博士論文

Doctoral Dissertation

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

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

令和2年12月博士(理学)申請

A Dissertation Submitted for the Degree of Doctor of Philosophy

December 22nd, 2020

東京大学大学院理学系研究科物理学専攻

Department of Physics, Graduate School of Science,

The University of Tokyo

万 琨

Wan Kun

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

by

Kun Wan

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

Abstract

Belle II experiment is a next-generation B-factory experiment that is aimed to search for New Physics. Most of data will be collected at the $\Upsilon(4S)$ resonance using SuperKEKB facility. It's designed at luminosity of 8×10^{35} cm^{-2}s^{-1} which is 40 times higher than its predecessor KEKB.

The thesis is based on the time dependent CP violation study of $B^0 \to K_S^0 K_S^0 K_S^0$ decay. The purpose is to precisely measure the CP parameters $\mathcal{S}$ and $\mathcal{A}$ in penguin-dominated $b \to s$ transition. It's sensitive to New Physics effects and quite interesting compared to other modes with tree-level contamination. Any undisputed deviation on CP parameters could be a signal beyond the Standard Model. Such precise measurements mainly require clean signal extraction, B^0 vertex reconstruction, flavor tagging and proper decay time resolution modeling. This thesis covers the development and optimization of analysis tools on the four aspects above. The blind fit and toy MC study are also included before using data, which show a reasonably good consistence in CP parameter measurements. By using data from Belle II 2019 and 2020 (Spring and Summer) operation at about 62.8 fb^{-1} integral luminosity, the measurement results of $\mathcal{S}$ and $\mathcal{A}$ are: $\mathcal{S} = -\sin(2\phi_1) = -0.82 \pm 0.85$ (stat) $\pm$ 0.07 (syst) and $\mathcal{A} = -0.21 \pm 0.28$ (stat) $\pm$ 0.06 (syst) are obtained. The result is dominated by statistical uncertainty and currently consistent with the Standard Model and the previous results in Babar and Belle.

Thesis Supervisor: Hiroaki Aihara
Title: Professor

Contents

Chapter 1

Introduction

1.1 The Standard Model

In the classical view of our physical world, even from the very early phase of human's history of understanding the universe, people have been convinced that all matter is made of some fundamental elements, and studying the elementary structure of the matter along with the interaction between those elements help us better understand the world and innovates the new technologies that deeply changes human society. Compared to the rather long history we start to think about question : what's the matter made? We only approached a mostly correct yet not perfectly precise answer of this old puzzle in the recent decades. Thanks to a few generation marvelously talented physicists' efforts, the Standard Model (SM) was built in the late 70th of 20th century, which describes the matter and interactions using an incredibly nice-looking table that contains a bunch of fundamental particles in two categories - fermions and bosons.

The fermions are the ones that assemble the matter and the bosons are for mediating the force between the fermions. They also have an essential difference in its spin number, which presents the different statistics rules they have to follow when describing their field functions.

In the fermion family, there are six leptons, classified based on their charge (Q), electron number (L_e), muon number (L_μ) and tau number (L_τ). Similarly, there are also six "flavors" of quarks, which are classified based on their charge, strangeness (S), charm(C), beauty(B) and truth (T). The last two can also be called as "bottom" and "top" number. Unlike leptons, quarks are all fractional charged, which naturally divides them into two sides. All up-side quarks have $+\frac{2}{3}$ unit charge and all down-side quarks have $-\frac{1}{3}$. Besides, all quarks are colored, meaning each one of them has an inner quantum number presenting 3 kinds of colors. One thing obvious is that all these fermions have its own antimatter correspondences, with all these quantum number reverse-signed but leaving mass identical valued. This comes to a fact that there are 36 different particles of them in total. According to the leptons and quarks flavors, charge and lepton numbers, they nicely fit in 3 generations, as Figure 1-1 shows.

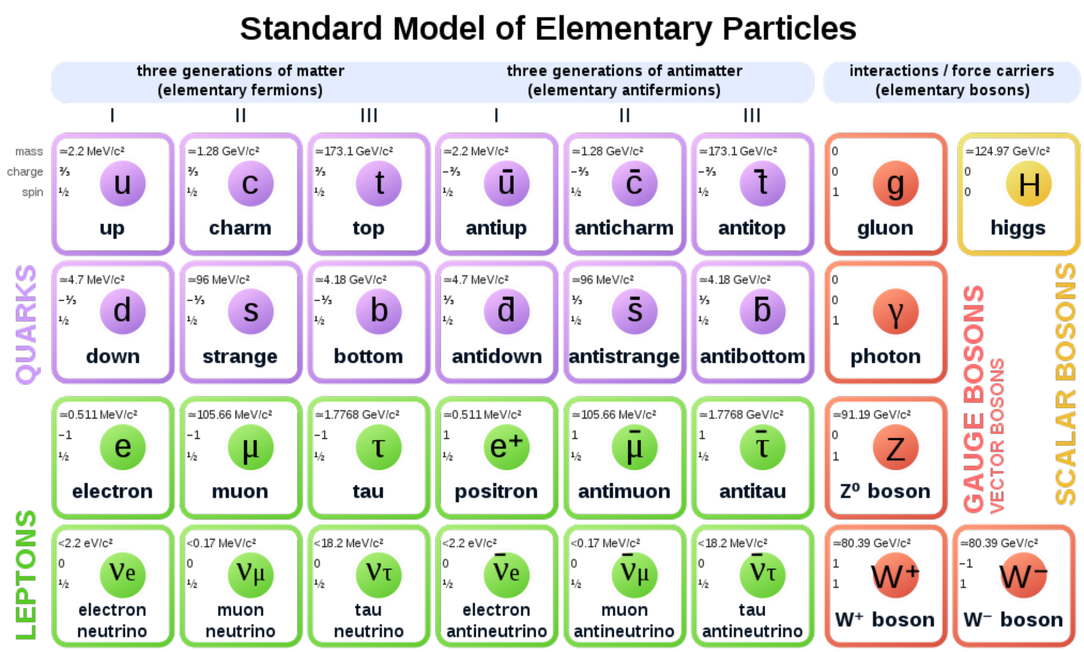


Figure 1-1: Elementary particles in the Standard Model.[1]

The SM not only accounts for the matter composition, but also explains the interaction of fermions and bosons in a picture where they interact through exchange of certain force carriers which are bosons as well. More specifically, each set of these carriers mediate one type of fundamental interaction. The force in which most of classical objects interact is electromagnetic interaction and it's mediated by photons.

The force that plays a role in the β decay is weak interaction, of which the mediators are W and Z bosons. When a proton and a neutron interact within a nuclei transforming into each other, the actual force carriers are 8 types of gluons (or π mesons between nucleus), and it's called strong force. Last but not least, in the quantum field theory, all of these particles has zero mass if there's no spontaneous symmetry breaking by introducing one another boson - Higgs boson, in the Lagrangian formalism of all interactions. With this being said, the "mass" is taken as the weight of how strong all these particles interact with Higgs field.

This complete set of elementary particles describes the whole picture of the Standard Model. It leads to a comprehensive and symmetric theory of fundamentals in particle dynamics. However, the SM prediction does not perfectly match with experimental observations. Since the day theory was built, generations of particle physics experiments have been searching for evidences beyond the SM, as known as New Physics (NP). New Physics is expected to unfold a more profound truth of nature which hopefully explains these observed mismatches. The studies on these fields naturally draw a large attention from modern physicists, focusing on discovering and explaining the mismatches between the SM predictions and experiments. Among these research fields, the studies of symmetry violations (or called as asymmetries) plays an important role. The studies of symmetries was once the driving force for physicists when building the modern theory about particle physics. It is no wonder that now the violation of symmetries, which physicists didn't expect to happen , has become the cutting edge research topics for New Physics.

1.2 Symmetry Violation

Symmetries have been one of the focuses in modern physics research since physicists found the internal link between symmetries and conservation laws. The invariance of physical system under infinitesimal transformation is regarded as "continuous symmetry". For instance, any infinitesimal shift on space and time holds physics law in the same form, which implies the physical processes should hold the conservation law of momentum and energy. Such statement is stated as "Noether's theorem"[2]. Symmetries become a powerful tool in discovering physics laws due to this reason. When the known symmetries are found to be broken, it usually leads to a discovery of the new theory.

There are three types of discrete symmetric operations which play important roles in particle physics. Charge-conjugation C is the operation that turns particle to its anti-particle. Parity transformation P is the one that puts a negative sign before all the spatial related vector such as $\overrightarrow{r} \to -\overrightarrow{r}$, feeling like look at a physical process in a mirror. The time-reversing operation T is to reversely proceed a physical process backward time. Physicists are convinced that each of these three kinds of symmetric operations makes no change to any physical system. In 1950s, Lee and Yang [3] first questioned that parity symmetry might be broken in weak interactions. They offered a few possible ways to test it and then by Wu [4], an observation on the β decay of ^{60}Co was presented that the electrons emitted from ^{60}Co decay prefer to flying against the direction of nuclear spin that can be control by the external magmatic field, see Figure 1-2.

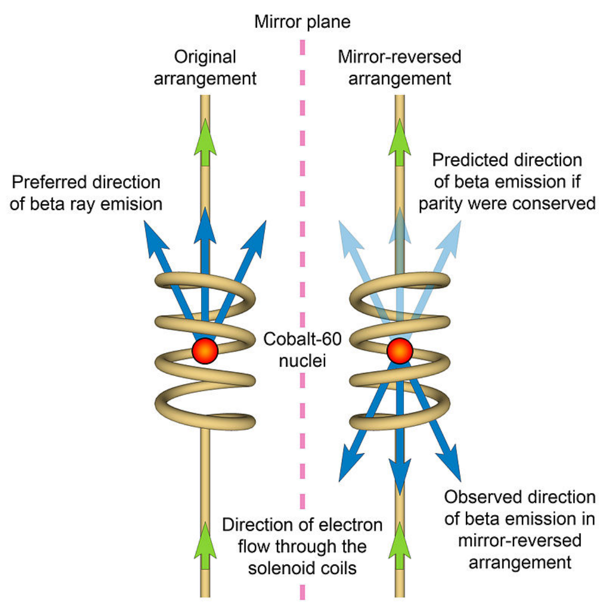


Figure 1-2: ^{60}Co decay violates the parity because of the unbalance of electron emissions.

Soon, the fact that coupling of weak interaction only involves either neutrinos (left-handed) or anti-neutrinos (right-handed) was realized. So C symmetry is also violated since no left-handed anti-neutrino was found which should be theocratically possible in weak interaction. But if one performs CP operation on such weak interaction involving neutrinos, then such process "seems" to be equally possible, indicating the conservation of CP in weak interaction for a moment.

The first evidence of CP violation was found in the Kaon system. Charged Kaon identification was called mystery of "$\tau - \theta$", because they yield different decay modes with different parity, but have identical mass between τ and θ. It turns out that they are both charged Kaon meson K^+, whose decay violates the conservation of CP. In the neutral Kaon system, Cronin and Fitch's experiment was the first experiment that proves the violation of CP. They measured the decay products at 57 foot of a neutral kaon beamline assuming all the particle in the beam should be long lifetime K_L, nearly no K_S. But 0.002% of them decay through the $K_L \rightarrow \pi^+\pi^-$ ($CP = 1$ in final states, yet K_L has $CP = -1$). Given that the expected distance to have 0.002% of K_S at about speed of light is no more than 1 meter, such deviation at 57 foot is an obvious evidence that CP conservation is also violated in the neutral Kaon system.

In 1973, Kobayashi and Maskawa introduced the third generation of quarks, and the mixing of flavor eigenstates and weak eigenstates is described by a 3×3 unitary matrix, which is called Cabibbo-Kobayashi-Maskawa (CKM) matrix[5]. The theory allows a free 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 Collaborations. 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 outstandingly clearness in theoratical prediction and has relatively large branching fraction, thus it's called the "golden mode". In 2008, Kobayashi and Maskawa were rewarded the Nobel Prize to highly value their contribution in the discovery of the mechanism of the CP violation. The Belle experiment was regarded as a great success for validating the theory. Later in 2010, the upgrade of Belle, Belle II and the upgrade of KEK accelerator, SuperKEKB, were both approved by Japan to further push the understanding of CP violation in

b sector along with other excitingly interesting topics in New Phsyics.

1.3 CKM mechanism

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

Equation 1.1 is the Higgs potential doublets. H's value is 174 GeV as the expected Higgs potential for vacuum and ϕ and χ are the pseduo-Goldstone fields which are appearing when introducing Higgs field ϕ without breaking the gauge symmetry. The idea of introducing Higgs is to naturally solve the mass origin of particles under the gauge symmetry in the SM, as Equation 1.2 presents.

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

Here the primed fields stand for the weak eigenstates of quarks. ϵ is a $2\times$ 2 matrix. $Q^\dagger$ is the left-handed doublets that stand for weak eigenstates of up and down types quarks.

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

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

Yukawa matrix is an arbitrary 3×3 complex matrix $Y^{u,d}$ which give the rise of two types of massive quark field $M^{u,d} = Y^{u,d}\nu$. The choosing the weak eigenstates makes these matrices un-diagnosed. So the "rotation" of weak eigenstates to diagnose them is done by:

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

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

In Equation 1.5 and 1.6, $S^{u,d}_{L,R}$ are all unitary matrices since they are actually generated by the normalized eigenstate vectors of the Yukawa matrix, which means ${S^{u,d}_{L,R}}^\dagger S^{u,d}_{L,R} = I$. The mass items M_q, arised after the "rotation" can be presented as:

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

where $q = (q_L + q_R)$ is four-component Dirac field, and $q^\dagger_L q_L = q^\dagger_R q_R = 0$. The matrix $S^{u,d}_{L,R}$ are contributing to the weak interactions as Equation 1.8 shows.

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

The Lagrangian hereby clearly declair the transition of different charged quarks through the coupling of charged current $W^{+/-}$. Interestingly, such coupling only applies for the left-handed quarks. For example, a left-handed c quark only transits to left-handed s quark by giving out a W boson. If the Parity conjugation is applied to the Lagrangian, the left-handed quarks becomes right-handed. The same thing happens if the Charge conjugation is applied, cause it changes the chirality of the 4 components quark fields and charge sign at the same time. This means the weak interaction will not conserve the Parity and the Charge symmetries. However, if the CP conjugation is applied, all left-handed quark fields' chirality are unchanged and only the conjugation of quark fields are made, as Equation 1.9 shows.

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

Comparing Equation 1.9 and 1.8, the CP symmetry requires:

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

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

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

On the one hand, if the CKM matrix is real, CP will be conserved in the weak interaction in the SM. On the other hand, from Equation 1.10, it's still possible to make Lagrangian unchanged even if V_{CKM} is not real. 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:

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

Assuming the complex phase of CKM kj-th element is θ_{kj}, it's obviously required that:

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

Historically, before the 3rd generation of quarks is discovered, k, j can only be either 1 or 2, and V_{CKM} is 2 dimensional unitary matrix. With two generations of quarks, there are 8 real parameters in CKM matrix, 4 describe amplitudes and 4 describe complex phases. Unitary condition $V_{CKM} \dagger V_{CKM} = 1$ gives 4 equations, containing 2 real equations and 2 complex equations. The 2 complex equations are identical so only 3 independent equations arise. 2 complex equations leads to 2 equations about phases plus 2 equations are about amplitude. Based on this, the degree of freedom of phases on 2 dimensional CKM matrix is $4 - 2 = 2$. Thus it's always possible to make CKM real by properly render the non-physical phases. This means with only existence of 2 generations of quarks, there is no spontaneous explanation for CP violation in the SM. When Kobayashi and Maskawa realized that this is no longer true if 3rd generation of quarks exists, they presented CKM matrix as a 3×3 matrix. In such case, there will always be one irreducible complex phase parameter in the CKM matrix, which means CP symmetry is no longer conserved in the weak interactions.

The 3×3 unitary CKM matrix can be written as Equation 1.14.

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

It can be parameterized:

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

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

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

and the following relations are often used to simplify CKM matrix presentation:

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

By using Equation 1.17, CKM matrix is parameterized as:

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

Using the unitary condition, the following equation is obtained.

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

Plotting the relation in a complex 2D plane using Equation 1.19,

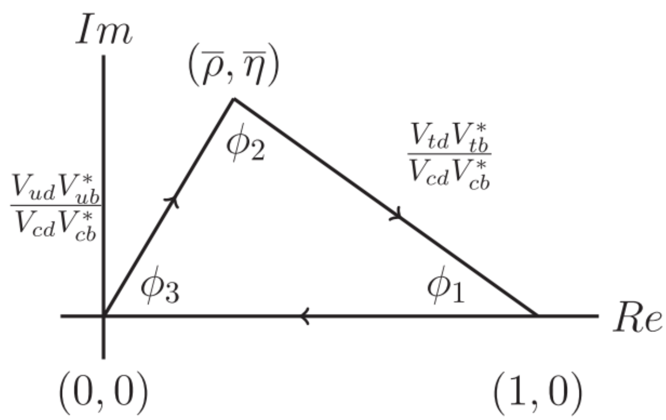


Figure 1-3: The unitary angles of CKM

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

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

These angles are obtained by drawing the $(\bar{\rho}, \bar{\eta})$ on the complex coordinates, and they are also well-known in the names as: $\phi_1 = \beta, \phi_2 = \alpha, \phi_3 = \gamma$. The results presenting the measurement of CKM angles or $(\bar{\rho}, \bar{\eta})$ in 2019 is shown in Figure 1-4.

The measurement of ϕ_1 and ϕ_2 are coming from the time-dependent CP violations (TDCPV). ϕ_1 in the tree-level dominated decays has been precisely measured thanks to the small hadronic uncertainties. FCNC process rises through the B_d^0–$\bar{B}_d^0$ mixing in

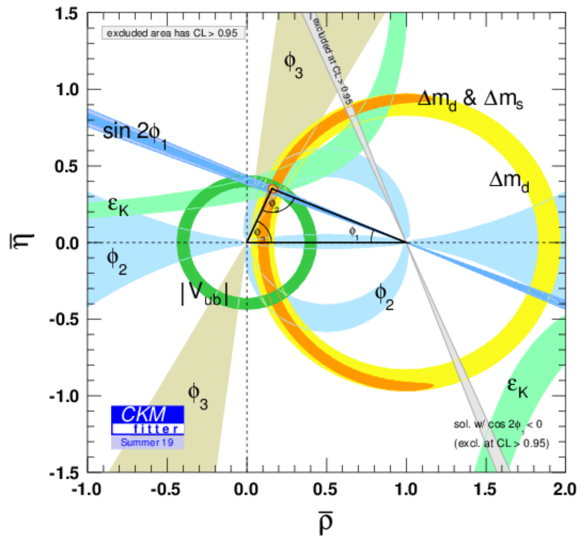


Figure 1-4: The CKM angles fit in the complex plane of $\bar{\rho} - \bar{\eta}$.[6]

box diagram, and it's believed that potential New Physics processes might contribute to the difference in between results of CKM angles measured from experiments. For example, angle ϕ_1 that is observed from tree-level dominated process like $B^0 \rightarrow J/\psi K_s^0$ could be different from penguin-dominated $b \rightarrow s$ transition. The Standard Model provides an approximated correction on how large the deviation might be, where it requires much precise measurement on each decay channel to validate the potential New Physics effect. The prospective large Belle II data and more precise time resolution performance will be much helpful to clear the tension in future.

1.4 Time Dependent CP violation

1.4.1 CP violation in neutral B system

The ϕ_1,ϕ_2 and ϕ_3 are essentially measuring the CKM CP violating phase since there's only one complex phase in CKM and it can be determined by these 3 angles. And $\phi_{1,2}$ are related to mostly in the time dependent decay rate difference. From Figure

1-3, one can obtain:

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

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

The general strategy to extract the $\phi_{1,2}$ is to measure the time-dependent CP violation. The mass eigenstates which are driving the propagation of B meson states in the mixing are: $|B\rangle_{H,L} = p|B\rangle \pm q|\overline{B}\rangle$, where H and L stand for the larger and smaller eigenvalues corresponded states. $|B\rangle$ and $|\overline{B}\rangle$ are the flavor eigenstates. The Hamiltonian matrix can be written using flavor eigenstates:

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

and the time evolution of mass eigenstates are defined as (using the notation of $B_{H,L}$ as physical states at $t = 0$):

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

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

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

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

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

and replacing $g_\pm(t) = \frac{1}{2}(e^{-im_H t - \Gamma_H/2t} \pm e^{-im_L t - \Gamma_L/2t})$, above two equations become:

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

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

Considering all the phases space of the decay from flavor eigenstates to final states $f(\bar{f})$ are included in the amplitudes $\mathcal{A}_f(\bar{\mathcal{A}}_f)$, one needs to expand the flavor eigenstates using such amplitude to have the differential decay rate $\Gamma(B \to f, t)$. From $B(t) \propto \mathcal{A}_f\psi_f + h.c$ and $(\bar{B}(t) \propto \bar{\mathcal{A}}_f\psi_{\bar{f}} + h.c)$:

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

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

where the parameter λ_f is

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

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

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

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

where the M_{12} and Γ_{12} stands for the contribution of non-diagnosed term in the Hamiltonian matirx. $\Delta M = M_H - M_L$ and $\Delta\Gamma = \Gamma_H - \Gamma_L$ are the difference of mass and decay width for two mass eigenstates. It's obvious that if $|A_f| \neq |\bar{A}_{\bar{f}}|$, direct CP violation will occur.

The time-dependent decay rate difference is defined as Equation 1.36:

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

,where

$$\mathcal{S} = \frac{2Im(\lambda_f)}{1 + |\lambda_f|^2}; \mathcal{A} = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}; A^f_{\Delta\Gamma} = -\frac{2Re(\lambda_f)}{1 + |\lambda_f|^2} \tag{1.37}$$

The origin of time is set to the flavor tagged moment. The time-dependent CP violation parameters are $\mathcal{S}$ and $\mathcal{A}$, which are determined by λ_f.

1.4.2 ϕ_1 from $B^0 \to J/\psi K^0_S$

If final states are CP definite, the amplitude then equals to: $\mathcal{A}_f \equiv \langle f|H|B\rangle$ and $\bar{\mathcal{A}_f} \equiv \langle f|H|\bar{B}\rangle$. In $B^0_d - \bar{B}^0_d$, q/p can be treated as $e^{i\phi_d}$ safely as pure phase term. This relative phase accounts the transition from b to up-type quarks to 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 the golden mode $B^0 \to J/\psi K^0_S$, considering $\Delta\Gamma$ can be treated as zero safely in the Standard Model[7] in this case, Equation 1.36 can be reduced to:

$$A_{CP}(t) = \mathcal{S}\text{sin}(\Delta M t) - \mathcal{A}\text{cos}(\Delta M t) \tag{1.38}$$

For decay amplitude, which recieve contribution from tree-level and loop-level processes shown in Figure **??** ,

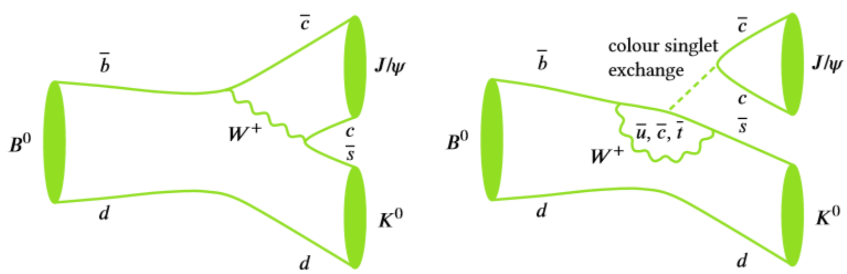


Figure 1-5: The dominated tree-level (left) and the suppressed loop-level (right) of $B \to J/\psi K^0$, in which K^0 particles are detected K^0_S. [8]

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 surppressed in the Standard Model. Defining η_f as the CP eigenvalue for CP definite final states,

Given $\eta_f = 1$ and $|\lambda_f| = 1$ in $B^0 \to J/\psi K^0_S$, from 1.37, CP parameters can be presented as:

$$\mathcal{S} = Im(\lambda_f) = -sin(\phi_d)\eta_f = -sin(2\phi_1);\ \mathcal{A} = 0 \tag{1.39}$$

From Equation 1.39 , ϕ_1 can be accessed pretty precisely in the measurement of time-dependent CP violation in $B^0 \to J/\psi K^0_S$, whose branching fraction is relatively high.

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

Compared to $B^0 \to J/\psi K^0_S$ channel, the measurement of $\mathcal{S}$ and $\mathcal{A}$ from penguin-dominated channels through $b \to q\bar{q}s$ where q is u, d, s can be different due to the varied tree-to-penguin amplitude ratio. Furthermore, they are quite interesting for the following reasons[9]. First, they can probe $B^0 - \bar{B}^0$ through different short-distance vertices than the tree-level dominated decays. Second, the relatively small penguin amplitude may be more sensitive to the NP effects than tree-dominated modes. Last but not least, they comprise a large number of different final states, which can help disentangling non-perturbation long-distance physics from short-distance information, such as ϕ_1 or NP contributions to the weak Hamiltonian.

Considering possible New Physics contribution as A^{NP}_f, the decay amplitude can be rendered as:

$$\mathcal{A}_f = \lambda^s_u T_f + \lambda^s_c P_f + A^{NP}_f \tag{1.40}$$

where T_f and P_f are tree-level and penguin-level amplitude. The coefficients are determined from CKM matrix elements: $\lambda^q_i \equiv V^*_{ib} V_{iq}$. Note that compared to the $B^0 \to J/\psi K^0_S$, the tree level amplitude T_f is suppressed and penguin amplitude P_f is dominated in $b \to q\bar{q}s$. It is worth noting that T_f contains both tree-level W exchange, QCD and electroweak penguin contributions. These carry the combination of CKM matrix elements $\lambda^s_t = V_{ts}V^*_{tb} = -(1 + \epsilon_{uc})\lambda^s_c$ where $\epsilon_{uc} \equiv \lambda^s_u/\lambda^s_c = \mathcal{O}(\lambda^2)$. In the Standard Model with neglected ϵ, $b \to q\bar{q}s$ modes are pure penguin with the same weak phase as $B^0 \to J/\psi K^0_S$ has. Thus, direct CP violation vanishes and time-dependent CP violation reflects $\mathcal{S}$ in the same way as $B^0 \to J/\psi K^0_S$ does.

Departures from this limit, non-neglected tree amplitude T_f (often called "tree pollution"), as well as possible NP effects, could give different results on ϕ_1. Introducing the tree-penguin ratio $r^T_f = T_f/P_f$ and NP-to-SM ratio $r^{NP}_f = \mathcal{A}^{NP}_f/(\lambda^s_c P_f)$, the following statements usually used[9]:

- Branching ratios are affected at $\mathcal{O}(|\epsilon_{uc} r^T_f|, |r^{NP}_f|)$

- Direct CPV in the SM are of $\mathcal{O}(\epsilon_{uc}\text{Im}(r_f^T))$
- $-n_f^{CP}\mathcal{S}_f = \sin(2\phi_1) + \Delta\mathcal{S}_f$, where $\Delta\mathcal{S}_f = 2\cos 2\phi_1 \sin\phi_3 |\epsilon_{uc}| \text{Re}(r_f^t) + \Delta\mathcal{S}_f^{NP}$

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

Since for $B^0 \to K_S^0 K_S^0 K_S^0$ decay, there is no appearance of top side flavor in the final states so the worries about tree-level contribution is absent. The possible contribution of the NP may lead to the deviation of ϕ_1 measurement.

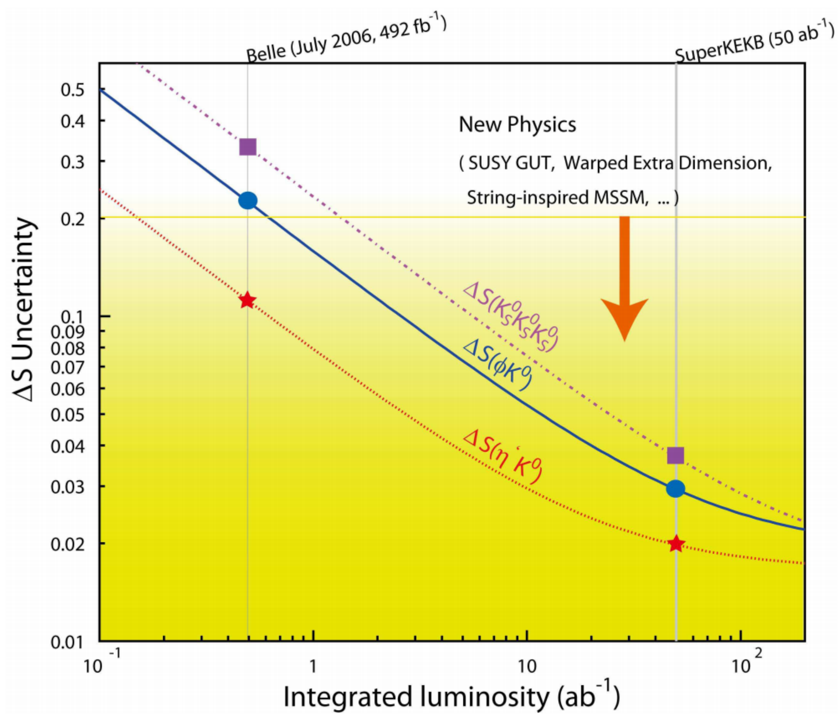


Figure 1-6: Expected precision of ΔS_f as the function of integrated luminosity.

It's clear the clean measurement of ϕ_1 from the S_f is highly depends on the control of r_f^t in Eq(1.49). If the small correction regarding this ΔS_f is estimated as proportional to the small strong phase difference in r_f^t and also $\mathcal{O}(\lambda^2)$, the average value of ΔS is $0.03\pm^{0.01}_{0.04}$[11]. So the study of decay $b \to \bar{s}ss$ can probe the difference using the time-dependent decay rate to extract $sin(2\phi_1)$. If the NP effects can not be explained by the small correction ΔS_f, this will be a clear sign of BSM. Here list a set of penguin-dominated mode that are sensitive to BSM on the expected results of

ΔS_f, along with their improvement with the accumulation of dataset available [11].

Table 1.1: ΔS_f expected sensitivity with integral luminosity

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

As the main decay process that have been studied in this thesis, $B_d^0 \to K_S^0 K_S^0 K_S^0$ propagates through $b \to \bar{s}ss$ as well. There are more than one process that will yield the same final states ($3K_S^0$). The difference here is that $3K_S^0$ final states can be produced not only from the ss intermediate state like $f_0(980)$ through penguin mode, but also viable from charm decay (see Fig(1-10)), which $\bar{c}c$ comes from tree level transition. The contribution from tree level process is estimated to be small but must be aware of using Dalitz analysis in future. Similarly, the $\bar{s}s$ intermediate decay, such as $B^0 \to f_0(980)K_S^0$, would also give the same CP-eigenvalue in final states, which is also considered as signal because of CP-even), while $b \to c \to s$ tree level is considered as background that yields CP-odd final states. The previous result in this channel from Belle full dataset [12] is: $\mathcal{S}$ = -sin2ϕ_1 = $-0.72 \pm 0.23(stats) \pm 0.05(syst)$ and $\mathcal{A} = 0.12 \pm 0.16(stats) \pm 0.05(syst)$. The earlier results from Babar[13] is: $\mathcal{S}$ = -sin2ϕ_1 = $-0.94^{+0.21}_{-0.24}(stats) \pm 0.06(syst)$. The results are mostly limited by the statistical uncertainty with minor difference. With the expected luminosity from Belle II in future, we are able to cut off the margin in between the results and the Standard Model. This thesis, as the first trial to perform the CP measurement on this channel for Belle II, is inevitably dominated by the low data size in the early operation, but will show the good capability and potential in the analysis tools and methods towards the promising future.

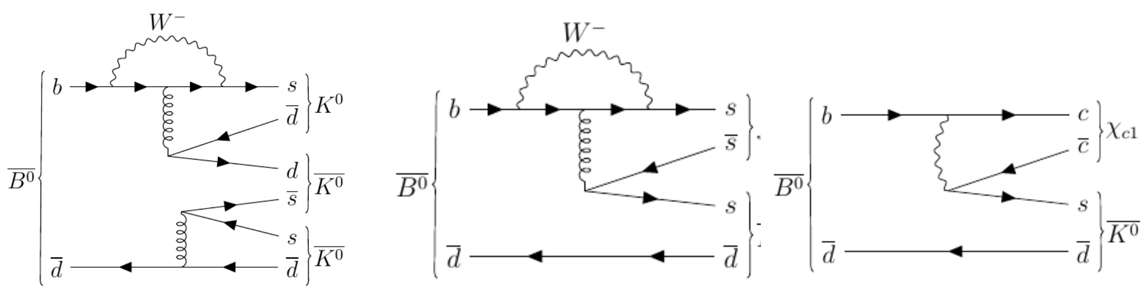


Figure 1-7: non-resonant signal (left),resonant signal (middle) and tree-level resonant backgrounds (right) $B_d^0 \to 3K_S^0$ diagrams

.

Chapter 2

Belle II Experiment

2.1 Belle II and SuperKEKB overview

The fundamental goals of Belle II experiment are to search for evidence of New Physics in the luminosity frontier, and to improve the precision of the measurement of the SM parameters, such as CP parameters.[9] It takes SuperKEKB accelerator as its particle collider at the center-of-mass energy in the region of υ resonances. Majority of the production will at the $\upsilon(4S)$ resonance that is slightly above the mass of two B meson.The electron and positron beams are designed at 7 GeV and 4 GeV respectively, with boost factor 0.28. This creates an environment for measuring time-dependent CP violation by displacing the decay vertices of B mesons pair in a measurable distance in boosted direction. SuperKEKB has a targeted luminosity at $8 \times 10^{35} cm^{-2} s^{-1}$, 40 times higher than its predecessor, KEKB at peak luminosity. The expected operation period will be around 8 years and over 5×10^{10} b, c, τ pairs will be produced. The facilities are located in KEK, Tsukuba City, around 70 km in the north of Tokyo, Japan. Some key parameters of SuperKEKB are listed in Table 2.1.

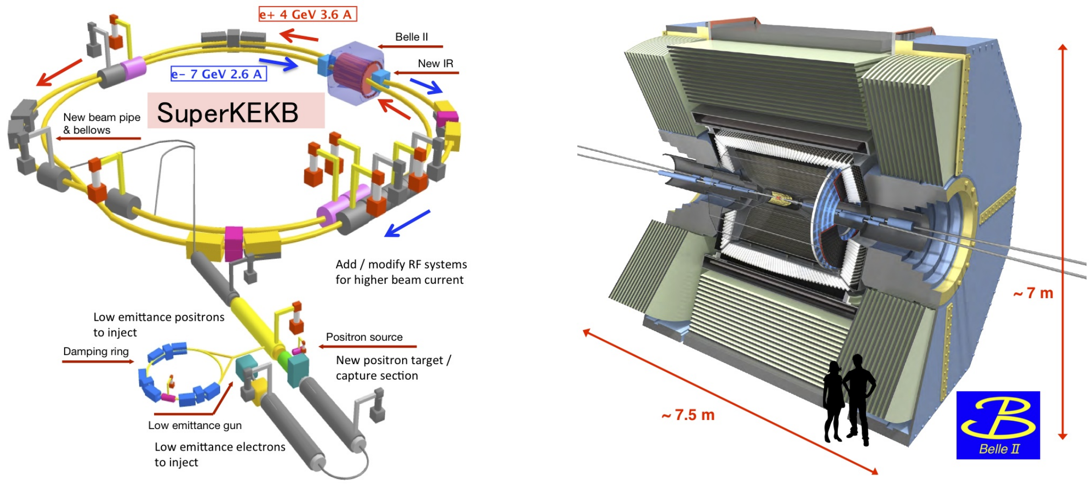


Figure 2-1: Overview of SuperKEKB and Belle II detector.

Table 2.1: SuperKEKB parameters for low energy (LER) and high energy (HER) rings.[9]

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

Belle II detector has a similar size as Belle so it fits in the previous shell, but all sub-detectors and components are either newly built or considerably upgraded. The advantage of SuperKEKB requires that Belle II has to be able to stably operate at 40 times higher events rates as well as 10 to 20 times higher beam background compared to Belle at its peak luminosity. This means mitigation of the effects caused by such high beam background is essential to the success of Belle II. Higher background level leads to high occupancy and radiation damage to the detectors at close range, along with more fake hits, pile-up noises in electromagnetic calorimeter and neutron-induced hits in muon detector. Data-acquisition system (DAQ) and trigger are also upgraded not only to adapted to higher luminosity but also for low multiplicity events sensitivity to support a broader search especially in dark sector. Overall, Belle II detector top view is shown as Figure 2-2, and expected performances are summarized as follows:

- vertex resolution of B mesons at $\sim 50\mu m$
- excellent reconstruction efficiency for charged tracks down to several 100 MeV and fairly good efficiency for charged tracks down to $\sim$ 50 MeV.
- excellent momentum resolution up to 8 GeV/c
- highly efficient particle identification to separate π, μ, e, $K^{+/-}$ and p^+ at full energy range of experiment.
- full cover of experimental acceptance solid angle.
- ultra fast and highly efficiency DAQ and trigger system to cope with large data quantities and fast triggering frequency.

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 covered in the following sections.

2.2 Vertex detector (VXD)

There are two components in VXD, the silicon based pixel detector (PXD) and silicon based vertex detector (SVD), which total 6 layers are placed in the inner-most region from interaction point (IP). As for PXD, two layers are placed at radii of $r = 14$ mm and $r = 22$ mm with DEPFET type pixel sensors respectively. And SVD sensors are made of "double-sided silicon strip sensors" (DSSD) with 4 layers at 39mm, 80mm, 104mm, 135mm. The geometry of VXD from the is shown in Fig(2-3).

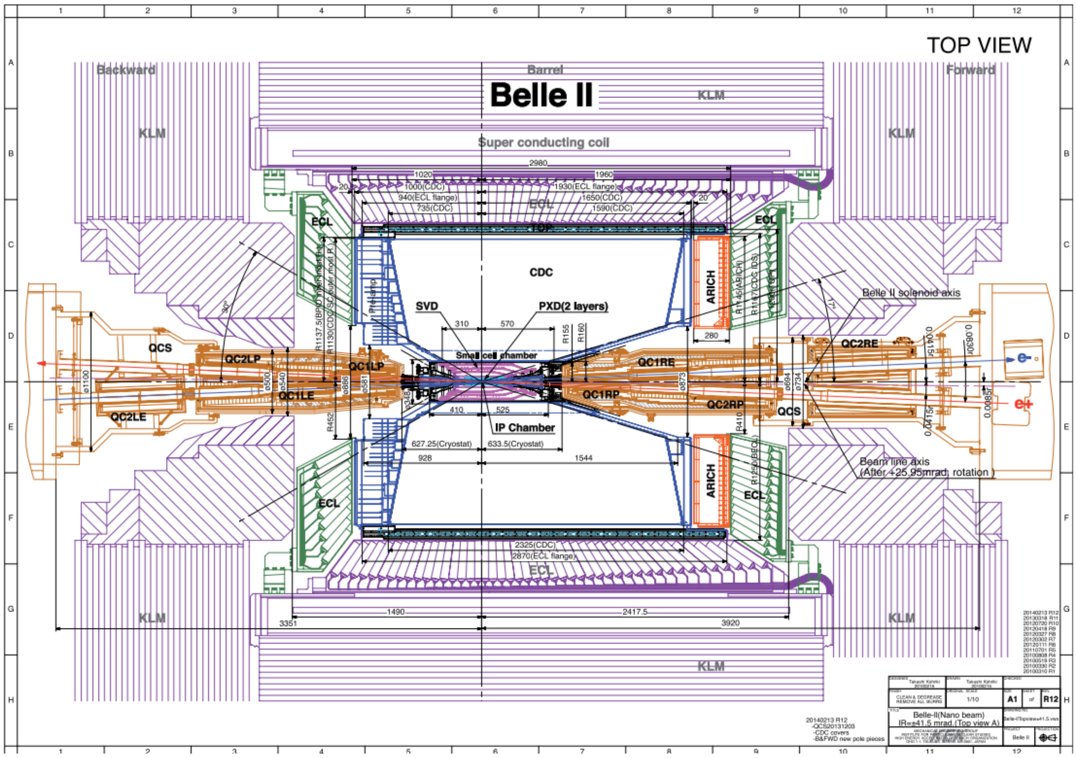


Figure 2-2: Belle II detector top view

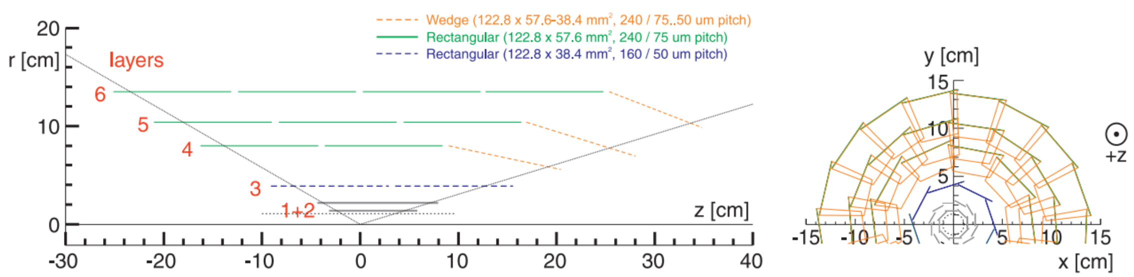


Figure 2-3: A schematic view of PXD (2 layers in gray) and SVD (4 layers in green and orange).

The PXD layers are the closest to Interaction point (IP) so the vertex resolution will be much improved. However, much higher events rate comes with much higher background level on PXD sensors. The overloaded occupancy leads to severe dead time and incredibly large data size from PXD if no data reduction scheme is implemented. In order to trim down data from PXD, a fast online tracking system is built up. When the DAQ system is triggered, the data from PXD will be first readout to a system called "ONSEN" which can store large size data in parallel up to 5 seconds. In this timing window, a fast online tracking system will perform a multiple track fitting for VXD (PXD + SVD) and CDC tracks, and extrapolate the fitted tracks backward to PXD plane so the region of interest (ROI) on PXD sensors can be defined. The signal outside of ROI will not be read out from ONSEN system to external tapes where offline data is written. Such system setup creates a buffer for the large data and efficient PXD data reduction is made. The outmost layer of SVD is also larger than Belle SVD1/2. This could be helpful for ensure the reconstruction efficiency for the decay like $K_S^0 \to \pi^+\pi^-$[11].

2.3 Central drift chamber (CDC)

The central tracking system 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 He-C_2H_6 gas. 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 or the stereo direction. Such design can utilize the information from axial and stereo wires to construct a full 3 dimensional hits which forms helix tracks. Thus, CDC is one of the key components for measuring the helix parameters for tracking system.

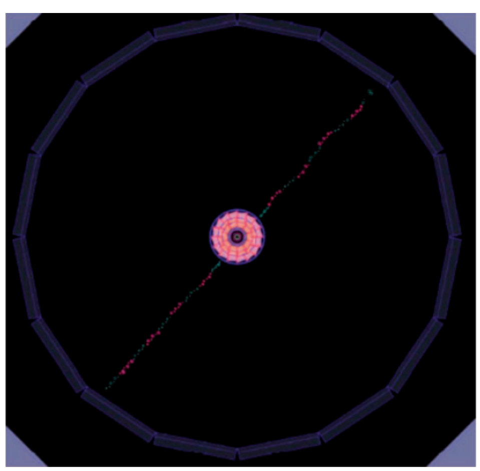


Figure 2-4: CDC tested with a cosmic ray event

2.4 TOP and ARICH detector

The particle identification system of Belle II mainly consists of two parts, time-of-propagation counter (TOP) and aerogal 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. Due to the ultra-fast flying time of photon, the TOP detectors has to achieve timing resolution at around 100 ps. A 16 channels micro-channel plate photon-multiplier (MCP-PMT) with custom-made waveform electronics of readout are used. The resolution of starting time is achieved about at 50 ps. [11].

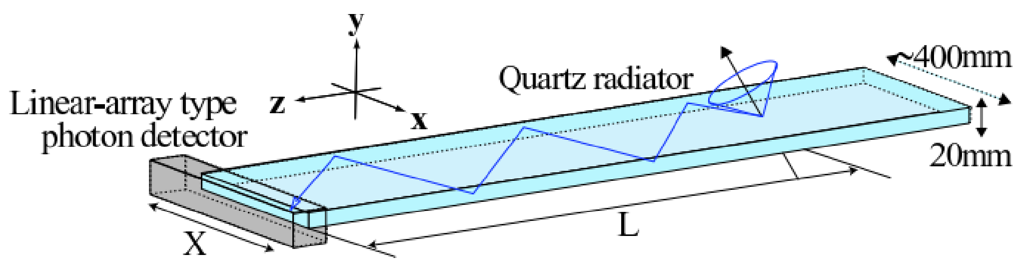


Figure 7.1: Conceptual overview of TOP counter.

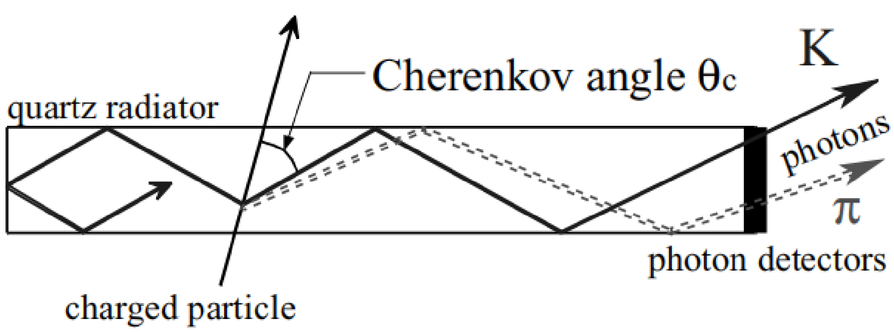


Figure 2-5: Schematic view of TOP counter (up) and its imaging process of $K^{+/-}$ and $\pi^{+/-}$ (down)

As for ARICH, it uses areogal as the sensitive material to approximately image the Cherenkov ring by a special focusing structure to identify charged particles. ARICH should be able to separate charged particles in a momentum range from 0.5 GeV/c to 4 GeV/c. ARICH requires single-photon-sensitive high-granularity sensor to reconstruct the Cherenkov angle with small photon yield. Hamamatsu Corporation, Japan, has developed a hybrid avalanche photon detector (HAPD, Figure 2-6) to meet the requirements. Each sensor is 73×73 mm^2 embedded with 144 channels to accelerate emitted electrons in a 8kV field. Avalanche photo-diodes (APD) are used for the detection of electrons at the end of electron acceleration.

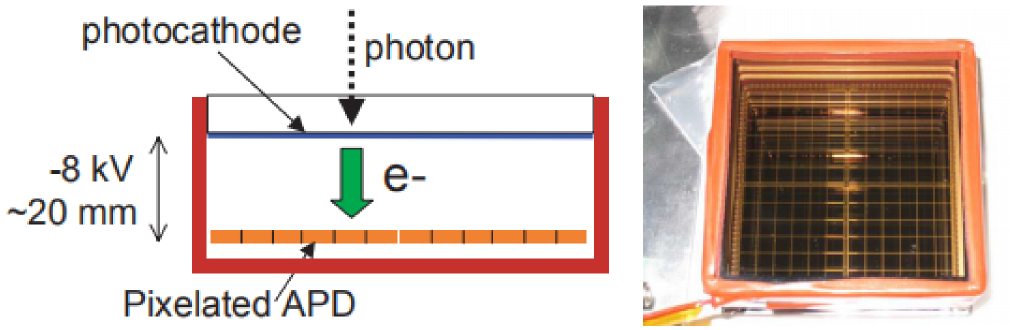


Figure 2-6: photon-electrons acceleration and pixelated APD at the end. (left) a picture of HAPD outlook [11]

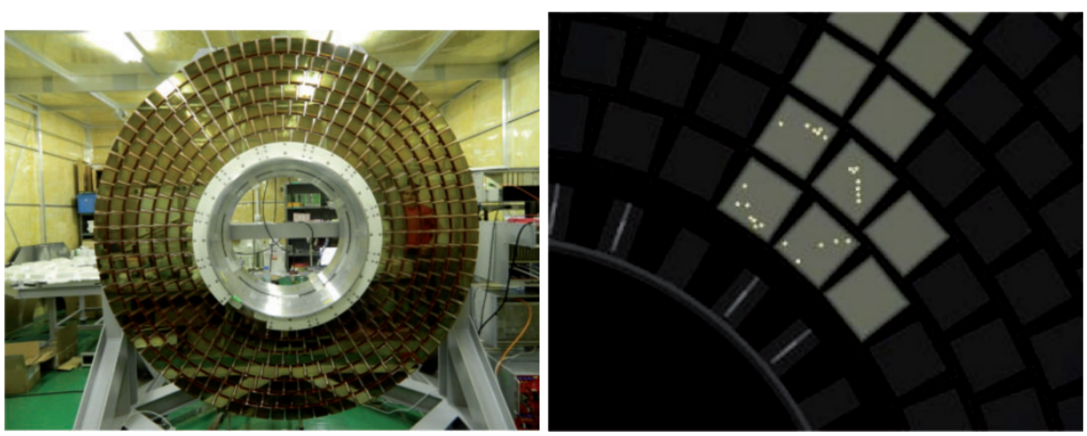


Figure 2-7: ARICH detector (left) and ring of cosmic μ on HAPD sensor[9]

2.5 Electromagnetic calorimeter (ECL)

The electromagnetic calorimeter of Belle II is mainly responsible for detection of γ radiation and electrons. The thallium doped caesium iodide CsI(Tl) crystals are assembled tightly in all three different regions, backward/forward end-caps and barrel region, as shown in Figure 2-2. Compared to the previous ECL in Belle crystal scintillation , 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 scintillation of CsI(TI). This will lead to the pile-up effect of readout noise. To compensate this effect, wave-form sampling electronics are embedded with the photon detectors. 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. Given this situation, 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(TI)

2.6 K_L^0 muon detector (KLM

KLM system of Belle II consists of a sandwich stacked iron plates and detectors at outside of the superconducting solenoid. The iron plates also serve as the interaction materials with 3.9 or more times interacting length of material compared to the ECL, allowing the K_L^0 can shower through hadronic processes. The Belle KLM material used glass-electrode resistivity plate chambers (RPC) which is not suitable for Belle II due to high background level. The structure of RPC layers is depicted in Fig(2-8). Neutrons dose is significantly larger because of much more electromagnetic radiation reaction on detector materials. The long dead time of RPC under such dose rate will reduce the efficiency of KLM. Besides, the mis-identification possibility would be raised so PID contribution from this part of detector will be meaningless.

To overcome this issue, the RPC is replaced with layers of scintillation strips covered by wave-length shifters. The readout sensors are PMT of Geiger mode operated APDs. By setting up proper working threshold, the damage of the SiPMs type APDs caused by the neutron fluxes can be reduced in the acceptable range.

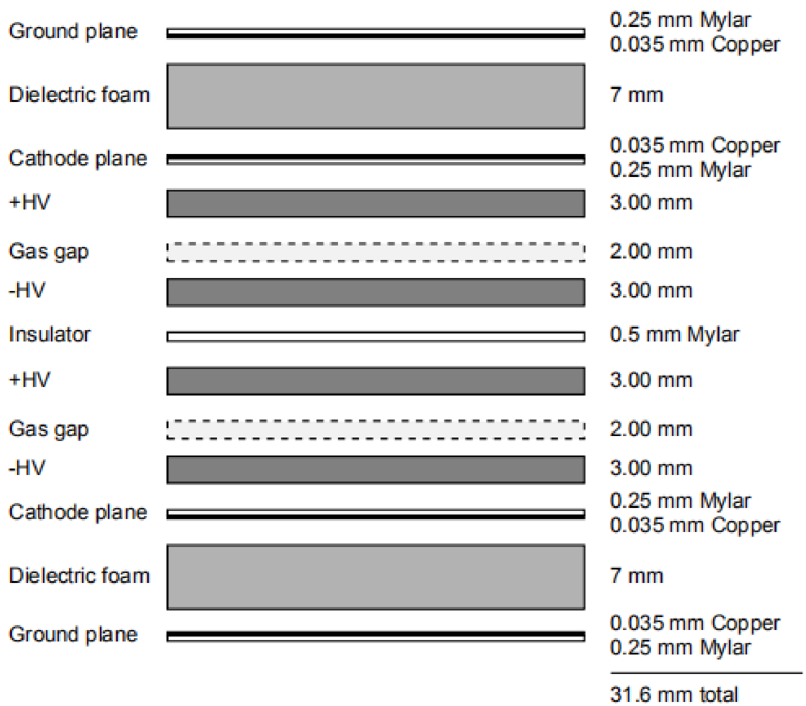


Figure 2-8: RPC layers structure[9]

2.7 Trigger and DAQ system

The interested topics in Belle II physics analysis highly depends on the trigger system to collect the co-responding events. With the updated capabilities to study a broader range of physics analysis under 10 to 20 times higher background level, the trigger system has to be able to work properly while the relatively low event recording rate of DAQ limits the pursuit of speed-only design concept of trigger system.

The beam-induced background is called beam background in general. 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. The impacts from such varieties of sources depend on many factors such as beam current, luminosity and vacuum conditions,etc. One of the featured topology of those processes is a combination of two charged tracks in CDC and one or two clusters in ECL. They are assembled with some low multiplicity events from primary collision, which is main focus of dark sector studies. Thus it's quite important to distinguish such low multiplicity events from various beam backgrounds.

Table 2.2: Simulated beam background rate (12th BG campaign)[9]

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

Since the primary goal of Belle II will still be focusing on B physics studies, it is natural that the trigger system should be able to operate over all of the interested B physics conditions, with normally 3 or more CDC tracks and large energy deposition in ECL. By offline reconstructing these events and studying the efficiency, almost of 100% for B decays are recorded by Belle II trigger.

However, the extensive capabilities of studying a large range of physics not only in b sector brings a challenge to Belle II DAQ system. Belle II has announced its excellence at performing measurement on other important topics such as τ physics, dark sector studies and initial state radiation processes (ISR). These topics and B physics studies are mutually beneficial in many ways. The reconstruction of them, as mentioned above, suffer from a large beam background, thus online algorithms must be considered in addition to offline reconstruction.

Based on the reasons discussed above, Belle II trigger has been designed to have 2 separated levels of triggers. Low level trigger, also called as L1 trigger, is hardware-based trigger. and high level trigger (HLT) is the software based trigger. The L1 trigger rate can go up to 30kHz that is also the up-limit of DAQ read-in rate. The latency of L1 is control to be 5 μs, improved from Belle trigger. And yet 30kHz is still to high for writing out the data to tape, so the HLT must be implemented to reduce the trigger rate to about 10kHz and it has to be able to select ROI on the PXD to reduce the data flux limited by bandwidth of read-out cables. To do that, HLT utilize the full offline reconstruction algorithms to allow the access of full-granularity event reconstruction using all detectors except for the PXD.

2.8 Detector simulation

As partially described before, the simulation of Belle II make a use of GEANT4 software. GEANT4 package can accept the event created by module called "particle gun" which directly injects particles to detector volume. Or it takes in software simulated data, which in general is called "event generator". Belle II Analysis Framework (BASF2) software (see the next section for the details of BASF2) provides the interface for createing simulated data from event generator to GEANT4. Most of the primary particles are simulated by event generator and sent to GEANT4 for simulation between detectors' components. The out-flying particles that has relatively long life time compared to the primary interaction such as $K_S^0 \to \pi^+\pi^-$ are simulated in GEANT4 after the event generator does its job. Exchanged bosons and primary electrons(positrons) will not be feed into GEANT4. Then GEANT4 creates secondary particles during the particle interaction and detector material, such as the radiations from charged tracks and also the scattering processes with detector materials. The hits digitization are generated by other BASF2 modules using primary and secondary particles together. Finally, the response from detectors are sent to the persistent data storage (called "DataStore" as C++ objects, detail in next section.) to be used in the analysis chain of BASF2.

For each type of the particles and each type of detector material, the interaction is varied in different processes. The co-responding process of physics should be specified by the users or using the provided list from GEANT4 developer group. In Belle II simulation, the Fritiof quark–gluon string model at high energy and the Bertini intranuclear cascade model at low energy are used by default from GEANT4 list.

The simulation of the beam background is done by a software called SAD, as external part of BASF2. It simulates the flux of particles from the beamline of the SuperKEKB accelerator. Whenever a particle trajectory deviates from the beamline region and hit the Belle II detector part, its momentum and position vectors will be saved into a configuration file. Then such configuration file will provide the initial information for GEANT4 simulation software to simulate the interaction between the given particle and Belle II detector, which is eventually analyzed as normal particles by BASF2. The output of BASF2 is standard ROOT format and it presents how a beam induced particle interacts with detector material to create simulated hits as beam backgrounds.

The mixing of background is then implemented to provide a realistic view of physical events and beam background overlay. Since the format of beam background is simulated hits, thus adding the background events is done by injecting the simulated hits, then move to the digitization of hits to detector responses. In a event time window Δt, assuming the given type background has a average rate of R, the mixing number of background hits in such event is:

$$\bar{N} = sR\Delta t \tag{2.1}$$

s is optional scaling factor which can be used to study the influence of given type background in different level. Because R is averaged value, in the actual mixing, the number of $\bar{N}$ is used as the expected value of Poisson distribution, which presents the number of observed events when many trials of such events is made with certain small possibility per event. In order to simulate the effect of timing different of background and physical events, the mixing timing window over Δt is randomly shift according to the physical events. With the real experimental data comes in handy, the method of adding background events to physics events is slightly different since using real beam background can provide a more precise result than simulation. By setting a random trigger for beam background, the hits digitization from real beam background will be collected and add to simulated physics events. Although the pile-up noise collected in this method is not very precise because of the threshold set for detectors allowing only part of noise to be added, the non-recorded noise can still contribute to the pile-up noise for physics events, and they are not included in this method. Yet overall it provides a more realistic evaluation of beam background overlay.

2.9 Belle II Analysis Software Framework

The data acquired by the Belle II experiment or simulation can be processed by Belle II Analysis Software Framework, as known as BASF2. It has a good capability to handle the tasks of sophisticated algorithms for simulation, reconstruction, visualization, and analysis. The official BASF2 is developed in different release versions, light-versions and featured-versions. For this analysis in this thesis, we use release-05-01-01 version.

2.9.1 Core Structure

The core structure of BASF2 contains three major parts: the analysis codes specifically required by the needs of Belle II data (called Belle II codes), the external libraries which utilizes the third-party software that Belle II use, and the tools for configuring and installing the BASF2 software.

The Belle II codes consists of many packages. They are categorized based on the different levels of Belle II detector components, like the packages of base-level system control called "framework", the package for track reconstruction called "tracking", and the one for post-reconstruction data analysis called "analysis". Codes are written in C++ for these packages and share a directory called "lib" on the top level of path. Usually users can work either with compiled binary version of BASF2 installed centrally on the working server or build from the source by their own need.

As for the externals, it contains various types of the code that installing or running BASF2 requires. For example, some basic packages, like gcc compiler, cmake, tar, wget, Python and git are included. In particular, due to the dependence of the analysis tools that may be frequently used by Python, around 100 additional Python packages are also installed as externals. The complexity of building all of these external software could be tough for users so the compiled versions that covers common user-based platform are available from BASF2 official repository.

Tools are collections of shell or Python scripts for setting up BASF2 and externals environment. It can easily handle the need of setting up an environment of specific BASF2 version and the externals tied to that version. It also provides a function to setting up the environment of developing BASF2, which developers can get one developing copy of BASF2 and write the additional codes as the modification so the compatibility of BASF2 could be easily maintained by building release version from the developing branches.

2.9.2 Event Processing Workflow

The data from Belle II detector or from the simulation, are organized into a set of variable-duration runs, and each run contains a sequence of independent events. Every event is recorded as the measurement of by-products of an electron-positron collision or a cosmic ray injection. A set of runs which presents a similar detector conditions is packed as an "experiment". Such experiment-run-event structure is the basic data structure for BASF2. Thus BASF2 processes the data by a set of modules that contains the following phases as follows:

- initialize: called at the start of a event to properly set up this module
- beginRun: called at the start of each run to handle the run-independent data.
- event: called at the start of each event processing.
- endRun: called at the end of a sequence of events in a run.
- terminate: called at the end of the processing of all events.

BASF2 executes a series of modules loaded dynamically to process the data set according to above sequence, where each module will have above phases inside their python interface. The selection, configuration and executed order of the modules are defined by a file called "steering file" written in Python as well. The modules parameters are attributes which can be set during the runtime using steering file. For example, the "Path" object declared in the "steering file" stores the sequence of modules to be executed, to which allow other modules such as "mdstInput" or "reconstructDecay" to be added. An integer result set in "event" phase can be used for a conditional branching of a "Path" in case that one event needs to be processed with different set of modules based on its features. BASF2 starts running when it checks in the "steering file" there is at least one module specify the number of events to be processed in a "path" and return the time and number of events as information printed in standard output or stored in ROOT file.

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

# Generate 100 events with event numbers 0 to 99↩
     that contain only the event meta data.

import basf2
main = basf2.create_path()
main.add_module('EventInfoSetter', evtNumList↩
    =[100])
basf2.process(main)
```

Figure 2-9: Example of BASF2 steering file, setting up a processing of 100 events in path called "main".

The object that interacts with BASF2 I/O is called Data Store as mentioned before.This implementation doesn't rely on the event data model. The only mandatory component is called EventMetaData which presents the event,run and experiment index and production unique ID to distinguish events. The raw format of data depends on the readout of detector data. Unpacker module of BASF2 converts the raw format into digits object. In simulation, digitization is done by module called digitizer using energy deposition from Geant4. The simulated hits are from common based SimHits object, which allows for adding machine-induced background to an actual physics energy deposition process.

2.9.3 mDST structure

The output from the reconstruction will contain several detector-specific objects, which are concluded as mini data summary table (mDST) type file. For instance, the object called “RecoTracks” will be created for track pattern recognition and perform the track fit using the hits from different detectors. Typically for a reconstruction of recorded detector response of a physical event, mDST could consist of following classes:

• Track: object presenting any charged particle trajectory, and 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 vertex detector and CDC.

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

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

• ECLCluster: : reconstructed cluster in the electromagnetic calorimeter. It consists of energy deposition and hit position. Hit shape related variables. If a cluster is matched with an extrapolated track, a relation between them will also be created.

• reconstructed cluster from K_L^0 and muon detector. It consists of momentum and position measurement. If a cluster is matched with an extrapolated track, a relation between them will also be created.

• KLId: K_L^0 candidates with the particle identification as related to KLM and ECL clusters.

• TRGSummary:L1 trigger information.

• SoftwareTriggerResult: high level trigger information mapped by trigger names to trigger results.

• MCParticle: if data is from simulation, a particle from simulation containing the momentum, production and decay vertex, relations to mother and daughter particles, and information about traversed detector components. Particle-Detectors relations are created if simulated particles are reconstructed as tracks or clusters from Belle II.

The size of mDST level data is very important to processing performance. Thus, mDST level data is restricted from contain non-physical analysis related information such as raw detector response digits or calibration constants. For detailed detector or reconstruction algorithm performance studies, and also for calibration tasks, a dedicated format higher than mDST, called cDST, (calibration data summary table), is provided.

2.9.4 Conditional Database

In addition to the physics data, analysis relies on various condition data that defines different calibration of detector, weight files for multi-variate analysis usage and more. This part of data is stored in a central data based called central Conditional Database (CDB).

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), and GT is subject to change so the assignment of IoVs or payloads data can be modified. Except for the central database, a local database backend that reads global tag information from GT and uses a local database is implemented. It automatically download the database files that are needed for a BASF2 execution and store them in a local folder. This means even if the computer is offline or the CDB is not reachable, one can still run BASF2 as long as the local folder was there.

User access to conditions objects is provided by two interface classes, one for single objects called DBObjPtr and one for arrays of objects called DBArray. To facilitate easy creation of new conditions data – for example, during calibration – we provide two payload creation classes, DBImportObj and DBImportArray. They have an interface very similar to DBObjPtr and DBArray. [14] Users instantiate one of the creation classes, add objects to them and commit them to the configured database

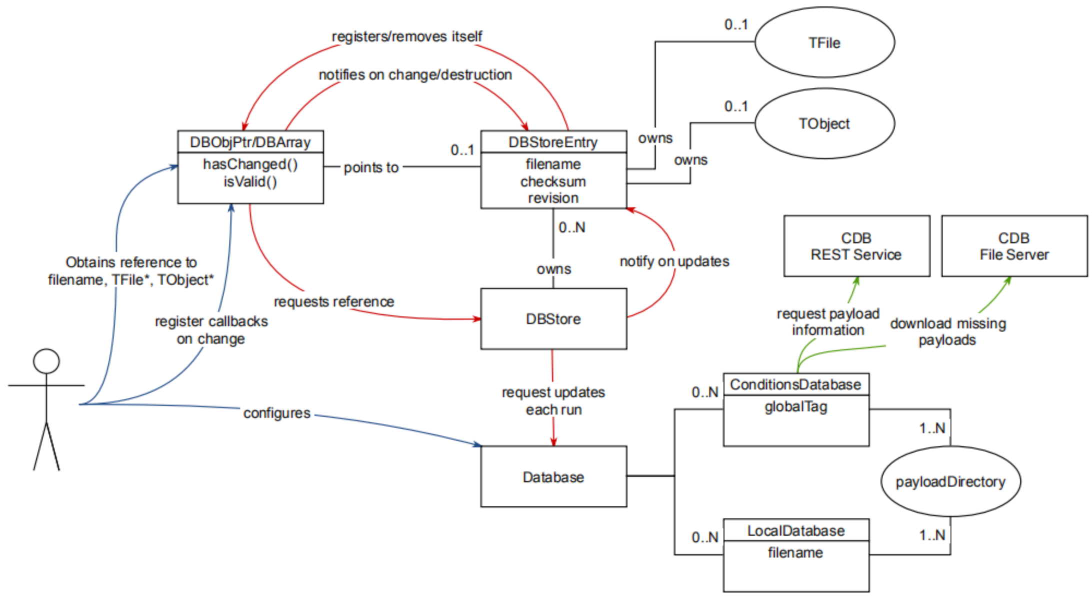


Figure 2-10: Relations of all entities in CDB. [14]

with a user-supplied IoV. This includes support for run dependency. 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. The scheme of this entities and how users interact with CDB object is demonstrated in Fig 2-10.

2.9.5 Summary

BASF2 has been developed for an emphasis on providing reliable and high quality performance for Belle II analysis. It satisfies the most of demanding requirements of data taking, simulation, reconstruction, and offline analysis.

Chapter 3

K_S^0 Reconstruction Study

The final states of $B^0 \to K_S^0 K_S^0 K_S^0$ only depends on the decay of K_S^0. The main decay channels of K_S^0 is to either $\pi^+\pi^-$ at branching fraction of about 0.692, or to $\pi^0\pi^0$ at branching fraction of 0.307. The characteristics of these two decays are much different in terms of the response from Belle II detector. The charged decay that yields $\pi^+\pi^-$ leaves two tracks (track as data object) originating from VXD or CDC volumes with opposite charge. While the π^0 main decay channel is $\pi^0 \to \gamma\gamma$. This creates a bunch of clusters on ECL (ECLClusters as data object). The reconstruction of the decay channel is performed through $\pi^+\pi^-$. There are mainly two reasons for not selecting π^0 as final states. Firstly, $\pi^0 \to \gamma\gamma$ comes with a big load of background as fake K_S^0. The reconstruction of ECLClusters provides no constrain on K_S^0 vertex so it's almost impossible to suppress the combination background using vertexing quality in this case. The only reliable selection will be the mass of K_S^0 which typically varied around its nominal mass with a few hundred of keV. The γ however, could be originating from many other resource, such as beam background, and charged track radiation. Using mass window of K_S^0 could not effectively reject the noticeable fraction of fake rate. The MC study shows the fraction of true K_S^0 among all accepted candidates using mass cut only is about 3% using 2 neutral pions. Secondly, with K_S^0 from neutral pions included, the events of B^0 that consist of one or more such K_S^0 will have poorly reconstructed vertices. The precise measurement of time-dependent CP violation emphasizes vertexing quality. Even with $B^0 \to K_S^0 K_S^0 K_S^0$ that only uses charged pions for final states, it already has no primary vertex from B^0 which leads to the worse resolution of vertex position compared to the channel like $B^0 \to J/\psi K_S^0$, which has two direct charged tracks of e^+e^- or $\mu^+\mu^-$ because of the very short flight distance of J/ψ. If one or more of K_S^0 has bad vertexing quality from its decay products, it will further reduce the precision of vertexing of B^0. This leads to a large uncertainties in defining the decay time of signal B^0 so as to the decay time difference which is the key observable to TDCPV measurement. Based on above, we only reconstruct K_S^0 from its charged decay products to improve the vertexing quality of B^0.

3.1 Cut-based K_S^0 Reconstruction

The K_S^0 has average life time at $(8.954 \pm 0.004) \times 10^{-11}s$ so the flight length of K_S^0 before it decays could be compared with the scale of detector size. In the typical Belle II energy scale, K_S^0 mainly decays inside of VXD volume, which covers from a few micrometer inside first layer of PXD, to the even outside of last layer of SVD ladder that is placed at 14cm away from the interaction point(IP). The flight length of K_S^0 from B^0 generic decay and $B^0 \to K_S^0 K_S^0 K_S^0$ are shown in Fig 3-1 from MC13:

Due to the different topology of B^0 decay, the average momentum of K_S^0 in generic decay is smaller than the ones from generic B^0 decay, as well as the average flight length. It's clear that the fraction of K_S^0 decaying outside the IP is majority and there's no reason to constrain the origin of K_S^0 decay vertex to IP. So the basic cut-based reconstruction of K_S^0 is performed by the selection of invariant mass from its decay products then putting requirements on its vertexing fit quality without geometric constrain. The cu-based selection of K_S^0 is done by using BASF2 "stdKshort:merged" list. In this "stdKshort:merged" list, we first take all the "V0" objects from BASF2 which use 2 online reconstructed charged tracks with opposite charge and a converged fitted vertex. The invariant mass between $0.45 < M < 0.55$ GeV is applied, where M is calculated from daughters' 4-vector. Then the reconstruction of K_S^0 is also performed by offline reconstruction using BASF2 which loads all tracks as charged pions

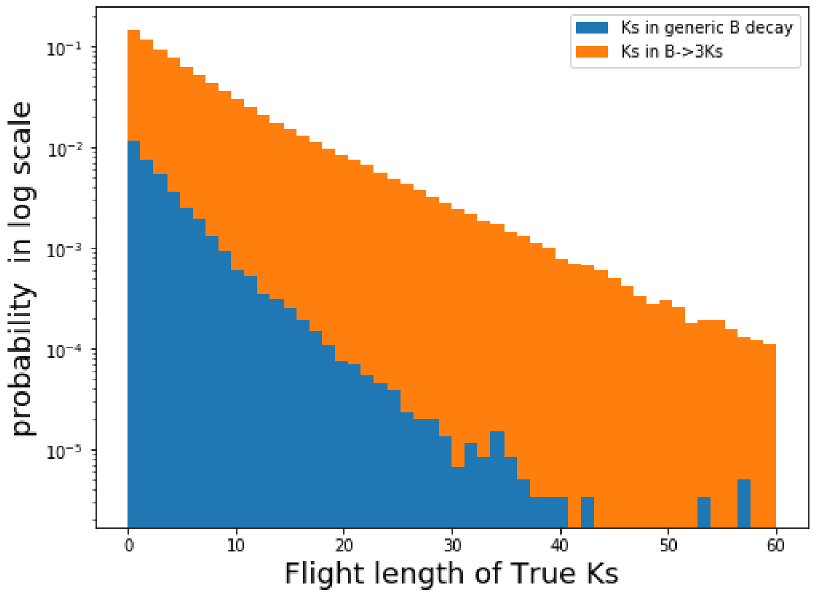


Figure 3-1: Flight length of K_S^0 from MC, Blue is from generic B^0 decay and orange is from $B^0 \to K_S^0 K_S^0 K_S^0$. Both are from MC13.

and also requires the invariant mass between $0.45 < M < 0.55$ GeV and a converged fitted vertex. In both cases, the vertex fit is performed using TreeFit. The reconstruction efficiency B^0 is highly sensitive to the efficiency of charged pions, so we use charged pions with no extra cuts but only the M and converged vertex. Compared the K_S^0 from V0 and offline reconstructions, we can remove the duplicated K_S^0 when the two pions are from two same tracks, to form the K_S^0 candidates that still have many fake ones, see Fig 3-2.

Table 3.1: Pre-selection criteria of $\pi^+\pi^-$

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

The K_S^0 candidates from "stdKshort:merged" is the default way to obtain K_S^0 in BASF2, however, the limitation of this cut-based K_S^0 reconstruction is the pollution from fake K_S^0. When using these K_S^0 candidates to reconstruct $B^0 \to K_S^0 K_S^0 K_S^0$, each of the fake K_S^0 has the chance to propagate to form a fake B^0 since K_S^0 is only intermediate states in this channel. This creates high fake rates in B^0 candidates and costs a large extra processing time for computing the kinematics and vertex

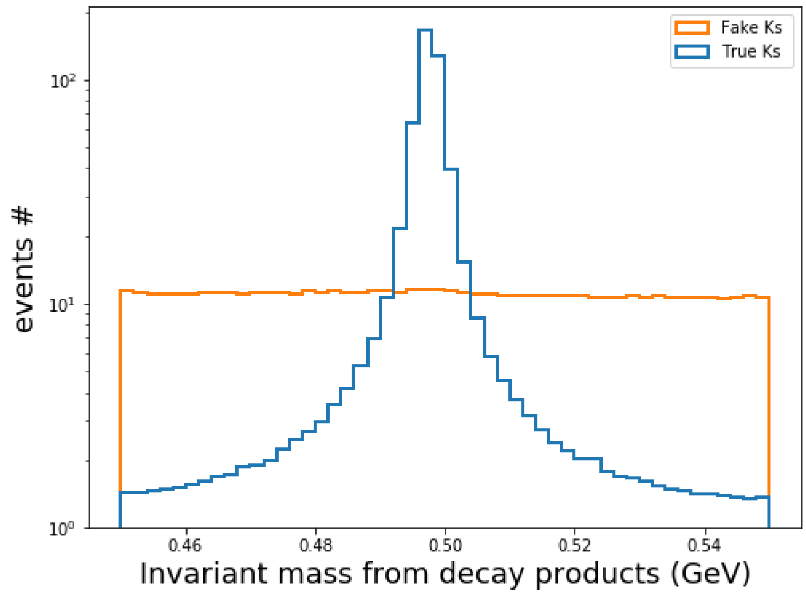


Figure 3-2: The distribution of invariant mass of K_S^0 candidates, blue line is the MC truth matched K_S^0 and the orange is the fake K_S^0 passing the selection.

fit of B^0 which could've been avoid. Even by then, the number of combinatorial backgrounds in B^0 from $3K_S^0$ is still high. Thus, a multi-variate analysis (MVA) based K_S^0 identification package is developed.

3.2 MVA-based K_S^0 Identification: KsFinder

3.2.1 Experience from Belle

In order to improve the reconstruction performance of K_S^0, a multi-variate analysis based (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 observable that describes the characteristics of $K_S^0 \to \pi^+\pi^-$ decay. The targeted output is the signal or background flag from the MC truth-matching called "isSignal" where isSignal = 1 (0) stands for being a true(fake) K_S^0. From the experience of Belle, the K_S^0 reconstruction was first done by using cut-based method to select primary candidates, then MVA-based classifier was implemented by assigning two likelihood indicators to each K_S^0 candidates. The package used by Belle is called “nisKsFinder”[9]. It classifiers the two likelihood variables based on NeuroBayes algorithm, which defines the goodness of K^0_S. A good candidate from “nisKsFinder” is the one with low likelihood of being Λ particle and the high likelihood of being a V0-like particle. The two variables used here are called “nb_nolam” and “nb_vlike”. By putting cuts on these two variables, a purification of K^0_S can be made.

3.2.2 FastBDT algorithm

In Belle II, such tool is of missing from the current BASF2 framework. Since “NeuroBayes” algorithm is a third-party commercial product and is not being supported by its producer, the usage and maintenance of it could face unexpected issue with no warranty of fixing ensured. So it’s not going to be the default option for K^0_S classification in Belle II. Instead, stochastic gradient-boosted decision trees are widely employed for multivariate classification and regression tasks in modern high energy physics field. Particularly, a speed-optimized and cache-friendly implementation for MVA classification called FastBDT (FBDT) is popularly used. Compared to other popular classification algorithm in software framework, such as TMVA, scikit-learn and XGBoost, FastBDT method is proven to be one order of magnitude faster during the fitting and applying phase.

A general DT (decision tree) performs classification using a number of consecutive cuts. The maximum number of cuts are called “depth of tree” and it’s a hyper-parameter of DT. For each data point, there are known labels called “feature” and data points passing the previous cuts on one feature are sent to the next label. At each nodes, a cumulative probability histogram can be defined by counting the signal and background regarding a certain label. A gain presenting the separation power at each node is calculated using all possible labels. The cut on that label is determined by making a maximum of gain. This process locally maximizes the separation of signal and background at each node. At the end layer of tree, which is called “terminal nodes” of tree, the whole data set are spitted into different groups by DT. In each group, a fraction of signal is calculated using all data points in the group. Such process is illustrated as Fig 3-4.

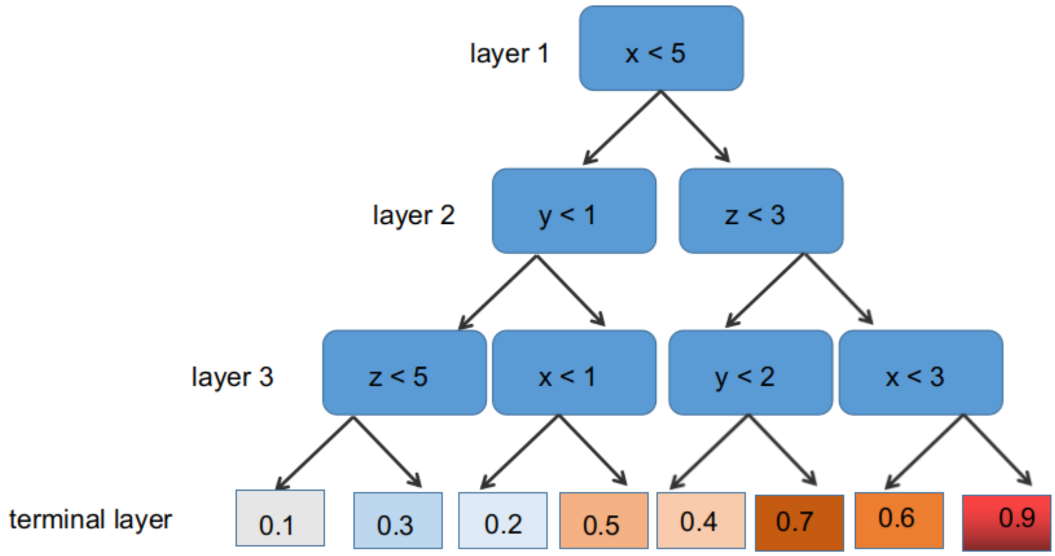
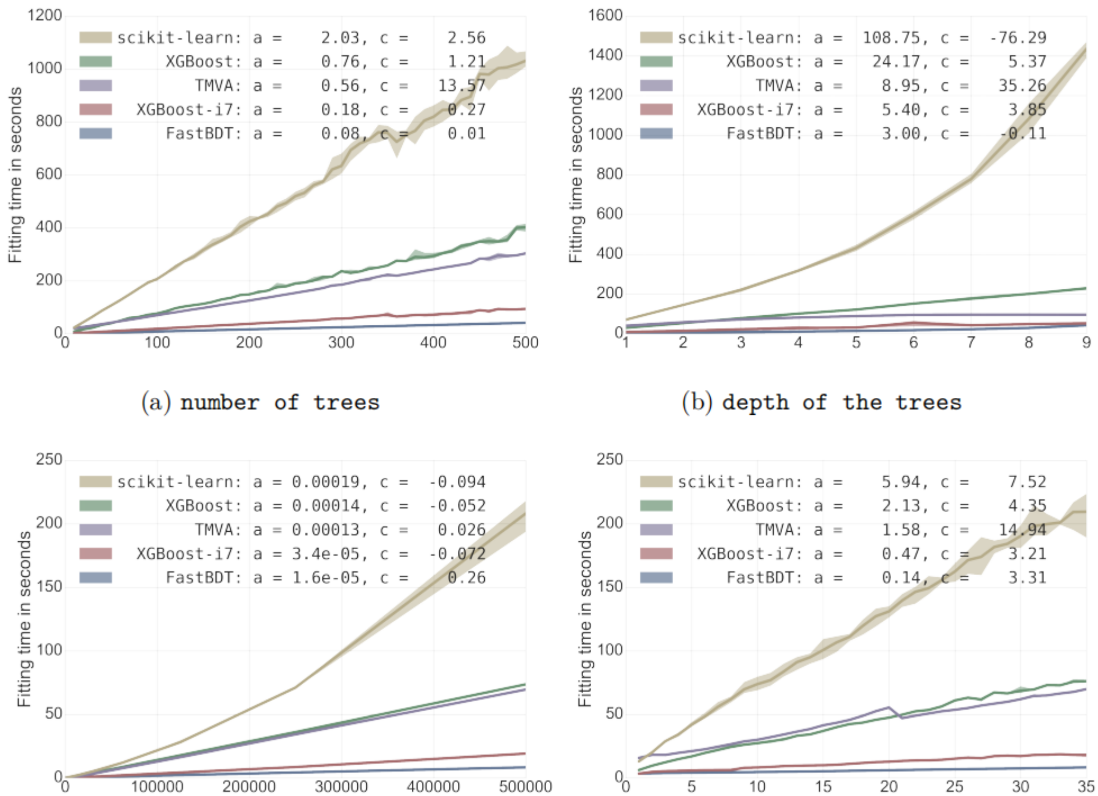


Figure 3-3: Basic structure of a DT with depth = 3 and label of x,y,z. The last layer is terminal node and the number is signal fraction. The number (color demonstrated) is the signal fraction of data points in the nodes

The mathematical idea behind this method is to treat the data points as a data set defined on a multi-dimension hyper-space. As long as the signal/background data points show certain concentration in a sub-region of the hyper-space, it's possible to locally increase the signal fraction by consecutively cutting on the edge where signal and background separate. The cut on labels at each node is the edge of the sub-region. The deeper a DT is, the more edges of the hyper-space will be cut. So if a DT has too many layers (too deep), the data points in the sub-region can achieve a over-fitted signal fraction and won't represent a true distribution of whole data set because of the statistical fluctuation in the small sub-region. As a result, the classifier is over-fitted and performs poorly on new data points. There are pruning algorithms which automatically remove cuts prone to over-fitting from the DT, details can be found here [15].

Avoiding the over-fitting of DT limits the depth of a tree strongly while a single shallow DT can only roughly separate signal and background. For a problem of K_S^0 classification, number of observables is much more than the usual tree depth (a few layers). So during the fitting phase, a sequence of many shallow DT is formed. For all the DTs, a negative binomial log-likelihood loss-function is minimized. The generation of DTs in this step is called “Boost”. Combining many weak-learners (single DT), a classifier with large separation power is constructed. The number of trees N (or boosting steps) is the additional hyper-parameter of the model. The FastBDT implements a optimized algorithm from a derived Gradient BDT method (GBDT)[16] and gain an order of magnitude faster execution time. The comprehensive comparison between FastBDT and other popular methods in Belle II scope is described in here [17].



(a) `number of trees` (b) `depth of the trees`

(c) `number of training data-points` (d) `number of features`

Figure 3-4: Runtime in fitting phase with different hyper-parameters comparison among FastBDT and XGBT,TMVA,scikit-learn.[17]

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

The first step for developing K_S^0 MVA classification is to determine the input variables for FastBDT method which should represent the decay feature of K_S^0. The remaining background of $K_S^0 \to \pi^+\pi^-$ after the cut-based reconstruction comes from different resources. In these resources, the main contributions are false combination of tracks, V0-like particle mis-identification, and looped tracks.

The false combination of tracks includes two major cases. First is when one of the two daughter is $\pi^{+/-}$ and the other is not. This often happens when a charged pions from a neutral mother with a capability of decaying into multiple charged tracks, like $D^+ \to \mu^-\nu_\mu K^-\pi^+$. On the other hand, it's also possible that both of two tracks are correctly reconstructed from $\pi^{+/-}$ but they are not from the same mother, or the mother is not a K_S^0 particle due to the missing of other daughters, such as $D^+ \to K_S^0(\to \pi^+\pi^-)\pi^+$. The decay shape resembled the above cases are illustrated as the following:

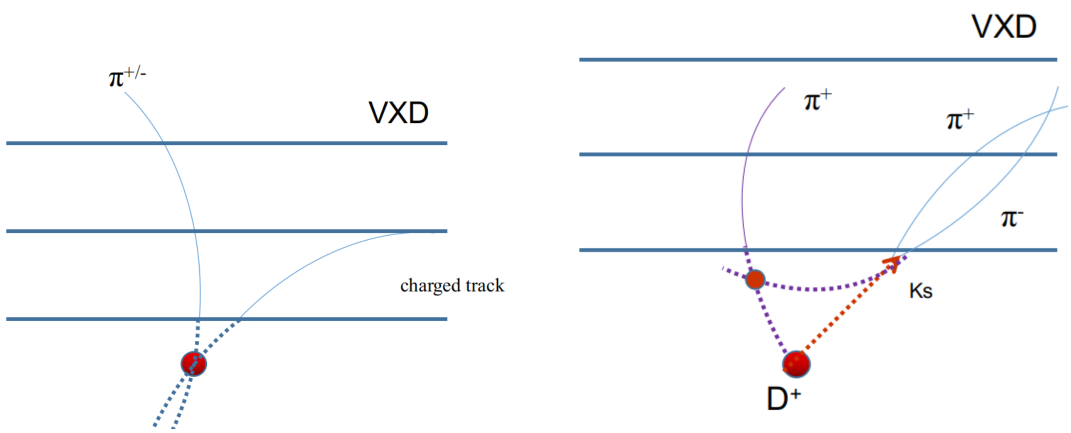


Figure 3-5: The left shows the case when a charged track (not a pion) combined with a true pion as a fake K_S^0, the right shows the case when two daughters are correctly reconstructed as pion but not from the correct mother.

The V0-like particles mainly refer to K_S^0, Λ and γ. $\gamma \to e^+e^-$ yield is significantly lower than the previous two types and the mass different between pion and electron is very large, so the PID values can be used to well-distinguished them. As for the contribution of $\Lambda \to p^+\pi^-$, it's happens when the positive charged tracks (proton track) is wrongly identified as π^+, see Fig 3-6 left. The key observable to distinguish this background is the invariant mass of mother particle, which Λ is at 1.115 GeV, much larger than the K_S^0. The left-over Λ after the cut-based reconstruction is minimal and can be further reduced by checking the PID information of the positive charged daughter.

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 $r - \phi$ plane. 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. Thus it also gives a potential fake K_S^0, see Fig 3-6 right.

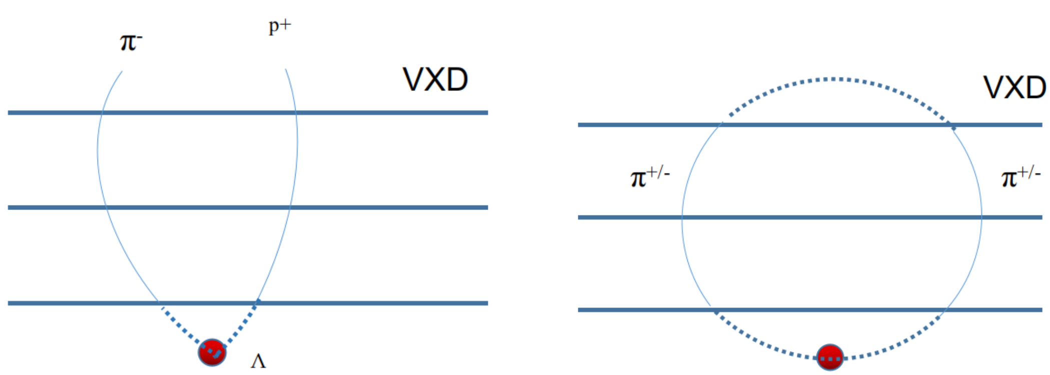


Figure 3-6: The left shows the $\Lambda \to p^+\pi^-$ decay shape that can be treated as K_S^0, the right shows a self-loop formed by a low p_T charged pion reconstructed as two separated tracks with a vertex

3.2.4 Determination of training observables from K_S^0 decay

Given the characteristics of $K_S^0 \to \pi^+\pi^-$ discussed in the previous section, a set of observables as training features of FastBDT classifier can be constructed. The set includes categories of observables: kinematics, decay shape parameters , particle identifications and detector hits information. Observables are listed below with a sketch showing the shape parameters of K_S^0 decay in Fig 3-8.

- Kinematics

  - Invariant mass of K_S^0 before and after fitting vertex
  - momentum of K_S^0 and $\pi^{+/-}$, vectors and magnitudes.

- Decay shape parameters
  - cosine angle between K_S^0 vertex and momentum.
  - helicity angle of two daughters in reference of K_S^0 momentum.
  - decay angle of two daughters in the mother's frame.
  - flight distance projection on K_S^0 momentum direction.
  - significance of flight distance, defined by ratio of flight length and its uncertainties.
  - distance on z-axis of two daughters helix
  - impact parameters on K_S^0 vertex

- Particle identifications
  - pion-ID for K_S^0 daughters.
  - muon-ID for K_S^0 daughters.
  - proton-ID for K_S^0 positive charged daughter.

- Hits information
  - the number of PXD hits for each K_S^0 daughter, up to 2.
  - the number of SVD hits for each K_S^0 daughter.

The decay shape category is of the most importance because it demonstrates the best separation power. For instance, if a false combination is made of two tracks, it's likely that the momentum direction of reconstructed fake K_S^0 is not aligned with the vertex position. So the projection of flight length on the momentum could be negative value for background. While in case of a true K_S^0, such projection is almost always a positive value.

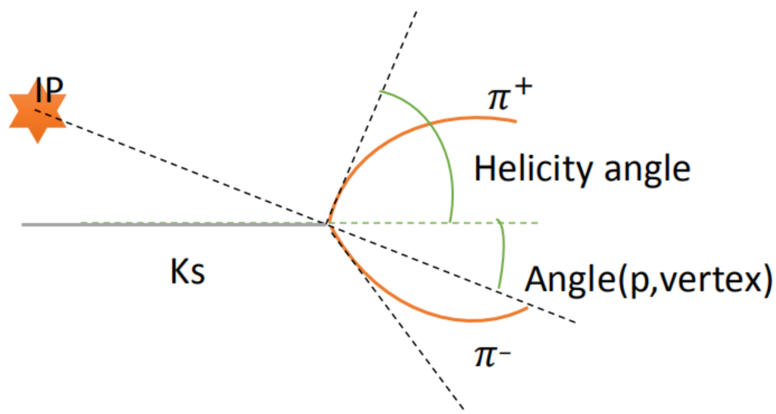


Figure 3-7: The decay shape parameters, vertex vector takes IP as origin.

There are a few points to be checked for using FastBDT classification, given the nature of the algorithm. First, the distribution of the observables should be different in true K_S^0 and background, so the FastBDT classifier can perform distinguish the true and the fake at each nodes to maximize the separation gain, just as Section 3.2.2 discussed. Secondly, there will a correlation among the training observables and they should also be different in signal and background. The boosting phase 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 by 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 VXD hits, without clear correlation. At last, one should also avoid using many observables with too strong correlations since the classifier won't gain much improved separation power when put a cut on each of them. The correlation of the observables in signal and background samples are shown in Fig 3-9. The selected observables meets the requirements for classification.

3.2.5 Training, Testing and Application of KsFinder

Training samples of K_S^0 are first extracted based on "stdKshort:merged" from generic B^0 decay and $B^0 \to K_S^0 K_S^0 K_S^0$ using MC13 samples, respectively. For the hyper-

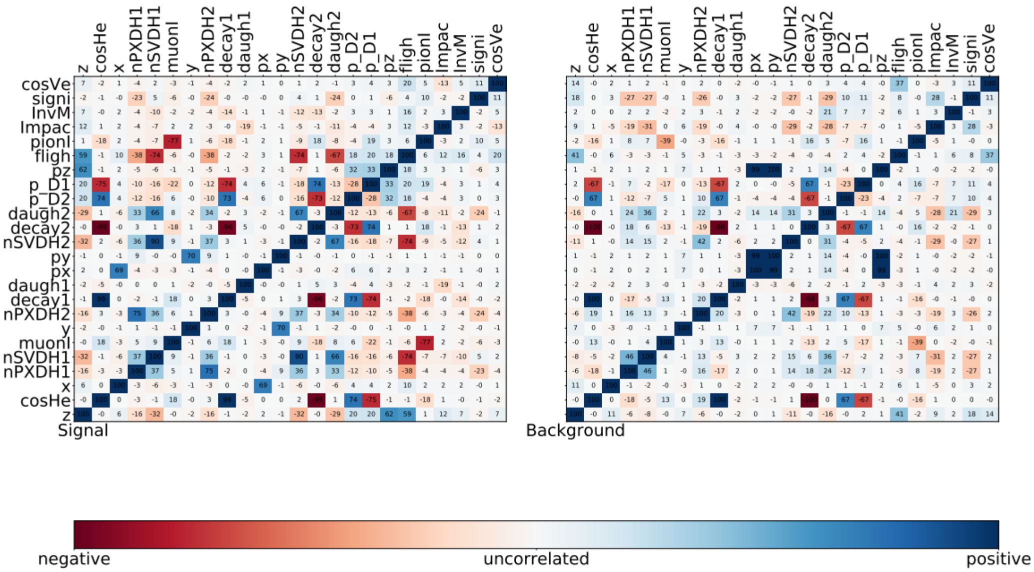


Figure 3-8: The correlation matrix of the chosen observables. It shows different correlation between observables in signal and background. And there's no single observable showing strong correlation with all other members.

[arameters of FastBDT method, the depth of each DT is 3 and the number of total trees (boosting steps) is 200, with a balance of computing time and performance. The training target variable is "isSignal" flag. The ratio between the true and fake K_S^0 is 1:1 in both training and tesing sample. The training sample and tesing sample are separately prepared using different input file from MC13. The distribution of observables in true and fake samples (from $B^0 \to K_S^0 K_S^0 K_S^0$) are shown in the Appendix A. The abbreviations of those variables and their importance rank are shown in Table 3.2 and 3.3.

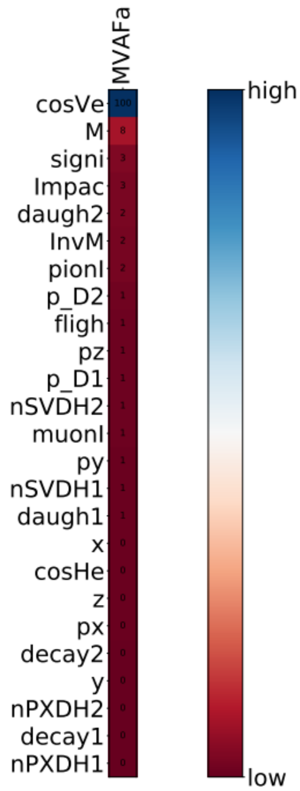
3.2.6 The Performance and Over-fitting check

The performance of classifier is evaluated by applying it on a independent data sample. So in the accordance of training data used, the same mount of events are prepared. Since the training is performed on three different types samples, three classifiers are actually obtained. Signal efficiency and background rejection are calculated using the output of trained classifier that presents the likelihood of probability of being a true K_S^0, as defined in Eq (3.1) and Eq (3.2).

Table 3.2: The Abbreviations.

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

Table 3.3: Importance rank



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

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

Using the weight file created by the training process, the output between 0 and 1 is assigned to every candidates as a quality index standing for the likelihood to be true K_S^0 that can be used as a cut. In order to check the performance of the classification, the ROC (receiver operating characteristics) curve is plotted, which shows the dependence of rejection power regarding the signal purity. The larger area a ROC curve is covered, meaning that background rejection drops slower when increasing classifier cut, the better performance is achieved. Three classifiers are tested with generic B^0 decay, and $B^0 \to K_S^0 K_S^0 K_S^0$ to check the robustness of classifier. The ROC, signal efficiency and purity are shown:

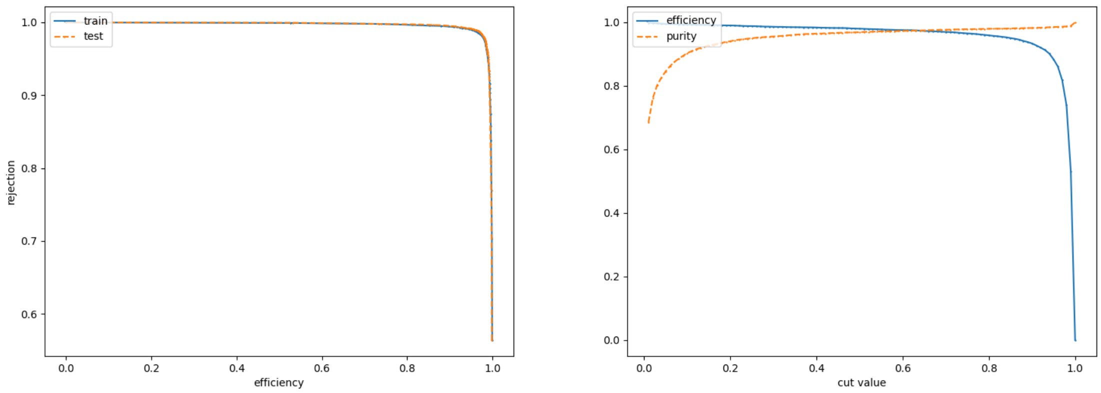


Figure 3-9: The left is ROC curve(blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of classifier output. Results are from $B^0 \to K_S^0 K_S^0 K_S^0$ sample.

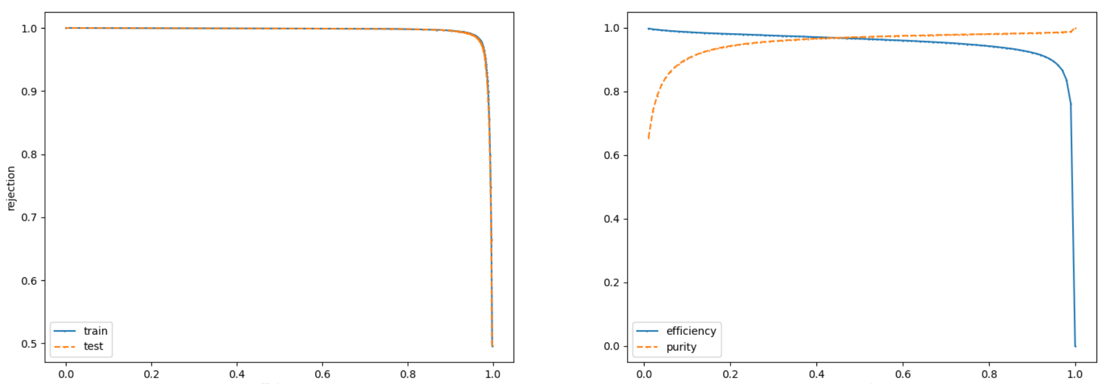
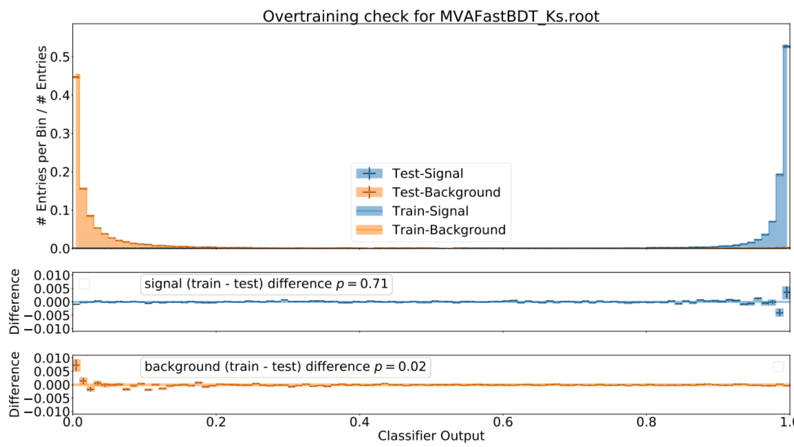
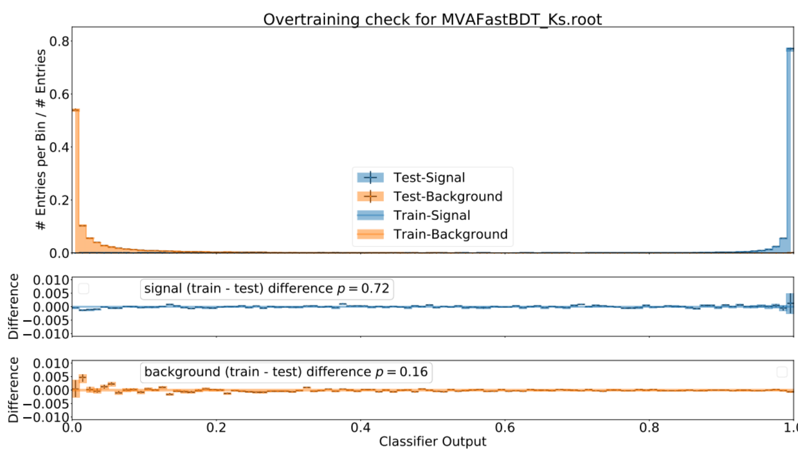


Figure 3-10: The left is ROC curve (blue for training and orange for testing) and the right is efficiency and purity (blue for efficiency and orange for purity) depending on cut of classifier output. Results are from B^0 generic decay sample.

The ROC curves show a good rejection power on classifiers. To be noted, the curves are consistent in training and testing samples. While the ROC curve has shown the absence of noticeable over-fitting in classification, the detailed check can be made by comparing the distribution of classifier output on true and fake K_S^0 respectively in training and testing. In the classifiers we obtained, results from training and testing are very much close thus no over-fitting is spotted, as shown in Fig 3-13.



a) Over-fitting check for $B^0 \to K_S^0 K_S^0 K_S^0$.



b) Over-fitting check for B^0 generic decay.

In order to use the output of KsFinder, a cut value must be chosen. The output of KsFinder is named “FBDT_Ks”. Here we can define a “Figure of Merit” (FOM) to determine the cut value. S and B is the number of true and fake K_S^0 after the cut,respectively. The value of 0.74 of “FBDT_Ks” maximize the KsFinder output in $B^0 \to K_S^0 K_S^0 K_S^0$, so we will use this value as only K_S^0 with larger “FBDT_Ks” will be kept from cut-based selection.

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

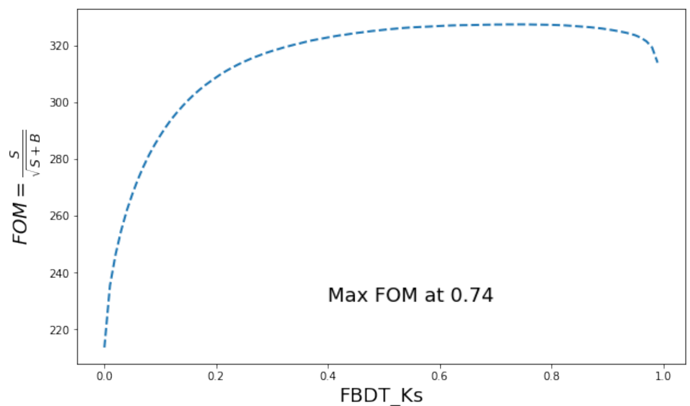


Figure 3-12: FOM of classifier output (FBDT_Ks) in $B^0 \to K_S^0 K_S^0 K_S^0$, maximum value is obtained at 0.74 and curve is almost flat after 0.5.

By comparing the fitted invariant mass of K_S^0 before and after the application of this cut, it's clear that the fraction of background has been largely reduced and most of the signal remains. The true K_S^0 fraction in the pre-selected K_S^0 before applying KsFinder cut is 39%, and 95.3% of them are remained after KsFinder applied. The fake K_S^0 fraction before applying KsFinder cut is 61%, and 97.6% of them are rejected after KsFinder applied. A much cleaner K_S^0 candidates is created as shown in Fig 3-15.

3.2.7 Data Validation for Classifier

The results from MC studies show an excellentperformance of KsFinder. However, such classification is based on the observables reconstructed from MC samples, and FastBDT algorithm is depending on the training features apparently. As a result, the validation of such tool on the real experiment data is necessary. This would justify the usage of classifier on data and is also essentially helpful to check the potential discrepancy between MC and data.

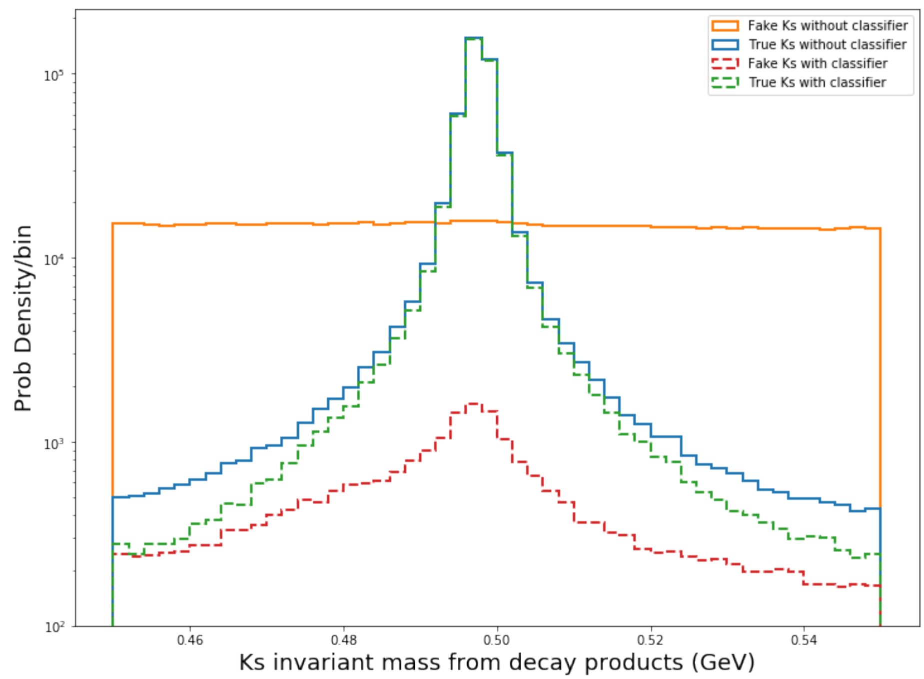


Figure 3-13: K_S^0 purity improvement with cut value of FBDT_Ks at 0.74 applied. Blue solid line is true K_S^0 before KsFinder and green dashed line is the true K_S^0 after. The orange solid line is fake K_S^0 before KsFinder and red dashed line is fake K_S^0 after. 95.3% of true K_S^0 are kept while 97.6% the fake are rejected by the classification.

The validation comes from the following aspects. First of all, since there's no "isSignal" truth in real data, there's no way to direct check performance on data. Since the FastBDT method is based on the distribution of training variables, if these variables shows close distribution among MC and data, then the classification performance is expected to be close.

Thus, variables must be compared between data and MC to ensure the consistence. Then, the expected performance represented by the distribution of selected K_S^0 should be similar between data and MC. Particularly, since K_S^0 candidates will be used for further reconstruction of B^0, its kinematics such mass and momentum may change after the classifier application, so the validation that approves no clear bias on B^0's M_{bc} and ΔE which are used for B^0 reconstruction is also needed.

We take the small data sample from Belle II early phase 3 operation experiment 7 and 8 in 2019 for comparison. The integral luminosity at $\Upsilon(4S)$ resonance is about 5.17 fb^{-1}. MC13 sample is extracted from generic B^0 decay with equivalent events number. There are two campaigns of MC included (MC12b and MC13, later one is the latest). Fig 3-16 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.

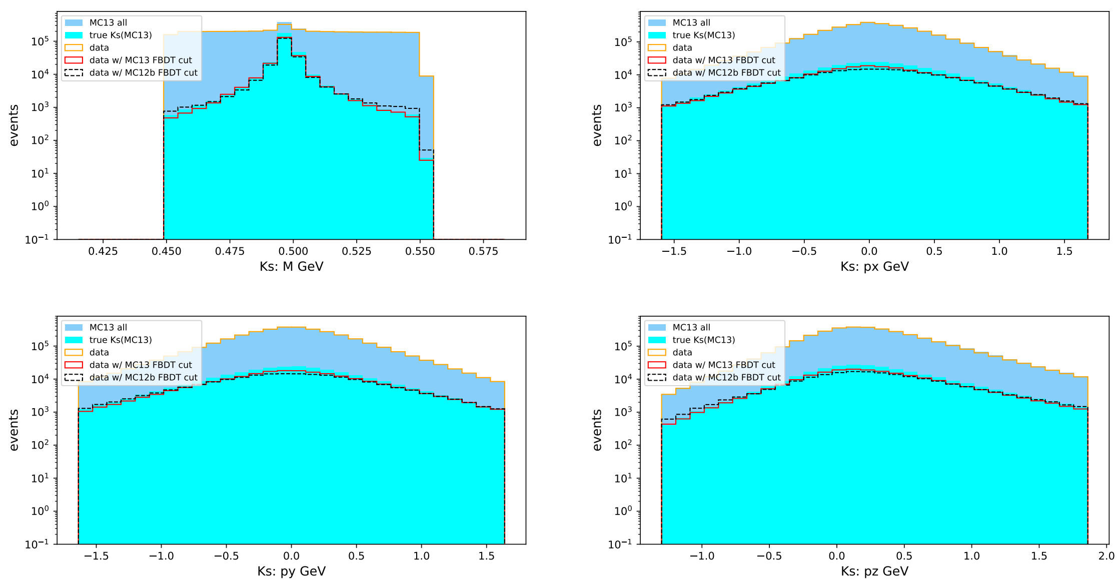


Figure 3-14: The distribution of invariant mass from daughters and the momentum of x,y,z direction. Blue bar is from all MC13, cyan bar is the true K_S^0 in it. Yellow step histogram is data with no cut, solid red data with MC13 trained cut, and the dashed is with MC12b cut. Experimental data has a good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

3.2.8 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 direct 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 check 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 define 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:

$$R = \frac{S_{MC}}{S_{data}} \tag{3.4}$$

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:

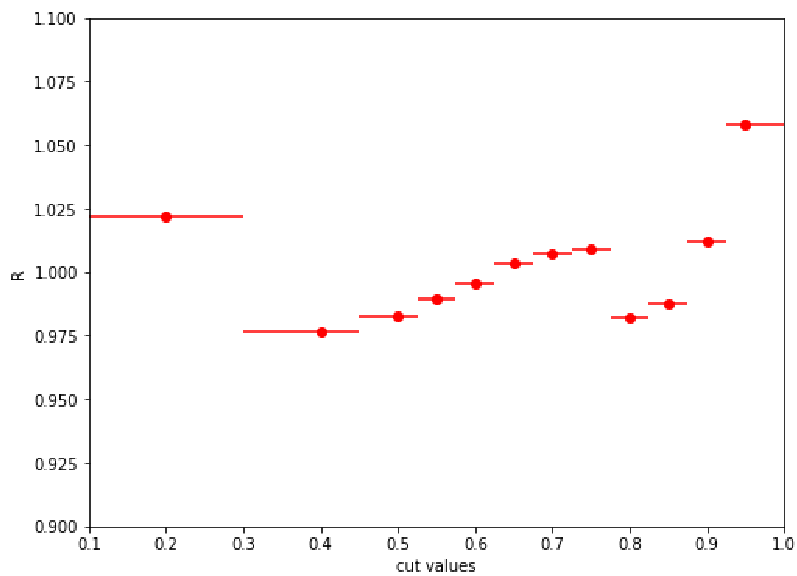


Figure 3-16: Data MC correction induced by K_S^0 classifier.

The R, for example in cut value 0.74 of maximum FOM, is $R = 1.009 \pm 0.081$. In $B^0 \to K_S^0 K_S^0 K_S^0$, 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.

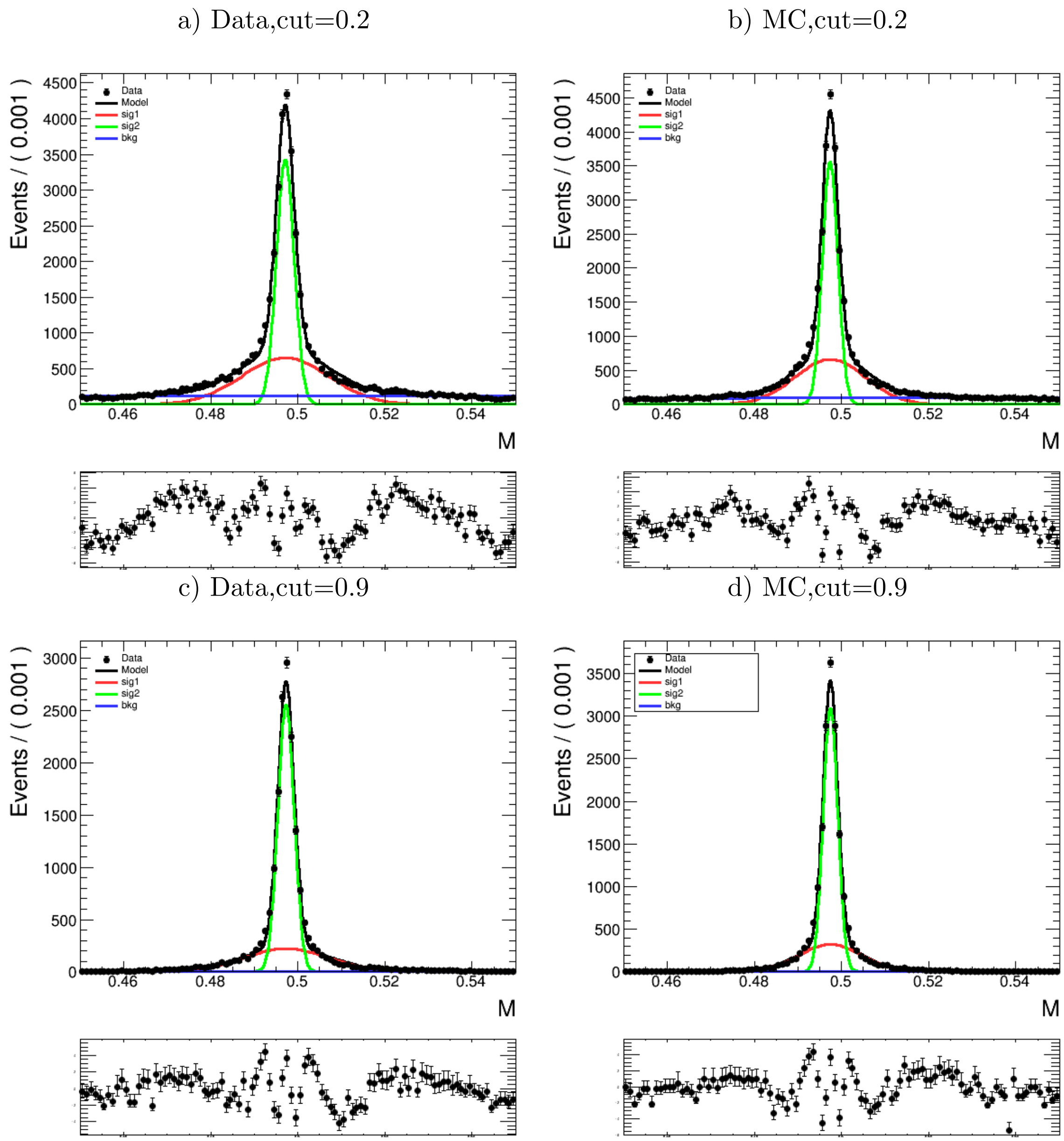


Figure 3-15: Invariant mass fit of K_S^0 using cut at 0.2(loose) and 0.9(tight) to calculate $S_{data/MC}$.

3.2.9 Summary

The development of Belle II K_S^0 classifier is enlighten by the experience from Belle. A comprehensive study of training observables from K_S^0 decay characteristics has been exploited. It takes the advantage of FastBDT algorithm to achieve a high fake rejection power. As a result, classifier is able to give a output which can be used as a cut to select good K_S^0 candidates with high purity. The classifier is validated with real experimental data as well. A primary data validation study of KsFinder is conducted with implementing correction on data and MC along with its contribution to B^0. The performance of KsFinder is in a good shape and no clear bias is found on the yield of the number of K_S^0. For the reconstruction of $B^0 \to K_S^0 K_S^0 K_S^0$, the development of KsFinder is critical to suppress large fraction of combination background from fake K_S^0.

Chapter 4

Analysis Strategies

4.1 Data Sample and Event Selection

The simulation data (MC) is taken from Belle II official MC campaign 13, named MC13. Both signal MC and generic MC are produced. In signal MC, one of the B^0 from $\Upsilon(4S)$ decays into final states $3K_S^0$ then into 6 charged pions, while the other B^0 decays generically using all possible channels. In generic MC, B^0 from $\Upsilon(4S)$ all decay generically using all possible channels.

The signal MC sample is produced by EvtGen package with provided "decay.dec" file that describes the required decay mode and branching fraction. As $K_S^0 \to \pi^0\pi^0$ leads to large background with poor vertex quality, only the final states to 6 charged pions is used. The corresponding branching fraction used is $\mathcal{B}(B^0 \to K_S^0 K_S^0 K_S^0) = 6.0 \times 10^{-6}$. The simulation takes the $\Upsilon(4S)$ as the mother particle and uses decay mode "VSSBMIX" to generate its decay process to two scalar B^0 mesons with mixing. Then the $B^0 \to K_S^0 K_S^0 K_S^0$ is simulated based on only the possible phase-space of kinematics that final states could have, which no CP violation parameters is used, so the generated sample in both generic and signal MC13 has zero input for $\sin 2\phi_1$ and $\mathcal{A}$. In this analysis of $B^0 \to K_S^0 K_S^0 K_S^0$, 1 million signal MC events with and without overlay of beam background are generated separately independent from physics runs, and 1ab^{-1} generic MC is also generated.

4.1.1 K_S^0 Selection

Benefiting from the KsFinder, K_S^0 reconstruction is first based on the cut-based selection, then candidates is purified largely by cutting on the KsFinder output. Here two major additional selections are used in addition to the cut-based method in section 3.1. Firstly, in order to reduce the impact from slow pions from D^* decays, K_S^0 momentum is set to larger than 0.05 GeV/c, see Fig 4-1. Secondly, regarding the different tracking quality of daughter $\pi^{+/-}$, the distribution of fitted invariant mass named "InvM" of K_S^0 are varied. So based on the SVD hits number pions from K_S^0 have, different cuts of fitted invariant mass "InvM" are used on the K_S^0 to further improve the reconstruction purity, see the discussion in detail as Fig 4-5 shows.

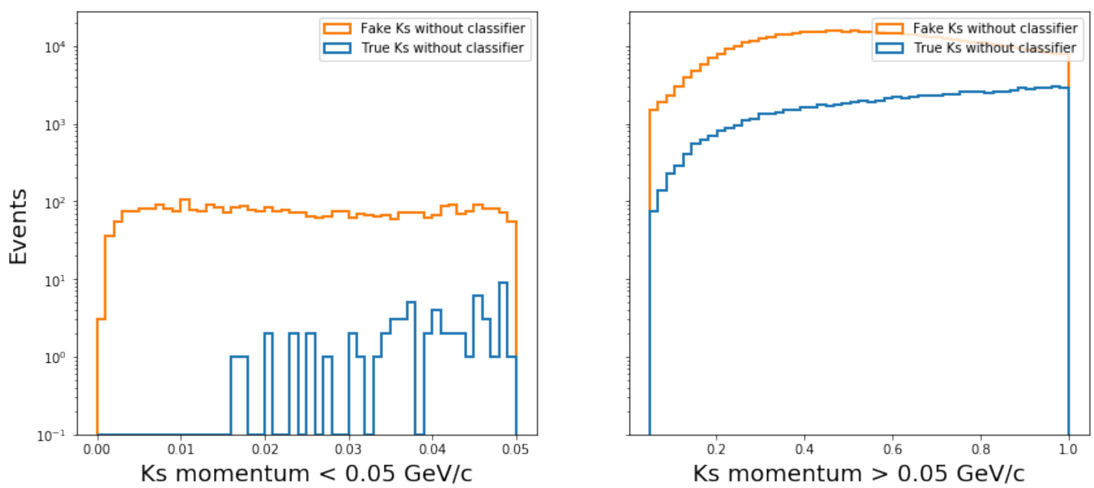


Figure 4-1: The distribution of K_S^0 momentum. Candidates smaller than 0.05GeV/c are rejected.

The reconstruction quality of K_S^0 highly 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. An essential difference between Belle and Belle II is the level of beam background which Belle II is around $10 \sim 20$ times higher regarding the luminosity. So the CDC of Belle II becomes much more sensitive to this effect when doing track finding. Track finding is performed on VXD and CDC separately. The track candidates from different sub-detectors then merged into a same track by extrapolating the trajectory and pattern matching.

However, for certain fraction of K_S^0, they decay outside of layer 5 of SVD, and if these tracks are not passing the windmill structure of SVD, there will be no SVD information recorded. This is due to the feature of SVD track finding, where a track candidate needs either at least 3 SVD hits to form a SVD-only track, or 2 hits to form a hit double-lets, to be used for combining tracks with other track candidates. Single hit case is filtered out to suppress the large fraction of beam background induced by fake hits on a single layer. This effect is shown in Fig 4-2. Similar to Fig 3-1, K_S^0 are categorized based on how many SVD hits their daughters have. SVD10 and SVD01 stands for K_S^0 that only π^+ and π^- has non-zero SVD hits number. SVD11 and SVD00 stands for K_S^0 that both or neither charged pions have SVD hit non-zero SVD hits number. This is related to the track quality of K_S^0 such that SVD11 K_S^0 has the best quality and SVD00 has the worst. It's clear that SVD00 K_S^0 show up at about 10 cm where SVD layer 5 is placed at $r = 10.5$ cm. And most of SVD10(SVD01) K_S^0 also show at same range because one of the pions is passing the overlapping area. The structure of SVD is shown in Fig 4-3 and the fraction for each types of K_S^0 in $B^0 \to K_S^0 K_S^0 K_S^0$ is listed in Table 4.1.

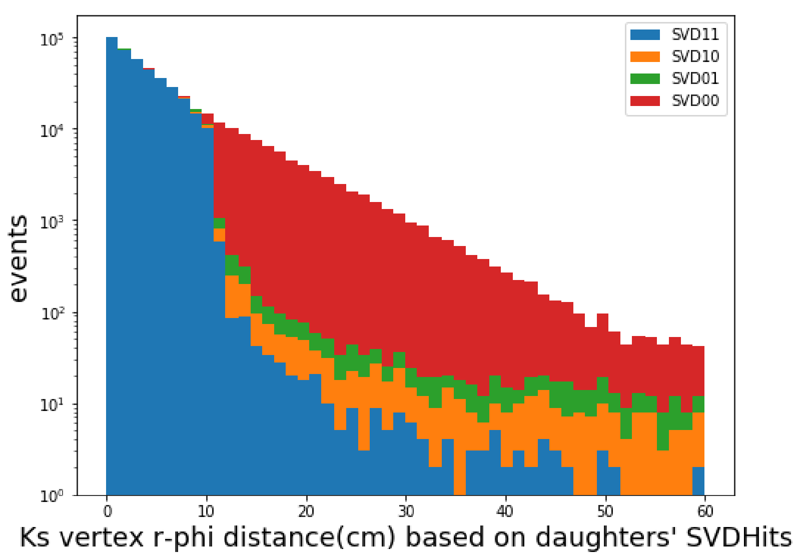


Figure 4-2: K_S^0 flight length on r-ϕ plane based on pions SVD hits, SVD11: both pions have SVD hits, SVD10(SVD01), positive(negative) pions have SVD hits, and SVD00: no SVD hits from pions. The result is from MC13 signal of $B^0 \to K_S^0 K_S^0 K_S^0$.

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

Table 4.1: The fraction of each types K_S^0 based on pions SVD hits in $B^0 \to K_S^0 K_S^0 K_S^0$ MC in Belle I & II.

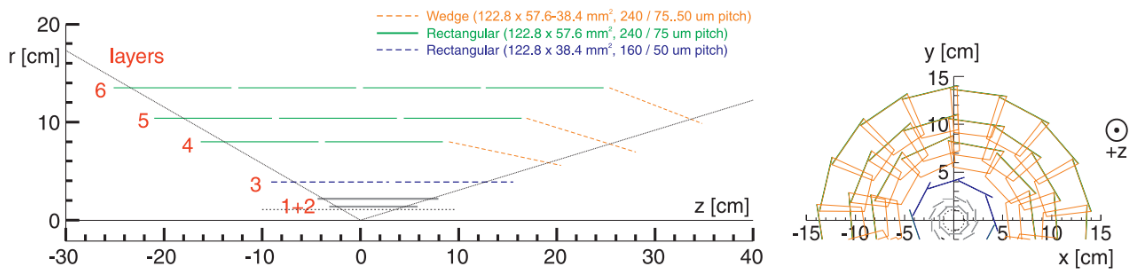


Figure 4-3: Layer 5 is at $r = 11$cm and K_S^0 that decay outside are very likely to lose SVD hits information. Right figure shows the windmill structure with overlapping area at edges.

The vertex fit of K_S^0 is performed using TreeFit package. [18] Using fitted K_S^0 momentum and energy as four-vector, fitted invariant mass (called InvM) is different from the one obtained directly using daughters' four-momentum (calle M). This quantity often receive impact of the measurement uncertainties. If we check out the distribution of fitted invariant mass based on daughters' SVD hits. The dispersion if much severe in $B^0 \to K_S^0 K_S^0 K_S^0$ than it is in generic sample due to the larger fraction of long flight length K_S^0, see Fig 4-4.

Using the fitted invariant mass (the bottom left in Fig 4-4), a slight tuned mass window can be set on each K_S^0 type and improve the purity. We apply the cut on InvM based on K_S^0 types to reduce background in mass sideband. Details of the InvM window are in Table 4.2.

In summary, we first select K_S^0 using "stdKshort:merged" from BASF2 library, which the internal cut on M is 0.45 to 0.55 GeV/c and 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, InvM cut on each K_S^0 types are applied and finally the candidates are selected by KsFinder cut at 0.74.

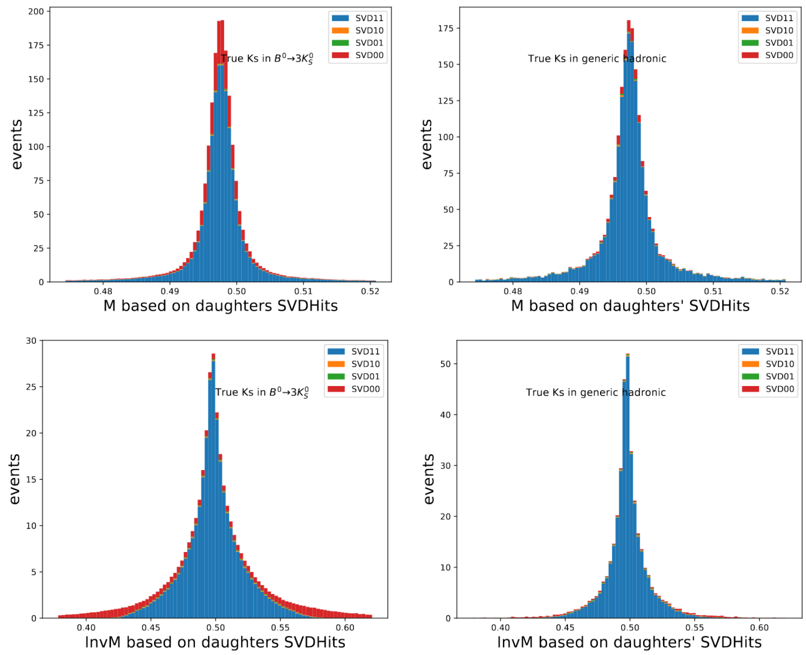


Figure 4-4: (Top two plots)The invariant mass from daughters' 4-vector named M, the left is $B^0 \to K_S^0 K_S^0 K_S^0$ and the right is generic MC; (bottom two plots)The invariant mass from fitted 4-vector named InvM, the left is $B^0 \to K_S^0 K_S^0 K_S^0$ and the right is generic MC; In both cases, the red shows the clear dispersion on SVD00 type K_S^0.

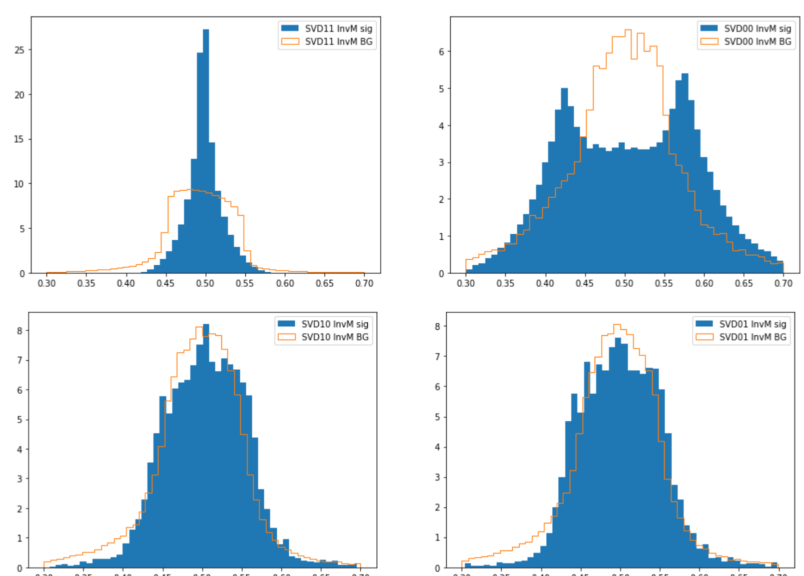


Figure 4-5: Fitted invariant mass on K_S^0 types, sideband is excluded in each distribution to further reject fake K_S^0

K_S^0 type	SVD11	SVD10	SVD01	SVD00
InvM window (GeV)	(0.45,0.55)	(0.38,0.7)	(0.38,0.7)	(0.3,0.7)

Table 4.2: fitted invariant mass window used for K_S^0 based on Fig 4-6.

4.1.2 B^0 Signal 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 ΔE are used to extract signal. The B^0 candidates with $M_{bc} > 5.2$ GeV and $|\Delta E| < 0.2$ GeV are accepted. Then the vertex fit using TreeFit is performed on each B^0 candidate and we keep ones with chiProb > 0.001. When multiple B^0 candidates show up in a single event, a rank by their vertexing quality is performed using χ^2 of the vertex fit, where the smallest have the highest rank. The best candidates are selected by the highest rank. 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 χ_2 from TreeFit before the BCS is used, as shown in Fig 4-6 top right. The distribution of candidates number per event without BCS is shown in top left of Fig 4-6 as well, showing small difference between data and generic MC. The distribution of candidates per event is consistent with the one from signal MC (bottom left of Fig 4-6). The 2D distribution of M_{bc} and ΔE from $B^0 \to K_S^0 K_S^0 K_S^0$ signal MC13 is shown in Fig 4-6 bottom right.

The summary of B^0 selections for is:

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

Table 4.3: B^0 selection criteria

where FBDT_CS is for continuum suppression which is covered in the next section.

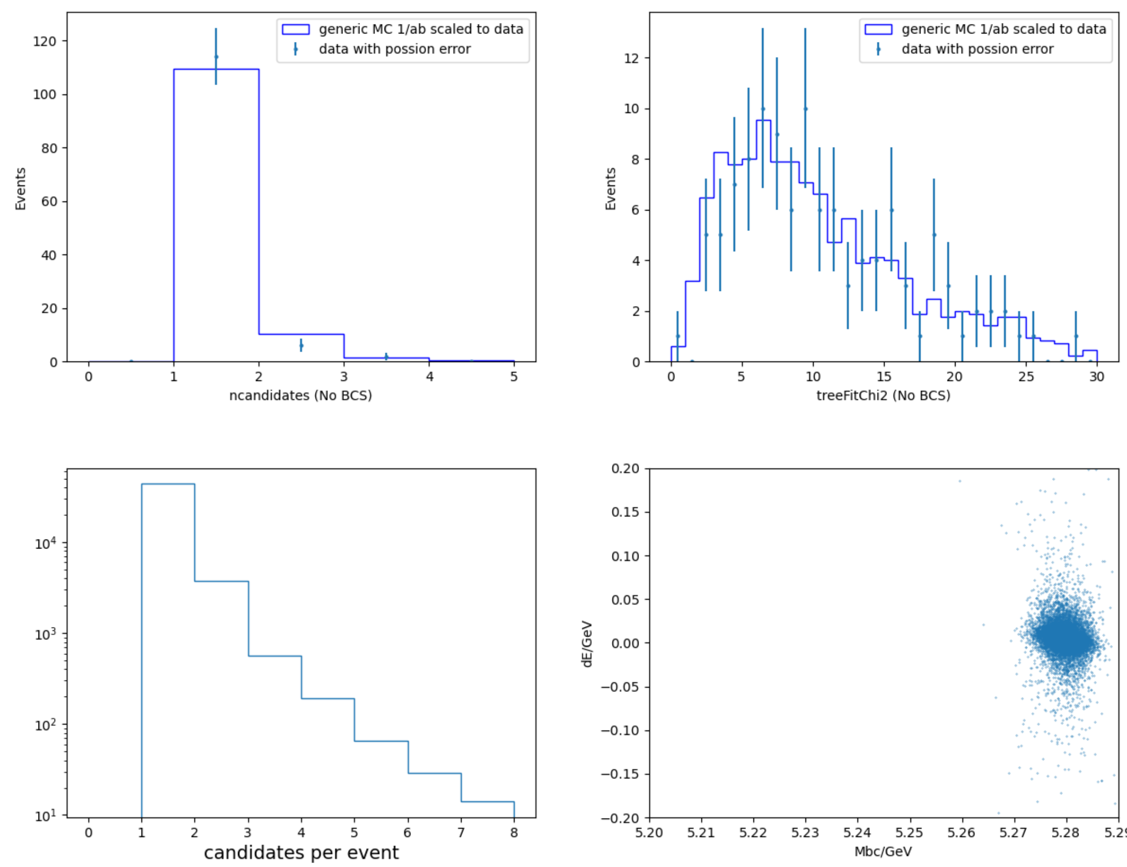


Figure 4-6: Top left is the candidates per event in data and generic MC before the BCS. Top Right is the χ_2 for data and generic MC before BCS. Bottom left is the number of candidates B^0 per event from signal MC. Bottom right is the 2D M_{bc} and ΔE distribution from signal MC.

The Table 4.4 shows that, given the newly developed K_S^0 selection, efficiency of B^0 reconstruction as well as purity 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 4.4: 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 signal. 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 $|\Delta E| < 0.1$, or called "signal region" where efficiency is lower but purity is higher.

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 $e^+e^- \to q\bar{q}$, so called continuum suppression. The background shape projection on spectrum of M_{bc} is a continuum shape. The rejection of this background is essential since they are the dominated background. In the case of $b \to s$ charmless decay like $B^0 \to K_S^0 K_S^0 K_S^0$, the number of continuum background can exceed the signals by a few orders of magnitudes. They can be reduced by taking advantage of the difference on kinematics of their decay products. It's also useful to perform a decay products angular distribution study to further reduce the continuum background. In a $B\bar{B}$ event, two mesons ae produced almost at rest in the CMS frame since the resonance state $\Upsilon(4S)$ is just slightly lighter than beam energy. As a results, decay products fly out more isotopically, compared to other lighter flavor decay which flying-out particles trajectories are more back-to-back distributed. Thanks to experience from ARGUS and CLEO collaboration[19], a set of practically useful variables has been implemented into BASF2 framework that can be used in continuum suppression. CLEO cones momentum can be presented as Eq (4.1), where p_i is momentum of particle i in Rest Of Event (particles in a event except for the ones used to reconstructed CP-side B^0), $\theta_i$is angle against thrust of reconstructed CP-side B meson.

$$L_n = \sum_{i \in ROE} p_i \times |cos\theta_i| \tag{4.1}$$

Not only the variables defined by CLEO cones of momentum are used, in fact, the modified Super Fox-wolfram momentum named KSFW momentum are calculated based on event-shape based variables from every event in Belle II case as well. KSFW momentum are defined as:

$$KSFW = \sum_{l=0}^{4}(R_l^{so} + R_l^{oo}) + \gamma \sum_{n=1}^{N_t} |P(t)_n| \tag{4.2}$$

The first term is:

$$R_l^{so} = \frac{\alpha_{cl} H_{cl}^{so} + \alpha_{nl} H_{nl}^{so} + \alpha_{ml} H_{ml}^{so}}{E_{beam}^{*} - \Delta E} \tag{4.3}$$

when l is odd:

$$H_{nl}^{so} = H_{ml}^{so} = 0 \tag{4.4}$$

and in the meanwhile:

$$H_{cl}^{so} = \sum_i \sum_{jx} Q_i Q_{jx} |p_{jx}| P_l(cos\theta_{i,jx}) \tag{4.5}$$

i runs over B daughters particles and jx for other particles in ROE. Q is charge and p is momentum for each particle. $P_l(cos\theta_{i,jx}$ is the l-th order Legendre polynomial of cosine of i and jx-th particles. On the other hand, for l is even,

$$H_{xl}^{so} = \sum_i \sum_{jx} |p_{jx}| P_l(cos\theta_{i,jx}) \tag{4.6}$$

The second term in Eq (4.2), when l is odd:

$$R_l^{oo} = \sum_j \sum_k \beta_l Q_j Q_k |p_j||p_k| P_l(cos\theta_j, k) \tag{4.7}$$

j and k runs over ROE particles and others are same as Eq(4.8). For even l:

$$R_l^{oo} = \sum_j \sum_k \beta_l |p_j||p_k| P_l(cos\theta_j, k) \tag{4.8}$$

β is five Fisher coefficient to determine in training. Using above definitions, we can form the possibility density functions for KSFW, cosine angle against B meson thrust and ΔZ of two side vertex. Then based on each event's variables' value, we can calculate a ratio as Eq(4.9), where the Likelihood L of signal and background are obtained from the possibility density functions, see Eq(4.10). The $\mathcal{R}$ is:

$$\mathcal{R} = \frac{L_S}{L_S + L_B} \tag{4.9}$$

$$L_{S/B} = P(KSFW)_{S/B} \times P(cos\theta_B)_{S/B} \times P(\Delta Z)_{S/B} \tag{4.10}$$

where P is probability density function for signal and continuum, regarding the discriminating variables in the parentheses. For example, the distribution of a variable called R_2 that is the most important one shown in Fig(4-7) which possibility density function is varied in signal and background.

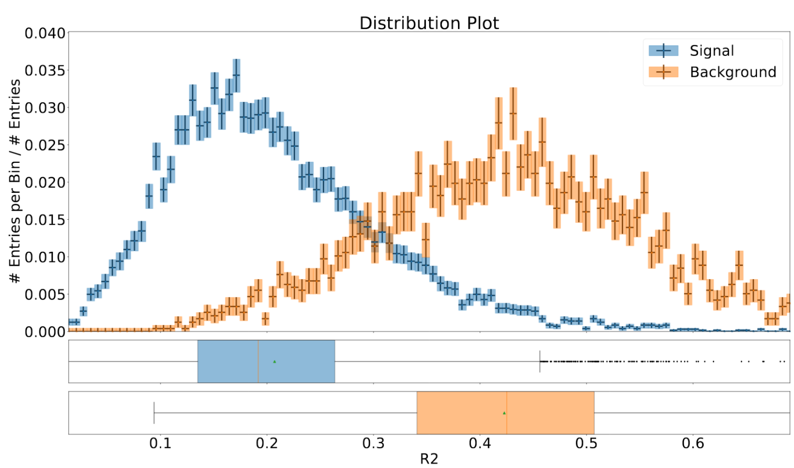


Figure 4-7: R_2 serves as the highest weight as a variable in discriminating the continuum events, having a drastically different distribution between signal and background. R_2 is the ratio of the second to the zeroth KSFW momentum in Eq(4.2)

In order to maximize the performance for this analysis, these categories of variables

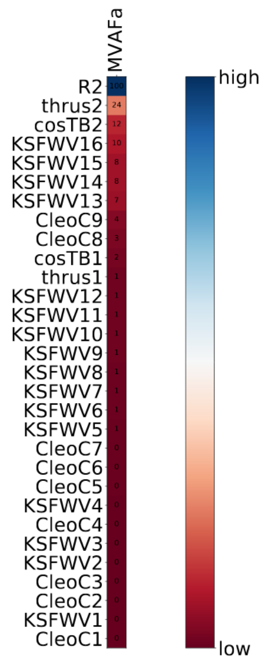
(KSFW, CLEO cone momentum and angular distributions) are combined as a training input for FastBDT classifier, and targeted variables is the continuum event flag. The working principles of this framework is much similar to KsFinder. The MC samples using signal $B\bar{B}$ events and off-resonance in each of the flavor $(q\bar{q})$ are prepared in a scaled ratio of their cross-section in $\Upsilon(4S)$ energy. The same events reconstruction procedures for B^0 is applied as well. Then the events passing the reconstruction for B^0 are used for training the continuum suppression. The fraction of signal and background is set to 1:1 during the training. Noted that the ratio does affect the choice of cut value when applying the suppression. The 1:1 ratio is chosen to make classifier learns about signal and background equally. Then we determine the cut based on the curve of FOM like KsFinder. The variables used in training are listed in Table 4.5 with their abbreviations and importance rank is in Table 4.6.

The correlation between these training variables are shown in Figure 4-8. The correlation among the variables are varied in signal and background. Also, no strong correlation for one variable over the others is found, which shows a good selection of using these variables. The output of continuum suppression classifier is called "FBDT_CS".

Table 4.5: The Abbreviations.

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 4.6: Importance rank



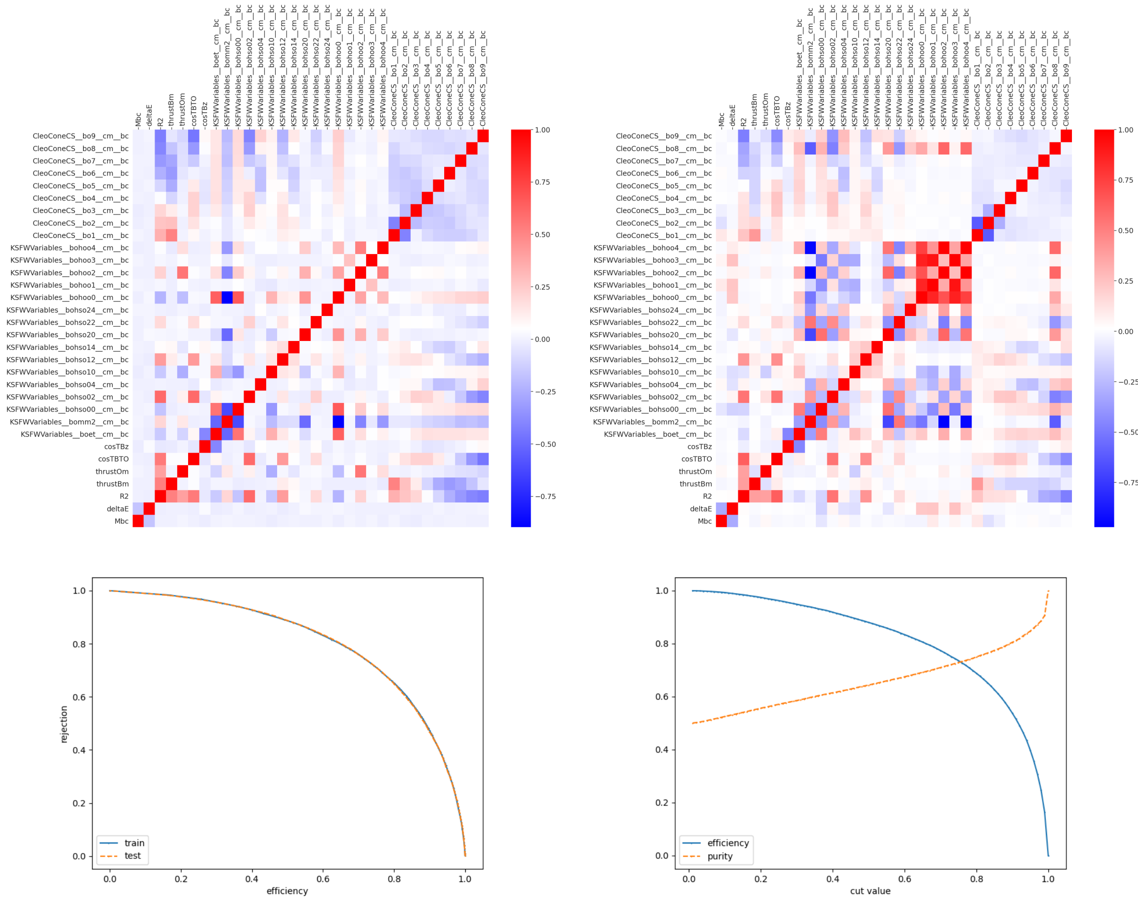


Figure 4-8: The top is correlation in variables for continuum suppression. Top left is for signal and top right is for background. The bottom left is the ROC curve (blue for training and orange for testing, in good consistence) The bottom right is the efficiency(blue) and purity(orange) regarding the cut value “FBDT_CS”

Overtraining check is made by comparing the distribution of signal and background regarding the classifier output in both training and testing samples, which no clear difference is found thus no evidence of over-training found.

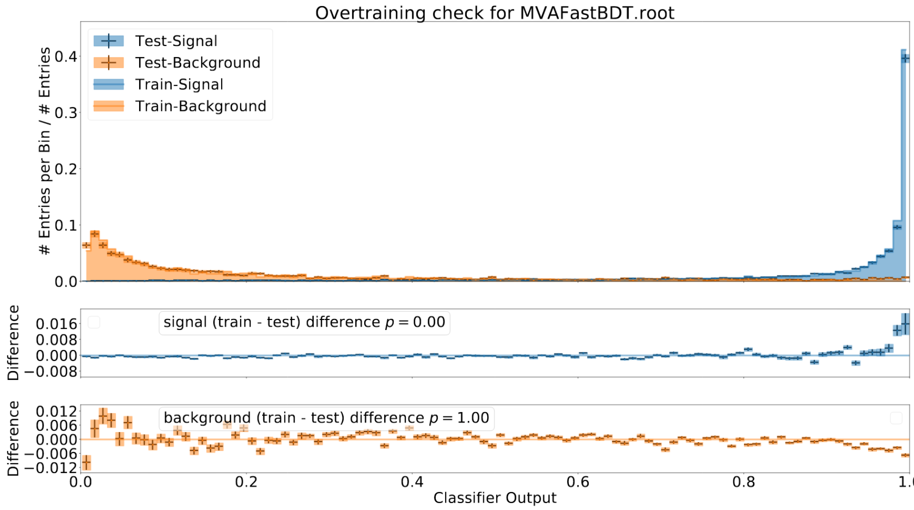


Figure 4-9: Over-training check of continuum classifier, where a very small difference in training and testing (1%) is found.

We implemented the continuum suppression by defining the cut at maximum of FOM using classifier output like Fig(4-10) shows.

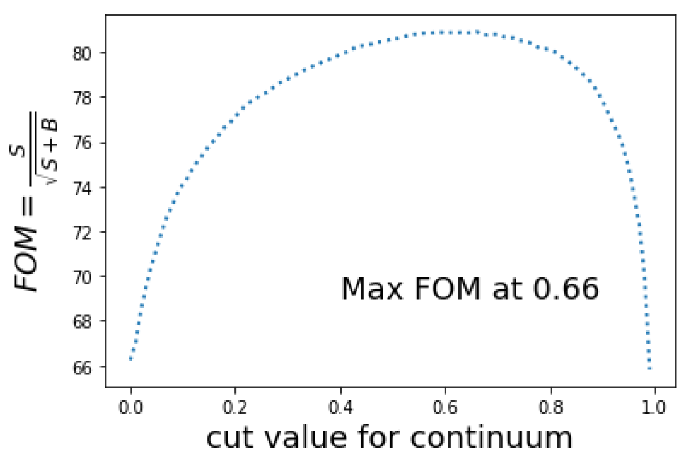


Figure 4-10: FOM depending on the cut value of continuum classifier output, cut value at 0.66 is used for continuum suppression.

4.1.4 Resonance Background

Besides the major contribution from continuum background, charmonium resonance that mediates through $b \to c$ transition brings odd CP eigenvalue in the final states as same as $B^0 \to K_S^0 K_S^0 K_S^0$. 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 $B^0 \to X(K_S^0 K_S^0)K_S^0$, there are two types of resonant events that give out same final states, one is resonant signal and the other is resonant background. For $b \to s$ 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 $f_2(2010)$. And for $b \to c$ transition as resonance background because of CP-odd final states, X could be D^0, J/ψ, $\psi(2S)$, χ_{c0}, χ_{c1}, and χ_{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. Some of these resonance have not been implemented inside generic MC production for Belle II. Given the very limited statistics of data accumulation we used in this analysis, we only present the expected number of these resonances in $400fb^{-1}$ luminosity 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 used for rejecting the resonant background, which should be fine given the estimated background number is about 1 event.

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

Another possible contribution of backgrounds are the background from $B\bar{B}$ events including the charged and the neutral ones. 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 events from $B\bar{B}$ is very limited. Self-cross feed backgrounds stands for the events from the signal-like events but the tag-side particles help form a fake signal. The combined contributions from $B\bar{B}$ background and self-cross feed is about 3% in the channel and we don't perform special treatment on them at current luminosity.

Table 4.7: Expected yield for signal and background resonances $2.14 \times 10^8 B\bar{B}$ in generic MC. The branching fraction of $B \to XK_S$ and $X \to 2K_S$ are listed for both PDG value and value in Belle II DEC. file. The number expected from CP-odd contamination is very low at current luminosity.

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

4.1.6 Signal Extraction

After determining the contribution of all kinds of background, the event selections defined in Table 4.3 is applied to signal MC, generic MC and also the data. The integral luminosity from generic MC is 1ab^{-1} and the data used in this analysis is the full data set from 2019 and 2020 spring and summer. The official processing of data used is proc11 and bucket 9,10,11,13,14,15 and the integral luminosity is about 62.8fb^{-1}. To be noticed, the good runs from data has slightly lower luminosity.

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

$$\mathcal{P}(M_{bc}, \Delta E) = f_{signal} \times \mathcal{P}_{sig}^{M_{bc}} \times \mathcal{P}_{sig}^{\Delta E} + (1 - f_{signal})\mathcal{P}_{bkg}^{M_{bc}} \times \mathcal{P}_{bkg}^{\Delta E} \tag{4.11}$$

where $\mathcal{P}_{sig}^{M_{bc}}$ and $\mathcal{P}_{sig}^{\Delta E}$ are Gaussian and three Gaussians function. f_{signal} is fraction of signal events. $\mathcal{P}_{bkg}^{M_{bc}}$ is primarily continuum events, and presented as Argus distribution like Eq(4.12) shows, with a preset mass threshold at $c = 5.29$ GeV.

$$f(x; \chi, c) = \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} \cdot \frac{x}{c^2}\sqrt{1 - \frac{x^2}{c^2}} \cdot exp\left\{-\tfrac{1}{2}\chi^2(1 - \tfrac{x^2}{c^2})\right\} \tag{4.12}$$

x is defined in $0 < x < c$. χ 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. $\mathcal{P}_{bkg}^{\Delta E}$ is modeled by 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 2D on M_{bc} and ΔE for generic sample and data. Fitting results on signal MC are show in Fig 4-11.

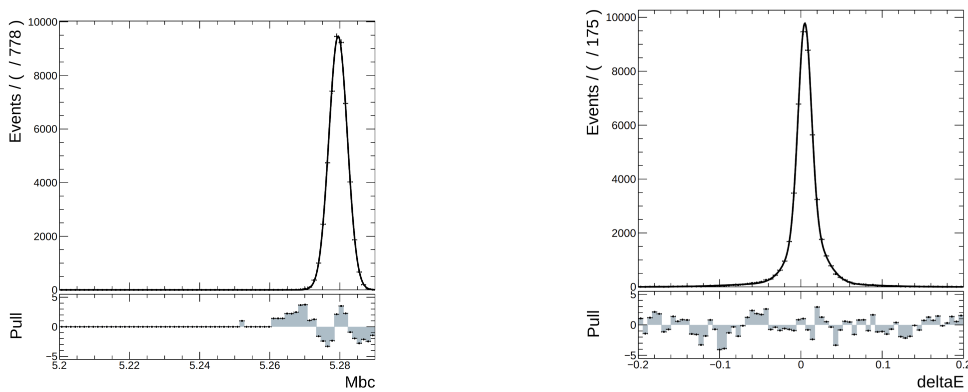


Figure 4-11: The distribution of M_{bc} and ΔE of signal MC13 of $B^0 \to K_S^0 K_S^0 K_S^0$ fitted with single and triple gaussian respectively.

The continuum background is fitted also first on off-resonance generic samples to determine the shapes then fix them as constants for 2D fit, shown as follows:

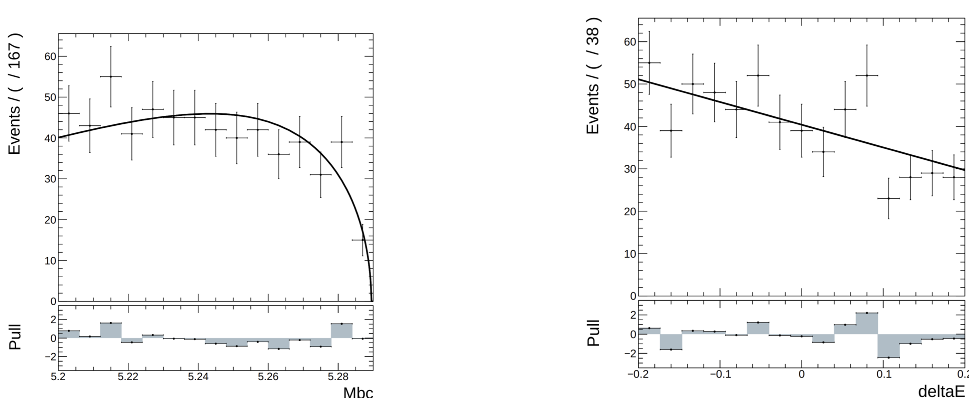


Figure 4-12: The distribution of M_{bc} and ΔE of continuum events fitted with Argus and Chebyshev polynomial respectively.

Then we set the events number for signal and background as the only parameter in float, to use Eq(4.11) as 2D fit functions on 1ab^{-1} generic sample and data, which the fit is also done using unbinned maximum likelihood fit. For B^0 in generic MC sample, the 2D fit result projected on M_{bc} and ΔE is:

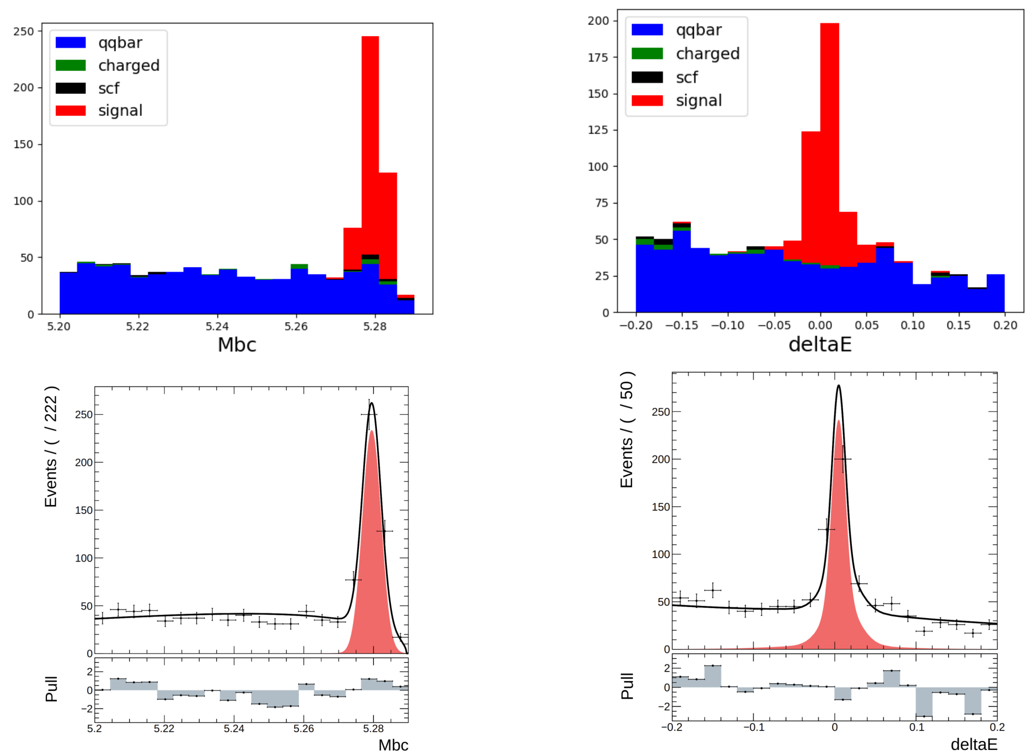


Figure 4-13: Top is the stacked plots for generic MC of M_{bc} and ΔE, where each background components are stacked with signal. The bottom is the 2D fit on generic MC13, the red is signal PDF. The integral luminosity is 1 ab^{-1}

For the invariant mass of K_S^0 from B^0, we compared the distribution of “InvM” and “M” for all K_S^0 that have been used for B^0 reconstruction in data and generic MC. The distributions are shown in Fig 4-14 bottom, which the generic MC is scaled to the luminosity of data and agreed well. For data, the 2D fit result projected on M_{bc} and ΔE is in Fig 4-14 top.

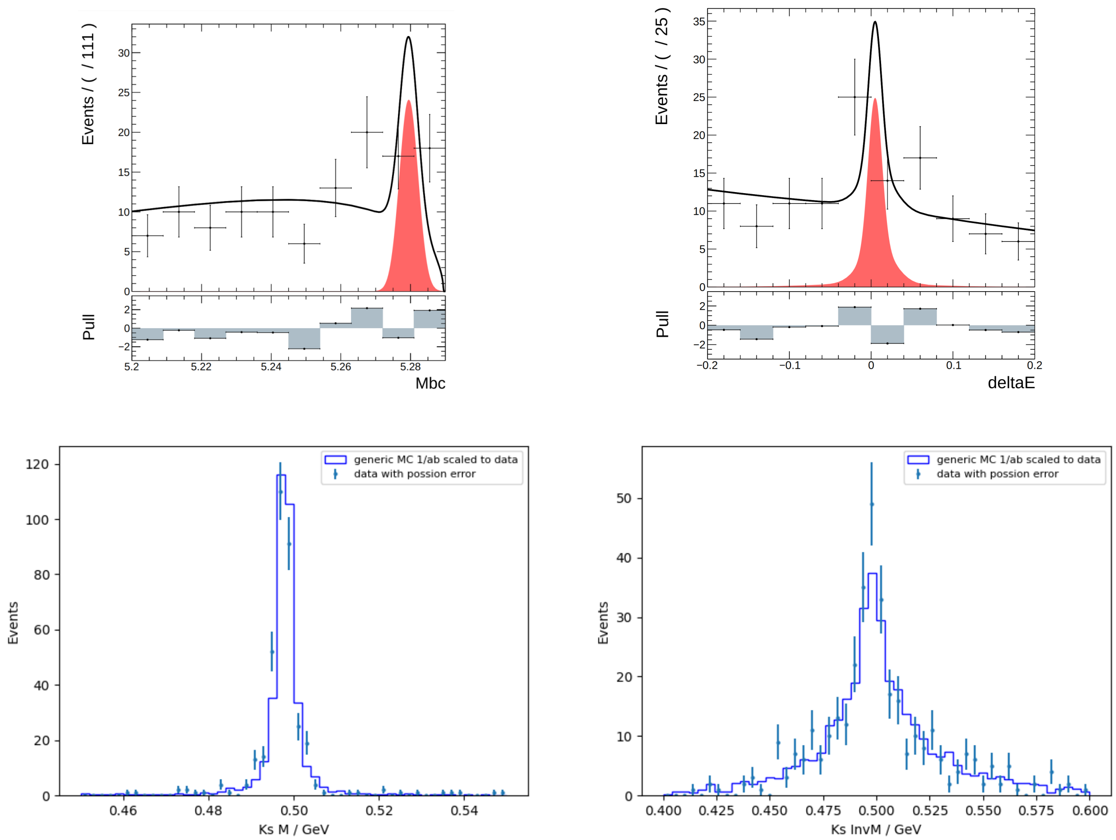
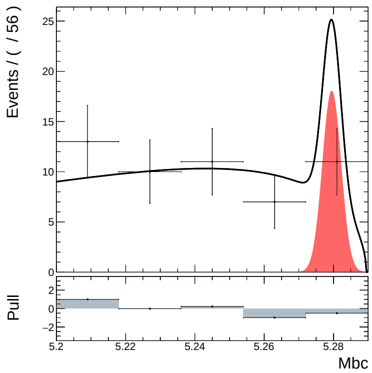


Figure 4-14: Top: M_{bc} and ΔE 2D fit using 62.8 fb^{-1}data, the red is signal PDF. Bottom: "M" and "InvM" from data and generic MC which is scaled to data luminosity. The distributions is in a reasonable consistency.

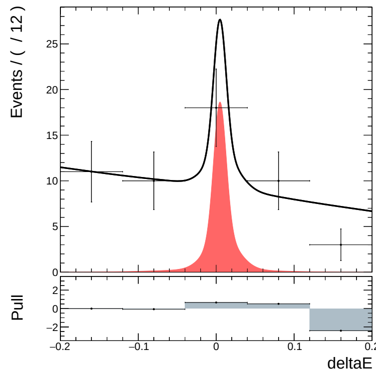
The number of signal events is extracted by integral of fit function over the signal region which is defined as $5.27 < M_{bc} < 5.29$ GeV and $-0.1 < \Delta E < 0.1$ GeV. The expected signal events with 35% efficiency in this analysis is calculated as:

$$\mathcal{B}(B^0 \to K_S^0 K_S^0 K_S^0) = \frac{N_{sig}}{\mathcal{B}(K_S^0 \to \pi^+\pi^-)^3 \times \epsilon_{rec} \times N_{B\bar{B}}} \tag{4.13}$$

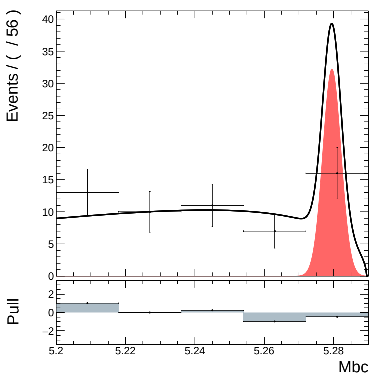
In 1 ab^{-1} generic MC13, $7.7 \times 10^8 \times 6 \times 10^{-6} (BR : B^0 \to 3K_S^0) \times 21\% (BR : K_S^0 \to \pi^+\pi^-) \times 35\% \simeq 339$. The fit result from $M_{bc}/\Delta E$ yields 341 ± 20 events which agrees with expectation. The event number in sideband $M_{bc} < 5.26$ GeV in generic MC is 507. Compared to Belle result with 772×10^6 $B\overline{B}$ pairs used, signal from data yields 327 ± 19, which is also consistent. In 62.8 fb^{-1} data from Belle II, we extract $N_{sig} = 17.4 \pm 4.2$ in this region. Considering the good runs luminosity will be slightly lower than the recorded, the number of signal is in a good agreement with expectation. The sideband region $M_{bc} < 5.26$ GeV contains 60 events in data. Both the results of expected event number for signals are consistent in generic MC and data. To check linearity of the event number fitted from the M_{bc} and ΔE in this low statistics case, we extract the fraction of continuum backgrounds from generic MC13 sample rescaled to the experimental data luminosity, which yields about 46.5 background events in 62.8fb^{-1}. Then the series of (5,10,15,20,25,30) signal MC13 events is injected into the background, to perform the $M_{bc}/\Delta E$ fit to check the output signal events number. The fitted yield from $M_{bc}/\Delta E$ shows a good agreement on both signal and background events number in data equivalent luminosity. M_{bc} and ΔE distribution in each injection test are shown as well. Details about the linearity of signal events are in Fig 4-16.



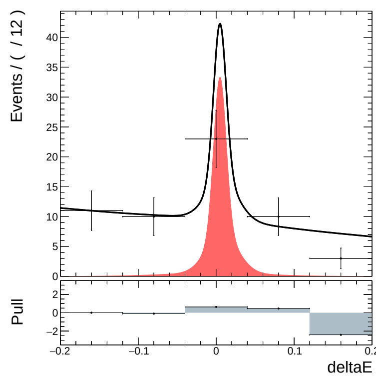
a) signal injected: 5



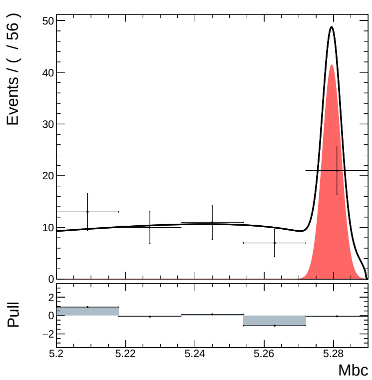
b) signal injected: 5



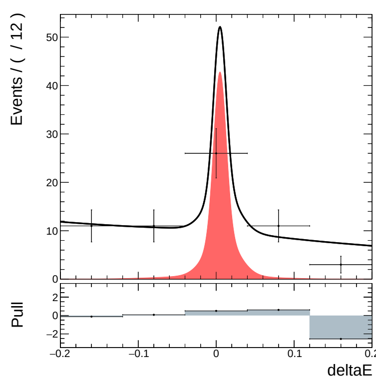
c) signal injected: 10



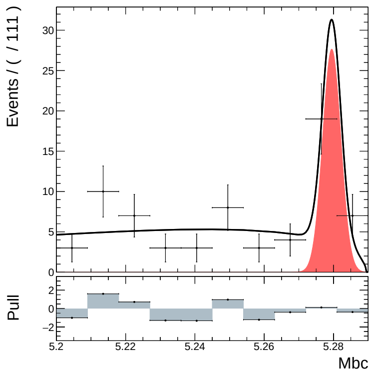
d) signal injected: 10



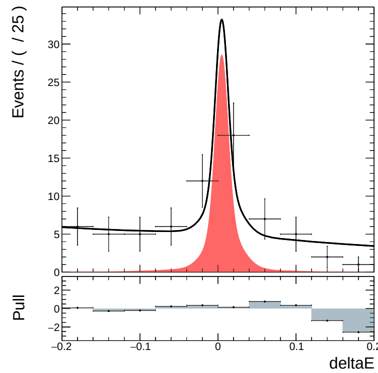
e) signal injected: 15



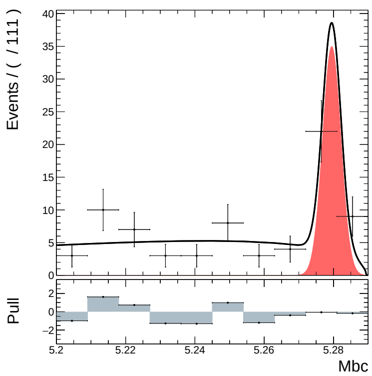
f) signal injected: 15



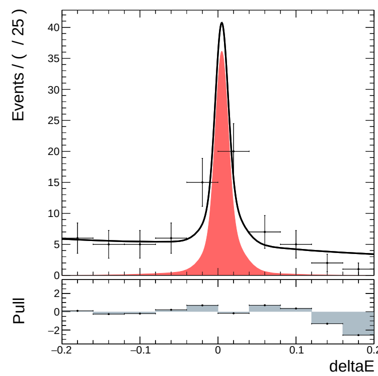
g) signal injected: 20



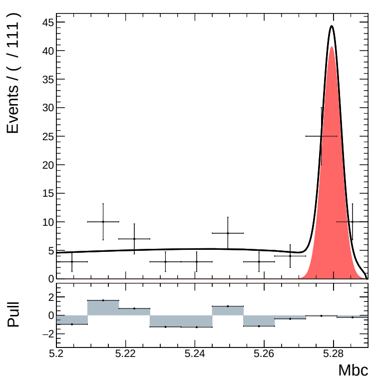
h) signal injected: 20



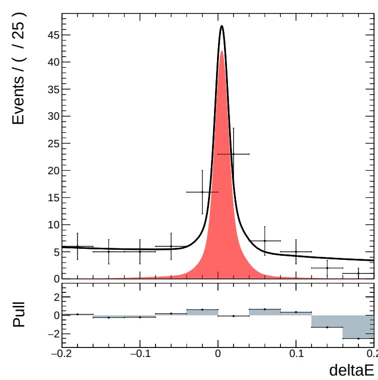
i) signal injected: 25



j) signal injected: 25



k) signal injected: 30



l) signal injected: 30

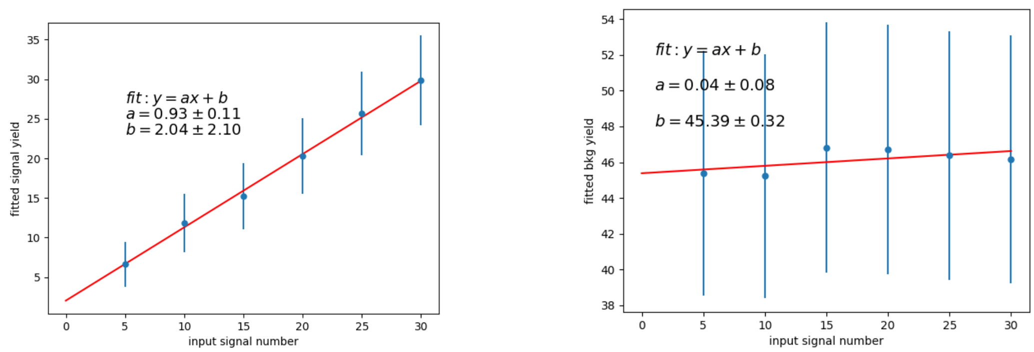


Figure 4-16: Injection test for signal extraction. The linearity is clear between input and output signal events number.

4.1.7 Kinematics and Vertexing Dependence on KsFinder

KsFinder largely reduce the combinatorial background of B^0. The previous section shows a good reconstruction performance at low statistics in early phase 3 data. Without the power of rejection provided by K_S^0 finder, rediscovery of $B^0 \to K_S^0 K_S^0 K_S^0$ in early phase 3 of Belle II won't be feasible.

However, it's essential to check the potential impact on kinematics and vertexing of K_S^0 regarding the implementation of KsFinder. The K_S^0 classification uses information such as invariant mass and decay vertex which may propagate bias into B^0 vertexing, eventually may affect CP parameters. Besides, kinematics bias might tune the shape of fit function and contribute bias in defining signal fraction used in CP fit. At last, since we have different types of K_S^0 based on their SVD hit numbers, a check with B^0 based on different K_S^0 types are also worthy of doing.

Given each type of B^0 based on how many CDC-only tracks it has in final states, the check on M_{bc} and ΔE with or without KsFinder is done by fitting the distribution in signal. M_{bc} and ΔE are modeled by double-Gaussian. Compared corresponding fit results, no clear bias on M_{bc} and ΔE are found by using KsFinder or not. Results are plotted in Fig 4-17 and 4-18. The main distribution of M_{bc} and ΔE show very trivial difference.

Similar to the comparison of M_{bc} and ΔE, Z direction vertex position along with

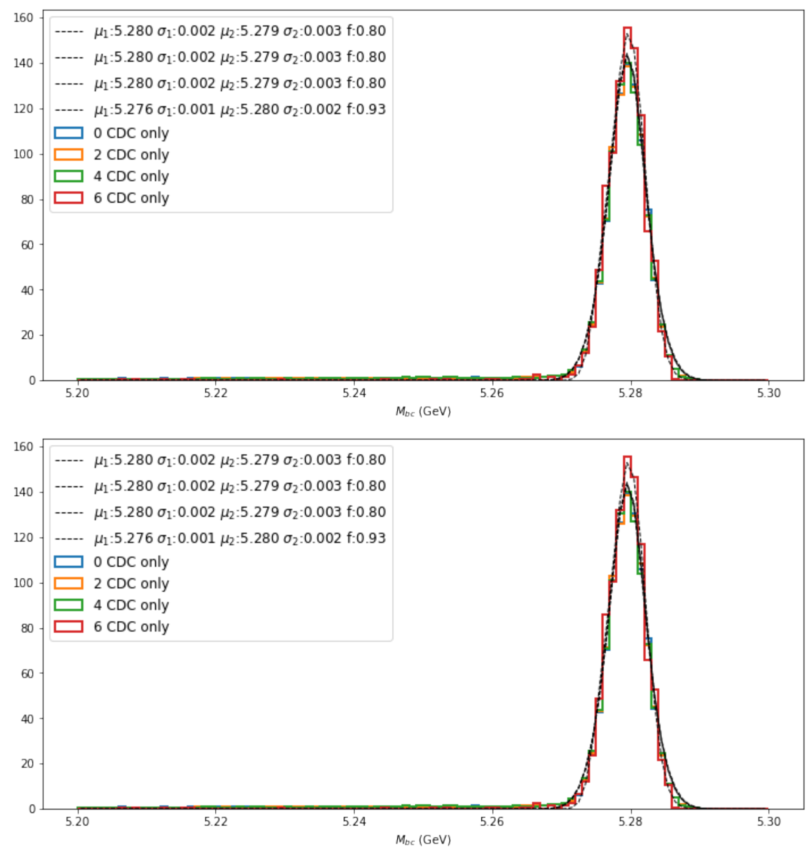


Figure 4-17: M_{bc} distribution based number of CDC-only tracks in final states. Up: no K_S^0 finder; Down: K_S^0 finder used. f is the fraction for the first Gaussian

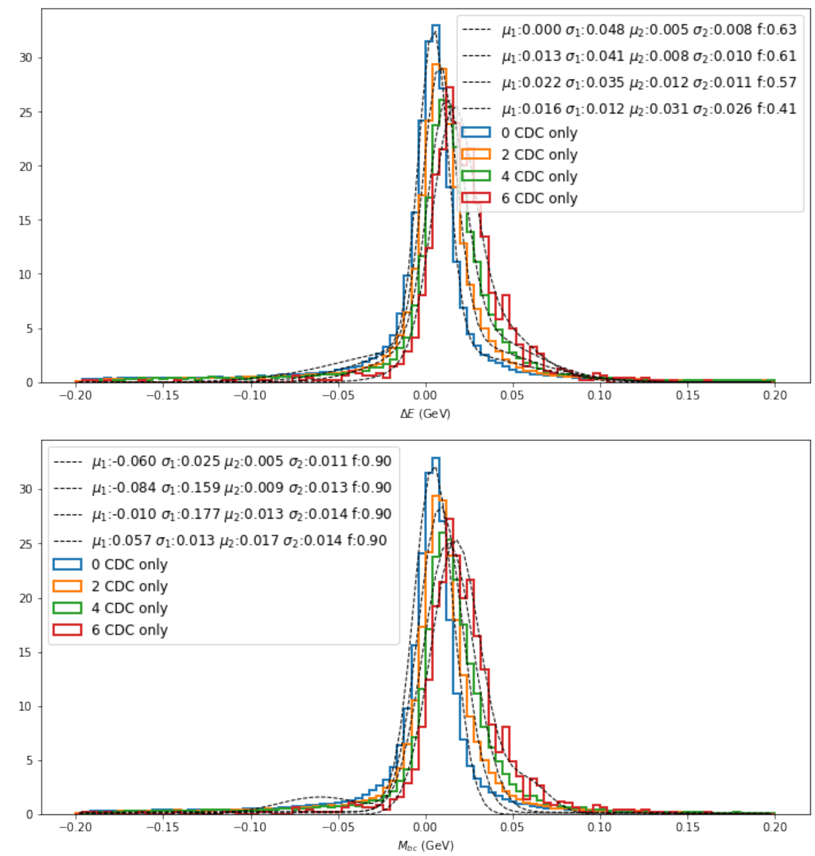


Figure 4-18: ΔE distribution based number of CDC-only tracks in final states. Up: no K_S^0 finder; Down: K_S^0 finder used. f is the fraction for the first Gaussian

its uncertainties are also checked. No clear bias on vertex position and uncertainties are spotted either. Results are plotted in Fig 4.19 and 4.20. And it obivious that the resolution of vertex on z-axis is much worse when the final states of B^0 only have CDC tracks.

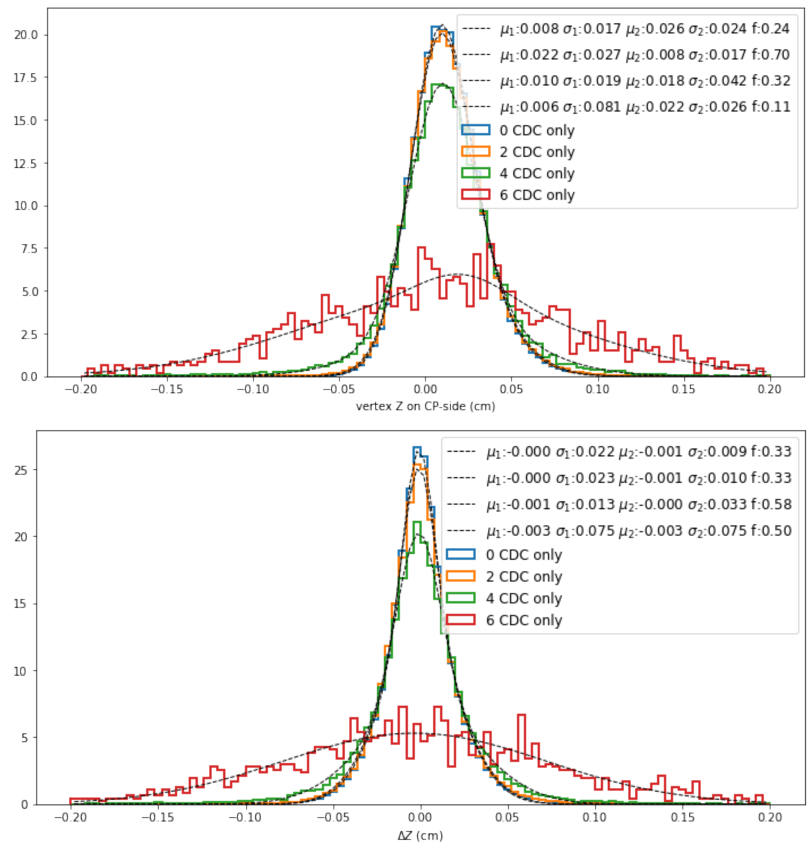


Figure 4-19: Δz distribution based number of CDC-only tracks in final states. Up: no K_S^0 finder; Down: K_S^0 finder used. f is the fraction for the first Gaussian

Above all, no clear sign of bias on both kinematics and vertex position from using K_S^0 finder has been found, KsFinder implements a small shift on the vertex positions which is trivial compared to the large statistical uncertainty. For the moment, there's no correction on these variables are considered based on the neglect able difference from using KsFinder. And further effects will be watched in future improvement of KsFinder.

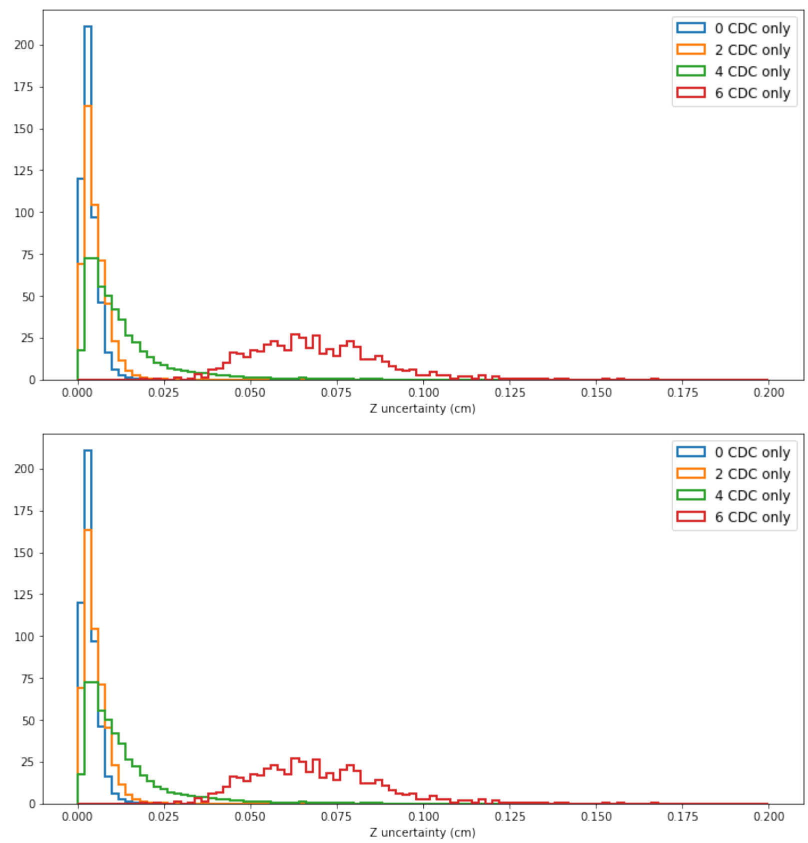


Figure 4-20: $\delta\Delta z$ distribution based number of CDC-only tracks in final states. Up: no K_S^0 finder; Down: K_S^0 finder used.

4.1.8 Summary

To properly reconstruct B^0, a typical cut-based loose selection of K_S^0 is first done and then developed KsFinder using FastBDT classification is applied. Due to the tracking issue regarding the K_S^0 daughters missing SVD hits, fit mass with varied cuts are used in addition. Combining K_S^0 candidates and implementing continuum suppression from event-shape analysis yield signal B^0 with relatively low background in full range. 2D fit on M_{bc} and ΔE are performed to extract signal and background number inside defined signal box. The results are in a good agreement in data and MC, as well as Belle result. Potential data MC correction of using KsFinder has been studied. The potential bias on kinematics and vertexing are also studied, which are in a trivial level under the current luminosity.

Chapter 5

CP Parameters measurement

The measurement of CP parameters requires the fit to the time-dependent distribution of events based on the decay time difference, flavor, and CP parameters, in which Δt is decay time difference and q is flavor of tag-side meson. $q = \pm 1$ when tag-side is B^0 or $\bar{B}^0$. The physical distribution of events satisfies the function in Eq(5.1), where we use the notations that is slightly different than Chapter 1. The S_f and A_f which stand for mixing induced and direct CP violation that are noted as $\mathcal{S}$ and $\mathcal{A}$.

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

In above function, it's parameterized by $\mathcal{S}$ and $\mathcal{A}$. A complete model that includes the overlay of background components and outlier bands can be defined as :

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

where $\mathcal{P}_{bkg}$ and $\mathcal{P}_{ol}$ are defined by:

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

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

$\delta(\Delta t_i - \mu_{bkg}^{\delta})$ is Dirac type δ function. Eq(5.2) is using Δt_i as i-th event's decay time difference $\Delta t_i = \Delta z/\beta\gamma c$ in data set, where $\Delta z = z_{cp} - z_{tag}$.

5.1 Vertex Resolution Model

5.1.1 *CP*-side resolution function

Eq(5.2) presents an ideal distribution of Δt_i without considering the difference between measured vertex position and the physical ones. The difference can be described by introducing resolution functions. If considering resolution function, Eq(5.2) turns into:

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

R_{sig} stands as resolution function for signal component, which receives smearing effect from *CP* and tag side, 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 reconstruction of B^0 so vertex fit is performed by geometric fitting with selected charged tracks. The background events has its own resolution model which is independent from any *CP* violation parameters. The outlier is used to compensate and smooth the long tails when Δt becomes very large. In this analysis, we don't include the outlier to have a more realistic model under very low statistics.

Considered only the detector effect on vertex difference along z-axis Δz.

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

where the primed ones stands for physical truth of position and the non-primed is measured. Footnotes are from either CP side or tag side position. The resolution function clearly comes from CP and tag-side, so for signal events, we determine the resolution models from CP and tag-side separately. The resolutions on both sides also depends on the fitting method and the applied constraint, so to generalize the options used, there's no IP constraint from both side when the vertex is fitted, which avoids potential bias from IP profile under this low statistical situation.

CP-side vertex is fitted with TreeFit as mentioned in the previous section where all tracks from a reconstructed event is used, thus the resolution models only depends on detectors' performance. For the CP-side vertex from TreeFit, the vertex resolution has relation towards the reconstruction quality, or to be more specific, the χ_2/N value, which χ_2 represents the goodness of fit and N is the fitting's degree of freedom. The distribution of χ_2/N (called as reduced χ_2) in selected events are shown below:

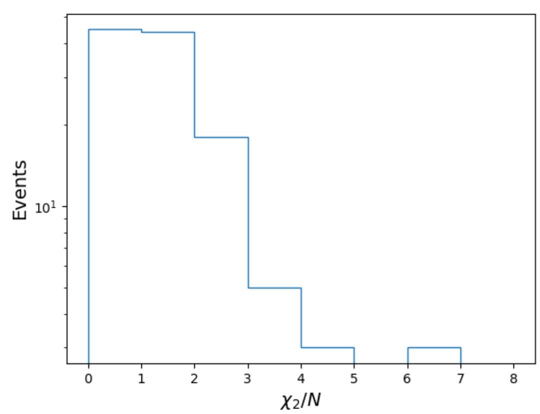


Figure 5-1: χ_2/N of selected events from data.

Based on this, we model the resolution functions on CP-side vertexing using a double-gaussian function, whose mean is set to zero and the standard deviation is scaled by χ_2/N and the error of vertexing.

$$R_{cp}(\delta z_{cp}) = (1 - f_{cp}^{tail})G(0, s_{cp}^{main}) + f_{cp}^{tail}G(0, s_{cp}^{tail}) \tag{5.7}$$

where:

$$
\begin{aligned}
s_{cp}^{main} &= (s_0^{main} + s_1^{main} \cdot \chi^2_{cp}/NDF) \cdot \sigma_{z_{cp}} \\
s_{cp}^{tail} &= (s_0^{tail} + s_1^{tail} \cdot \chi^2_{cp}/NDF) \cdot \sigma_{z_{cp}}
\end{aligned}
\tag{5.8}
$$

The dependence of resolution models on χ_2/N is shown as follows. Restrictively speaking, the CP-side resolution for $B^0 \to K_S^0 K_S^0 K_S^0$ is slight different from other channels such as $B^0 \to J/\psi K_S^0$, due to the lack of the direct charged tracks from the B^0 vertex. The way of scaling the resolution will be further studied along with more data available. Given the current low statistics, the above model works well as an approximation. By fitting the resolution functions from signal MC on CP-side, we could determine the parameters in Eq(5.7), and it's listed in Table 5.1.

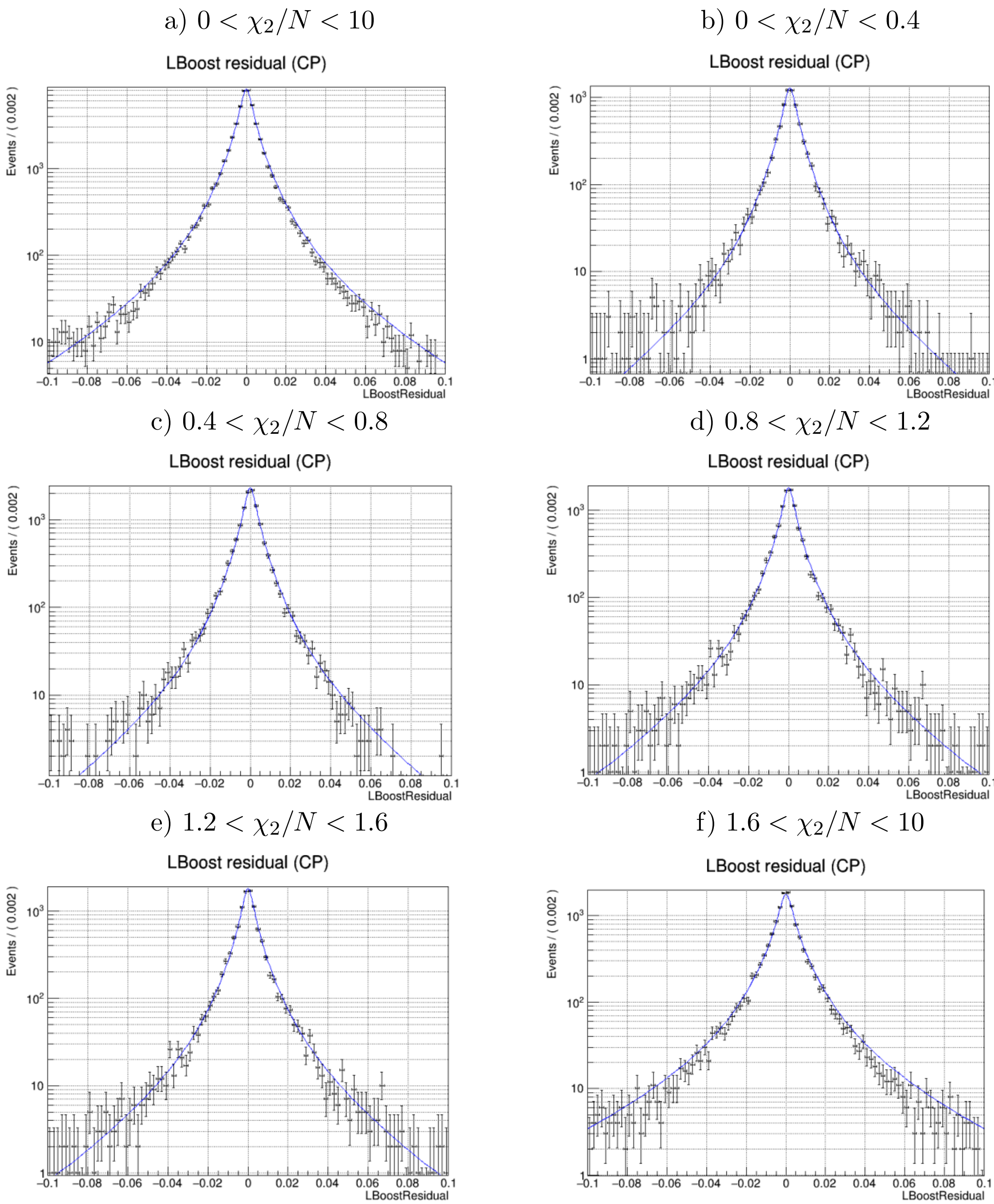


Figure 5-2: Resolution functions on CP-side, which shows dependence on the χ_2/N

Table 5.1: Parameters in R_{cp}.

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

5.1.2 Tag-side resolution function

For the tag-side, the vertexing is done by using KFit and No IP constraint. Since we don't reconstruct certain decay channel like CP-side, there's no guarantee that tracks used by KFit are all from the primary vertex. This leads to the resolution degradation on tag side vertexing called R_{np}. To the contrary, if all tracks that are used for tag vertexing are from primary vertex, then resolution will only be affected by the detector, and this is called R_{det}, similarly to R_{cp}. The model is demonstrated as below:

$$\begin{aligned} z_{tag} - z'_{tag} =& z'_{tag} + \delta z^{det}_{tag} + \delta z^{np}_{tag} - z'_{tag} \\ =& \delta z^{det}_{tag} + \delta z^{np}_{tag} \end{aligned} \tag{5.9}$$

$$R_{tag}(z_{tag} - z'_{tag}) = R^{tag}_{det}(\delta z^{det}_{tag}) \circledast R^{tag}_{np}(\delta z^{np}_{tag}) \tag{5.10}$$

$$R^{tag}_{det}(\delta z^{det}_{tag}) = (1 - f^{tail}_{tag}) G(0, s^{main}_{tag} \cdot \sigma_{z_{tag}}) + f^{tail}_{tag} G(0, s^{tail}_{tag} \cdot \sigma_{z_{tag}}) \tag{5.11}$$

where main and tail parts share same main value at zero, but deviation is defined by using uncertainties in each event χ^2_{tag}/N:

$$s^{main/tail}_{tag} = s^{main/tail}_0 + s^{main/tail}_1 \cdot \chi^2_{tag}/NDF \tag{5.12}$$

where χ^2_{tag}/N is returned by tag side vertexing tool on event-by-event basis. Technically R^{tag}_{det} can be fitted with MC of which tag side tracks are all from primary vertex. Then after fixing the fitted parameters of R^{tag}_{det} , R^{tag} will only be dependent on R^{tag}_{np}.

As for the fit of R_{np}, similar event-by-event fit is also implemented. The fit model of R_{np} is inherited Belle experience. It consists of three functions, including one Dirac type δ function and two single side exponential.

$$R^{np}_{tag}(\delta z^{np}_{tag}) = f_\delta \delta(\delta z^{np}_{tag}) + (1 - f_\delta)[f_p E_p(\delta z^{np}_{tag}, \tau_p \cdot \sigma_{z_{tag}}) + (1 - f_p) E_n(\delta z^{np}_{tag}, \tau_n \cdot \sigma_{z_{tag}})] \tag{5.13}$$

where $E_p(x, \tau_p) = (1/\tau_p)exp(-x/\tau_p)$ when $x > 0$ and $E_n(x, \tau_n) = (1/\tau_n)exp(x/\tau_n)$ when $x < 0$.

Also, since tag-side has no dependence of how CP-side decays, its resolution is 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[20]. The fit results for tag-side resolution is shown in Fig 5-3 and 5-4 and Table 5.3 and Table 5.4.

Table 5.2: Parameters in R_{cp}.

f^{tail}_{cp}	0.07424 ± 0.0008
s^{main}_0	0.9151 ± 0.0077
s^{main}_1	0.2142 ± 0.0064
s^{tail}_0	2.0477 ± 0.0779
s^{tail}_1	1.3470 ± 0.0720

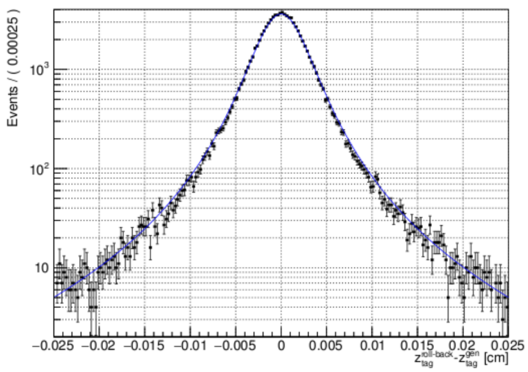


Figure 5-3: R_{det}^{tag} fit

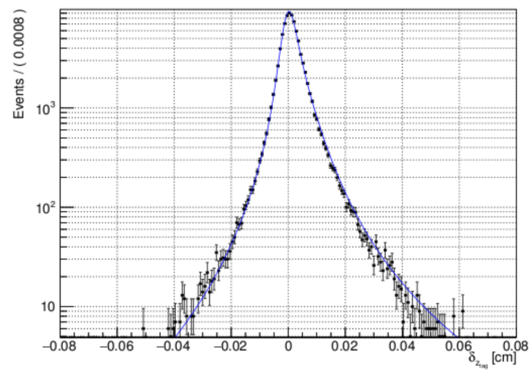


Figure 5-4: R_{np}^{tag} fit

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

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

Table 5.4: Parameters in R_{np}^{tag}

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

Last but not least, the boost direction of each event is not constant event-by-event, so the z-axis position of vertex may not be optimized by calculating $\Delta t_i = \Delta z/\beta\gamma c$. This effect can be reduced by replacing z position with the relative distance along the boosting direction, or introducing another resolution function called R_k. In the current modeling of Belle II tag side resolution, distance along the boost direction is available so R_k is not yet implemented. So we use the ΔZ projection on the boosted direction of each event to account for this effect when we define the parameters of resolution and also for the final fit to Eq(5.5).

5.1.3 Background events distribution

Instead, R_{bkg} is in a simpler shape and it's uncorrelated to vertexing method approximately. Because the source of background mainly comes from continuum events passing the selection, it's reasonable to model its resolution by Gaussian-like function. We define a double-Gaussian with its standard deviation scaled by the measured uncertainties from both sides as Eq(5.14) shows.

$$R_{bkg} = (1 - f_{tail}^{bkg})G(\Delta t_i, \sigma_{main}^{bkg}\sqrt{\sigma_{cp}^2 + \sigma_{tag}^2}) + f_{tail}^{bkg}G(\Delta t_i, \sigma_{tail}^{bkg}\sqrt{\sigma_{cp}^2 + \sigma_{tag}^2}) \quad (5.14)$$

It convolution integrals with $\mathcal{P}_{bkg}$:

$$\mathcal{P}_{bkg} = f_{\delta}^{bkg}\delta(\Delta t - \mu_{\delta}^{bkg}) + (1 - f_{\delta}^{bkg})\frac{1}{2\tau_{bkg}}e^{-|\Delta t - \mu_{\tau}^{bkg}|/\tau_{bkg}} \quad (5.15)$$

With this definition, one can make background component $\mathcal{P}_{bkg} \circledast R_{bkg}$ and fit to side-band data to obtained the parameters. There are total seven parameters in float to determine which are listed down below with primary fit results on sideband data, which according to the previous section, is totally 60 events in $M_{bc} < 5.26$ GeV.

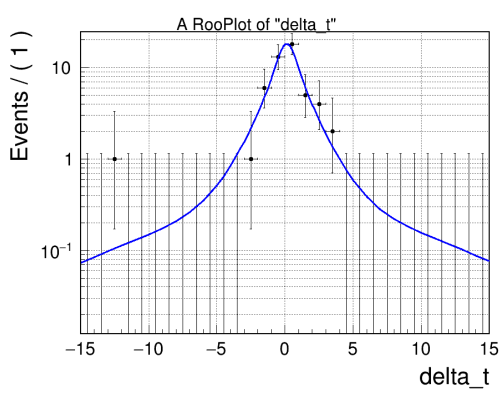


Figure 5-5: Background fit on data sideband with $M_{bc} < 5.26GeV$

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

Table 5.5: Parameters in Background Δt distribution.

5.2 Flavor Tagging

In order to determine the flavor of tag side B^0, flavor tagging algorithm is being developed. The tagger uses information from $\mu^\pm$, $\pi^\pm$,$K^\pm$ and Λ and categorized them into 13 different types as illustrated below.

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

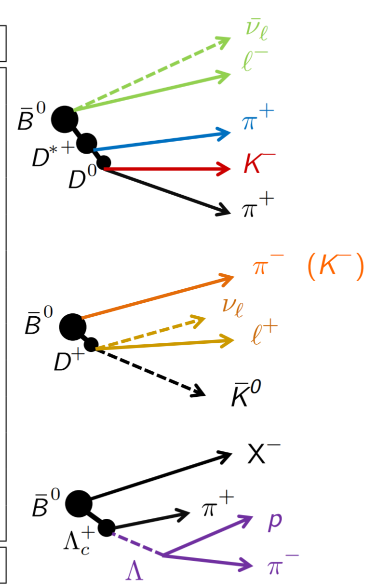


Figure 5-6: Particles and their categories used in flavor tagging algorithm.

For each particle that has been used from above categories, PID and kinematics (such as p_t,$cos\theta$,d_0,etc) information are extracted and feed to the combiner as training variables, to form a output corresponding each category. Then for all outputs from these categories, a total classifier is trained to present the likelihood of flavor q. This algorithm is called category-based algorithm which is used in our analysis. Similar tagger based on deep neural-network algorithm is also available but will not be further discussed here. After the reconstruction on the CP side is done, ROE objects that contains the tracks from tag side is formed, and tracks that will be used to fill the above particle lists are selected. Just like K_S^0 finder and continuum suppression, we choose FastBDT as classifier back-end algorithm. Targeted variable is q of tag side in MC truth. The ROE selection is affected by the cleanness of CP reconstruction, but flavor tagging should be almost mode-independent. To minimize impact of the reconstruction performance, MC sample of $B^0 \rightarrow \nu\nu$ is used for training the tagger where CP side is completely invisible. Then the trained classifier can be applied to signal sample.

The statistical uncertainty of $\mathcal{S}$ now receives contribution from ϵ and w: $\delta(\mathcal{S}) \propto \frac{1}{\epsilon(1-2w)N_{rec}}$, which make it crucial to correctly measure tagging efficiency and wrong tag fraction. w can be estimate using MC sample which the bias on individual events could propagate to the fit of $\mathcal{S}$. By measuring w from flavor specific decay using real data, bias from MC can be avoid. The validation of flavor tagger using flavor specific control sample is summarized here: [21]. For flavor specific decay, the Eq(5.17) and Eq(5.18) could be used to fit the wrong tagging fraction and efficiency.

$$\mathcal{P}^{obs}_{SF/OF}(\Delta t, q_{cp}, q_{tag}, \epsilon, w) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \epsilon \left[1 + q_{cp} \cdot q_{tag}(1-2w) \cdot cos(\Delta M_d \Delta t)\right] \quad (5.16)$$

so wrong tag fraction can be measured by:

$$\frac{\mathcal{P}^{obs}_{OF} - \mathcal{P}^{obs}_{SF}}{\mathcal{P}^{obs}_{OF} + \mathcal{P}^{obs}_{SF}} = (1-2w)cos(\Delta M_d \Delta t) \quad (5.17)$$

If w is extracted from MC, which is what we implemented in this analysis, bias could be reduced by using binned values of tagger output. To be more specific, saying the output of tagging classifier is α in range $[0, 1]$, wrong tag probability of being $q = +1$ is defined as $\beta = 1 - \alpha$, and flavor tagger will eventually gives tagged flavor as $\mathcal{F} = 1 - 2\beta = 2\alpha - 1$. It's obvious that $\mathcal{F} < 0$ means the tagged flavor is $q = -1$ instead of $+1$. The distribution of $\mathcal{F}$ is shown in Fig 5-7 (up) from signal MC. The average efficiency of among all tag-able event is measured to be $99.92 \pm 0.01\%$, and effective efficiency (percentage of correctly tagged events) is $31.31 \pm 0.52\%$ for B^0 and $31.13 \pm 0.51\%$ for $\overline{B^0}$. No clear bias on effective tagging efficiency is found considering the MC fraction of B^0 and $\overline{B^0}$ are 50.11% and 48.89% respectively.

The r-bins are intervals that splits the absolute value of $\mathcal{F}$, namely $r_{FBDT} = |\mathcal{F}|$, where the true flavor q is irrelevant to r-bins. Belle experience on defining the r-bins is inherited where bins are set as $[0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.875, 1]$.For all MC events that have been successful tagged, they are projected into histogram of $|\mathcal{F}|$, w can be calculated in each bin in the way dividing the number of events that its $\mathcal{F}$ is opposite to its MC flavor to all the events in the same bin. The binned tagging efficiency is calculated through dividing the number of successfully tagged events by all events in the same bin.

Besides, w can be different between B^0 and $\overline{B^0}$. Considering $w = (w_{B^0} + w_{\overline{B^0}})/2$ and $\Delta w = w_{B^0} - w_{\overline{B^0}}$, so for each event of flavor q, it may from a correct or incorrect tagged event when we look at the distribution of Eq(5.1).

From Eq(5.1), it's obvious that the distribution of Δt is independent from the flavor q, while this is not true considering the limit power of flavor tagging as we discussed. Only certain fraction of events can be flavor tagged and also only part of events are correctly tagged. Considering the effect from flavor tagging, the observed distribution of Δt is turning into:

$$\mathcal{P}^{obs}_{sig}(\Delta t, q, \epsilon, w) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \epsilon \left\{ 1 - q \cdot \Delta w + q(1-2w) \cdot \Big[\mathcal{S} sin(\Delta M_d \Delta t) + \mathcal{A} cos(\Delta M_d \Delta t) \Big] \right\} \tag{5.18}$$

compared to the original, the fluctuation is reduced by factor $r \equiv |1 - 2w|$, called as dilution factor.

As for efficiency ϵ, the difference $\mu = \epsilon_{B^0} - \epsilon_{\overline{B^0}}$ is about 1% to 2% in each bin, thus treated as zero. The average r is also calculated in r-bins. Fig 5-7 (down) shows

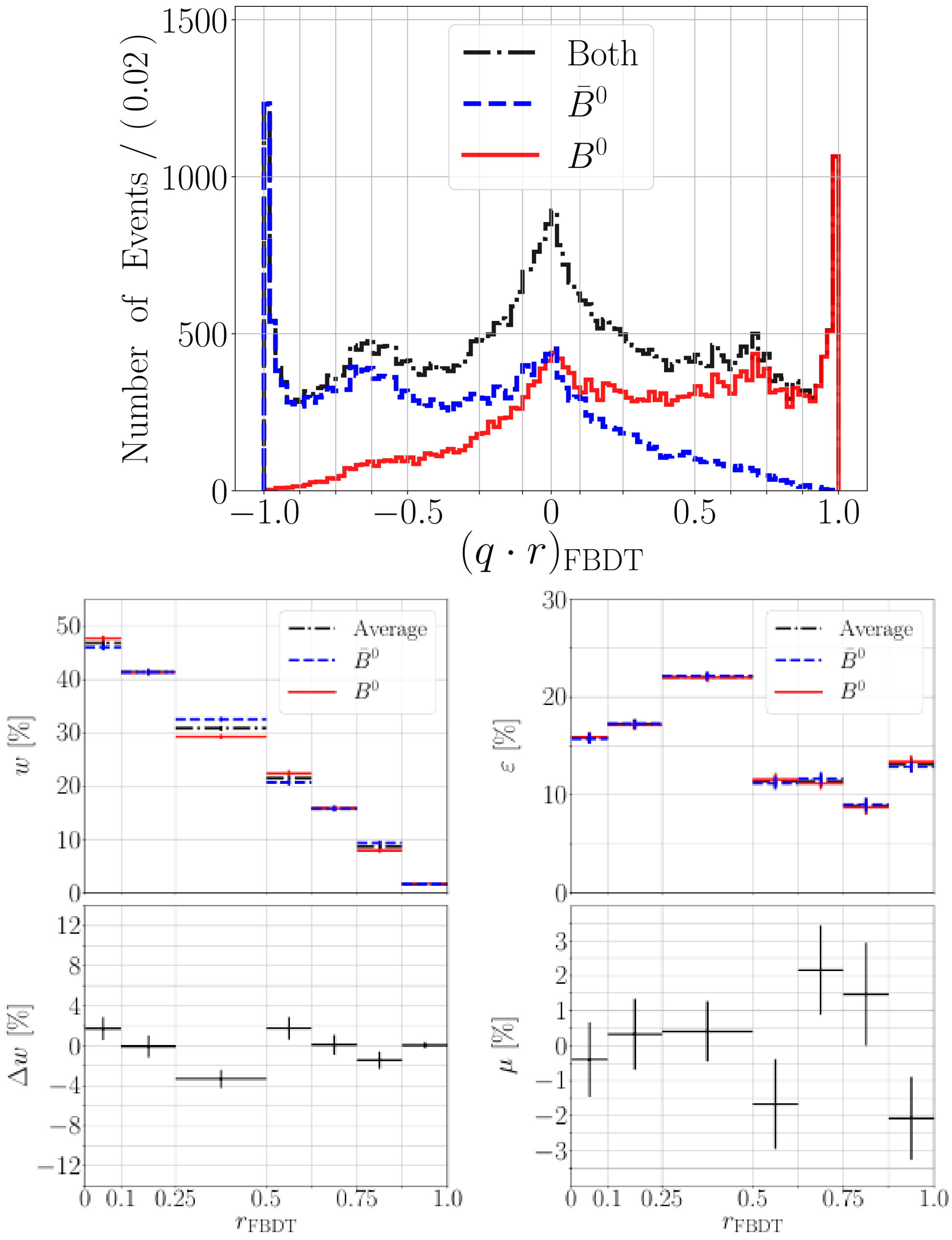


Figure 5-7: Top:the distribution of flavor tagger output $\mathcal{F} = (q \cdot r)_{FBDT}$ for both tag side of B^0 and $\overline{B^0}$. Bottom: flavor tagging efficiency, wrong tagging fraction, tagging difference in regard to rbin.

the flavor tagging results on w,Δw, ϵ and μ in each of the r-bin.

5.3 *CP* Fitter

The parameters that are needed for *CP* fit is basically all studied and obtainable for extracting $\mathcal{S}$ and $\mathcal{A}$ from Eq(5.2) with using observed Δt distribution. The fit is taking the Δt, signal fraction f_{sig}, the flavor charge q as observables, and in the meantime the vertexing error and χ_2/N are used as event-by-event conditional variables. RooFit is configured based on this setup and unbinned maximum likelihood fit is used as the fitting strategy. To actually perform the fit, the computation work load is an important factor to consider when the event number goes high in future. Especially, the multiple convolution integrals of resolution model and probability density functions are a set of computing intensive works. Belle legacy fitter utilize a highly customized library called "Tatami" to perform the convolution integrals. For Belle II, a new *CP* fitter is developed based on Python (version 3.7) and RooFit, which is naturally easy to use and maintain with BASF2 due to BASF2's analysis controls are mostly done by Python. The code management shows a modern, good consistency and readability, which will be aimed for providing a universal fitting tool to TDCPV analysis in Belle II. The fitter requires a configuration files which contains all the observables and parameters definition including their range, value, floating state. Data that is fed to the fitter will be converted into the RooDataSet object associated with each variables from configuration file. The models of fit such as each resolution functions, signal and background P.D.F is formed with the data set in the same workspace. The convolution of each components are handled by a customized shared object written in C++ from "SignalPdf.cxx". Combined these tools, the unbinned maximum likelihood fit is performed that can either be used to determine $\mathcal{S}$ and $\mathcal{A}$ or other parameters when they are floating. The result of the fit then restore as ROOT files as well as corresponding plots.

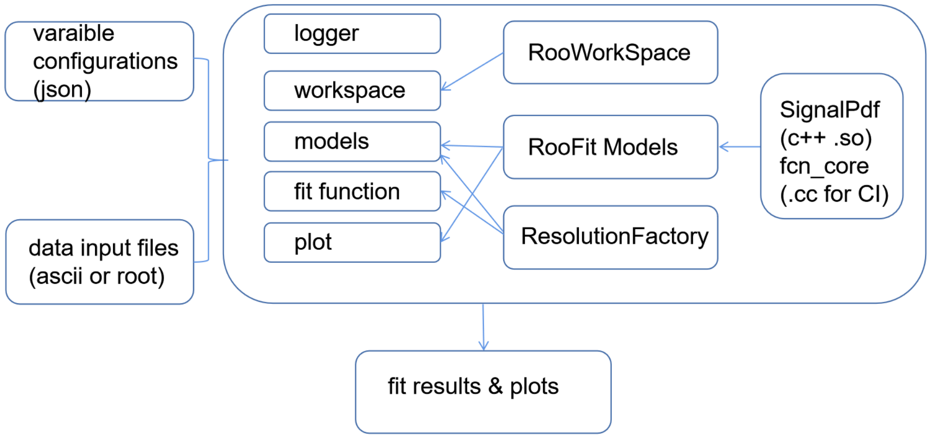


Figure 5-8: The structure of *CP* fitter for Belle II. Users provide configuration file containing the definitions of all parameters and observables in the *CP* fit. Data input files are used to create dataset for fitting. External library called “SignalPdf.cxx” are generated in runtime to calculate the integral of resolutions and physics distribution event by event, which the fit is performed by maximizing the “fcn_core” function, presenting the likelihood of the dataset to the fit model.

5.4 Blind Fit

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

5.4.1 *CP* fit on MC samples

Using *CP* fitter, we 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 ($\mathcal{S} = \mathcal{A} = 0$) in $B^0 \to K_S^0 K_S^0 K_S^0$ channel. The event with following cuts are selected for *CP* fit.

- $-70 < \Delta t < 70$ ps

- $0 < (\chi_2/N)_{cp} < 8$
- $0 < (\chi_2/N)_{tag} < 50$
- error on tag-side vertex < 0.1 cm
- $5.27 < M_{bc} < 5.29$ GeV and $|\Delta E| < 0.1$ GeV

We have 10000 (8873 passing selection) events from signal sample and 415 (373 passing selection) events from 1ab^{-1} generic sample to fit CP parameters. To mimic the events number expected in data sample, we randomly take 30 events from generic sample and perform the fit as well. The fit results are shown in Fig 5-9 to Fig 5-11.

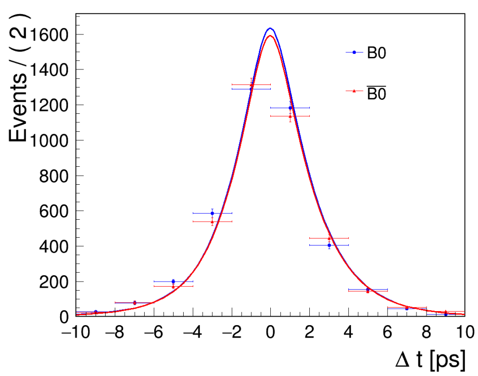


Figure 5-9: CP fit on 8873 signal MC.

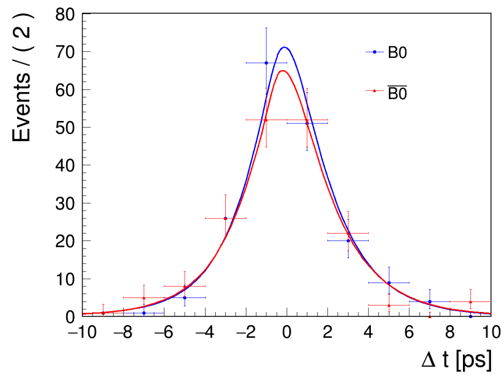


Figure 5-10: CP fit on 373 generic MC.

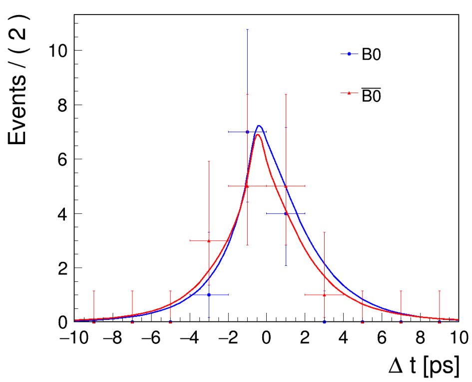


Figure 5-11: CP fit on 30 generic MC.

The fit result for 8873 signal MC is:

$$\begin{aligned} sin(2\phi_1) &= 0.00 \pm 0.04 \\ \mathcal{A} &= -0.01 \pm 0.02 \end{aligned} \tag{5.19}$$

the fit result for 373 generic MC is:

$$\begin{aligned} sin(2\phi_1) &= 0.00 \pm 0.21 \\ \mathcal{A} &= -0.05 \pm 0.07 \end{aligned} \tag{5.20}$$

the fit result for 30 events generic MC is:

$$\begin{aligned} sin(2\phi_1) &= 0.20 \pm 0.85 \\ \mathcal{A} &= -0.06 \pm 0.30 \end{aligned} \tag{5.21}$$

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. 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[20]. The fit result over 10000 events is shown in FIG.30, 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.

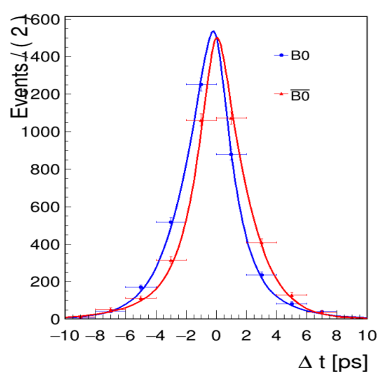


Figure 5-12: CP fit over 10000 $B^0 \rightarrow J/\psi K_S^0$ signal MC.

5.4.2 Linearity Test

To validate the CP fit linearity, we generate a series of toy MC sample, which the χ_2, N and vertex errors on CP and tag-side are sampled from distribution of signal MC. The resolution functions parameters are kept same as MC fit as previous section. 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 below. The linearity fit shows a good agreement between input and output.

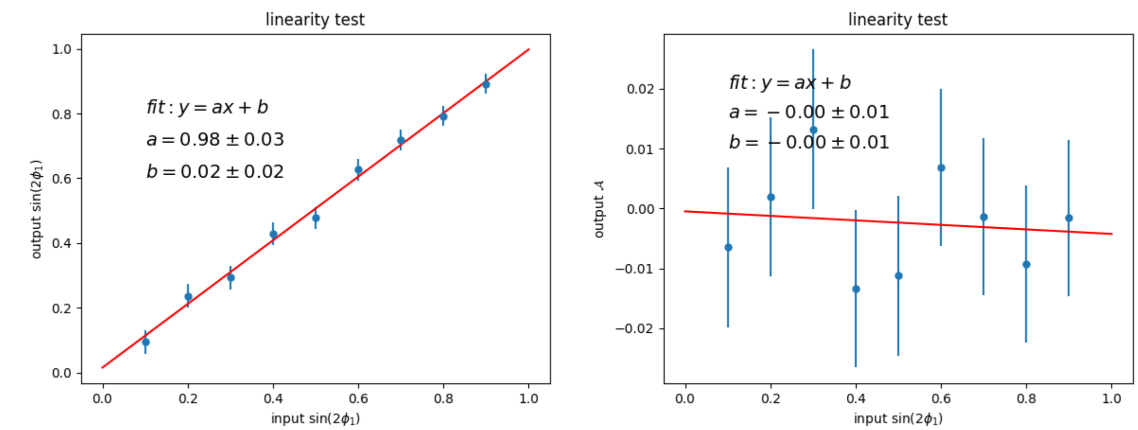


Figure 5-13: Linearity test of CP fit.

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

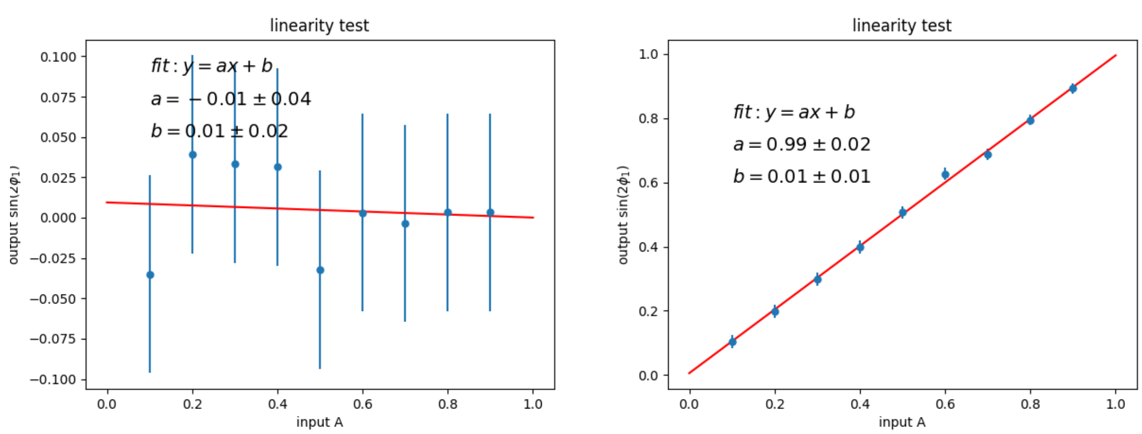


Figure 5-14: Linearity test of CP fit.

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. χ_2,

N and vertex errors on CP and tag-side are sampled from distribution of data. The fit to dataset has only $\sin(2\phi_1)$ and $\mathcal{A}$ as floating parameters, which input are both zero. We expect to use normal distribution to describe the pull of $\sin(2\phi_1)$ and $\mathcal{A}$.

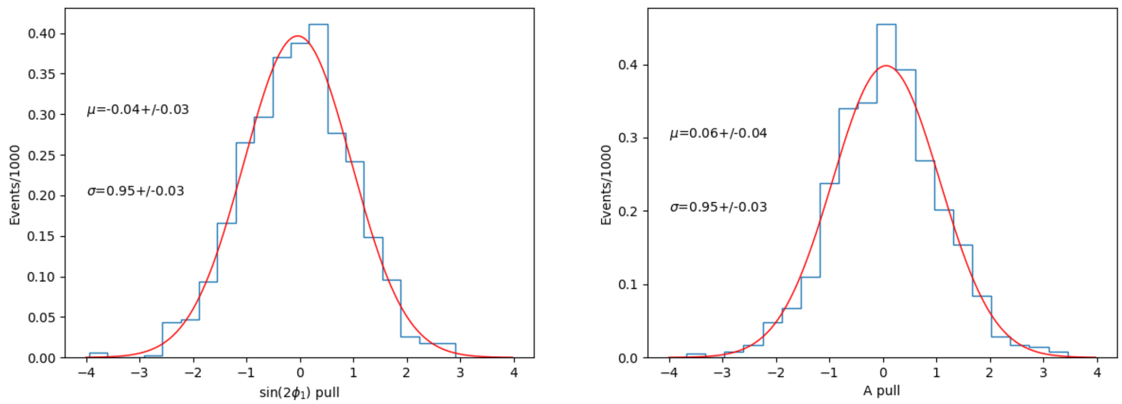


Figure 5-15: Pull of $\sin(2\phi_1)$ and $\mathcal{A}$ fitted.

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

5.4.4 Lifetime and Δm_d Fit

Before looking at CP parameters in data, we need to check if the physics parameters are consistent when setting the CP fitter to fit them in float. To test lifetime fit, first we use 10000 signal MC events which is generated by $\tau_{B^0} = 1.520$ from PDG value. The $\sin(2\phi_1)$ and $\mathcal{A}$ are fixed at zero during the fit, for which the generator level CP violation is zero. This is equivalent fit to:

$$\mathcal{P}(\Delta t, \tau_{B^0}) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \tag{5.22}$$

The fit result is 1.537 ± 0.024 ps which is consistent with the input. We perform the lifetime fit on data in signal region, the CP parameters are fixed based on PDG values to: $\sin(2\phi_1) = 0.69$ and $\mathcal{A} = 0$. The fitted lifetime from $B^0 \to K_S^0 K_S^0 K_S^0$ is 1.431 ± 0.382 ps. The result is consistent with PDG value. The distribution of Δt in lifetime fit is shown as Fig 5-16. The B^0 and B^+ lifetime fit using control sample is also performed and summarized in here[20]. The results are consistent with PDG values.

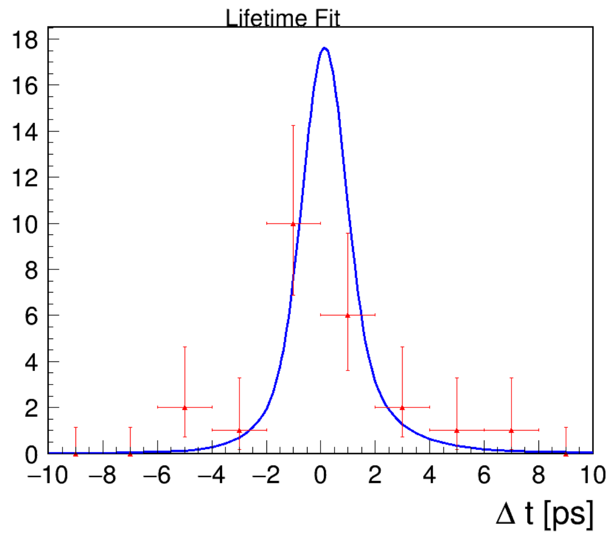


Figure 5-16: Lifetime fit on data

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

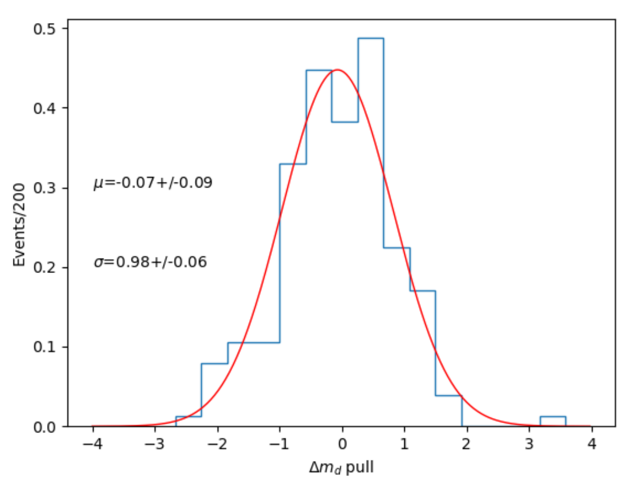


Figure 5-17: Pull of Δm_d

5.5 *CP* fit on data

After the *CP* fit procedures from the previous sections 5.1 to 5.5 is reviewed by Belle II collaboration, the permission of measuring *CP* parameters using Belle II 2019 and 2020 Spring/Summer data is granted by the review committee. The event number used for the *CP* fit is 26, and the fit result is shown as:

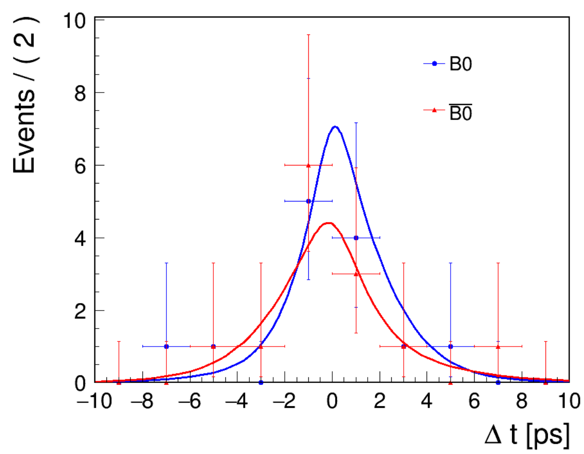


Figure 5-18: The *CP* fit from data.

The results of *CP* parameters are:

$$\begin{aligned} sin(2\phi_1) &= 0.82 \pm 0.85(stat) \\ \mathcal{A} &= -0.21 \pm 0.28(stat) \end{aligned} \tag{5.23}$$

5.6 Systematic Uncertainty

The systematic uncertainty that affects the fit results may come from many aspects of the measurement setup. In this measurement considering the options we used for signal extraction, vertexing and fit strategy, the major systematic uncertainty receives contribution from following points:

- resolution functions parameters
- signal fraction

- flavor tagging
- background Δt shapes
- fit bias
- physics parameters

For the above 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 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 5.6: systematic uncertainty from resolution models

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_δ	-0.007248	-0.000552	0.007231	0.000591
f_p	0.003037	0.004347	-0.003069	-0.004314
τ_n	-0.00101	-0.002841	0.000937	0.00294
τ_p	0.004497	0.002502	-0.004648	-0.002478

The background Δt shapes' parameters are determined from data sideband $M_{bc} <$ 5.26 GeV. The impact on fit results is summarized as follows:

Table 5.7: systematic uncertainty from background Δt shapes

source	$+\delta\mathcal{S}$	$+\delta\mathcal{A}$	$-\delta\mathcal{S}$	$-\delta\mathcal{A}$
μ_δ^{bkg}	-0.014294	-0.016581	0.006758	0.006537
μ_l^{bkg}	-0.002798	-0.012567	0.003789	0.012783
τ_{bkg}	0.001377	0.001689	-0.004159	0.000085
f_δ^{bkg}	-0.011315	0.001365	0.011187	-0.001395
f_{tail}^{bkg}	-0.002661	0.00153	0.00248	-0.001368
σ_{main}^{bkg}	0.0207015	0.022041	-0.0236175	-0.01569
σ_{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 5.8: 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 Δm_d and τ_{B^0} uncertainties are included using the PDG average value. The impact on fit results is summarized as follows:

Table 5.9: systematic uncertainty from physics parameters

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

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

Table 5.10: 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 $\delta\mathcal{S} = 0.0098$ and $\delta\mathcal{A} = 0.0057$.

5.7 Summary

The CP parameters measurement study targeting the extraction of asymmetry parameters $\mathcal{S}$ and $\mathcal{A}$ has been conducted. Based on physical Δt distribution, the model for CP fit have been properly built. Importantly, corresponding resolution functions for $\mathcal{P}_{sig}$ and $\mathcal{P}_{bkg}$ are built by separating smearing effects from CP and tag side. All the fits takes advantage of event-by-event conditional fit so the model has a good robustness in different measurement quality cases. As for flavor tagging, an algorithm based on FBDT classification is developed and implemented to give the estimated dilution factor r. To take a good use of these information, a Belle II CP fitter is developed to perform the fit on desired parameters. The linearity test and fit pull test shows a reliable performance of the fitter using toy MC study. The actual fit on lifetime using data is also consistent with the PDG value with a relatively large uncertainty due to very small amount of data.

After the blind fit and test for the CP parameters measurement is conduced, with the permission from Belle II collaboration, we unblind the data in signal box and perform the CP fit to extract $\mathcal{S}$ and $\mathcal{A}$. Also, the major sources of systematic uncertainty is estimated. The primary result of CP parameters in this low luminosity stage are:

$$\begin{aligned} \mathcal{S} = -sin(2\phi_1) = -0.82 \pm 0.85(stat) \pm 0.07(syst) \\ \mathcal{A} = -0.21 \pm 0.28(stat) \pm 0.06(syst) \end{aligned} \tag{5.24}$$

The results are dominated by the statistical uncertainty due to the very limited data sample in early stage of Belle II. Under such large uncertainty, the results is in an agreement with the Standard Model.

Chapter 6

Conclusion and Prospective

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.

$b \to s$ 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 $b \to c$ with a small tension. So far the precision of the measurement of CP parameter $\mathcal{S}$ in $b \to s$ 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 $b \to s$ are resonant decay such as $B^0 \to \eta' K_S^0$, $B^0 \to \phi K_S^0$ and decay like $B^0 \to K_S^0 K_S^0 K_S^0$, 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 $B^0 \to K_S^0 K_S^0 K_S^0$ 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 Δt 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 ΔE. For each event, the signal fraction is calculated based on the M_{bc} and ΔE 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.

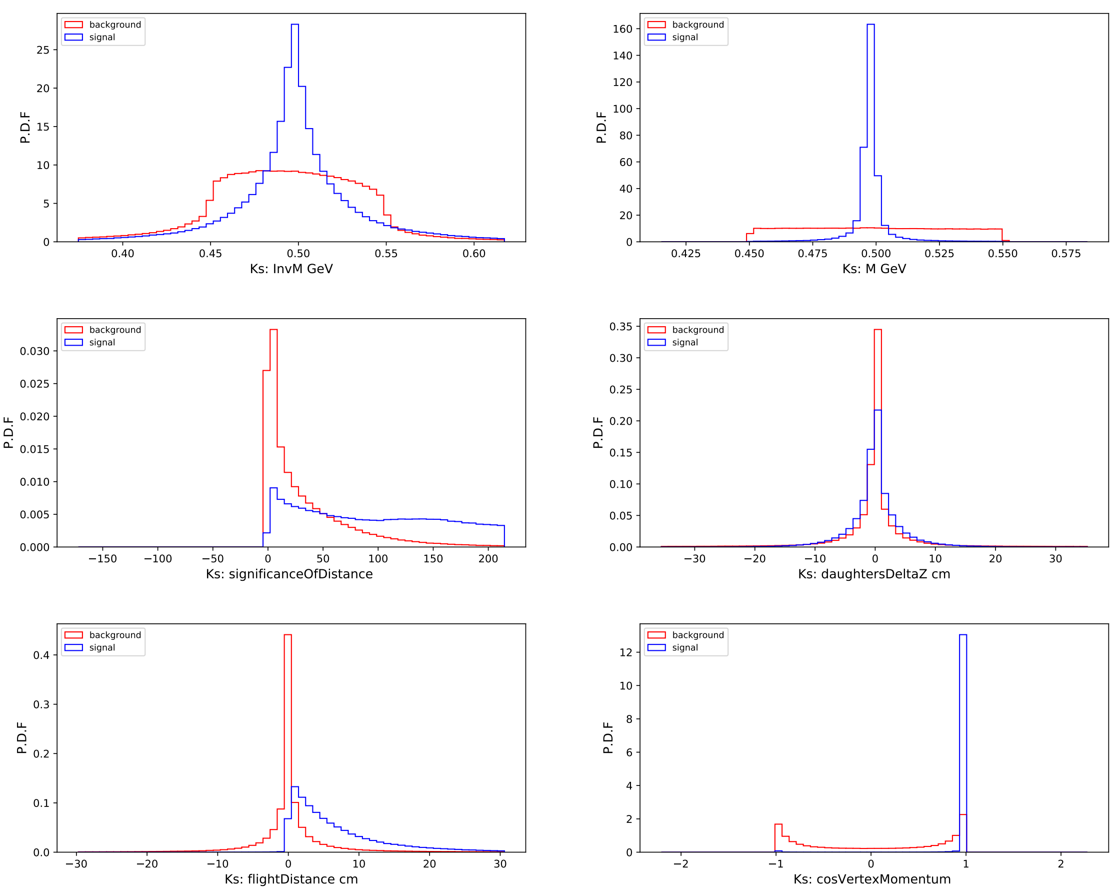
What is worth of noticing is, many analysis tools that are required by performing the CP measurement on this channel is in a good stage of development. The decay mode being successfully re-discovered under very low data collection mostly thank to the newly developed K_S^0 classification software, KsFinder, which will also be an essential asset in neutral particles dominated channels like $B^0 \to K_S^0 K_S^0 K_S^0$. 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 Δt 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.

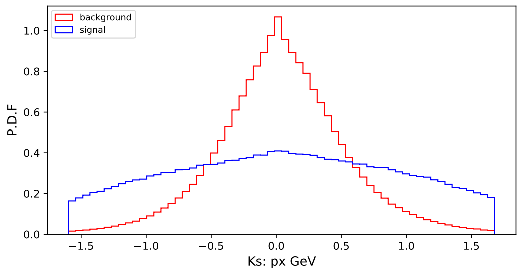
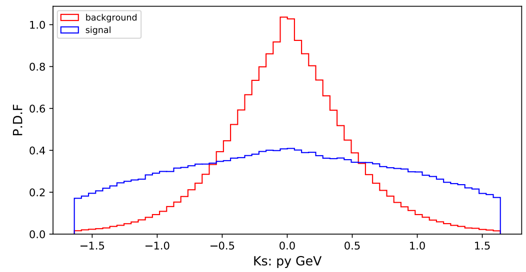
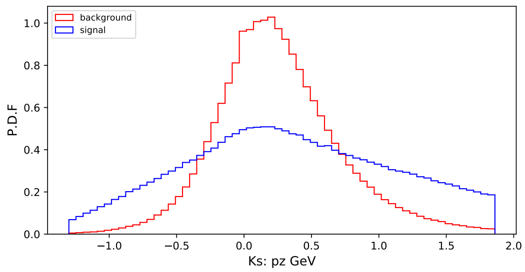
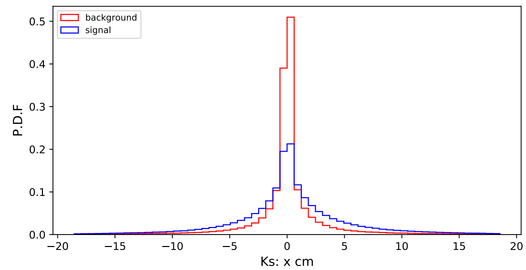
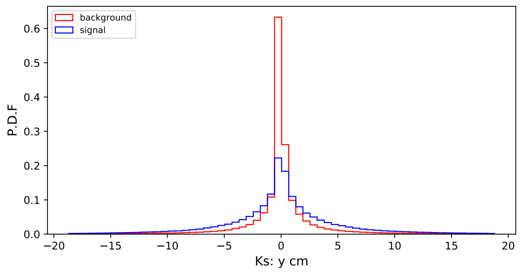
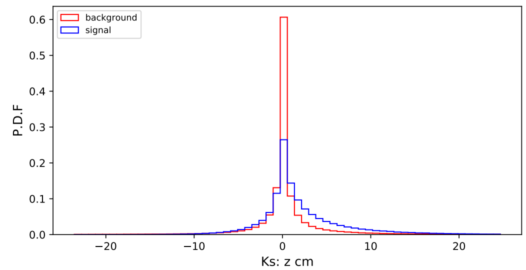
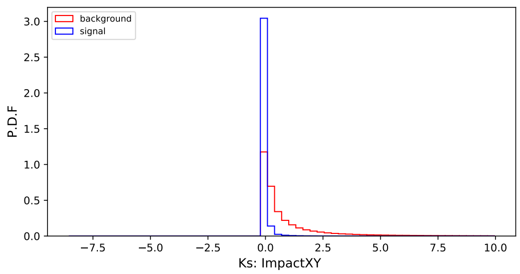
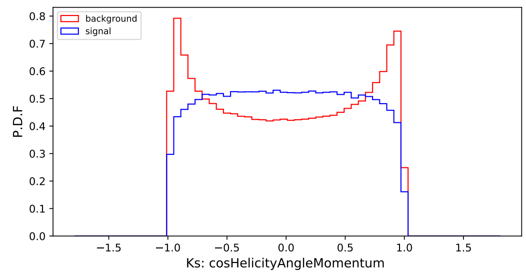
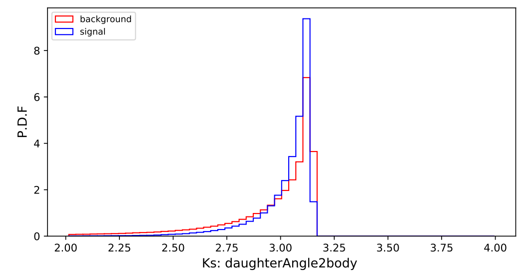
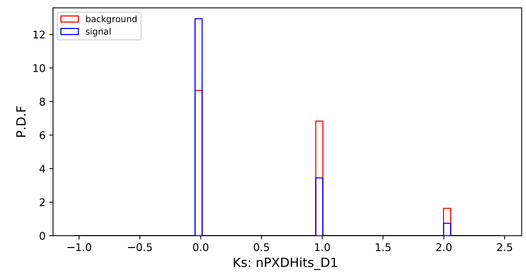
This study has shown a good potential of performing CP measurement in Belle II for the incoming years with more and more data recorded. The precision on $\mathcal{S}$ in $b \to s$ penguin-modes is highly depending on the large luminosity as Fig 1-9 shows, which the current precision fits in the expectation.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

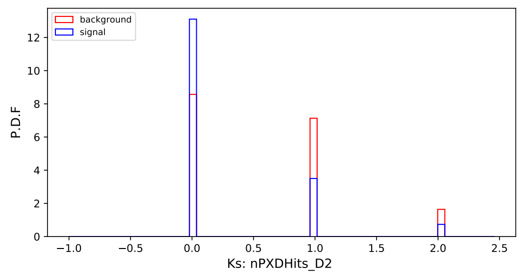
Appendix A

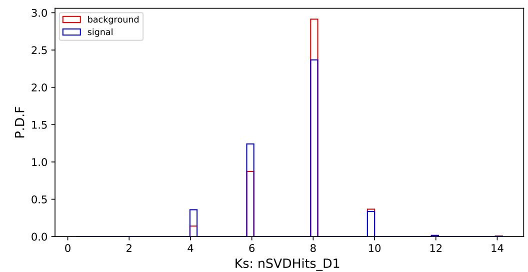
Training Observables for K_S^0 Classifier

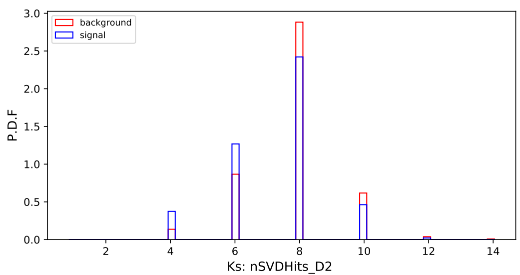
Figure A-1: The distribution of Ks training variables. The red is the from fake K_S^0 and the blue is from true K_S^0

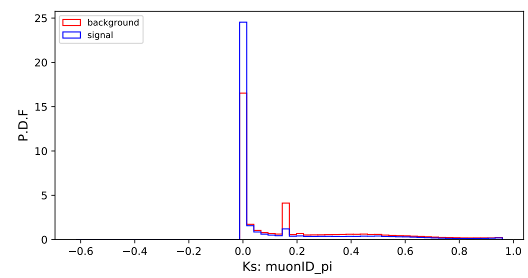


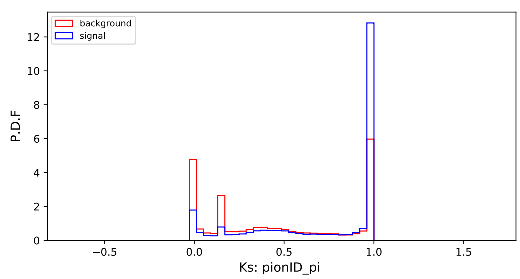

background
signal
P.D.F
Ks: px GeV

background
signal
P.D.F
Ks: py GeV

background
signal
P.D.F
Ks: pz GeV

background
signal
P.D.F
Ks: x cm

background
signal
P.D.F
Ks: y cm

background
signal
P.D.F
Ks: z cm

background
signal
P.D.F
Ks: ImpactXY

background
signal
P.D.F
Ks: cosHelicityAngleMomentum

background
signal
P.D.F
Ks: daughterAngle2body

background
signal
P.D.F
Ks: nPXDHits_D1

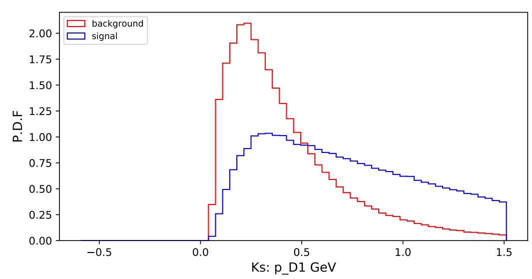

background
signal
P.D.F
Ks: nPXDHits_D2

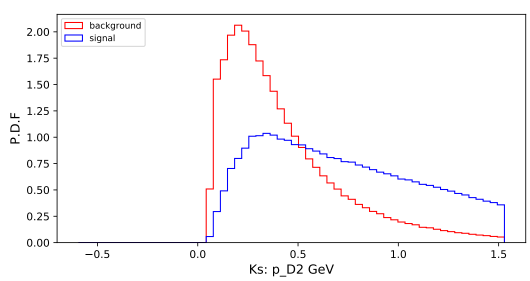

background
signal
P.D.F
Ks: nSVDHits_D1

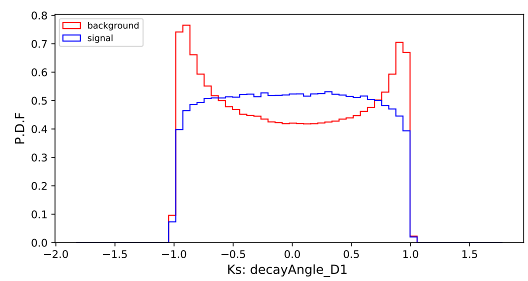

background
signal
P.D.F
Ks: nSVDHits_D2

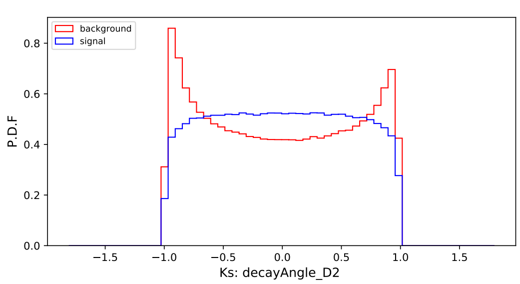

background
signal
P.D.F
Ks: muonID_pi


background
signal
P.D.F
Ks: pionID_pi


background
signal
P.D.F
Ks: p_D1 GeV


background
signal
P.D.F
Ks: p_D2 GeV

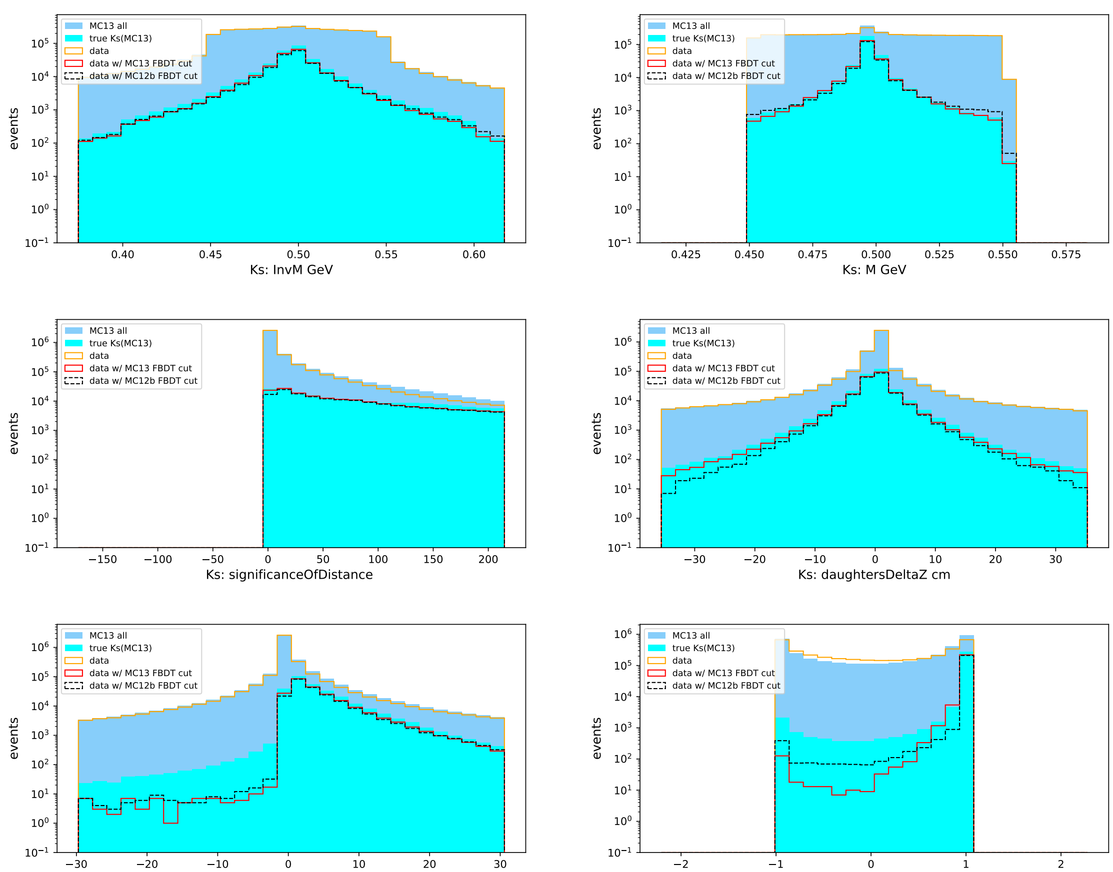

background
signal
P.D.F
Ks: decayAngle_D1

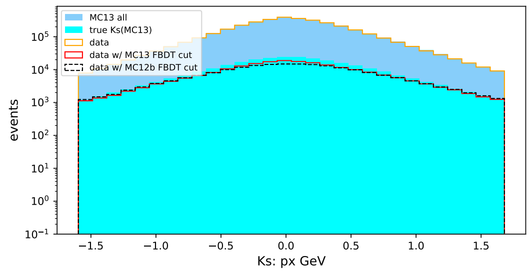

background
signal
P.D.F
Ks: decayAngle_D2

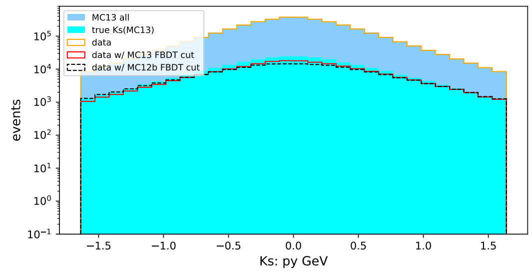
Appendix B

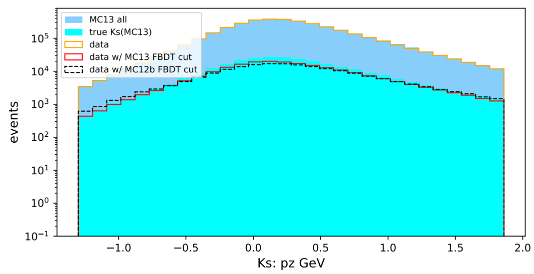
Data Validation Plots for K_S^0

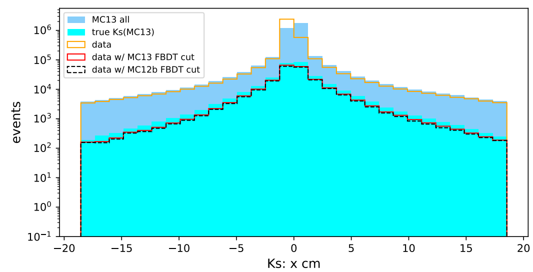
Figure B-1: The distribution of variables used in KsFinder classification. The variables abbreviation can be referenced from Table III. The yellow solid lines are K_S^0 from data without KsFinder, and red solid lines are K_S^0 with KsFinder using MC13a for training. The black dashed lines are K_S^0 with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

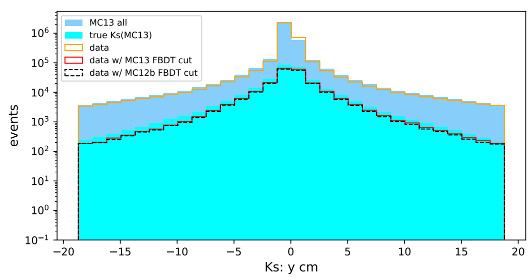


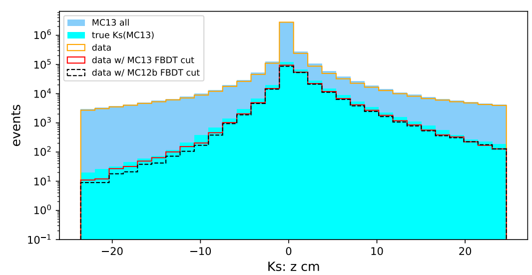

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: px GeV

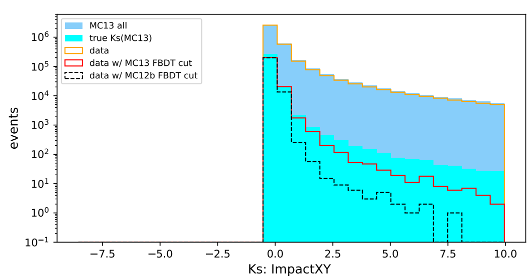

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: py GeV

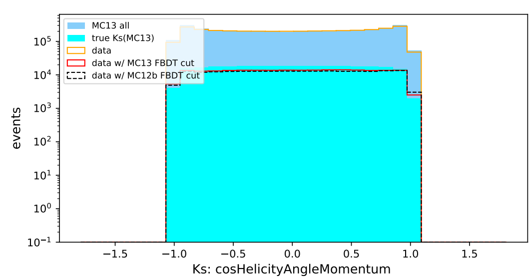

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: pz GeV

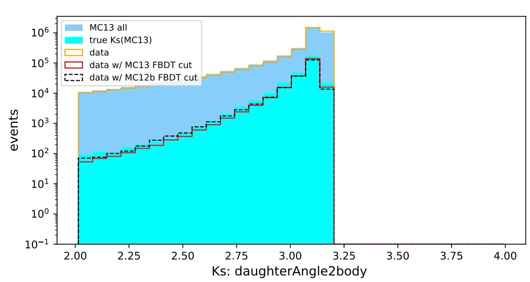

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: x cm

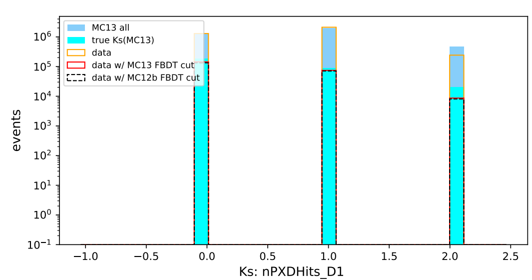

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: y cm

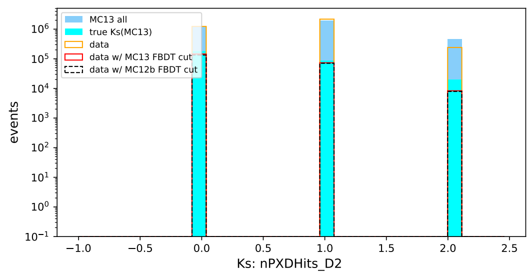

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: z cm

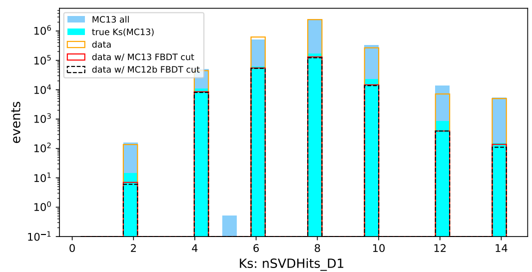

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: ImpactXY

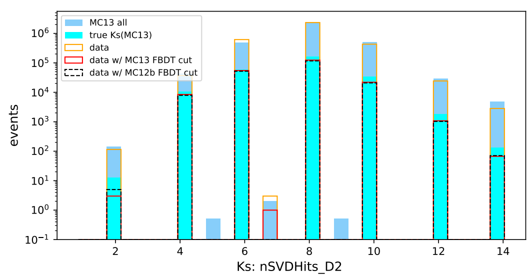

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: cosHelicityAngleMomentum

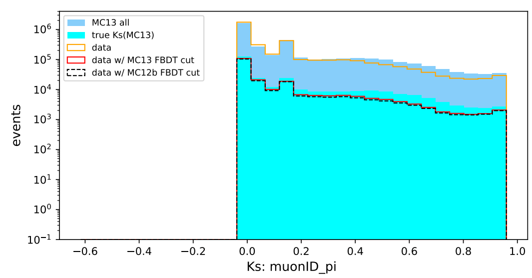

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: daughterAngle2body

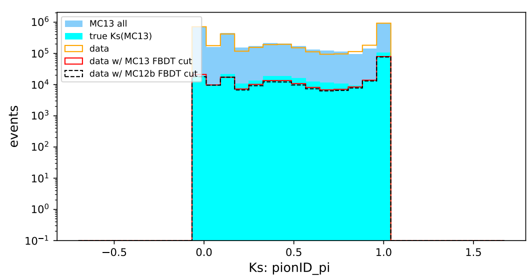

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: nPXDHits_D1

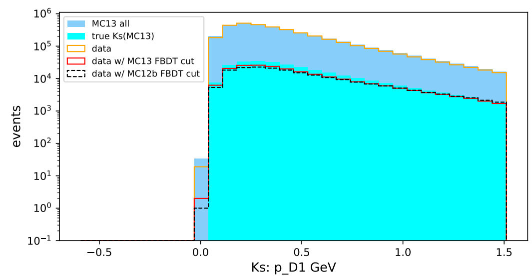

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: nPXDHits_D2

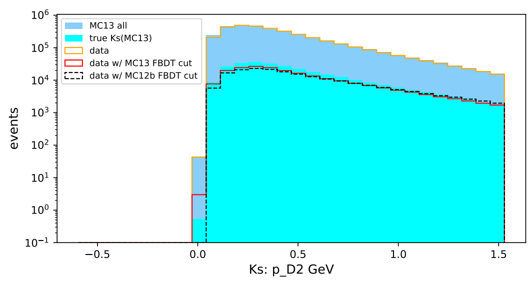

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: nSVDHits_D1

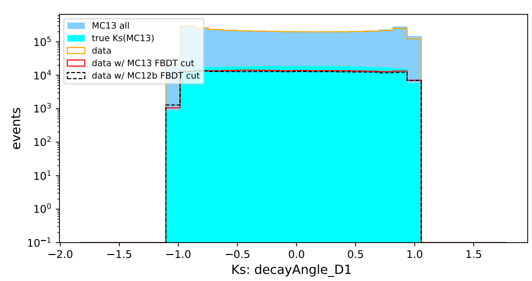

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: nSVDHits_D2

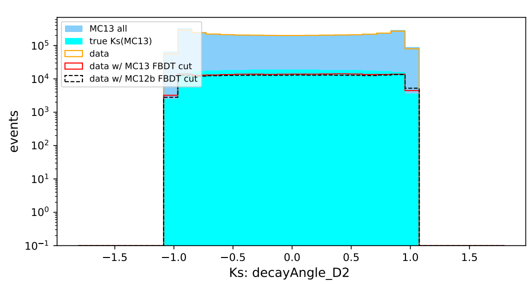

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: muonID_pi


MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: pionID_pi


MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: p_D1 GeV


MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: p_D2 GeV


MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: decayAngle_D1

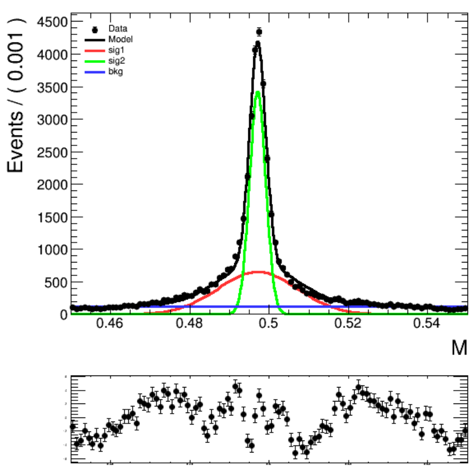

MC13 all
true Ks(MC13)
data
data w/ MC13 FBDT cut
data w/ MC12b FBDT cut
events
Ks: decayAngle_D2

Appendix C

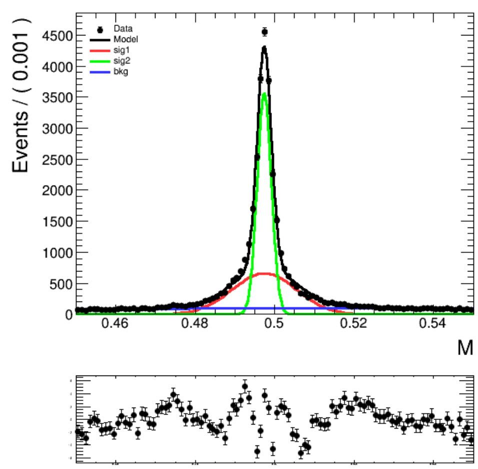
K_S^0 mass fit with varied cut value

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.

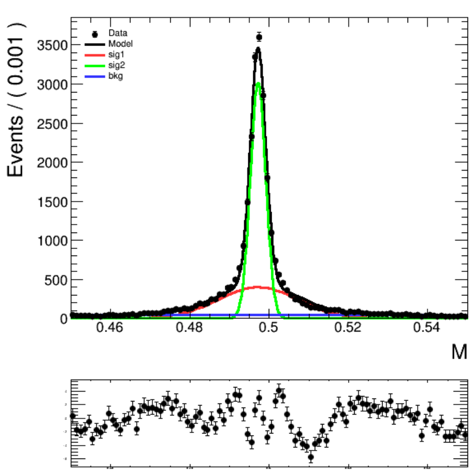
a) Data, cut=0.2



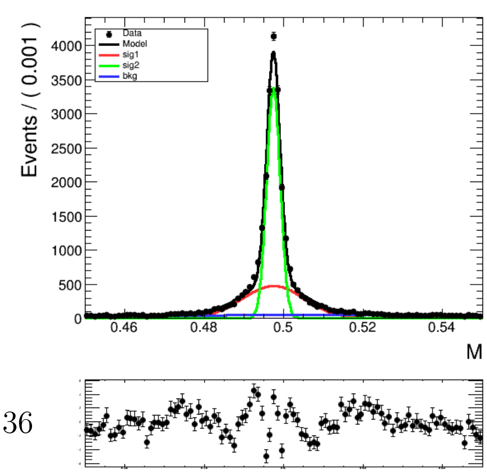
b) MC, cut=0.2



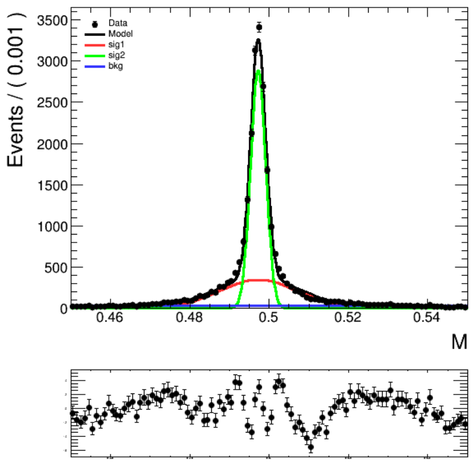
c) Data, cut=0.4



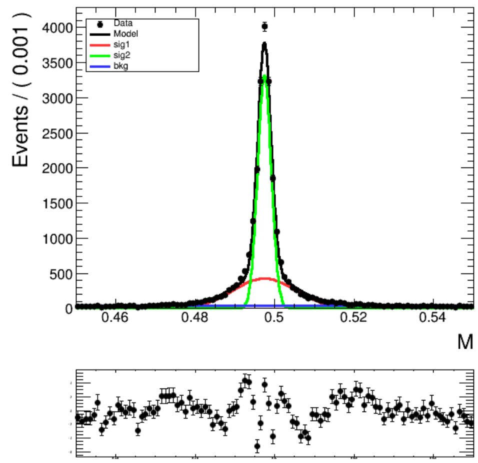
d) MC, cut=0.4



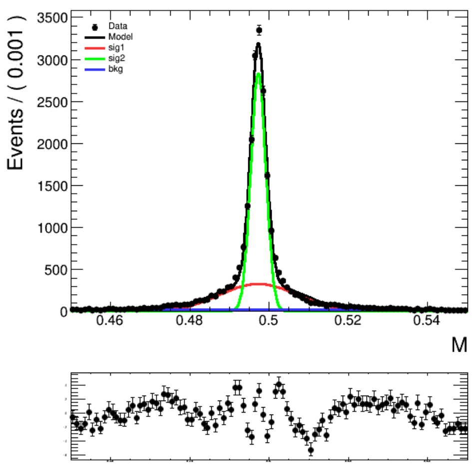
e) Data, cut=0.5



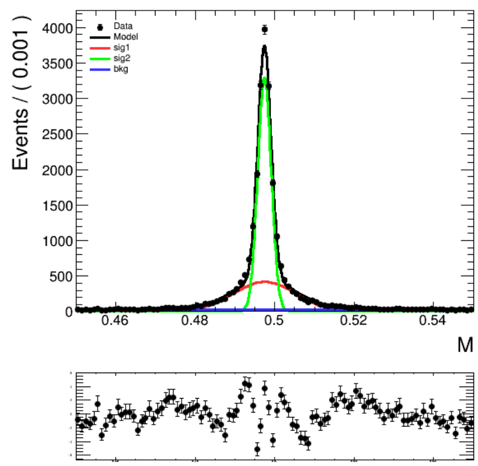
f) MC, cut=0.5



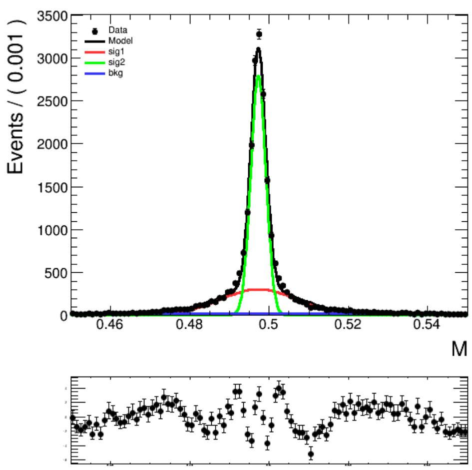
g) Data, cut=0.55



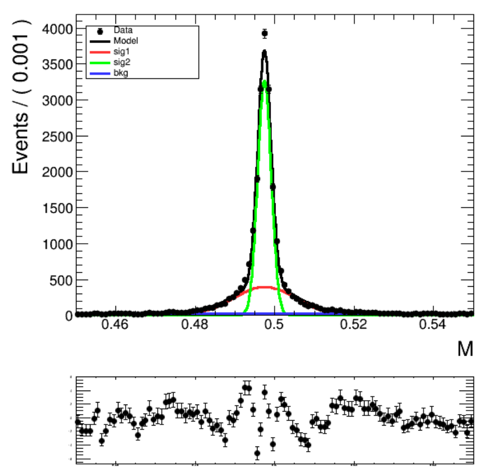
h) MC, cut=0.55



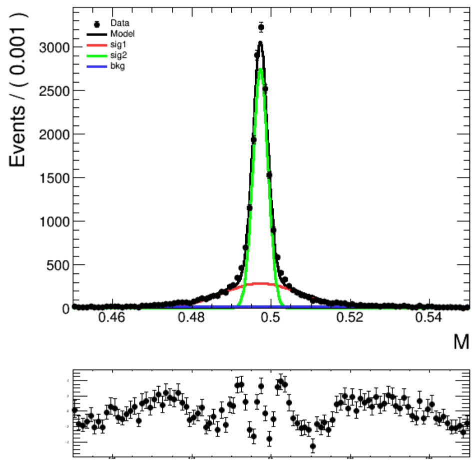
i) Data, cut=0.6



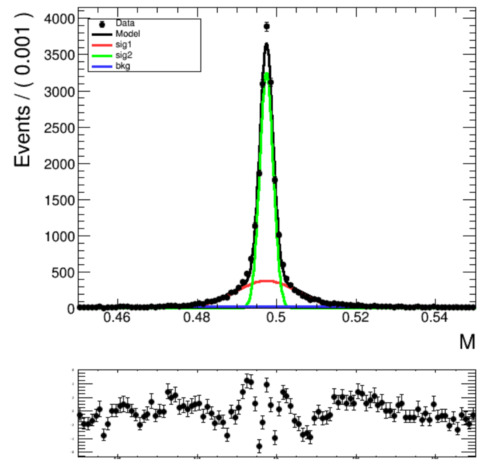
j) MC, cut=0.6



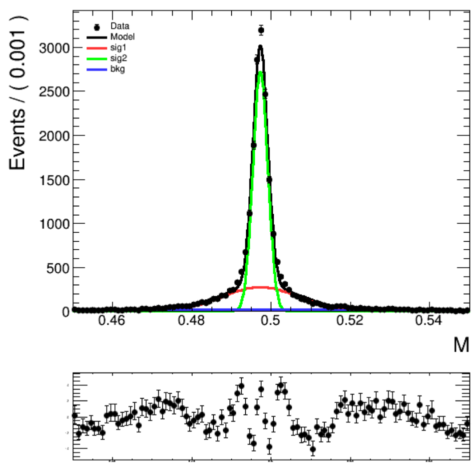
k) Data, cut=0.65



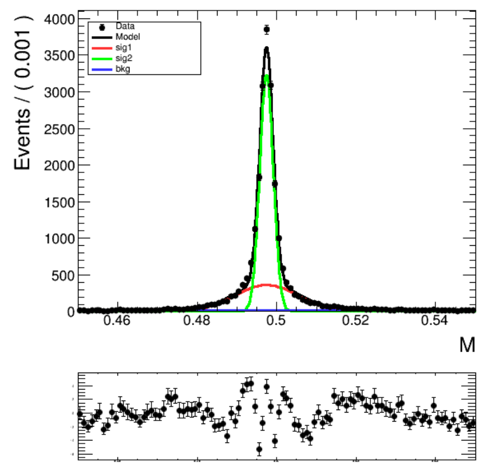
l) MC, cut=0.65



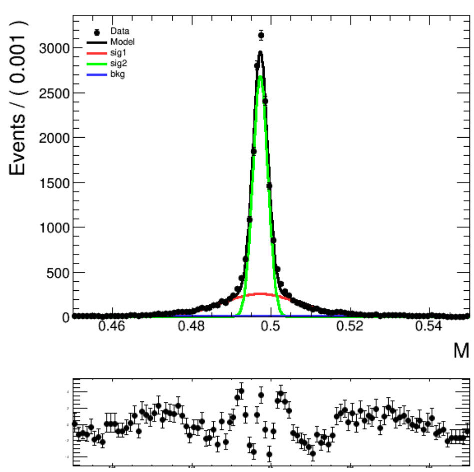
m) Data, cut=0.7



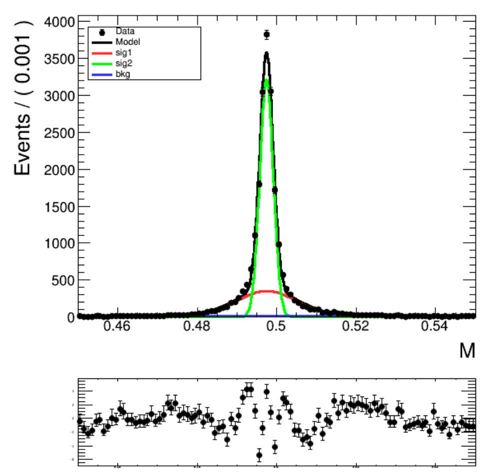
n) MC, cut=0.7



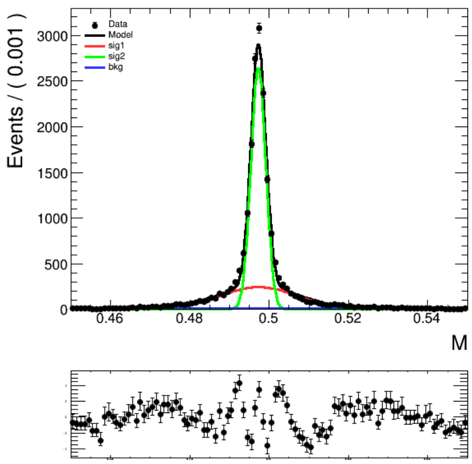
o) Data, cut=0.75



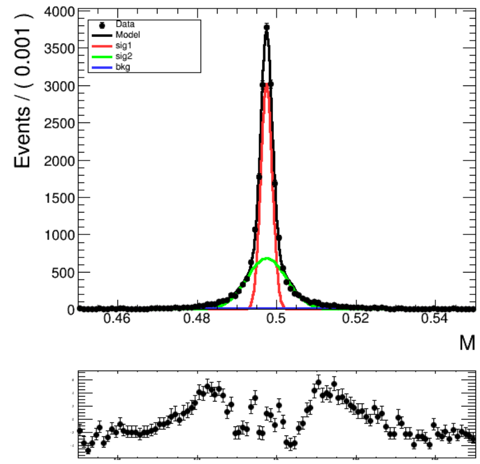
p) MC, cut=0.75



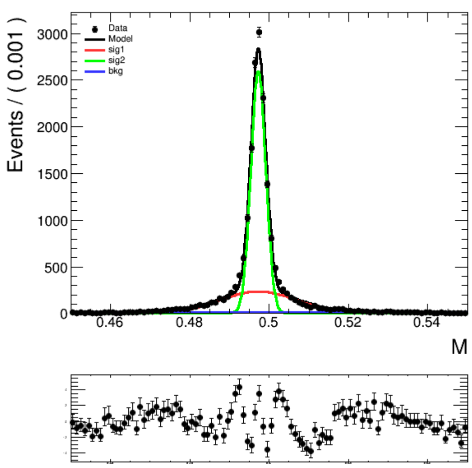
q) Data, cut=0.8



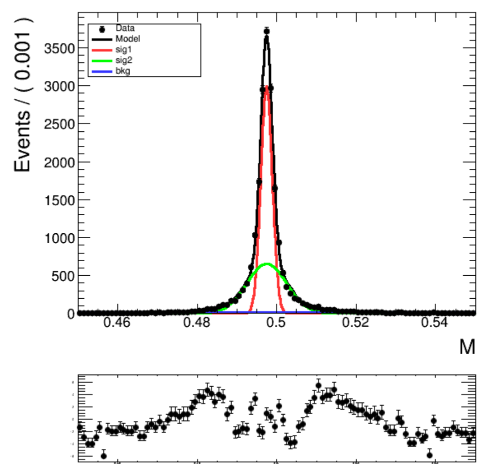
r) MC, cut=0.8



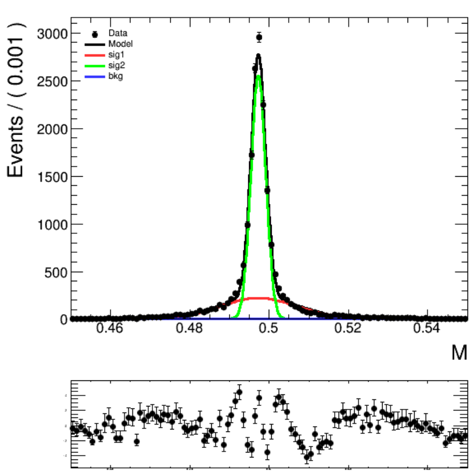
s) Data, cut=0.85



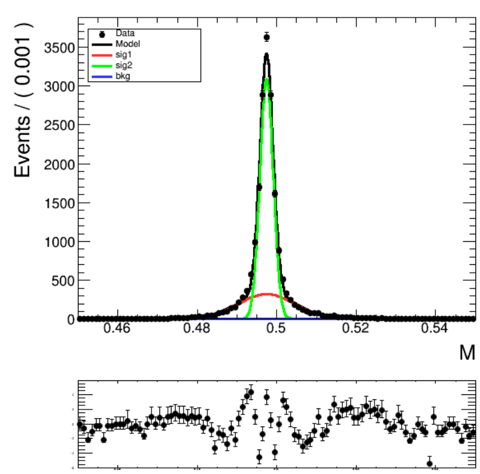
t) MC, cut=0.85



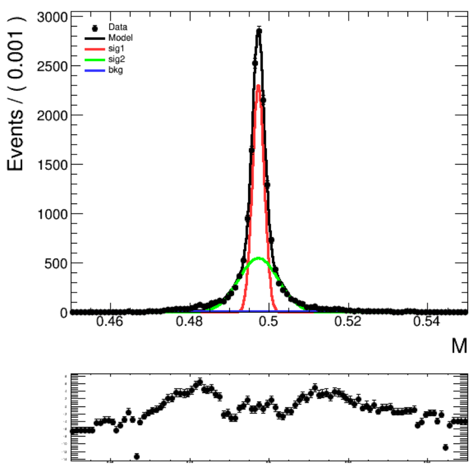
u) Data, cut=0.9



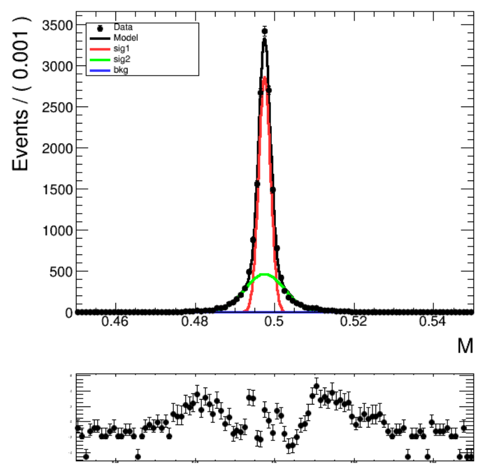
v) MC, cut=0.9



w) Data, cut=0.95



x) MC, cut=0.95



Bibliography

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

[2] Emmy Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, Jan 1971. doi:10.1080/00411457108231446.

[3] T. D. Lee and C. N. Yang. Question of parity conservation in weak interactions. *Phys. Rev.*, 104:254–258, Oct 1956. doi:10.1103/PhysRev.104.254. URL https://link.aps.org/doi/10.1103/PhysRev.104.254.

[4] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi:10.1103/PhysRev.105.1413. URL https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[5] 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.

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

[7] Amol Dighe, Tobias Hurth, Choong Sun Kim, and Tadashi Yoshikawa. The width difference of b_d mesons, 2001.

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

[9] 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.

[10] Michael Gronau, Yuval Grossman, and Jonathan L Rosner. Interpreting the time-dependent cp asymmetry in b0→pi0ks. *Physics Letters B*, 579(3-4):331–339,

Jan 2004. ISSN 0370-2693. doi:10.1016/j.physletb.2003.11.015. URL http://dx.doi.org/10.1016/j.physletb.2003.11.015.

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

[12] 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 b0-¿ksksks decays at belle. *arXiv preprint arXiv:2011.00793*, 2020.

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

[14] 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.

[15] L Breiman J Friedman RA Olshen and CJ Stone. Classification and regression trees. statistics/probability series, 1984.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

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