


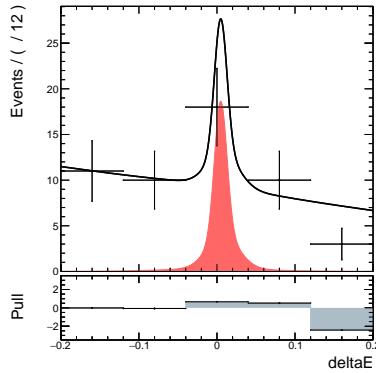
Figure 4-16: Injection test for signal extraction. The linearity is clear between input and output signal events number.

#### 4.1.7 Kinematics and Vertexing Dependence on KsFinder

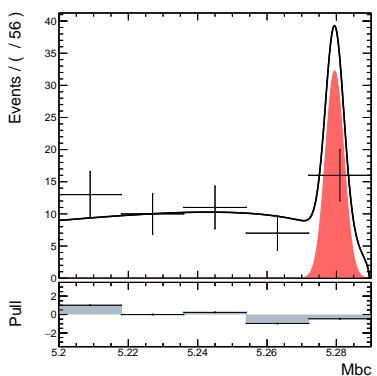
KsFinder largely reduce the combinatorial background of  $B^0$  by improving  $K_S^0$  purity. The previous section shows a good reconstruction performance at low statistics in



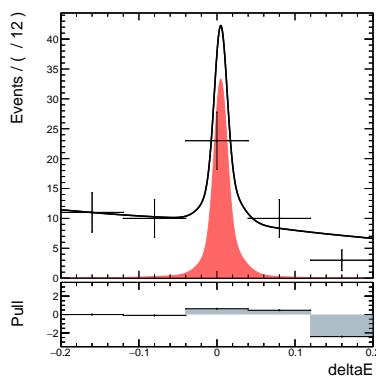
a) signal injected: 5



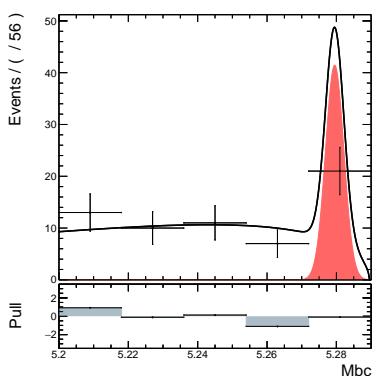
b) signal injected: 5



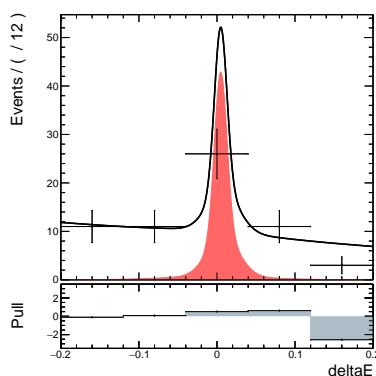
c) signal injected: 10



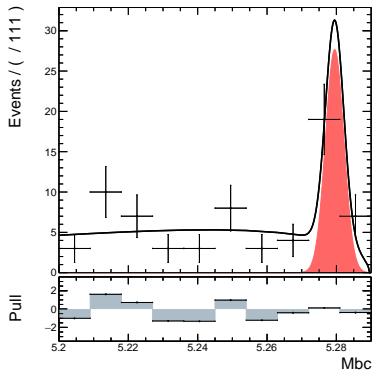
d) signal injected: 10



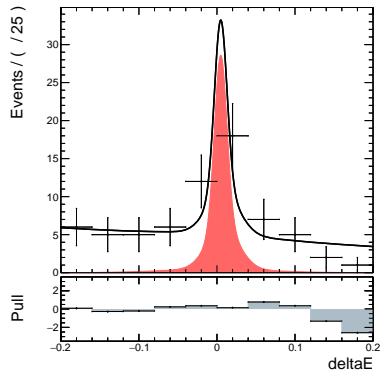
e) signal injected: 15



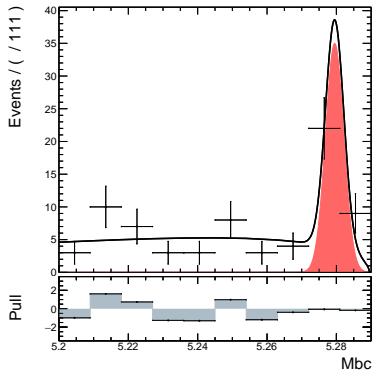
f) signal injected: 15



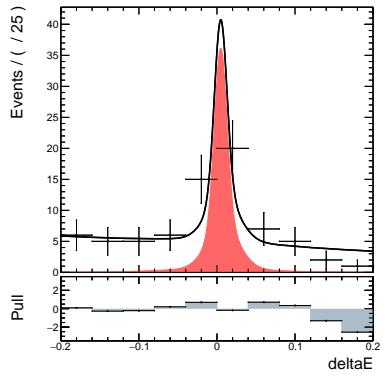
g) signal injected: 20



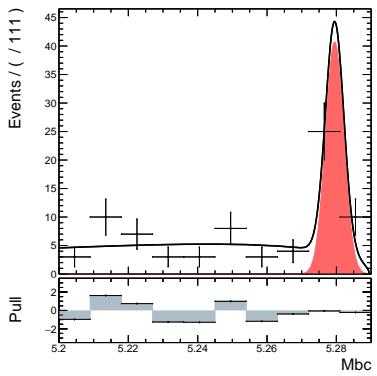
h) signal injected: 20



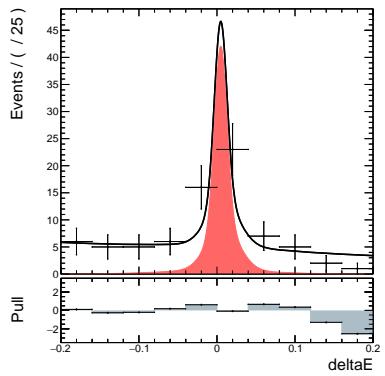
i) signal injected: 25



j) signal injected: 25



k) signal injected: 30



l) signal injected: 30

Figure 4-15: The fit results of  $M_{bc}$  and  $\Delta E$  in signal injection test, where signal events from 5 to 30 with 5 per step are injected with 46 continuum events.

early phase 3 data. Without the power of rejection provided by  $K_S^0$  finder, rediscovery of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  in early phase 3 of Belle II won't be feasible.

However, it's essential to check the potential impact on kinematics and vertex positions of  $B^0$  regarding the implementation of KsFinder. The  $K_S^0$  classification uses information such as invariant mass and decay vertex positions which may propagate bias into  $B^0$  kinematics and vertex information, eventually may affect the measurement of  $CP$  parameters. Considering that 4 different types of  $K_S^0$  based on their SVD hit numbers are used in  $B^0$  reconstruction, the estimation with  $B^0$  based on different  $K_S^0$  types are required as well.

Given each type of  $B^0$  based on how many CDC-only tracks (meaning  $B^0$  daughter  $K_S^0$  are *SVD00* type) it has in final states, the comparison on  $M_{bc}$  and  $\Delta E$  with or without KsFinder is done by fitting the distribution in signal MC.  $M_{bc}$  and  $\Delta E$  are both modeled by double Gaussian. The main Gaussian (with larger fraction) among the double Gaussian is used for checking the fit shape change by using KsFinder. Comparing corresponding fit results, no clear bias on  $M_{bc}$  and  $\Delta E$  is found by using KsFinder where the main Gaussian fit results are agreed well. The fit results are shown in Figure 4-17 and 4-18.

Similar to the comparison of  $M_{bc}$  and  $\Delta E$ , the Z direction vertex position and its uncertainties are also checked. No clear bias on vertex position and uncertainties are spotted either. The results are shown in Figure 4-19 and 4-20. It obvious that the resolution of vertex on z-axis is much worse when the final states of  $B^0$  only have CDC tracks.

Above all, no strong appearance of bias on kinematics and vertex positions from using KsFinder has been found, KsFinder may implement a small shift on the vertex position which is neglect-able compared to the large statistical uncertainty due to low luminosity. For the moment, there's no correction on these observables are applied.

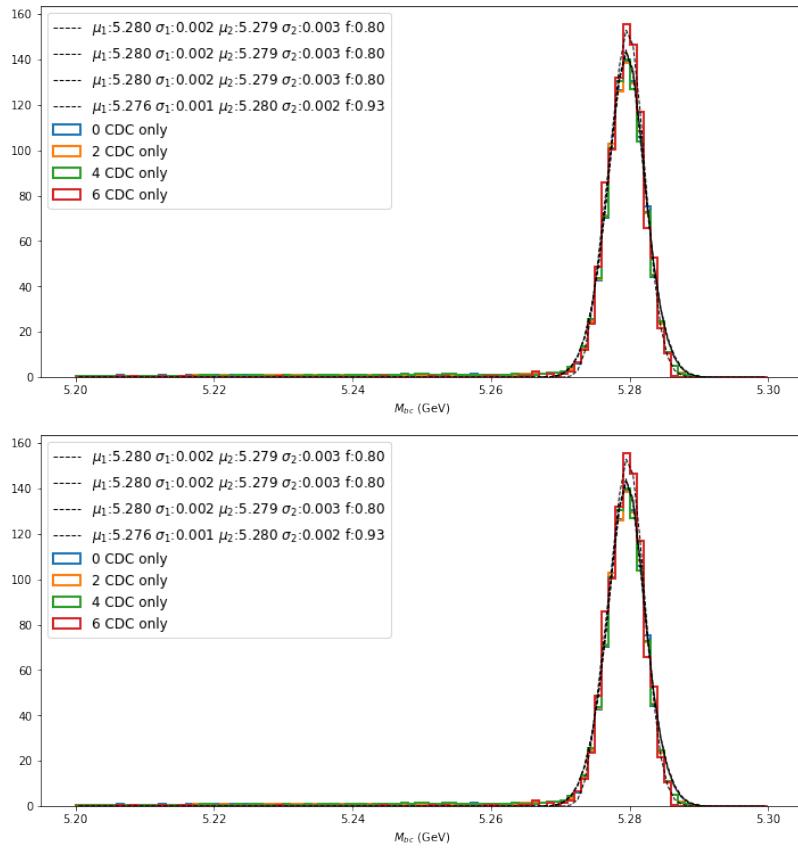


Figure 4-17:  $M_{bc}$  distribution based number of CDC-only tracks in final states. Top: no KsFinder; Bottom:  $K_S^0$  finder used.

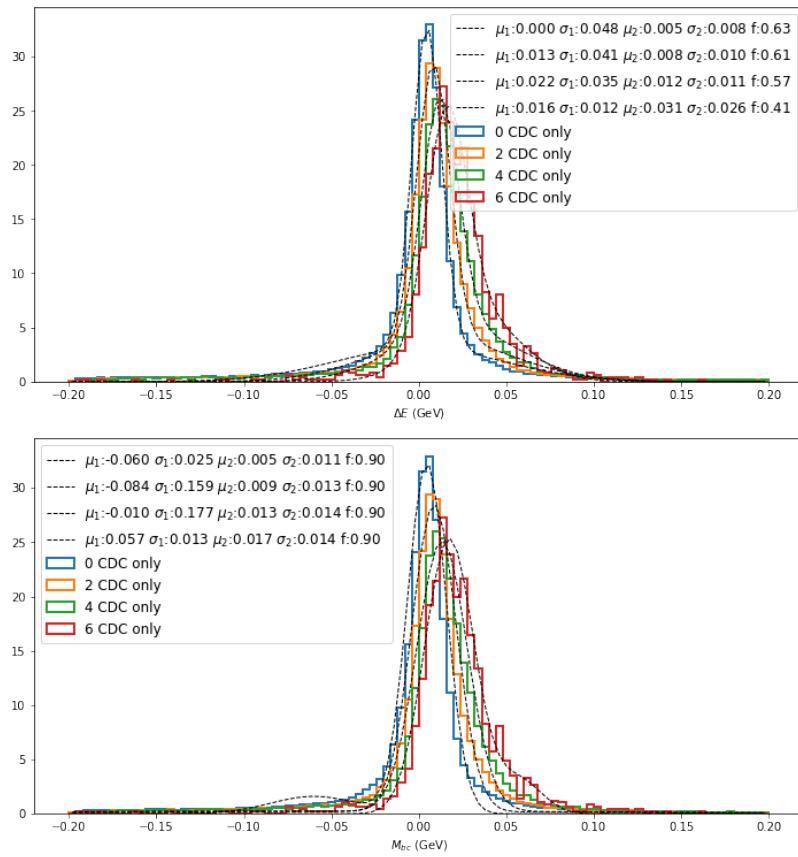


Figure 4-18:  $\Delta E$  distribution based number of CDC-only tracks in final states. Top: no KsFinder; Bottom: KsFinder used.

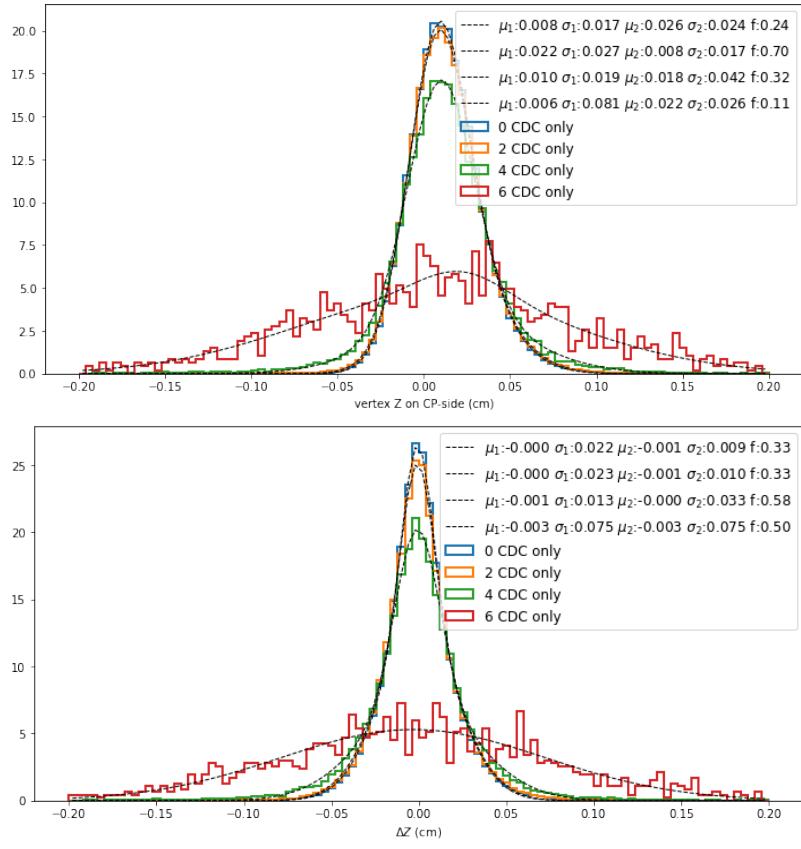


Figure 4-19:  $\Delta z$  distribution based number of CDC-only tracks in final states. Top: no KsFinder; Bottom: KsFinder used.

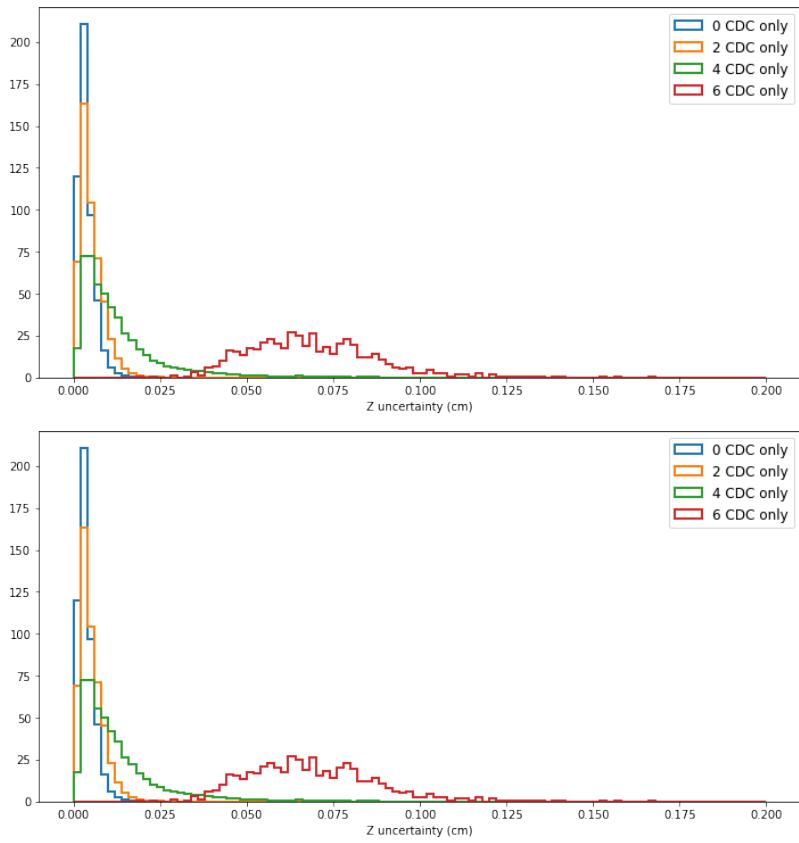


Figure 4-20:  $\delta\Delta z$  distribution based number of CDC-only tracks in final states. Top: no KsFinder; Bottom: KsFinder used.



# Chapter 5

## $CP$ parameters measurement

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

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

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.

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

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.

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

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

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

## 5.1 Vertex Resolution Model

The Equation 5.2 presents an ideal distribution of  $\Delta t_i$  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. This turns Equation 5.2 into Equation 5.5.

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

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 ROE. The background events have its own resolution model which is independent from  $CP$  violation parameters. The outlier is used to compensate and smooth the long tails when  $\Delta t$  becomes very large. In the early stage of Belle II, we don't include the outlier to have a more realistic model under very low statistics.

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  $\Delta z$  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  $\chi^2$  called  $\chi_2/N$  from *TreeFit*, which  $N$  is the degree of freedom of the fit. The distribution of  $\chi_2/N$  in data are shown in Figure 5-1.

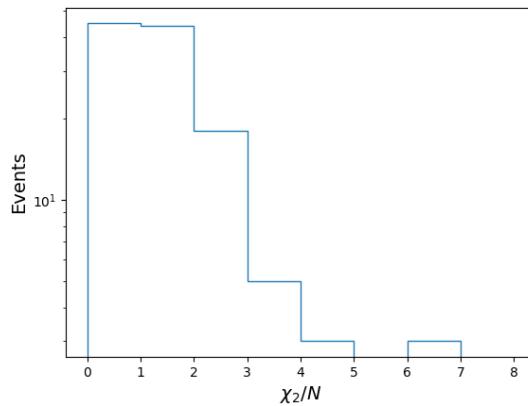


Figure 5-1:  $\chi_2/N$  of selected events from data.

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  $\chi_2/N$  and the error of reconstructed vertex  $\sigma_{z_{cp}}$ , as shown in Equation 5.8.

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

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

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

The dependence of resolution models on  $\chi_2/N$  is shown in Figure 5-2. Restrictively speaking, the  $CP$ -side resolution for  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is slight different from  $B^0 \rightarrow J/\psi K_S^0$ , 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, we can determine the parameters which are listed in Table 5.1.

Table 5.1: Parameters in  $R_{cp}$ .

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

### 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

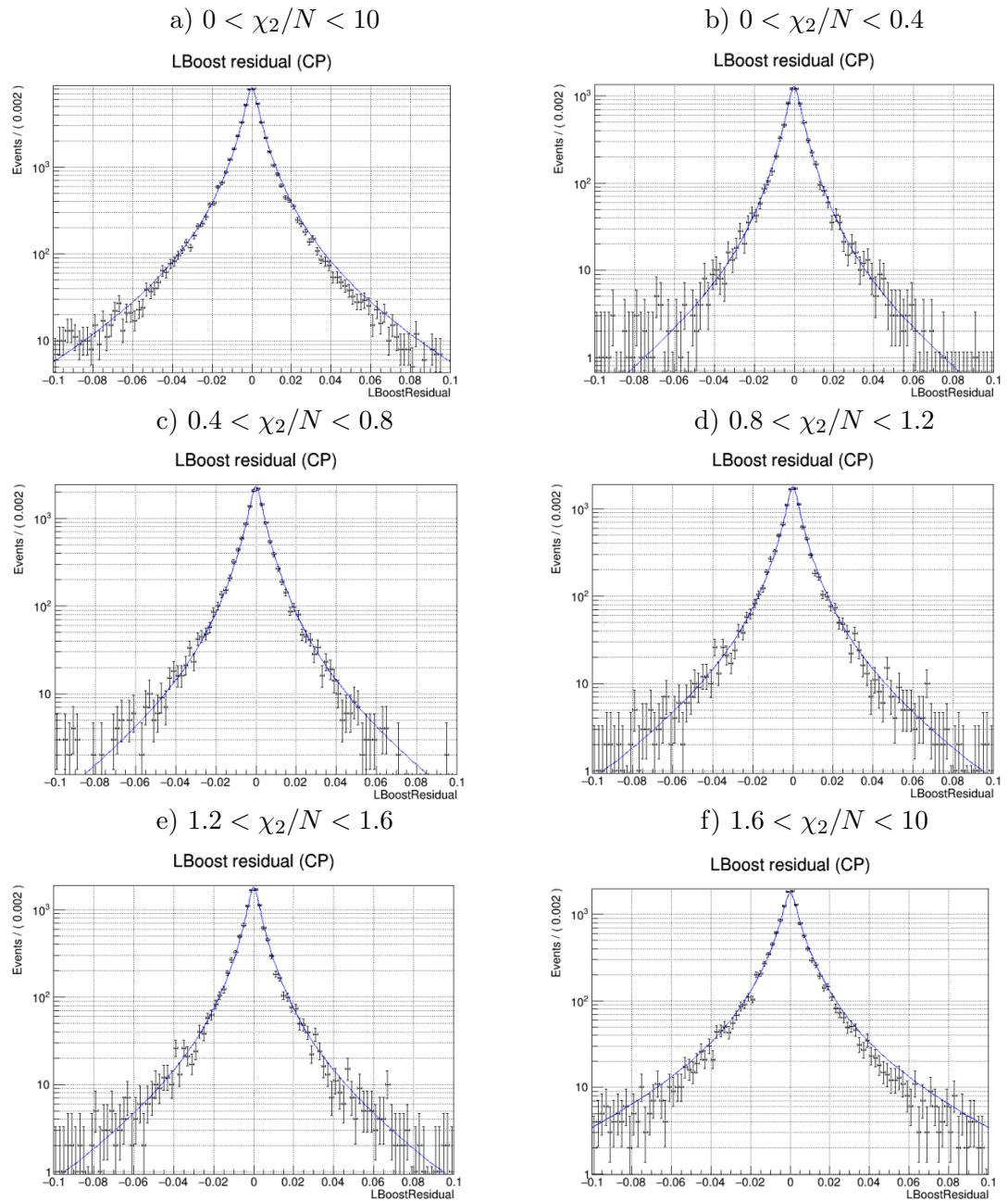


Figure 5-2: Resolution functions on *CP*-side, which shows dependence on the  $\chi_2/N$

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.

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

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

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

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

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

$$s_{tag}^{main/tail} = s_0^{main/tail} + s_1^{main/tail} \cdot \chi_{tag}^2/N \quad (5.13)$$

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  $\delta$  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  $\sigma_{z_{tag}}$ .

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

Also, since tag-side has no dependence on how  $CP$ -side is reconstructed, the res-

olution functions on tag-side are almost mode-independent. We use the parameters that is obtained by fitting to the control sample data. The details about control sample study are here[45]. 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.

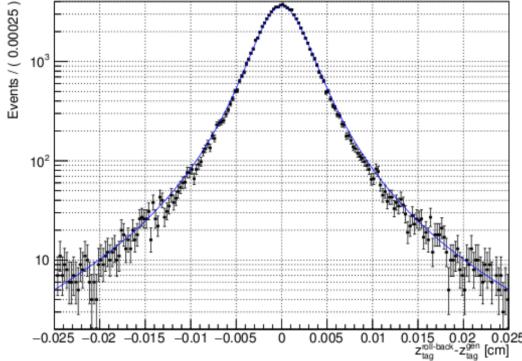


Figure 5-3:  $R_{det}^{tag}$  fit

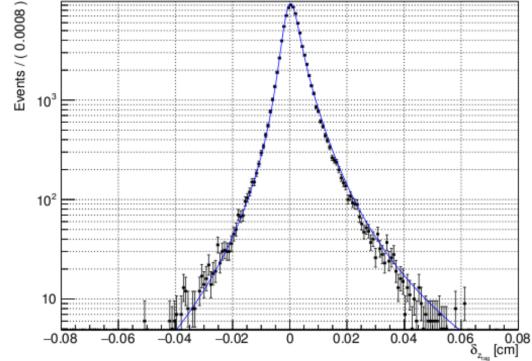


Figure 5-4:  $R_{np}^{tag}$  fit

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

$f_{tag}^{tail}$	$0.0523 \pm 0.0025$
$s_0^{main}$	$1.1446 \pm 0.0061$
$s_1^{main}$	$0.0443 \pm 0.0022$
$s_0^{tail}$	$3.4480 \pm 0.0897$
$s_1^{tail}$	$0.2666 \pm 0.0276$

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

$f_\delta$	$0.6256 \pm 0.0049$
$f_p$	$0.8316 \pm 0.0051$
$\tau_n$	$2.9141 \pm 0.0758$
$\tau_p$	$2.4846 \pm 0.0269$

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,  $\Delta z$  projection on the boosted direction of each event is used for reducing this kinematics effect on resolution function.

### 5.1.3 Background events $\Delta t$ 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  $\Delta t$  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

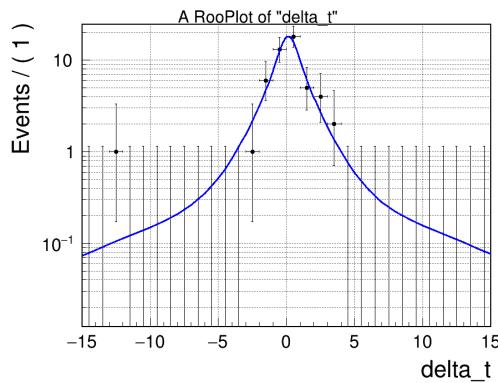


Figure 5-5:  $\mathcal{P}_{bkg} \otimes R_{bkg}$  fit using 60 sideband events at  $M_{bc} < 5.26$  GeV.

$\mu_{\delta}^{bkg}$	$0.1310 \pm 0.1902$
$\mu_l^{bkg}$	$0.1638 \pm 0.5030$
$\tau_{bkg}$	$1.0541 \pm 0.4370$
$f_{\delta}^{bkg}$	$0.5861 \pm 0.2570$
$f_{tail}^{bkg}$	$0.0417 \pm 0.0408$
$\sigma_{main}^{bkg}$	$1.4348 \pm 0.3940$
$\sigma_{tail}^{bkg}$	$28.0930 \pm 8.8221$

Table 5.4: Parameters in Background  $\Delta t$  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  $\Lambda$  which are categorized into 13 different types as illustrated in Figure 5-6.

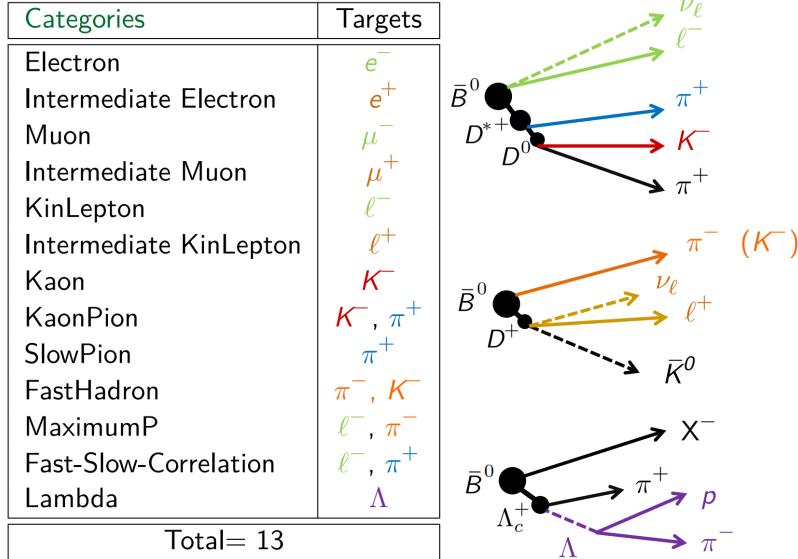


Figure 5-6: Particles and their categories used in flavor tagging algorithm[47].

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 which is used in this thesis. To be specific, after the reconstruction on the  $CP$ -side  $B^0$  is done, the rest-of-events tracks that are used to form the particle lists 5-6 are selected. 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  $B^0 \rightarrow \nu\nu$  is used as the training sample where the final states are completely invisible.

Considering the limited power of flavor tagging accuracy, there is a certain fraction of events that are wrongly tagged, among all events that can be flavor tagged based on the final states. Thus, the flavor tagging efficiency  $\epsilon$  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.

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

Compared to the original, the term with  $\mathcal{S}$  and  $\mathcal{A}$  is reduced by factor  $r \equiv |1 - 2w|$ , defined as the dilution factor.

The statistical uncertainty of  $\mathcal{S}$  now receives contribution from  $\epsilon$  and  $w$ :  $\delta(\mathcal{S}) \propto \frac{1}{\epsilon(1-2w)N_{rec}}$ , which means it's quite important to correctly measure  $\epsilon$  and  $w$ . The uncertainty of  $w$  is much larger than  $\epsilon$  which makes  $w$  an important source of systematic uncertainty. The validation of flavor tagger using flavor specific control sample is summarized here: [47]. The  $w$  for each single event is defined as a probability which can be presented by the average wrong tag fraction within a close range. The binned values of dilution factor  $r$  is defined for the calculation of  $w$  as  $[0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.875, 1]$ , which is called  $r$ -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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ .

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

$\Delta w = w_{B^0} - w_{\bar{B}^0}$ . Due to the small value of  $\Delta w$ , the contribution from  $\Delta w$  can be safely treated as zero in Equation 5.16. Similarly, for  $\epsilon$ , the values calculated based on each  $r$ -bin are used in Equation 5.16 as well. The difference  $\mu = \epsilon_{B^0} - \epsilon_{\bar{B}^0}$  is about 1% to 2% in each  $r$ -bin, thus treated as zero, too. The distributions of  $w$ ,  $\Delta w$ ,  $\epsilon$ , and  $\mu$  in each  $r$ -bin are shown in Figure 5-8.

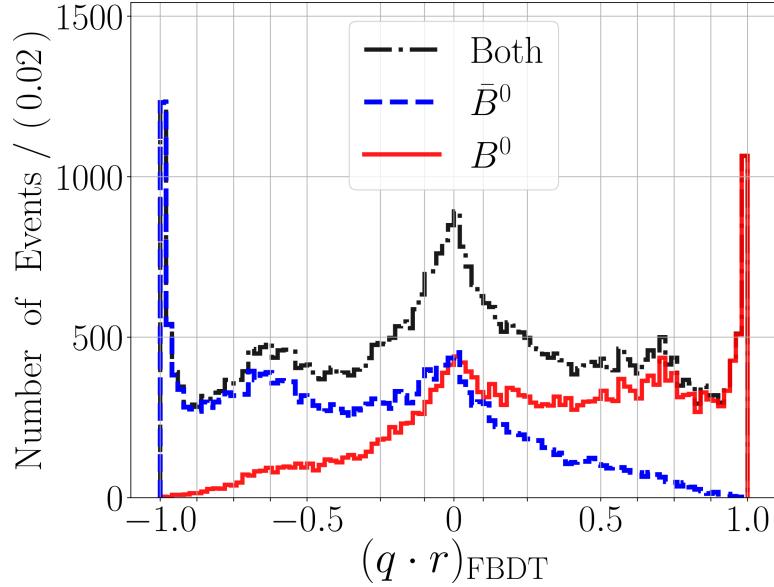


Figure 5-7: The distribution of flavor tagger output ( $q \cdot r$ ) for both tag-side of  $B^0$  and  $\bar{B}^0$

### 5.3 $CP$ Fitter

The parameters that are needed for measuring  $\mathcal{S}$  and  $\mathcal{A}$  are studied and obtainable. Using observed  $\Delta t$  distribution from selected events, Equation 5.5 can be fitted using unbinned maximum likelihood fit. The fit takes  $\Delta t$ , signal fraction  $f_{sig}$ , the flavor charge  $q$  as observables, and in the meantime the vertexing error  $\sigma_{z_{cp}}$ ,  $\sigma_{z_{tag}}$  and  $\chi_2/N$  are used as event-by-event conditional variables. For Belle II, a new  $CP$  fitter is developed based on Python (version 3.7) and RooFit, which is naturally easy 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.

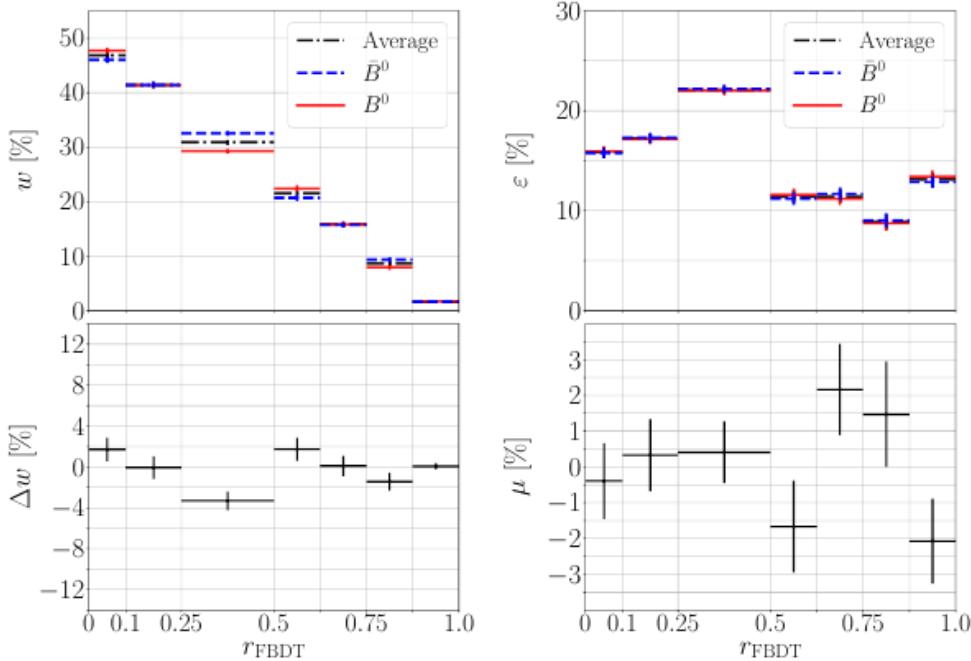


Figure 5-8: The flavor tagging efficiency, wrong tagging fraction, and their difference between different flavors in each of the  $r$ -bin.

## 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 reliability of fit result from  $CP$  fitter, a linearity test and toy MC study is also performed.

### 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 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

Observables	Selections
$\Delta t$	$-70 < \Delta t < 70$ ps
$CP\text{-side } \chi_2/N$	$0 < (\chi_2/N)_{cp} < 8$
tag-side $\chi_2/N$	$0 < (\chi_2/N)_{tag} < 50$
$\sigma_{z_{tag}}$	$\sigma_{z_{tag}} < 0.1$ cm
signal region	$5.27 < M_{bc} < 5.29$ GeV and $ \Delta E  < 0.1$ GeV

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

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

passing selections) events from  $1 \text{ ab}^{-1}$  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 fit results are shown in Figure 5-9, 5-10 and 5-11. The fit results of  $\mathcal{S}$  and  $\mathcal{A}$  are summarized in Table ??

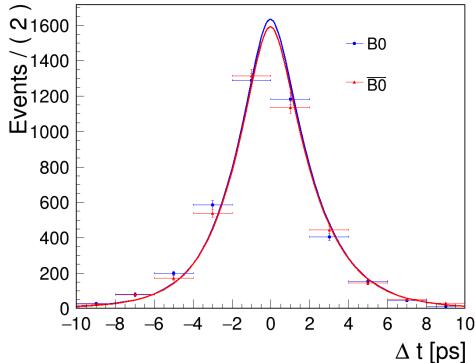


Figure 5-9:  $CP$  fit on 8873 signal MC.

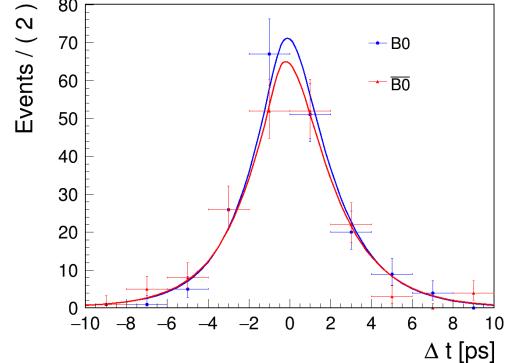


Figure 5-10:  $CP$  fit on 373 generic MC.

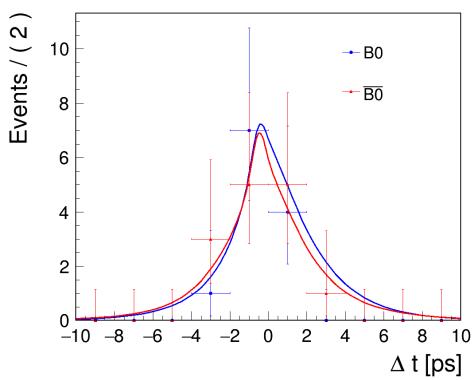


Figure 5-11:  $CP$  fit on 30 generic MC.

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  $B^0 \rightarrow J/\psi K_S^0$  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.

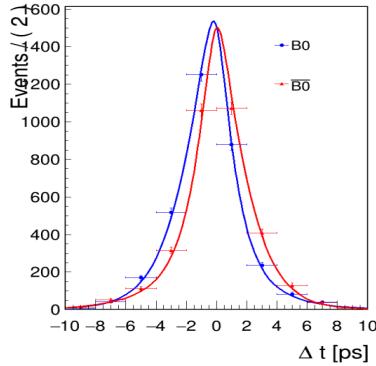


Figure 5-12:  $CP$  fit over 10000  $B^0 \rightarrow J/\psi K_S^0$  signal MC.

### 5.4.2 Linearity Test

To validate the  $CP$  fit linearity, a series of toy MC samples is generated, which the  $\chi_2$  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.

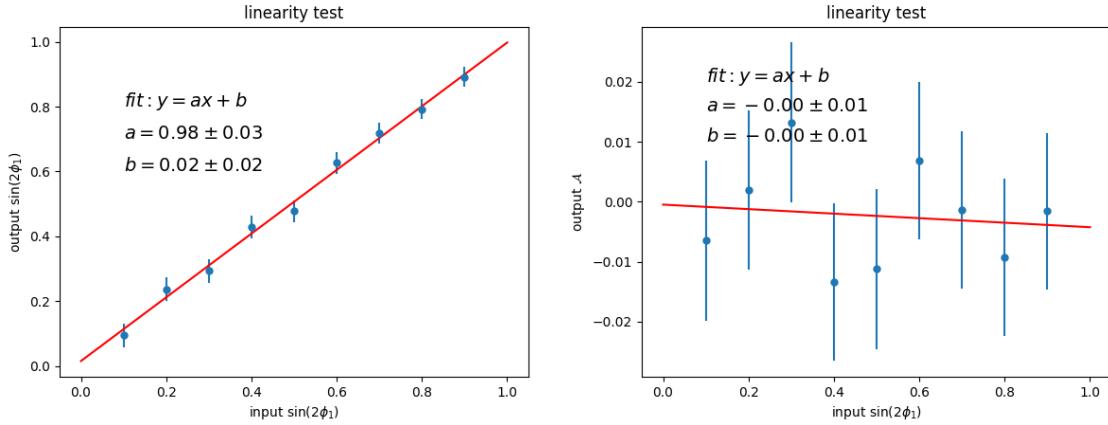


Figure 5-13: Linearity test of  $CP$  fit.

Also, we fix  $\sin(2\phi_1)$  at zero while floating  $A$  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.

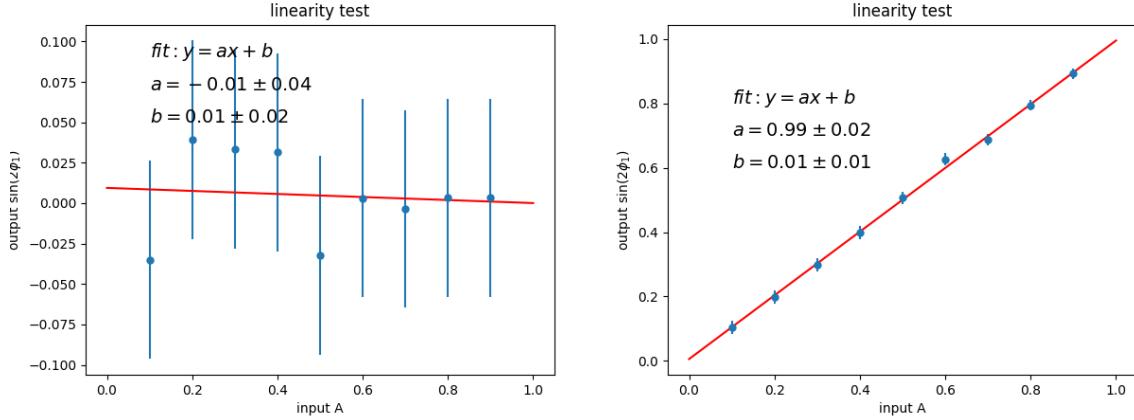


Figure 5-14: Linearity test of  $CP$  fit.

### 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  $\chi^2$  from vertex 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  $\mathcal{A}$ .

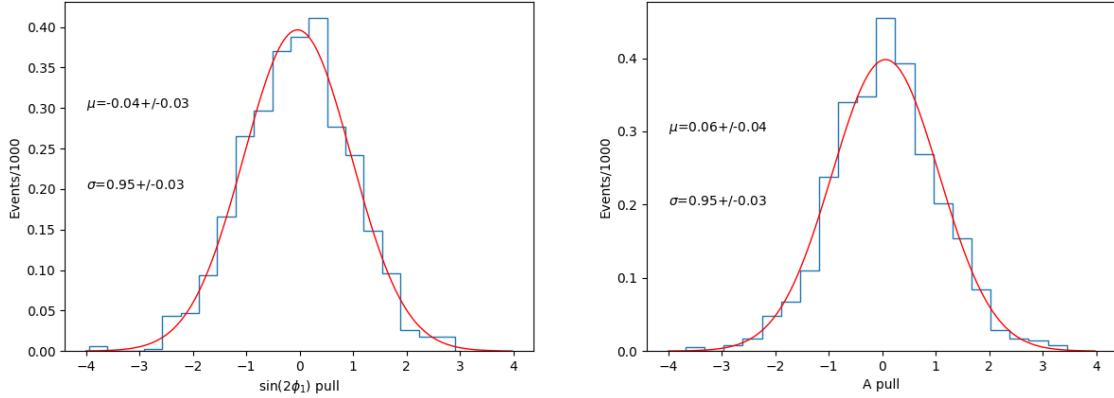


Figure 5-15: Pull of  $\sin(2\phi_1)$  and  $\mathcal{A}$  fitted with the standard normal distribution.

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 $\Delta m_d$ 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.

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

The fit result on signal MC is  $1.537 \pm 0.024$  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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is  $1.431 \pm 0.382$  ps. The result is consistent with PDG value. The distribution of  $\Delta t$  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.

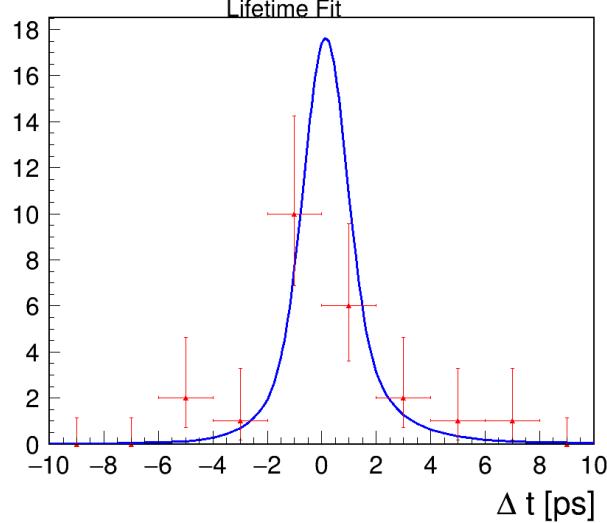


Figure 5-16: Lifetime fit on data

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

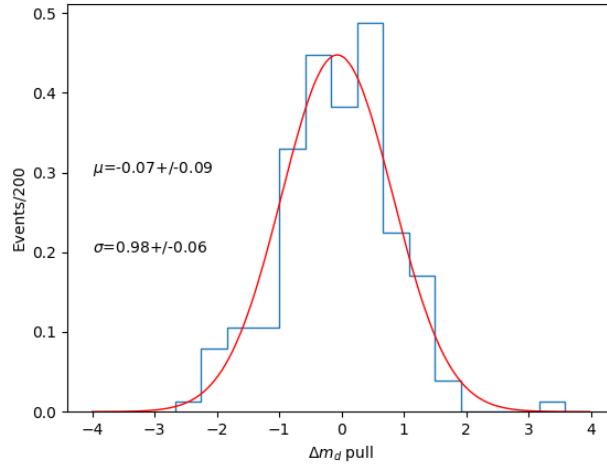


Figure 5-17: Pull of  $\Delta m_d$

## 5.5 $CP$ fit on data

After the  $CP$  fit procedures are reviewed by Belle II collaboration, the permission of measuring  $CP$  parameters using  $62.8 \text{ fb}^{-1}$  Belle II data is granted. The events number used for the  $CP$  fit is 26, and the fit result is shown Figure 5-18.

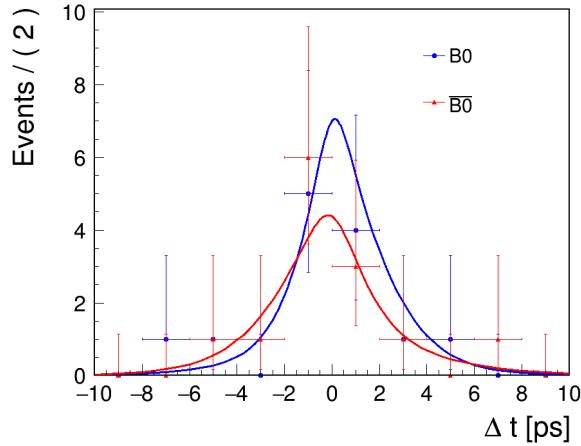


Figure 5-18: The  $CP$  fit from data.

The results of  $CP$  parameters are:

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

## 5.6 Systematic Uncertainty

The systematic uncertainty that affects the fit results may come from many aspects of the measurement setup. The contributions that matter at the current stage of Belle II are summarized as Table 5.6.

For the listed sources of systematic uncertainty, if the parameters are defined with MC study, we float the value by  $\pm 2\sigma$  of their uncertainty, and if the parameters are defined by data, we float the value by  $\pm 1\sigma$ , to have a more robust estimation of impact on fit results. The impact of  $CP$  parameters are separately estimated from each sources with positive and negative differentials. With sum of the quadrature of

Sources	$\delta\mathcal{S}$	$\delta\mathcal{A}$
signal fraction	0.04	0.04
background $\Delta t$ shapes (data)	0.04	0.04
signal $\Delta t$ resolution (MC)	0.03	0.01
fit bias	0.01	0.01
flavor tagging	0.004	0.004
physics parameters	0.007	0.001
KsFinder impact on data	0.004	0.001
vertex reconstruction	0.019	0.021
Total	0.07	0.06

Table 5.6: The contributions of each source of systematic uncertainty.

each term, the overall systematic uncertainty is obtained.

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.

Table 5.7: systematic uncertainty from signal  $\Delta t$  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
$s_0^{main}$	0.005443	0.001299	-0.005675	-0.001404
$s_1^{main}$	0.019934	-0.000903	-0.020204	0.000633
$s_0^{tail}$	-0.003233	-0.001623	0.00327	0.001596
$f_{tag}^{tail}$	0.00314	-0.001257	-0.003117	0.001266
$s_0^{main}$	0.002011	-0.001395	-0.001956	0.001398
$s_1^{main}$	0.005059	-0.00084	-0.004969	0.000825
$s_0^{tail}$	-0.000135	-0.000393	0.00010	0.000435
$s_1^{tail}$	0.000101	0.000027	-0.000472	0.000129
$f_\delta$	-0.007248	-0.000552	0.007231	0.000591
$f_p$	0.003037	0.004347	-0.003069	-0.004314
$\tau_n$	-0.00101	-0.002841	0.000937	0.00294
$\tau_p$	0.004497	0.002502	-0.004648	-0.002478

The background  $\Delta t$  shapes' parameters are determined from data sideband  $M_{bc} < 5.26$  GeV. The impact on fit results is summarized as follows:

Table 5.8: systematic uncertainty from background  $\Delta t$  shapes

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
$\mu_\delta^{bkg}$	-0.014294	-0.016581	0.006758	0.006537
$\mu_l^{bkg}$	-0.002798	-0.012567	0.003789	0.012783
$\tau_{bkg}$	0.001377	0.001689	-0.004159	0.000085
$f_\delta^{bkg}$	-0.011315	0.001365	0.011187	-0.001395
$f_{tail}^{bkg}$	-0.002661	0.00153	0.00248	-0.001368
$\sigma_{main}^{bkg}$	0.0207015	0.022041	-0.0236175	-0.01569
$\sigma_{tail}^{bkg}$	-0.000275	-0.000159	0.000179	0.000141

The flavor tagging parameters wrong tagging fraction  $w$  in each rbin is determined from signal MC. The impact in each rbin on fit results is summarized as follows:

Table 5.9: systematic uncertainty from wrong tagging fraction

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
$w_1$	-0.0018919	0.001911	0.0018549	-0.002004
$w_2$	-0.0016448	0.001104	0.0016085	-0.001155
$w_3$	-0.0004899	0.001344	0.0004726	-0.001341
$w_4$	0.0006556	0.000264	-0.0006542	-0.000255
$w_5$	-0.0001228	0.000204	0.0001225	-0.000195
$w_6$	0.0000948	0.000054	0.0000957	-0.000045
$w_7$	0.0001911	-0.000396	-0.0001907	0.000402

The physics parameters  $\Delta m_d$  and  $\tau_{B^0}$  uncertainties are included using the PDG average value. The impact on fit results is summarized as follows:

Table 5.10: systematic uncertainty from physics parameters

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
$\Delta m_d$	-0.001767	-0.000687	0.001778	0.000696
$\tau_{B^0}$	-0.004561	-0.000546	0.004565	0.000555

The signal fraction is determined using 2D fit results of  $M_{bc}$  and  $\Delta E$  from data. The impact on fit results is summarized as follows:

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	-0.0007965	0.003849
sigma1_mbc	0.0004755	0.008442	-0.000628	-0.008733
m0_argus	-0.000707	0.00414	0.001448	-0.005781
c_argus	-0.005544	0.001449	0.000922	-0.000078
f1_de	0.0278255	0.020589	-0.0192365	-0.008409
f2_de	0.020809	0.017649	-0.0161285	-0.007005
mu1_de	-0.000443	-0.000153	0.0004955	0.000088
mu2_de	-0.000563	0.001446	0.0005905	-0.001446
mu3_de	-0.0031635	-0.000834	0.003354	0.000981
sigma1_de	-0.0001715	-0.000966	0.000206	0.000906
sigma2_de	-0.0031495	0.002958	0.0026345	-0.002475
sigma3_de	-0.001926	-0.00255	0.0024695	0.002985
a0_cheb	0.0009515	0.000057	-0.0008925	-0.000102
N_sig_f	-0.0046395	0.003987	0.004922	-0.003504

The fit bias uncertainties is determined by the fit error of 100k signal MC events, which is  $\delta\mathcal{S} = 0.009817$  and  $\delta\mathcal{A} = 0.005702$ .

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 threshold 0.74, the  $R_B$  presenting MC and data signal yield ratio is  $R_B = 1.027 \pm 0.033$ ,

where the up and low limit for MC correction is 1.06 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.

Table 5.13: systematic uncertainty from KsFinder.

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
$R_B = 1.06$	0.004826	-0.000606
$R_B = 0.994$	-0.000508	0.000007

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 these values doesn't affect events collected from data so that systematic uncertainty can not be reflected correctly. Therefore,  $1 \text{ ab}^{-1}$  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.

Table 5.14: systematic uncertainty from vertex reconstruction

source	$\delta\mathcal{S}$	$\delta\mathcal{A}$
$\sigma_{z_{tag}} < 0.05 \text{ cm}$	0.004369	-0.003599
$\sigma_{z_{tag}} < 0.15 \text{ cm}$	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
$ \Delta t  < 50 \text{ ps}$	0.003325	-0.000396
$ \Delta t  < 90 \text{ ps}$	0.000000	0.000000
IP constraint	0.000000	0.000000

# Chapter 6

## Conclusion and Prospect

The Belle II experiment is built upon the success of its predecessor Belle and many other great efforts of exploring the mysteries of flavor physics, which have expand our knowledge and understanding of elementary particle physics. One of the most outstanding outcome of these efforts is the Standard Model, which it's capable of well describing a variety of experimental results in a large energy scale and fine precision for the past few decades. And yet open questions that still draws attention from particle physicists remain, wait to be discovered as New Physics. One of the most important question is that why the universe is mass-dominated while anti-matter seems to be vanished.

Belle II is aimed to search for New Physics through the precise measurements of related topics in heavy flavor physics at the world-record luminosity frontier. SuperKEKB accelerator is designed with asymmetric beam energies to provide a boost to the center-of-mass system and thereby allow for time-dependent  $CP$  symmetry violation measurements. The products of collision is in a very clean environment, with 40 times higher luminosity of peak at Belle. This create excellent opportunities for physicists to look for the undiscovered source of  $CP$  violation, for which the existing explanation from the complex phase of CKM matrix can't described the observed level of asymmetry in our universe.

$b \rightarrow s$  transition is an important flavor coupling process to be examined in search for New Physics. The  $CP$  violation in such process was first observed after the

precise measurement in  $b \rightarrow c$  with a small tension. So far the precision of the measurement of  $CP$  parameter  $\mathcal{S}$  in  $b \rightarrow s$  is still in an arguable difference with tree-level process considering the existing uncertainty, which allows a decent margin for New Physics. The representative processes of  $b \rightarrow s$  are resonant decay such as  $B^0 \rightarrow \eta' K_S^0$ ,  $B^0 \rightarrow \phi K_S^0$  and decay like  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ , on which Belle II experiment will have an excellent prospective sensitivity.

This thesis presents the first attempt to study the time-dependent  $CP$  violation in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  using early phase 3 data of Belle II and latest MC sample. In order to reconstruct clean signal sample of  $B^0$ ,  $K_S^0$  reconstruction performance is critical because of the unique characteristics of this decay. A KsFinder based on FastBDT classification algorithm is developed to offer a goodness indicator of traditional cut-based reconstruction of  $K_S^0$ . The performance of this new KsFinder is validated to have a great background rejection power at with small signal loss in the maximum FOM case.  $B^0$  are reconstructed with a good significance even with very low statistics with a good agreement with MC prediction and Belle experience. The overall efficiency of  $B^0$  is slightly improved than Belle with slightly higher beam background condition in current Belle II. The measurement of the  $CP$  fit is conducted based on the reconstruction. The  $CP$  fit using artificial model containing resolution functions from different sources are built with precise study of MC signal samples and the data sideband . As for flavor tagging information, wrong fraction as mandatory parameters in signal  $\Delta t$  distribution, are implemented too. The coefficient of signal and background in the  $CP$  fit model is determined by the signal extraction 2D fit over  $M_{bc}$  and  $\Delta E$ . For each event, the signal fraction is calculated based on the  $M_{bc}$  and  $\Delta E$  using the 2D fit model, which is used as a discrete observable in  $CP$  fit model.

Before performing  $CP$  fit on data, blind fit study and fit pull/linearity test are conducted to validate the fit model and procedures. In MC fit test, fit results for  $CP$  parameters are consistent with the generation level input. The linearity and fit pull test shows the reasonably good performance of extracting  $\mathcal{S}$  and  $\mathcal{A}$ . The validation of data using  $CP$  fitter to fit  $B^0$  lifetime and mass width are also in a good agreement with PDG value in this low statistics case.

After the  $CP$  fit procedures are validated and the permission from Belle II collaboration of fitting on data is given, the  $CP$  parameters  $\mathcal{S}$  and  $\mathcal{A}$  measurement using Belle II early phase 3 data (2019 and 2020 Spring/Summer) is performed with the result as below:

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

The result is in a consistent with the PDG value and the prediction of the Standard Model, and also with the previous results from Belle and Babar as Chapter 1 described. The systematics study is also performed considered on the majority of contributing sources at this moment. The result is primary limited in precision due to the large statistical uncertainties from very low data collection.

What is worth of noticing is, many analysis tools that are required by performing the  $CP$  measurement on this channel is in a good stage of development. The decay mode being successfully re-discovered under very low data collection mostly thank to the newly developed  $K_S^0$  classification software, KsFinder, which will also be an essential asset in neutral particles dominated channels like  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ . The Belle II experiment is crucial in these channels because of the cleaner background environment and better sensitivity compared with LHCb. Besides, an artificial model on the vertex resolution, or called  $\Delta t$  resolution has been finely studied using MC sample and sideband data. Several important dependence or behaviors of vertexing tools in the Belle II detector is understood to a good extension in the early phase of Belle II. Further, a new  $CP$  fitter which is based on these studies is built and being validated, which will provide a multi-functional analysis tool for Belle II time-dependent  $CP$  violation study in future.

This study has shown a good potential of performing  $CP$  measurement in Belle II for the incoming years with more and more data recorded. The precision on  $\mathcal{S}$  in  $b \rightarrow s$  penguin-modes is highly depending on the large luminosity as Fig 1-9 shows, which the current precision fits in the expectation. Along with the data collection

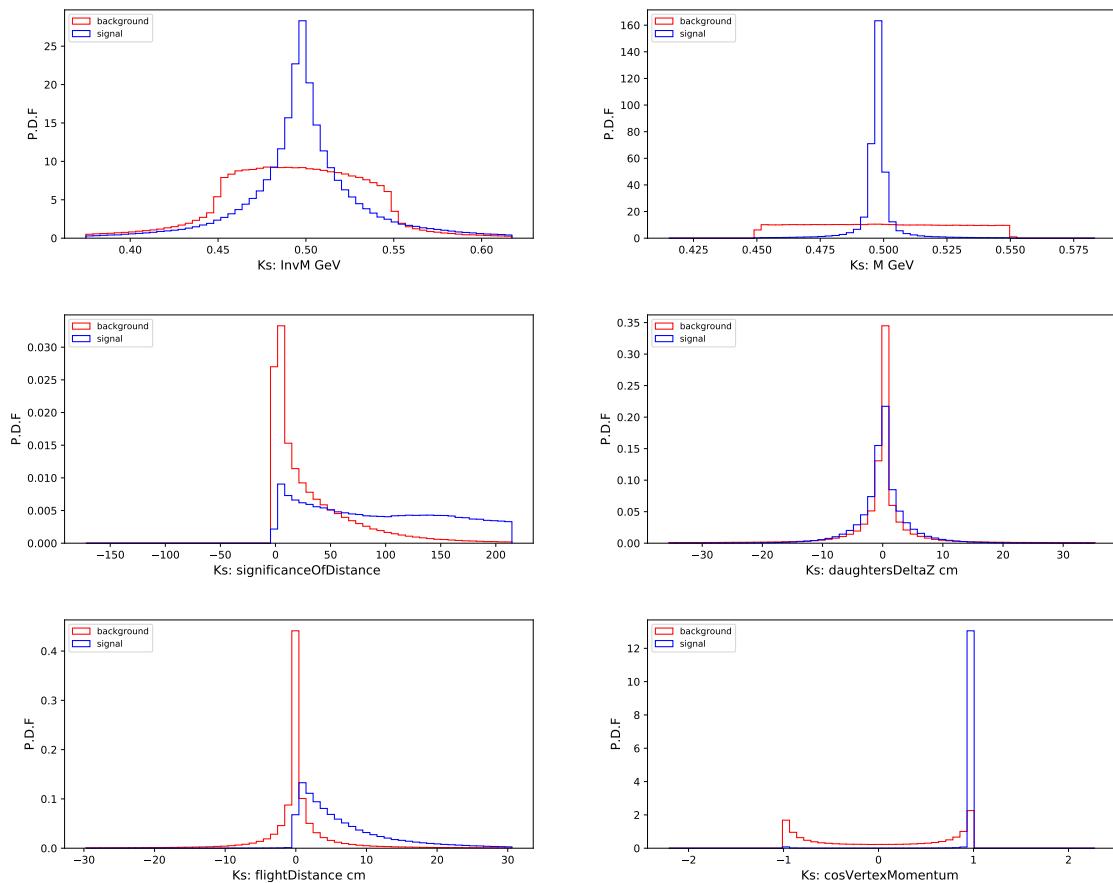
continues, we will be able to finely test and improve the analysis strategies and tools to a better stage using data, such as improving the reconstruction efficiency and purity of  $B^0$  when the luminosity ramps up to much higher level with much higher backgrounds. At integral luminosity at  $50 \text{ ab}^{-1}$  level, the statistical uncertainty of this decay on  $\Delta\mathcal{S}$  would be trimmed down to a comparable value around 0.03 which is close the Standard Model correction, offering a much better probe on whether New Physics is influential at this level of precision. The progress that has been made so far in this thesis paves a well-constructed and solid path towards future results. From the current result, the chance of having a much precised measurement in the next a few years on this channel is very promising and searching New Physics effect in penguin-mode  $b \rightarrow s$  transition from Belle II is proven to be an exciting and important topic.

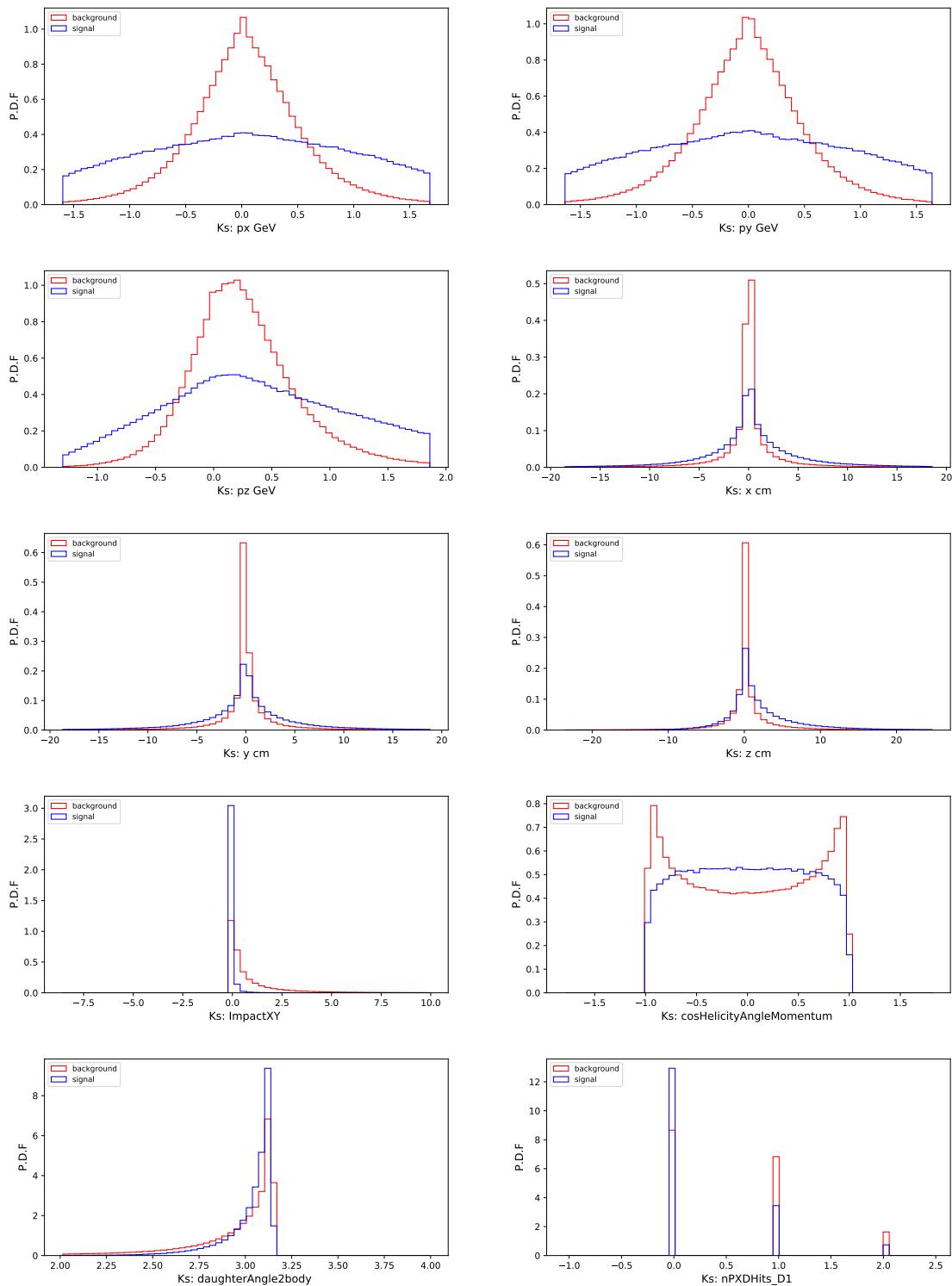


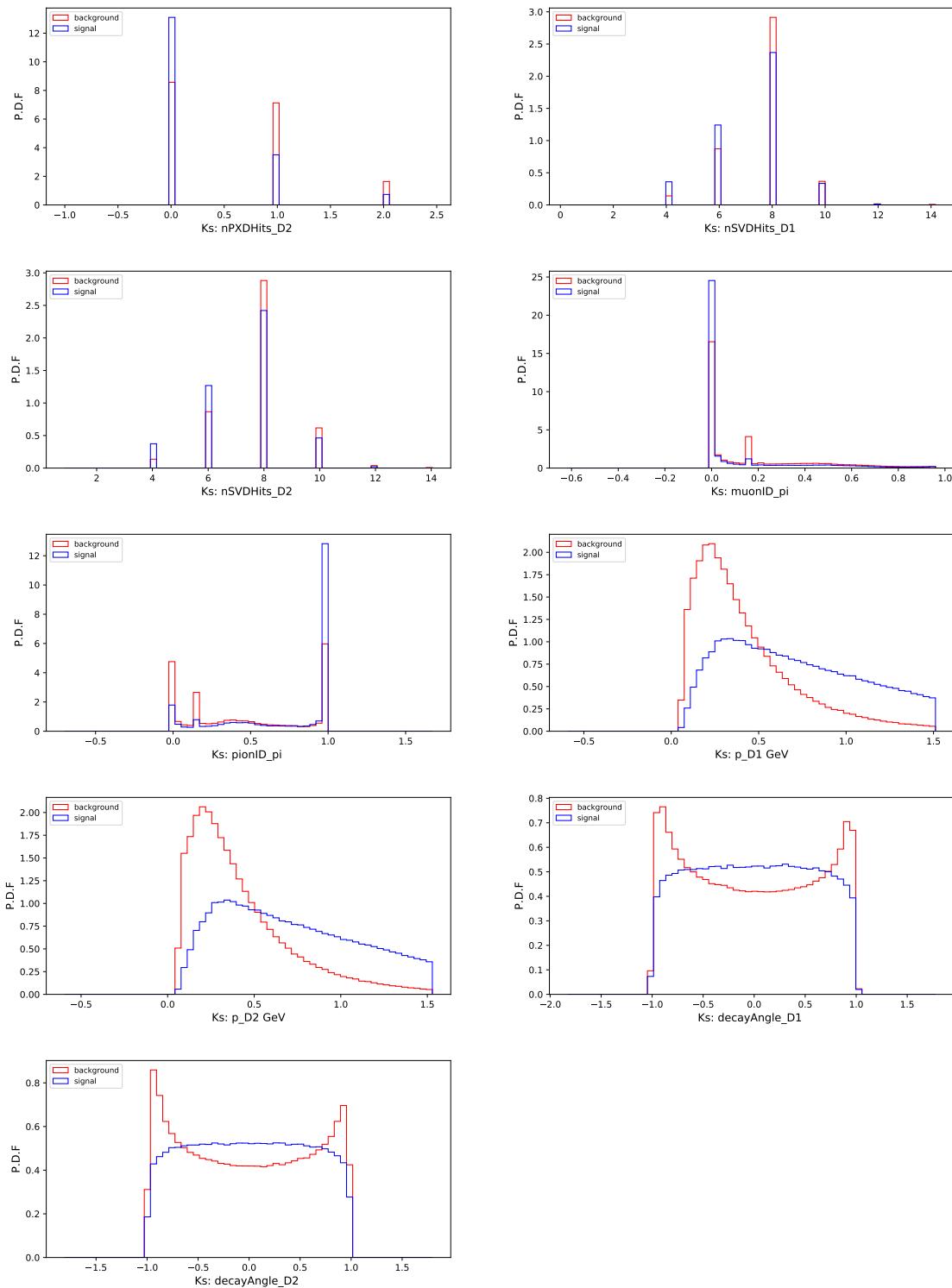
# Appendix A

## Training Observables for $K_S^0$ Classifier

Figure A-1: The distribution of input variables in signal MC for KsFinder. The red is the from fake  $K_S^0$  and the blue is from true  $K_S^0$





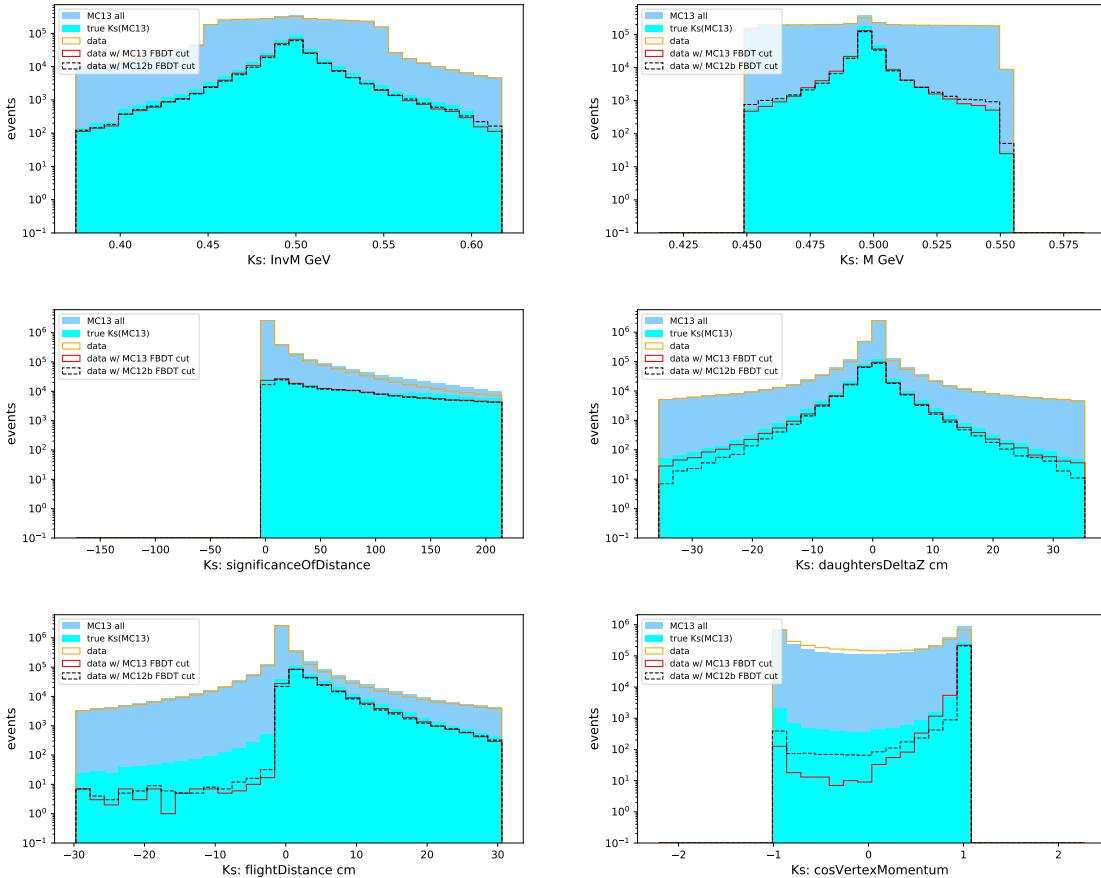


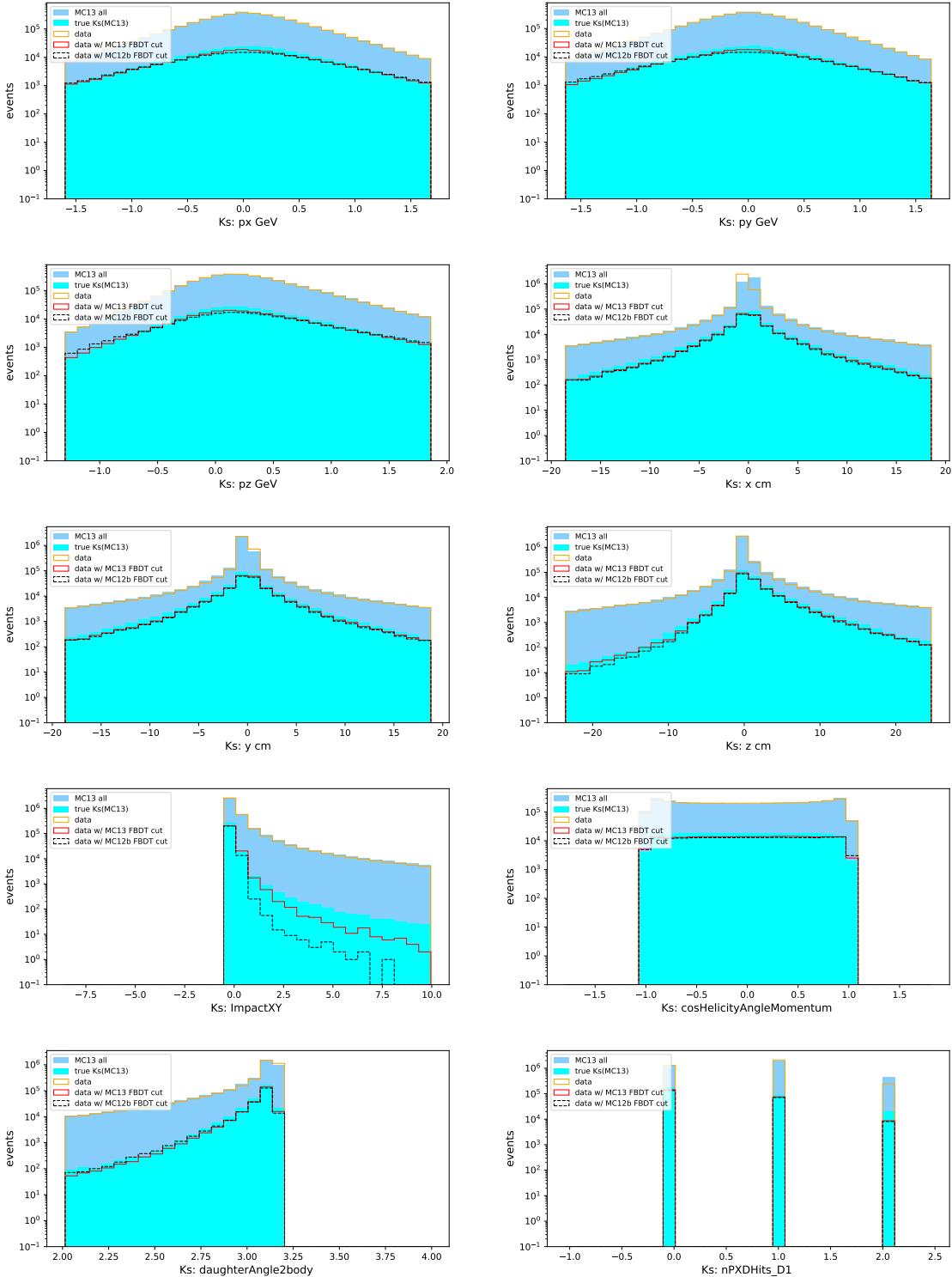


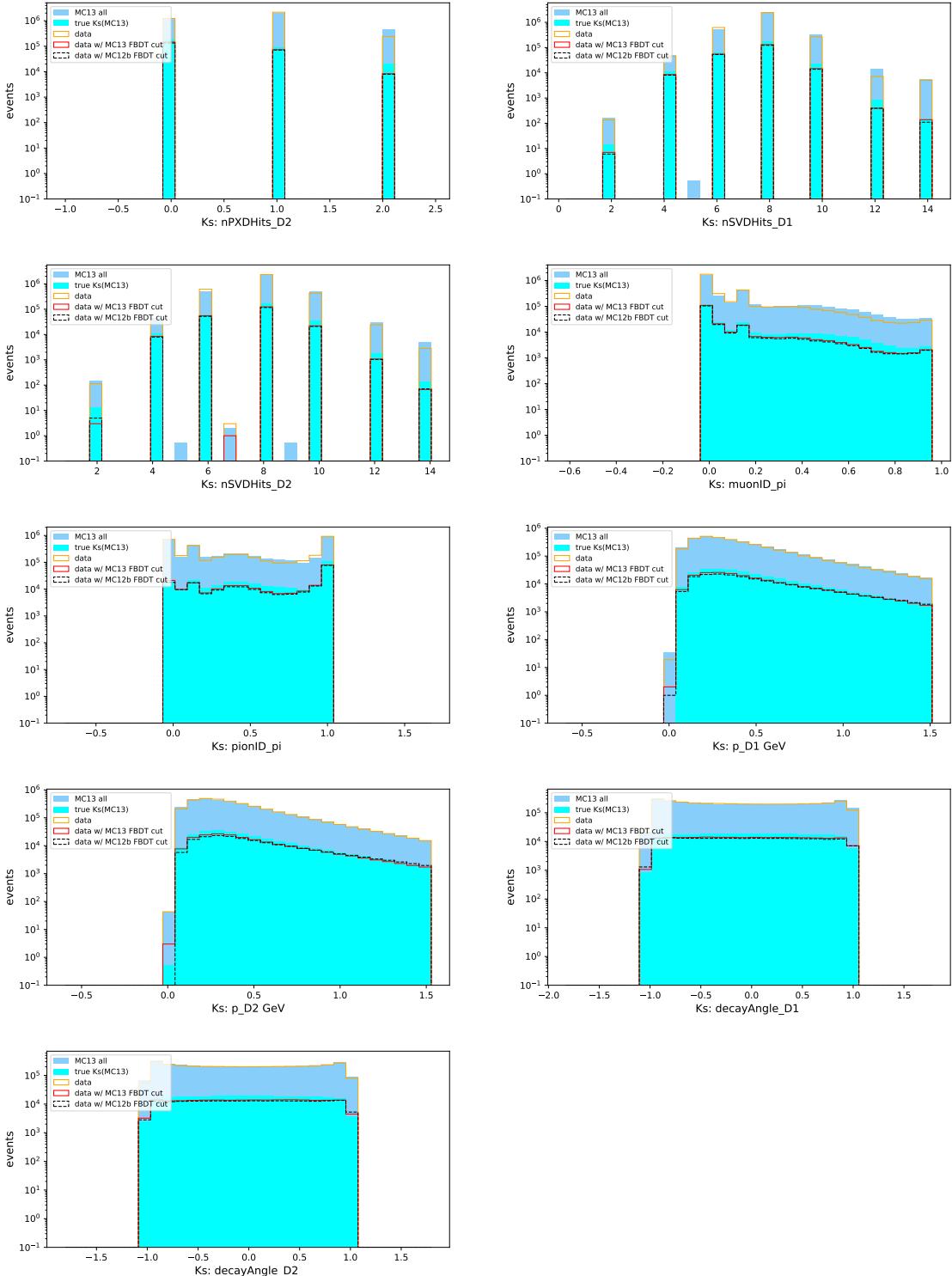
## Appendix B

# Data Validation Plots for $K_S^0$

Figure B-1: The distribution of variables used in KsFinder classification. The variables abbreviation can be referenced from Table III. The yellow solid lines are  $K_S^0$  from data without KsFinder, and red solid lines are  $K_S^0$  with KsFinder using MC13a for training. The black dashed lines are  $K_S^0$  with KsFinder using MC12b (run-dependent). Blue histogram is from MC13a without KsFinder and cyan histogram is from MC13a with KsFinder.







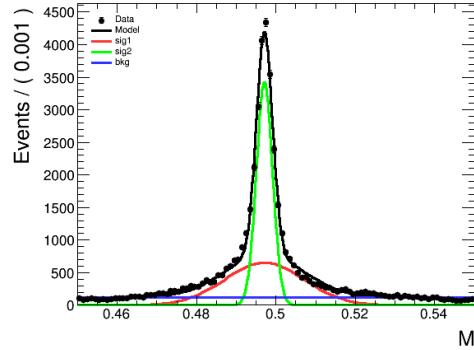


# Appendix C

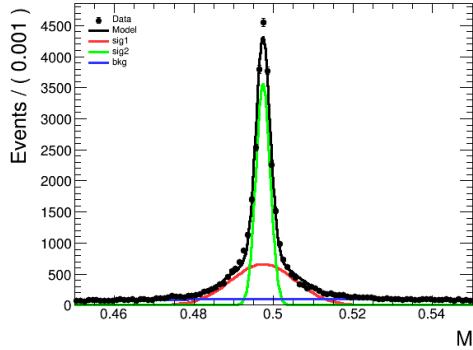
## $K_S^0$ mass fit with varied cut value

Figure C-1: Double-gaussian shape and 1st-poly are used for fitting signal and background  $K_S^0$  invariant mass under varies KsFinder cut values respectively. The data and MC are separately fitted to compared the yield in each cut, which defines the correction.

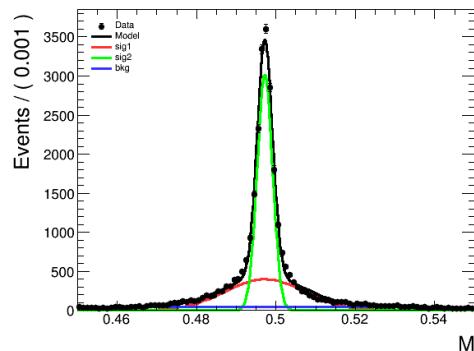
a) Data, cut=0.2



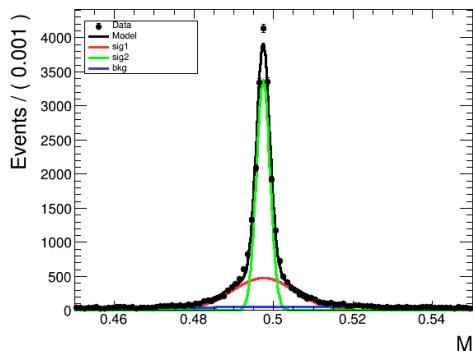
b) MC, cut=0.2



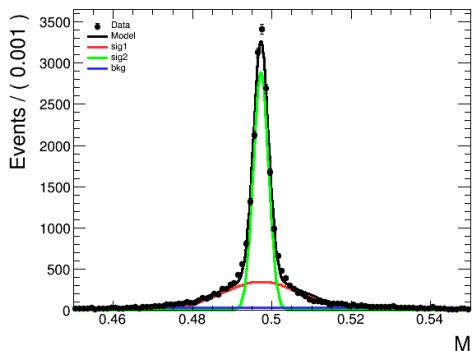
c) Data, cut=0.4



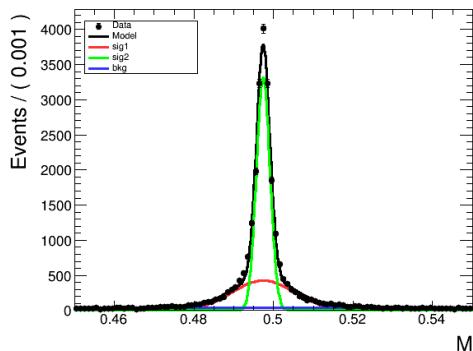
d) MC, cut=0.4



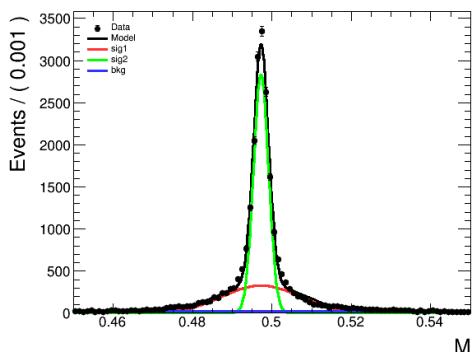
e) Data, cut=0.5



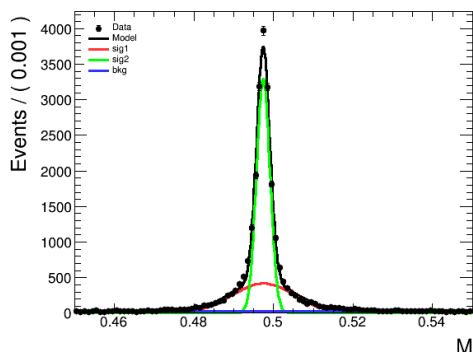
f) MC, cut=0.5



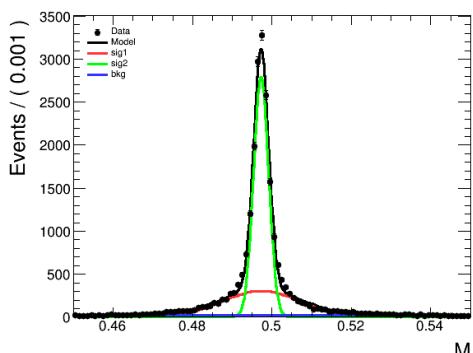
g) Data, cut=0.55



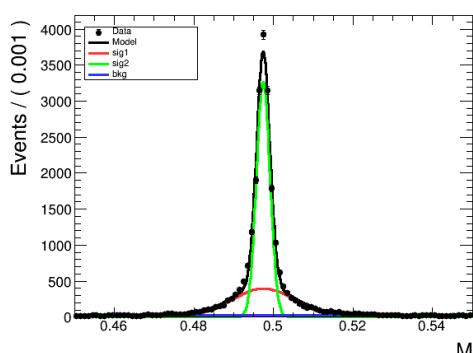
h) MC, cut=0.55



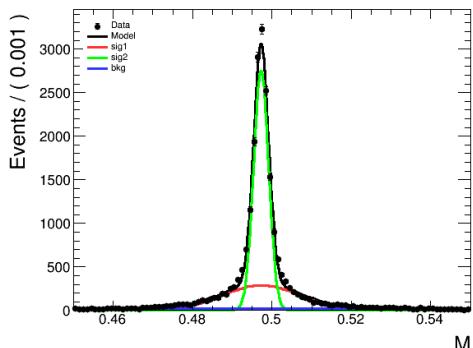
i) Data, cut=0.6



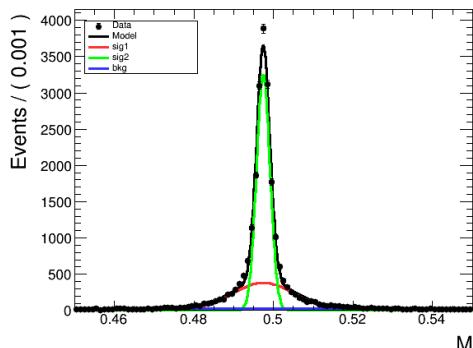
j) MC, cut=0.6



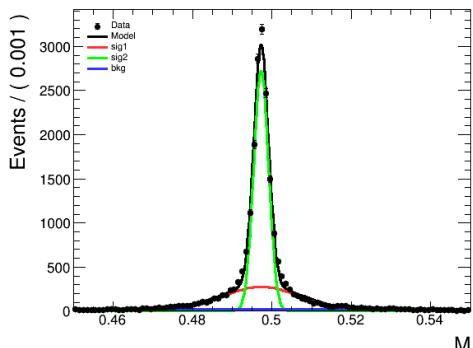
k) Data, cut=0.65



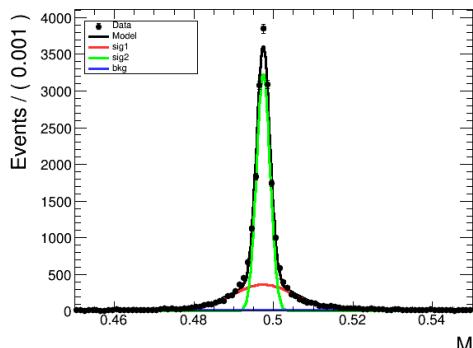
l) MC, cut=0.65



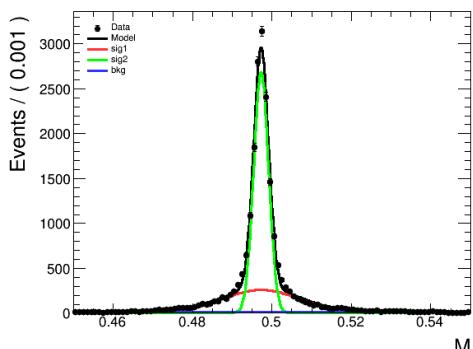
m) Data, cut=0.7



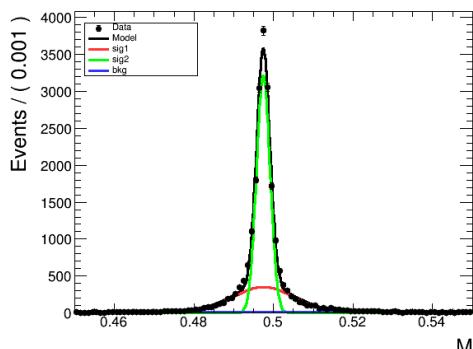
n) MC, cut=0.7



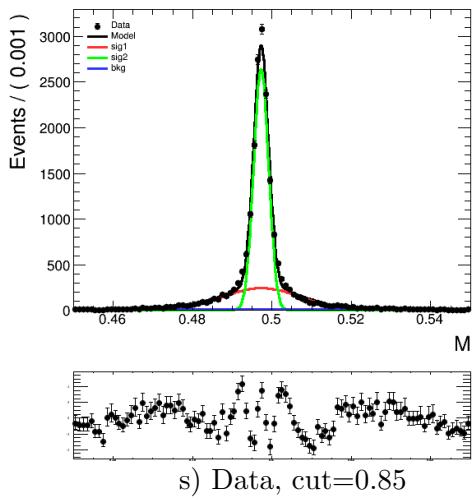
o) Data, cut=0.75



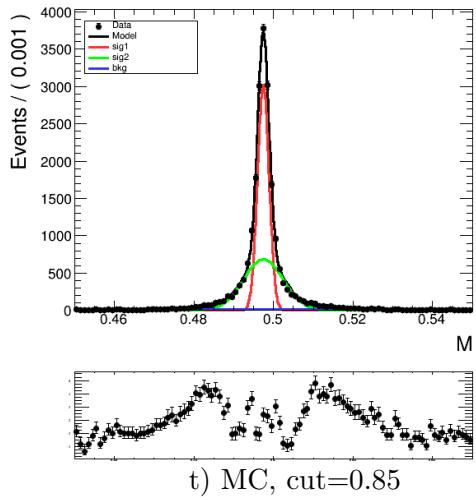
p) MC, cut=0.75



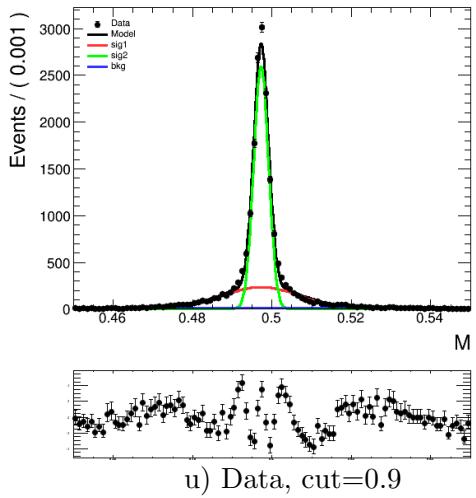
q) Data, cut=0.8



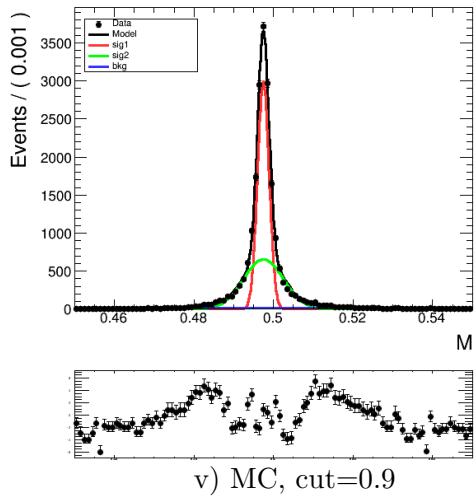
r) MC, cut=0.8



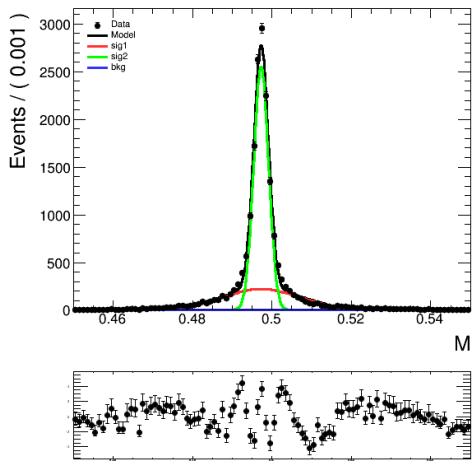
s) Data, cut=0.85



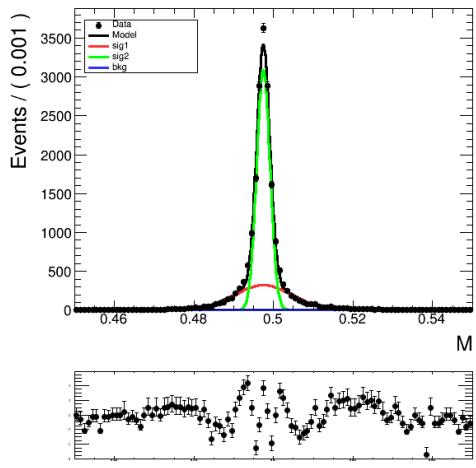
t) MC, cut=0.85



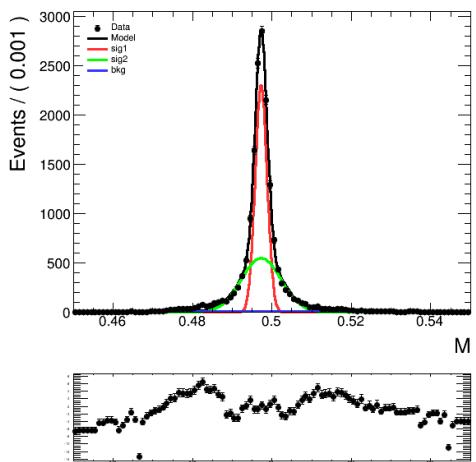
u) Data, cut=0.9



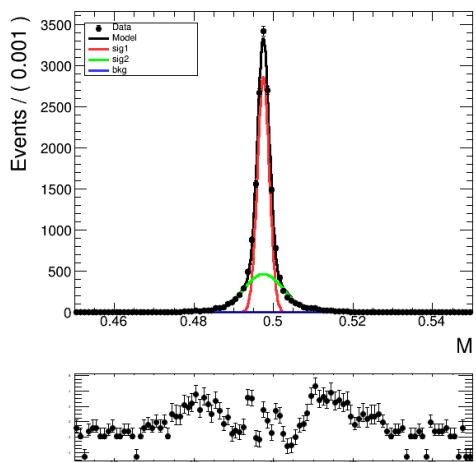
v) MC, cut=0.9



w) Data, cut=0.95



x) MC, cut=0.95



## Appendix D

# $2K_S^0$ invariant mass distribution

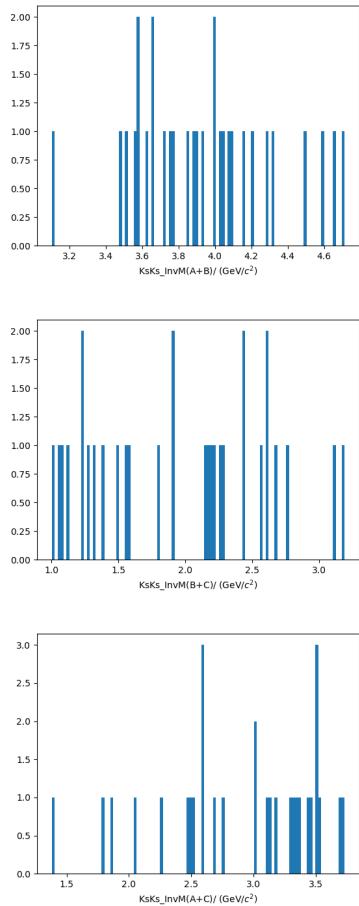


Figure D-1: Experiment data in signal region

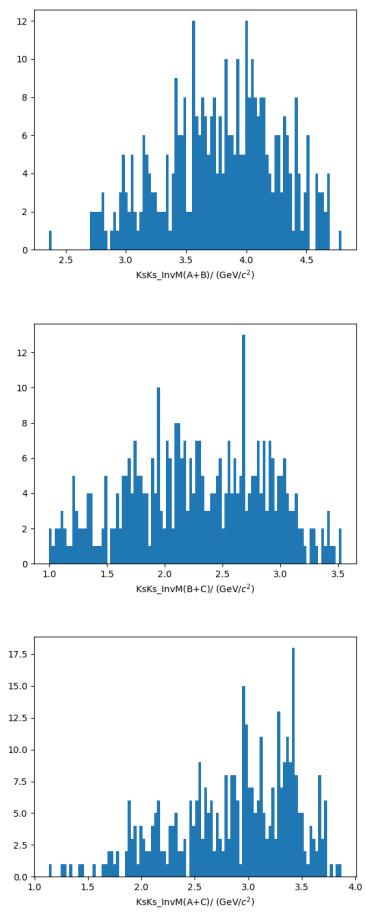


Figure D-2: Generic MC in signal region.



# 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](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](https://commons.wikimedia.org/wiki/File:ParityViolation_principle_Wu_experiment_(English).jpg).
- [5] James H Christenson, Jeremiah W Cronin, Val L Fitch, and René Turlay. Evidence for the  $2\pi$  decay of the  $K^0$  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](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  $B^0 \rightarrow J/\psi K_S^0$  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 cp violation in  $b^0 \rightarrow \eta' k^0$  decays and improved measurements of cp asymmetries in  $b^0 \rightarrow \varphi k^0$ ,  $k_s k_s k_s$  and  $b^0 \rightarrow j/\psi k^0$  decays. *Physical review letters*, 98(3):031802, 2007.
- [17] Tim Gershon and Masashi Hazumi. Time-dependent cp violation in  $b^0 \rightarrow p_0 p_0 x_0$  decays. *Physics Letters B*, 596(3-4):163–172, 2004.
- [18] Martin Beneke. Corrections to  $\sin(2\beta)$  from cp asymmetries in  $b^0 \rightarrow (\pi^0, \rho^0, \eta, \eta', \omega, \phi) k^0$  decays. *Physics Letters B*, 620(3-4):143–150, 2005.
- [19] KH Kang, H Park, T Higuchi, K Miyabayashi, K Sumisawa, I Adachi, JK Ahn, H Aihara, S Al Said, DM Asner, et al. Measurement of time-dependent cp violation parameters in  $b^0 \rightarrow k_s k_s k_s$  decays at belle. *arXiv preprint arXiv:2011.00793*, 2020.
- [20] J. P. Lees et al. Amplitude analysis and measurement of the time-dependent CP asymmetry of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  decays. *Phys. Rev. D*, 85:054023, 2012. doi:10.1103/PhysRevD.85.054023.
- [21] T. Abe et al. Belle II Technical Design Report. 2010. arXiv:1011.0352.
- [22] 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.
- [23] 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.

- [24] Jubin Mitra, Shuaib A Khan, Manoel Barros Marin, Jean-Pierre Cachemiche, Erno David, Frédéric Hachon, Frédéric Rethore, Tivadar Kiss, Sophie Baron, Alex Kluge, et al. Gbt link testing and performance measurement on pcie40 and amc40 custom design fpga boards. *Journal of Instrumentation*, 11(03):C03039, 2016.
- [25] URL <https://dirac.readthedocs.io/en/latest/DeveloperGuide/Overview/>.
- [26] 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.
- [27] URL <https://root.cern/>.
- [28] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The belle ii core software. *Computing and Software for Big Science*, 3(1), Nov 2018. ISSN 2510-2044. doi:10.1007/s41781-018-0017-9. URL <http://dx.doi.org/10.1007/s41781-018-0017-9>.
- [29] URL <https://evtgen.hepforge.org/doc/models.html>.
- [30] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G 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.
- [31] URL <https://pdglive.lbl.gov/Particle.action?init=0&node=S042&home=MXXX045>.
- [32] URL <https://confluence.desy.de/display/BI/Phase+3+luminosity+monitoring>.
- [33] J-F Krohn, F Tenchini, P Urquijo, F Abudinén, S Cunliffe, T Ferber, M Gelb, J Gemmler, P Goldenzweig, T Keck, et al. Global decay chain vertex fitting at belle ii. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 976:164269, 2020.
- [34] M Feindt and U Kerzel. The neurobayes neural network package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):190–194, 2006.
- [35] Thomas Keck. Fastbdt: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification. *arXiv preprint arXiv:1609.06119*, 2016. URL <https://arxiv.org/pdf/1609.06119>.
- [36] Jan Therhaag and TMVA Core Developer Team. Tmva-toolkit for multivariate data analysis. In *AIP Conference Proceedings*, volume 1504, pages 1013–1016. American Institute of Physics, 2012.

- [37] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [38] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [39] L Breiman J Friedman RA Olshen and CJ Stone. Classification and regression trees. statistics/probability series, 1984.
- [40] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.
- [41] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001. URL [https://projecteuclid.org/download/pdf\\_1/euclid-aos/1013203451](https://projecteuclid.org/download/pdf_1/euclid-aos/1013203451).
- [42] Tenchini, Francesco and Krohn, Jo-Frederik. Decay chain reconstruction at belle ii. *EPJ Web Conf.*, 214:06023, 2019. doi:10.1051/epjconf/201921406023. URL <https://doi.org/10.1051/epjconf/201921406023>.
- [43] 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>.
- [44] *BELLE-II-NOTE-TE-2019-010* (2020). URL [https://docs.belle2.org/record/1384/files/belle2-note-resol\\_belle.pdf](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](https://docs.belle2.org/record/1905/files/BELLE2-NOTE-PH-2020-013_v3.2.pdf?version=1).