



BELLE2-NOTE-PH-2020-070  
DRAFT Version 2.0  
December 17, 2020

## Measurement of the $CP$ Violation Parameter $\sin(2\phi_1)$ in $B^0 \rightarrow K_s^0 K_s^0 K_s^0$ in Belle II early operation

TDCPV WG  
(Dated: November, 2020)

### Abstract

In this note, the measurement of  $CP$  parameter  $\sin(2\phi_1)$  in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  decay using MC13a and data collected in 2019 and 2020 is presented. The content mainly covers the development of signal reconstruction and  $CP$  parameters fit.

## Contents

<b>1. Introduction</b>	2
<b>2. Data Sample and Event Selection</b>	3
2.1. Analysis Software	3
2.2. Data and simulation samples	4
2.3. Events selections Overview	4
2.4. $K_S^0$ pre-selection	5
2.5. KsFinder	8
2.6. $B^0$ Reconstruction	13
<b>3. Flavor Tagging</b>	21
<b>4. Vertex Reconstruction and <math>\Delta t</math> resolution Model</b>	23
4.1. Vertex Reconstruction	23
4.2. $\Delta t$ resolution Model	25
<b>5. Blind Fit</b>	27
5.1. $CP$ fit on MC samples	27
5.2. Linearity Test	29
5.3. Toy MC Fit Pull	30
5.4. Lifetime and $\Delta m_d$ Fit	30
5.5. Systematics	32
<b>References</b>	32
<b>A. Over-fitting Check for KsFinder</b>	33
<b>B. KsFinder and CS related information.</b>	34
B.1. KsFinder Variables comparison between data and MC	34
B.2. Continuum suppression over-training check	37
<b>C. Resonant Backgrounds in MC</b>	37
<b>D. Signal injection test</b>	39

### **1. INTRODUCTION**

In the Standard Model,  $CP$  violation arises from a complex phase in CKM quark mixing matrix[1]. Compared to tree-level dominated decay through  $b \rightarrow c \rightarrow s$  quark transition,  $B^0 \rightarrow K_s^0 K_s^0 K_s^0$  decay mediated by  $b \rightarrow sq\bar{q}$  transition which is loop-level dominated. The latest previous results from Belle in this channel[2] is:  $\mathcal{S} = -\sin 2\phi_1 = -0.72 \pm 0.23(\text{stats}) \pm 0.05(\text{syst})$  and  $\mathcal{A} = 0.12 \pm 0.16(\text{stats}) \pm 0.05(\text{syst})$ . The earlier results from Babar[3] is:  $\mathcal{S} = -\sin 2\phi_1 = -0.94^{+0.21}_{-0.24}(\text{stats}) \pm 0.06(\text{syst})$ . This three-body decay has no  $b \rightarrow u$  contribution but only from  $b \rightarrow s$  penguin diagram, which makes it free of tree-level contamination. Thus it's particularly interesting and sensitive to the potential New Physics effect.

10 The measurement of  $CP$  violation parameters in decay of  $B^0$  mesons to  $CP$ -even eigen-  
 11 state  $K_S^0 K_S^0 K_S^0$  requires the analysis of the time-dependent decay rates. The decay rate is  
 12 dependent on  $CP$  parameters  $\mathcal{S}$  and  $\mathcal{A}$ . The reconstructed  $B^0$  is called signal-side and the  
 13 rest of events forming another  $B^0$  is called tag-side. The probability density function of  
 14 finding a event that signal-side  $B^0$  decaying into  $K_S^0 K_S^0 K_S^0$  when the decay time difference  
 15 of signal-side and tag-side is  $\Delta t$ , can be written as Eq(1):

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

16 where  $\tau_{B^0}$  is  $B^0$  lifetime,  $\Delta M_d$  is the mass different between two mass eigenstates of  $B^0$ .  
 17 The  $\Delta t$  is observed decay time difference between  $B^0$  and  $\overline{B^0}$ . The  $q = \pm 1$  corresponds to  
 18 the flavor of tag-side  $B^0(\overline{B^0})$ .  $\mathcal{S}$  and  $\mathcal{A}$  are  $CP$  parameters representing mixing-induced and  
 19 direct  $CP$  violation respectively. For this channel, the final state is  $CP$ -even definite, thus  
 20 the expected  $\mathcal{S}$  is  $-\sin(2\phi_1)$  and  $\mathcal{A}$  is zero under the Standard Model prediction, where  $\phi_1$   
 21 is one of the CKM angles. It's worth noticed that this channel receives small contribution  
 22 from resonant backgrounds which yields same final states but with  $CP$ -odd eigenvalue, such  
 23 as  $B^0 \rightarrow \chi_{c0}(\rightarrow K_S^0 K_S^0) K_S^0$ . It should be rejected by checking the Dalitz-plots of two  $K_S^0$   
 24 invariant mass from data, then applying veto. However, the current MC doesn't have correct  
 25 implementation of branching fraction for all resonant backgrounds, and it's hard to be cross-  
 26 checked with data in the current luminosity. Then it might not be safe to just using MC  
 27 distribution of  $2K_S^0$  invariant mass for veto considering the possible data/MC discrepancy.  
 28 The branching fraction of resonant backgrounds in MC is summarized in TABLE XI in  
 29 Appendix , compared with PDG value. The decays with  $f_0(980)$ ,  $f_0(1500)$  are also considered  
 30 as signal because they give  $CP$ -even final states that is called resonant signal. The number  
 31 of expected resonant background events in current luminosity is low (about 1 events) and  
 32 we have no veto applied yet, which will be further studied in future.

## 33 2. DATA SAMPLE AND EVENT SELECTION

### 34 2.1. Analysis Software

35 The MC sample is produced by EvtGen package with provided decay.dec file that de-  
 36 scribes the required decay mode and branching fraction. In this analysis, the generic  
 37 sample and signal sample are both in MC13. For the software version, BASF2 (Belle II  
 38 Analysis Framework) release-05-01-01 is used. As  $K_S^0 \rightarrow \pi^0 \pi^0$  leads to large background  
 39 with poor vertex quality, only the final states to 6 charged pions is used. The branch-  
 40 ing fraction of  $K_S^0 \rightarrow \pi^+ \pi^-$  is 0.692. The corresponding branching fraction in EvtGen is  
 41  $\mathcal{B}(B^0 \rightarrow K_S^0 K_S^0 K_S^0) = 6.0 \times 10^{-6}$ . So the expected branching fraction from generic  $B^0$   
 42 decay of charged final states is  $1.99 \times 10^{-6}$ . The simulation takes the  $\Upsilon(4S)$  as the mother  
 43 particle and uses decay mode “VSSBMIX” to generate its decay process to two scalar  $B^0$   
 44 mesons. Then the  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is generated based on the phase-space without intrinsic  
 45  $CP$  violation.

46        **2.2. Data and simulation samples**

47        In MC13, 1 million signal events with and without overlay of beam background are  
48        generated separately. For generic decay sample,  $B^0$  generic decay equivalent to  $1 \text{ ab}^{-1}$   
49        luminosity with TDCPV skim (“bd\_qqs\_Channels”) is used. For data, 2019 and 2020 (Spring  
50        and Summer) is included. The TDCPV skim selections on particle lists are based on the  
51        script provided by skimming for TDCPV working group, where the  $B^0$  and particle list  
52        selections in skimming is summarized as follows:

- 53        • phi:loose  
54        • K\_S0:all  
55        • eta:loose  
56        • pi0:eff40\_Jan2020 and skim  
57        • pi+:all  
58        • omega:loose  
59        • J/psi: eeLoose and mumuLoose  
60        • psi(2S): eeLoose and mumuLoose  
61        • K\*0:loose  
62        •  $5.2 < M_{bc} < 5.29$   
63        •  $abs(deltaE) < 0.5$   
64        •  $ntracks > 4$

65        The link to the skim script (may be updated) can be found here[4].

66        **2.3. Events selections Overview**

67        The final states particles of this channel are only the charged pions produced by  $K_S^0$   
68        decay, thus the reconstruction performance of  $K_S^0$  is essential to extract signal candidates.

69        First, we reconstruct  $K_S^0$  using standard particle library in BASF2. Then we adjust the  
70        momentum selection by rejecting ones with less than 0.05 GeV. The distribution of  $K_S^0$   
71        momentum is shown in FIG.1. Regarding the hits number of each  $K_S^0$  daughter has, a tuned  
72        selection for  $K_S^0$  invariant mass after vertex fit (“InvM” in variable manager in BASF2) is  
73        applied to further improve  $K_S^0$  purity. The part is covered in section 2.4.

74        With these selection,  $K_S^0$  candidates still contains a noticeable fraction of fake particles.  
75        In order to reject remaining fake  $K_S^0$  more efficiently, a FastBDT based  $K_S^0$  selection package  
76        called ”KsFinder” is developed and applied to pre-selected  $K_S^0$  candidates, which is described  
77        in section 2.5.

78        After the  $K_S^0$  selection is applied, we perform a continuum suppression analysis for re-  
79        ducing the other flavors background. Combined above selections,  $B^0$  can be reconstructed  
80        with certain cuts. Then we use RooFit for fitting the signal yield from the obtained  $B^0$   
81        candidates. This part will be covered in section 2.6.

82      **2.4.  $K_S^0$  pre-selection**

83      The  $K_S^0$  particles are reconstructed using "stdKshort:merged"[5] list from BASF2. It  
 84 merges the V0 candidates as  $K_S^0$  and reconstructed  $K_S^0$  using charged pions. The pion par-  
 85 ticles used by "stdKshort:merged" is with "pi+/-:all". The invariant mass from daughters'  
 86 4-vector between  $0.45 \text{ GeV} < M < 0.55 \text{ GeV}$  is applied and vertex is fitted with "Tre-  
 87 eFit", with confidence level of zero used to take all converged results. After the vertex fit,  
 88 the invariant mass from fitted 4-vector of daughter pions is also calculated as "InvM" in  
 89 BASF2.

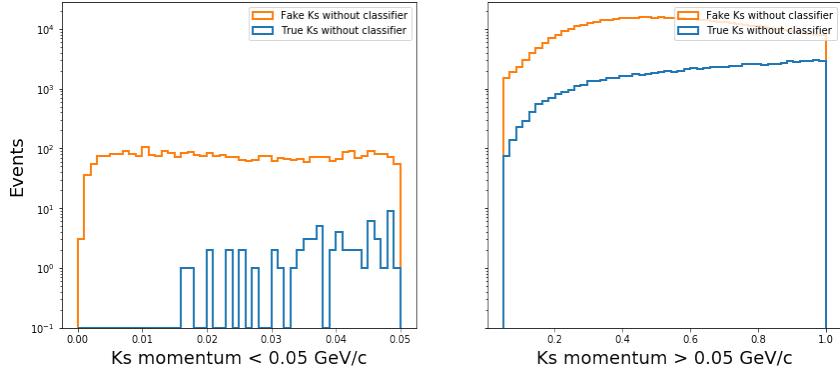


FIG. 1: The distribution of  $K_S^0$  momentum from stdKshort:merged. Candidates smaller than  $0.05 \text{ GeV}/c$  are rejected.

90      It's noticed that the  $K_S^0$  reconstruction performance is dependent on the flight length.  
 91 The distribution of flight length for true  $K_S^0$  in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  and generic decay are shown  
 92 in FIG.2. The efficiency is also reduced when the flight length becomes large[6]. The number  
 93 of SVD hits for each  $\pi^+$  and  $\pi^-$  correlates to different invariant mass ("InvM") distribution  
 94 of  $K_S^0$ . FIG.3 and TABLE I shows the vertex position of true  $K_S^0$  based on SVD hits in  $R - \phi$   
 95 plane, as well as the fraction of each type of  $K_S^0$ . The impact for SVD hits on invariant mass  
 96 after vertex fit is shown in FIG.4. The top two histograms shows the true  $K_S^0$  invariant mass  
 97 calculated directly from daughters' 4-vector, named "M" in BASF2, and as the stdKshort  
 98 list requires, between 0.45 and 0.55 GeV. The bottom 2 histograms in FIG.4 show the true  
 99  $K_S^0$  invariant mass using fitted 4-vector of daughters', named "InvM" in BASF2. It's clear  
 100 for  $K_S^0$  without SVD hits, the dispersion of fitted invariant mass (InvM) is larger. The lack  
 101 of SVD hits information in  $\pi^{+/-}$  leads to poor track fit quality which makes the invariant  
 102 mass after vertex fit dispersed, as the FIG.5 shows.

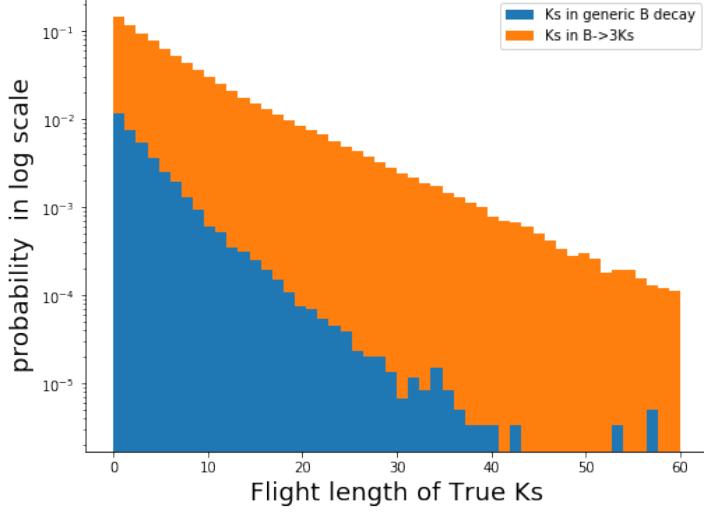


FIG. 2: Flight length of  $K_S^0$  in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  and generic decay. The orange is  $K_S^0$  from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  signal MC13 and blue is from generic MC13.

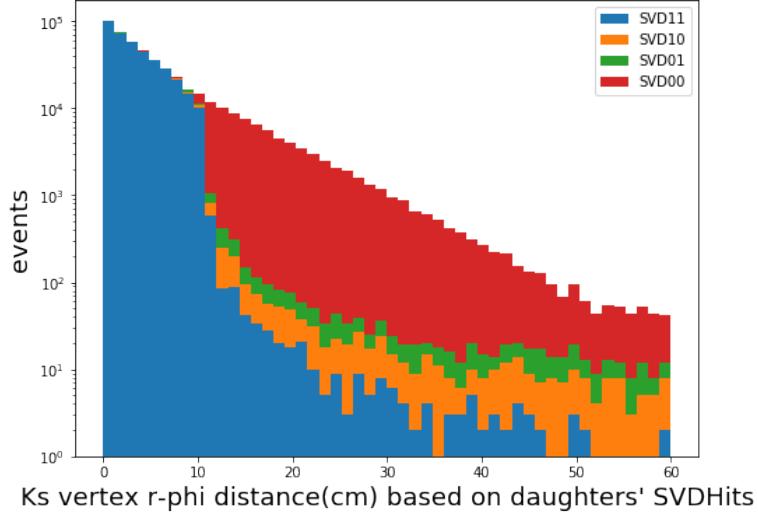


FIG. 3: Vertex position in R- $\phi$  plane for true  $K_S^0$  from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  signal MC. SVD11 stands for  $K_S^0$  which both daughters have at least 1 SVD hits, SVD00 stands for none SVD hits found from daughters. SVD10(SVD01) are  $K_S^0$  of which only  $\pi^+$  ( $\pi^-$ ) has SVD hits.

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

TABLE I: The fraction of each types  $K_S^0$  based on pions SVD hits in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  in signal MC.

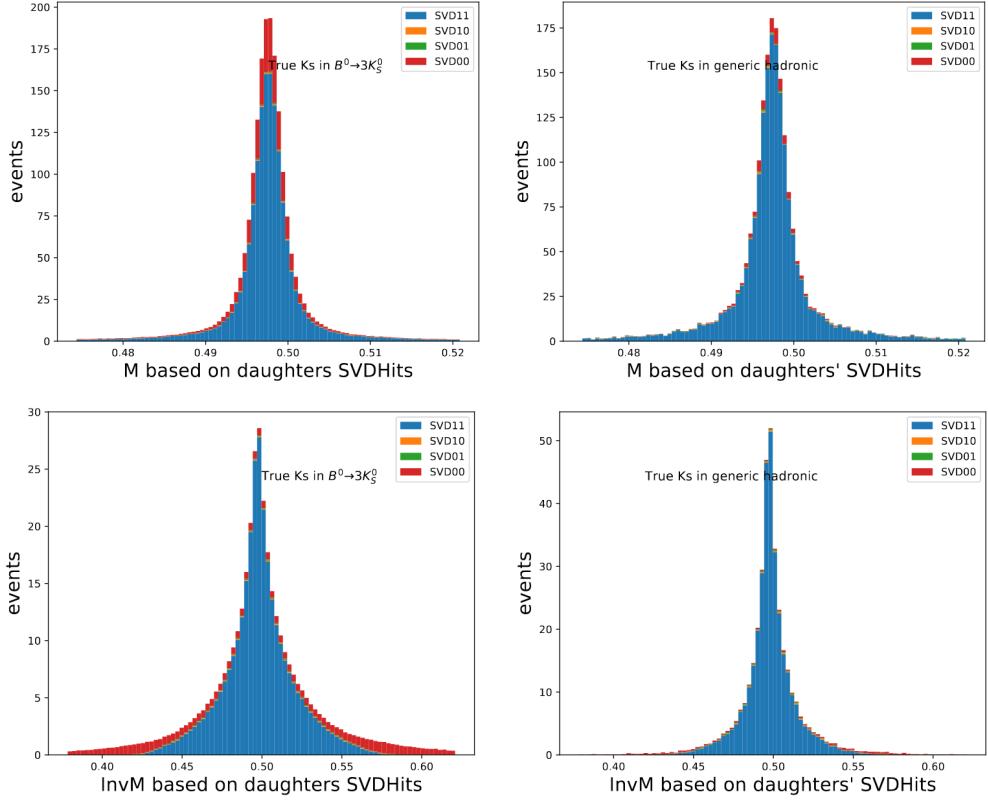


FIG. 4: (Top two) The invariant mass from daughters' 4-vector; (bottom two) The invariant mass from fitted 4-vector; The blue is SVD11, orange and green is SVD10 and SVD01, the red is SVD00  $K_S^0$ . The red part shows the clear dispersion on SVD00 type  $K_S^0$ .

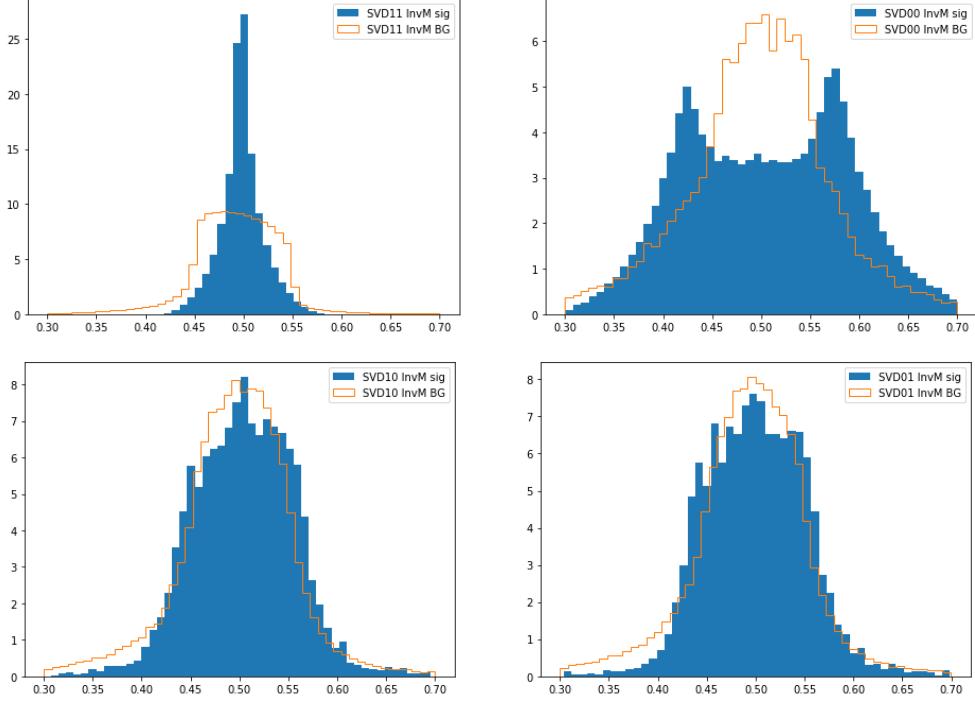


FIG. 5: Fitted invariant mass “InvM” on  $K_S^0$  types. The blue bar is for true  $K_S^0$  and the orange line is for fake  $K_S^0$ . A slightly tuned cut in TABLE II is used in each distribution to further reject fake  $K_S^0$

103      Besides the cut on “M” from stdKshort list, we choose the different invariant mass cut  
 104     on “InvM” for  $K_S^0$  in each category, to slightly reduce the fraction of fake  $K_S^0$  in ”InvM”  
 105     sideband. The cut window is summarized as TABLE II.

$K_S^0$ type	SVD11	SVD10	SVD01	SVD00
InvM window (GeV)	(0.43,0.57)	(0.38,0.7)	(0.38,0.7)	(0.3,0.7)

TABLE II: Invariant mass window “InvM” used for each  $K_S^0$  category according to FIG.5.

106      With the pre-selection from above, we collect the pre-selected  $K_S^0$  on which KsFinder  
 107     selection will be applied in the next section.

## 108      2.5. KsFinder

109      The pre-selected  $K_S^0$  candidates still contains a fraction of fake  $K_S^0$  that needs to be  
 110     purged. We developed a FastBDT based  $K_S^0$  selection package called KsFinder for this pur-  
 111     pose. It takes the characteristics of  $K_S^0$  decay including the information of kinematics, decay  
 112     shape parameters, particle identification and detector hits as input, with training target of  
 113     MCTruth of candidates  $K_S^0$  to achieve a better classification ability for  $K_S^0$  identification.  
 114     The choices of input parameters are considered based on typical fake  $K_S^0$  candidates, such  
 115     as the one demonstrated in FIG.6 and FIG.7. The output of KsFinder is the value between  
 116     0 and 1 which presents the likelihood of candidates being a true  $K_S^0$ .

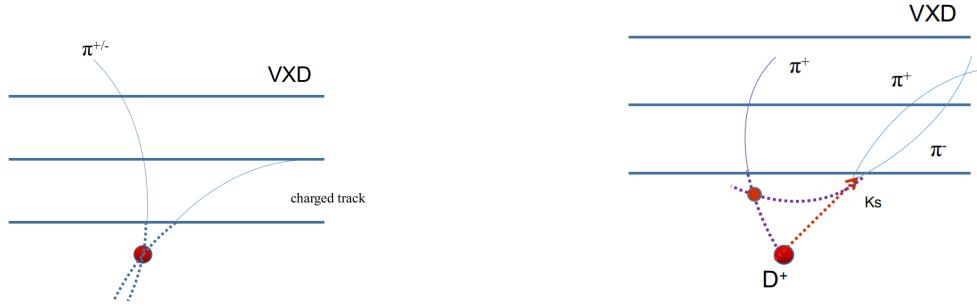


FIG. 6: The left shows the case when a charged track (not a pion) combined with a true pion as a fake  $K_S^0$ , the right shows the case when two daughters are correctly reconstructed as pion but not from the correct mother.



FIG. 7: The left shows the  $\Lambda \rightarrow p^+\pi^-$  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

117     The input variables for KsFinder is listed as follows:

- 118     ● Kinematics
  - 119       – Invariant mass of  $K_S^0$  before and after fitting vertex
  - 120       – momentum of  $K_S^0$  and  $\pi^{+/-}$ , vectors and magnitudes.
- 121     ● Decay shape parameters
  - 122       – cosine angle between  $K_S^0$  vertex and momentum.
  - 123       – helicity angle of two daughters in reference of  $K_S^0$  momentum.
  - 124       – decay angle of two daughters in the mother's frame.
  - 125       – flight distance projection on  $K_S^0$  momentum direction.
  - 126       – significance of flight distance.
  - 127       – distance on z-axis of two daughters helix
  - 128       – impact parameters on  $K_S^0$  vertex
- 129     ● Particle identifications
  - 130       – 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.

We checked the correlations between input variables to avoid including too many strongly correlated ones. The correlations are also supposed to be different in true and fake  $K_S^0$ . The correlations between input variables in signal and background are shown as FIG.8.

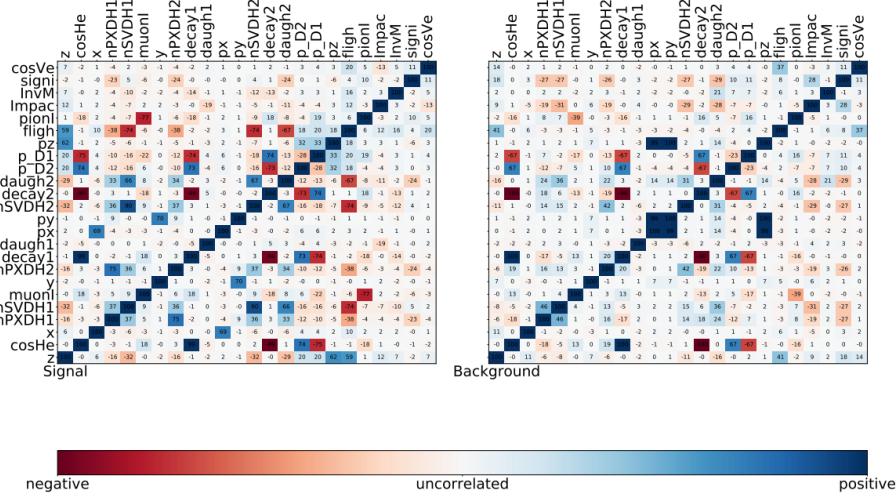


FIG. 8: The correlation matrix of the input variables. The left is obtained from true  $K_S^0$  and the right is from fake ones which shows differently. The correlations are obtained from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  signal MC13.

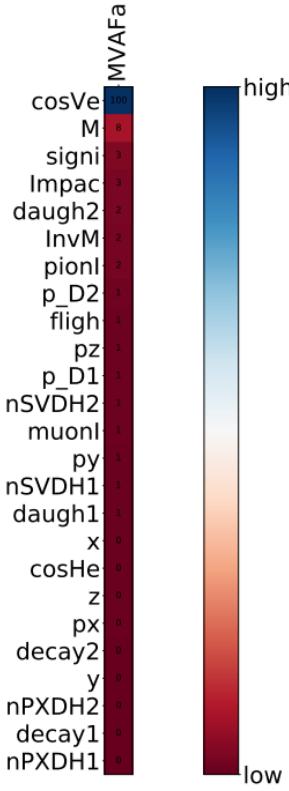
The training algorithm is FastBDT which is popularly used among various classification problems in high energy physics. For Belle II, the details of this method can be referenced from here[7].

As the preparation of training, we fill  $K_S^0$  particle list from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  and generic MC13 using stdKshort list, without further reconstruction on  $K_S^0$  mother. The pion particles are filled by “pi+/-:all” within the stdKshort list. The invariant mass “M” is constrain between 0.45 and 0.55 GeV, the vertex of  $K_S^0$  is fitted with TreeFit, according to the default selection inside stdKshort list, confidence level is set to zero to take all converged fit results. During the training, names of the variables and their importance are shown in TABLE III and TABLE IV.

TABLE III: 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 IV: Importance rank



149 In order to perform training and testing independently, training sample and testing sam-  
 150 ple are separately prepared using different input ROOT files from MC13. We train/test  
 151 KsFinder on both  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  and generic decay. The fraction of true and fake  $K_S^0$   
 152 in the sample is set to be 1 : 1. The target variable of training is the MCTruth of  $K_S^0$ .  
 153 To evaluate the performance, we use the ROC (receiver operating characteristics) curve to  
 154 check the background rejection in respect to signal efficiency. Background rejection and  
 155 signal efficiency are calculate as Eq(2) and Eq(3) by cutting on the output of KsFinder.  
 156 FIG.9 and FIG.10 shows the ROC, efficiency and purity from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  and generic  
 157 decay respectively. The testing results are in a good consistence with training result.

$$\text{signal efficiency} = \frac{\text{Number of true } K_S^0 \text{ with output} > \text{cut value}}{\text{Number of all true } K_S^0} \quad (2)$$

$$\text{background rejection} = \frac{\text{Number of fake } K_S^0 \text{ with output} < \text{cut value}}{\text{Number of fake true } K_S^0} \quad (3)$$

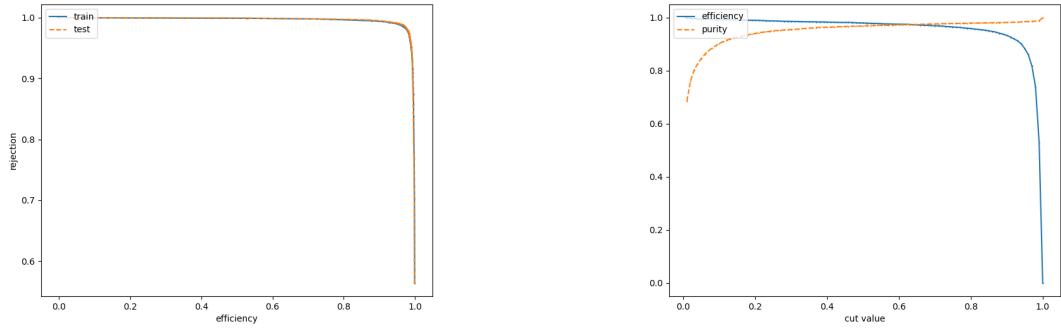


FIG. 9: The left:ROC curve, blue solid line is training and orange dashed line is testing, both are in a good consistence. The right: efficiency(blue) and purity(orange) depending on cut of classifier output. Results are from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  sample.

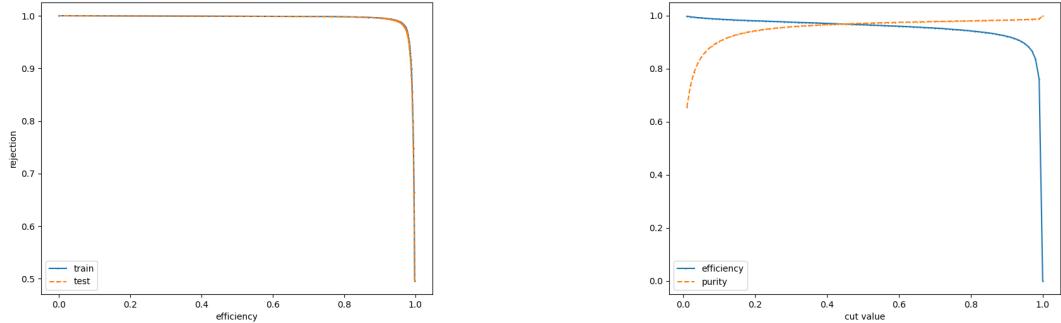


FIG. 10: The left:ROC curve, blue solid line is training and orange dashed line is testing, both are in a good consistence. The right: efficiency(blue) and purity(orange) depending on cut of classifier output. Results are from generic  $B^0$  decay.

158     The ROC curves show a good rejection power on fake  $K_S^0$ . The testing ROC should  
 159     not yield a noticeably better performance than it is in training, otherwise the classification  
 160     may be over-trained. The distribution of signal and background  $K_S^0$  respect to the cut  
 161     value are consistent between training and testing sample without over-fitting as proven in  
 162     the Appendix A. We check the maximized FOM ( $\frac{S}{\sqrt{S+B}}$ , S an B stands for number true  
 163     and fake  $K_S^0$  respectively) on signal MC of  $K_S^0$  from  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  to determine the cut  
 164     value at 0.74, see FIG.11. The true  $K_S^0$  fraction in the pre-selected  $K_S^0$  before applying  
 165     KsFinder cut is 39%, and 95.3% of them are remained after KsFinder applied. The fake  $K_S^0$   
 166     fraction before applying KsFinder cut is 61%, and 97.6% of them are rejected after KsFinder  
 167     applied. We use this cut on the  $K_S^0$  to improve the purity, which is particularly important  
 168     for  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ .

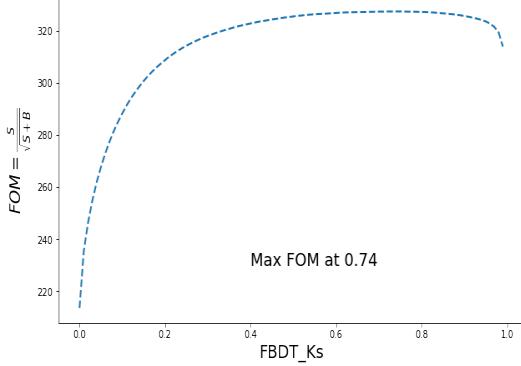


FIG. 11: FOM of classifier output (FBDT\_Ks) in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ , maximum value is obtained at 0.74.

To check the  $K_S^0$  classification on data, we compare the training variables distribution in both generic MC and data, which full plots are in Appendix B. It is shown that most of variables agrees well between MC and data. Data after KsFinder cut applied also agrees with MC signal in a reasonable degree.

## 2.6. $B^0$ Reconstruction

By combining three  $K_S^0$  particles from selected dataset, we can reconstruct  $B^0$ . The beam-constraint mass  $M_{bc}$  and energy difference  $\Delta E$  are used to extract signal. To suppress the continuum background from  $q\bar{q}$  events, a classification of  $B\bar{B}$  and  $q\bar{q}$  events are performed with a set of event-shape based variables, see here for details[8].

The continuum events are reconstructed from generic off-resonance ( $q\bar{q}$ ) sample in MC13. The ratio of continuum and  $\Upsilon(4S)$  is 1:1 in training and testing. The abbreviations of variables used and the importance rank can be found in FIG.13. The over-training check is listed in Appendix B.1, with about 1% difference in the train and test samples. The correlation between variables, ROC curve, efficiency and purity are all shown in FIG.12, which the correlation among  $M_{bc}$  and  $\Delta E$  are also included. We select the cut on the “FBDT\_CS” (output from classification) at 0.66 which maximizes the FOM ( $\frac{S}{\sqrt{S+B}}$ ) to suppress the continuum background, also see FIG.13 bottom right.

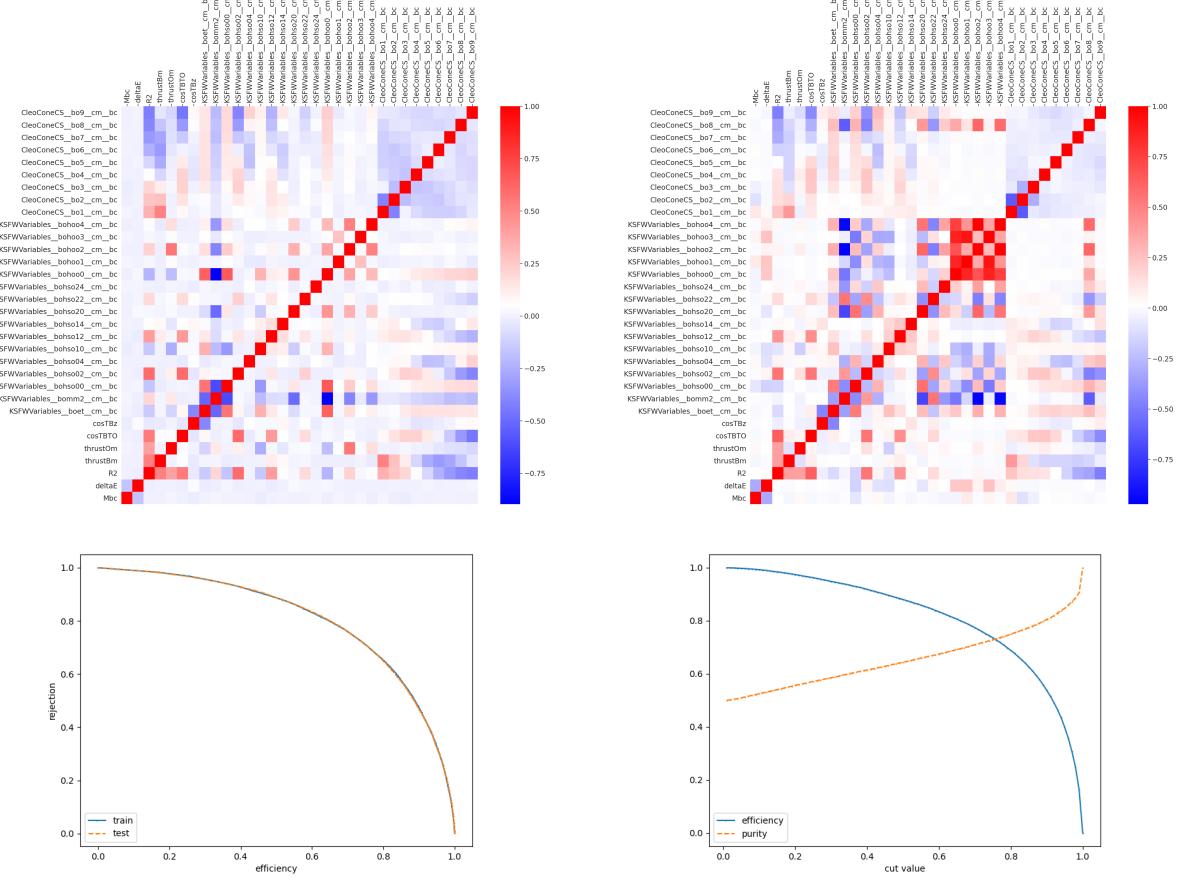


FIG. 12: The top is correlation in variables for continuum suppression, including  $M_{bc}$  and  $\Delta E$  (red for positive and blue for negative correlation). Top left is from signal and the top right is from background. The bottom left is the ROC curve (blue for training and orange for testing, in good consistence) The bottom right is the efficiency(blue) and purity(orange) regarding the cut value “FastBDT\_CS”

Observables	Abbreviations
CleoConeCS(9,)	CleoC1
KSFVWVariables(hoo1,)	KSFVW1
CleoConeCS(7,)	CleoC2
CleoConeCS(5,)	CleoC3
KSFVWVariables(hso22,)	KSFVW2
KSFVWVariables(hoo3,)	KSFVW3
CleoConeCS(4,)	CleoC4
KSFVWVariables(hoo4,)	KSFVW4
CleoConeCS(3,)	CleoC5
CleoConeCS(6,)	CleoC6
CleoConeCS(8,)	CleoC7
KSFVWVariables(hso14,)	KSFVW5
KSFVWVariables(hso00,)	KSFVW6
KSFVWVariables(et,)	KSFVW7
KSFVWVariables(hso24,)	KSFVW8
KSFVWVariables(hso04,)	KSFVW9
KSFVWVariables(hso20,)	KSFVW10
KSFVWVariables(mm2,)	KSFVW11
KSFVWVariables(hoo2,)	KSFVW12
thrustOm	thrus1
cosTBz	cosTB1
CleoConeCS(1,)	CleoC8
CleoConeCS(2,)	CleoC9
KSFVWVariables(hso02,)	KSFVW13
KSFVWVariables(hoo0,)	KSFVW14
KSFVWVariables(hso12,)	KSFVW15
KSFVWVariables(hso10,)	KSFVW16
cosTBTO	cosTB2
thrustBm	thrus2
R2	R2

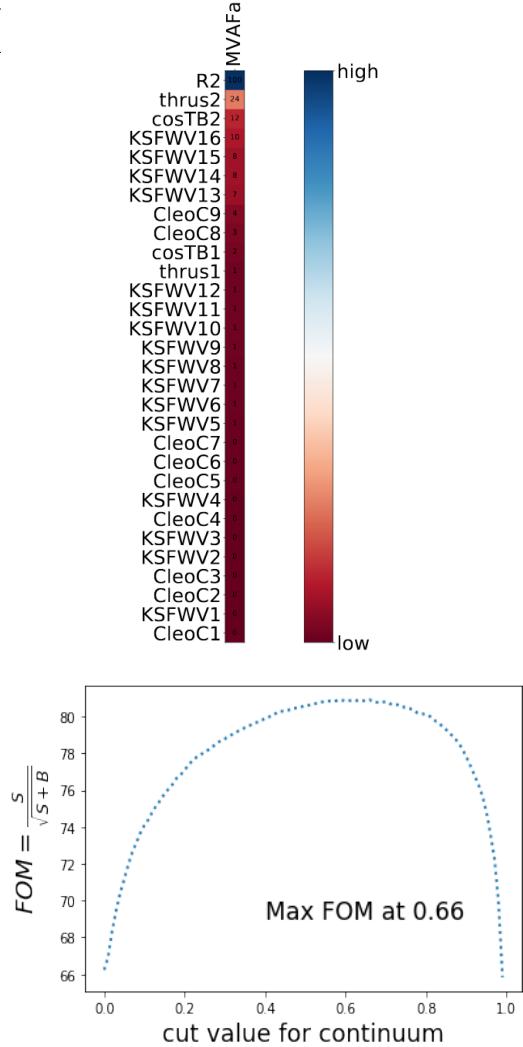


FIG. 13: Left: the abbreviation of variables used in continuum suppression. Right Up: the importance rank for train variables. Right Down: FOM depending on the cut value of continuum classifier output, cut value at 0.66 is used for continuum suppression.

186 We select  $B^0$  with the criteria as TABLE V. The vertex fit using TreeFit is performed on  
 187 each  $B^0$  candidate and we keep ones with  $\text{chiProb} > 0.001$ . When multiple  $B^0$  candidates  
 188 are found in a single event, a rank by their vertexing quality is performed using  $\chi^2$  of the  
 189 vertex fit, from which the smallest have the highest rank. The best candidates are selected  
 190 (BCS) with the highest rank. Since the BCS is based on the vertexing quality that might  
 191 introduce bias in the vertex positions for  $CP$  fit, we check the distribution of the vertex

192  $\chi_2$  from TreeFit before the BCS is used, as shown in FIG.14 top right. The distribution  
 193 of the number of candidates  $B^0$  per event in signal MC is shown as FIG.14(bottom left).  
 194 For  $M_{bc}$  and  $\Delta E$  which are used for defining signal region, we use  $5.20 < M_{bc} < 5.29$  GeV  
 195 and  $-0.2 < \Delta E < 0.2$  GeV. The 2D distribution of  $M_{bc}$  and  $\Delta E$  is in the bottom right of  
 196 FIG.14.

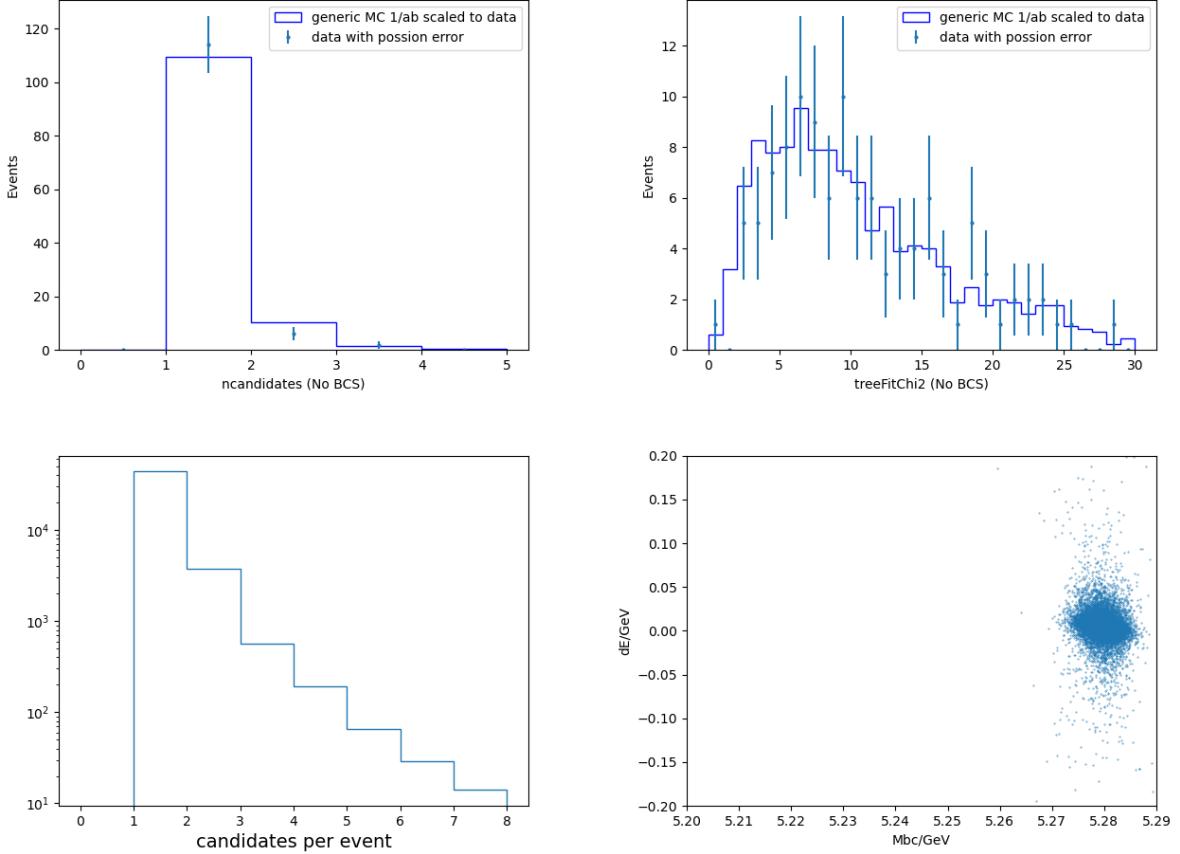


FIG. 14: Top left is the candidates per event in data and generic MC before the BCS. Top Right is the  $\chi_2$  for data and generic MC before BCS. Bottom left is the number of candidates  $B^0$  per event from signal MC. Bottom right is the 2D  $M_{bc}$  and  $\Delta E$  distribution from signal MC.

The  $B^0$  selection criteria is summarized TABLE V. The statistics of reconstruction using

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

TABLE V:  $B^0$  selection criteria

197 signal MC13 sample (1M events) is:  
 198

event selection	efficiency	purity	$f_{MB}$	BCS
Belle II (BG1)	36%(34%)	96%(98%)	(4%)(4%)	95%(96%)
Belle II (BG0)	40%(36%)	96%(99%)	(3%)(3%)	97%(97%)

TABLE VI: 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 the fraction of true  $B^0$  which comes from a event that has multiple candidates. BCS is the fraction of best candidates being a true signal. In each cell, the number in the parenthesis is calculated in  $5.27 < M_{bc} < 5.29$  GeV and  $|\Delta E| < 0.1$  GeV, so the efficiency is slightly dropped but purity is improved.

199 In order to extract  $B^0$  candidates, we use RooFit to fit to the  $M_{bc}$  and  $\Delta E$  distribution.  
200 The fit is firstly done on  $M_{bc}$  and  $\Delta E$  in 1D to obtain shapes, then use the parameters  
201 from 1D fit to form 2D  $M_{bc}/\Delta E$  model to extract  $B^0$ . For signal events, we use single and  
202 triple gaussian distribution to fit the distribution of  $M_{bc}$  and  $\Delta E$  respectively, which the fit  
203 method is unbinned maximum likelihood fit. The results are shown in below:

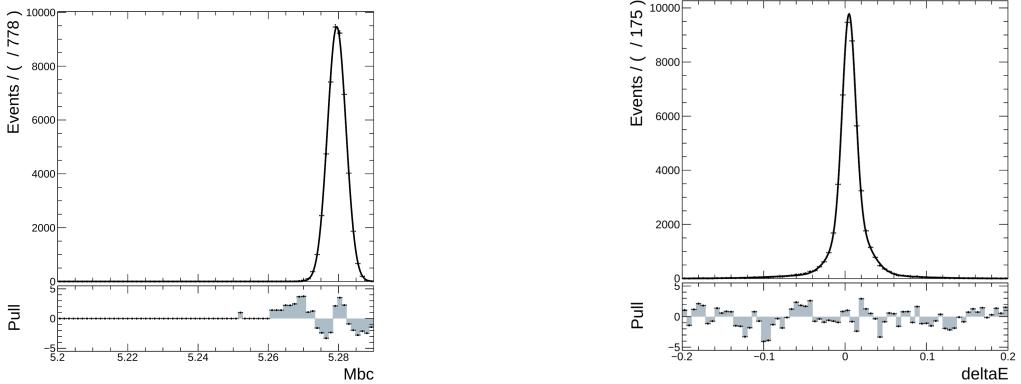


FIG. 15: The distribution of  $M_{bc}$  and  $\Delta E$  of signal MC13 of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  fitted with single and triple gaussian respectively.

204 This channel is much clean and the background contributions from self-cross feed(SCF)  
205 and  $B\bar{B}$  is expected to be about 3%. So the primary background is continuum background  
206 and we don't have stand-alone shape modeling for SCF and  $B\bar{B}$  background in  $M_{bc}/\Delta E$  fit.  
207 Also, as discussed in the introduction, the expected  $CP$ -odd resonant background is very  
208 low at current luminosity. No veto is applied on this concern and the Dalitz plot using  $2K_S^0$   
209 invariant mass in signal region is added in Appendix C.

210 For continuum background, we model the  $M_{bc}$  and  $\Delta E$  with Argus function and 1st order  
211 Chebyshev polynomial. The fit is performed on background events reconstructed from all  $q\bar{q}$   
212 generic sample after the continuum suppression. The fit method is also unbinned maximum  
213 likelihood fit.

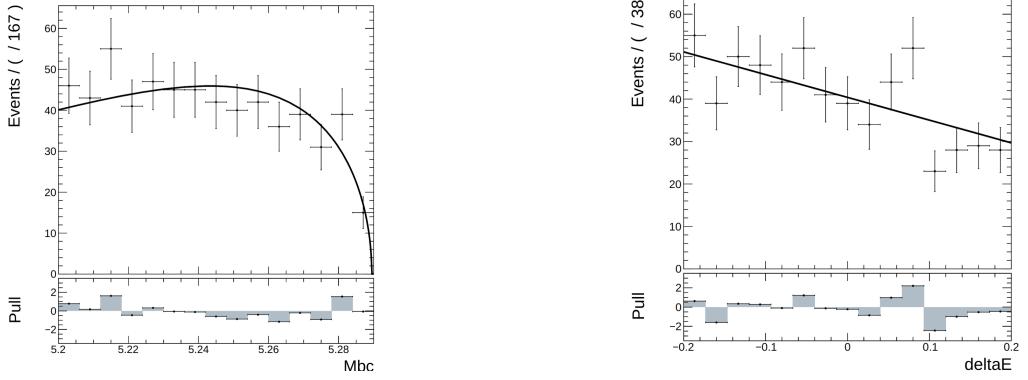


FIG. 16: The distribution of  $M_{bc}$  and  $\Delta E$  of continuum events fitted with Argus and Chebychev polynomial respectively.

The signal extraction is based on the 2D fit of  $M_{bc}$  and  $\Delta E$ . The shape parameters are fixed by  $M_{bc}$  and  $\Delta E$  fitted to signal and background sample in 1D fit. For 2D  $M_{bc}/\Delta E$  fit, the unbinned maximum likelihood fit is also used based on the Eq(4).

$$\mathcal{P}(M_{bc}, \Delta E) = N_{sig} \cdot \mathcal{P}_{sig}^{M_{bc}} \cdot \mathcal{P}_{sig}^{\Delta E} + N_{bkg} \cdot \mathcal{P}_{bkg}^{M_{bc}} \cdot \mathcal{P}_{bkg}^{\Delta E} \quad (4)$$

First, we perform 2D fit on  $1\text{ab}^{-1}$  generic MC sample to check the result in FIG.17. Then we perform 2D fit on reconstructed  $B^0$  from data collection of proc11 and bucket 9,10,11,13,14,15, which the integral luminosity is about  $62.8\text{fb}^{-1}$ . The event number for signal and background are floating in 2D maximum likelihood fit while shape parameters for  $M_{bc}$  and  $\Delta E$  are all fixed, see FIG.18.

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, where the  $B^0$  selection is listed in TABLE V. The distributions are shown in FIG.18, which the generic MC is scaled to the luminosity of data.

We also checked the distribution of  $\chi_2/N$  where  $\chi_2$  is from TreeFit and N is the degree of freedom in vertex fit, see Eq(8).The scatter plot of  $M_{bc}/\Delta E$  and  $\chi_2/N$  distribution in data are shown in FIG.19.

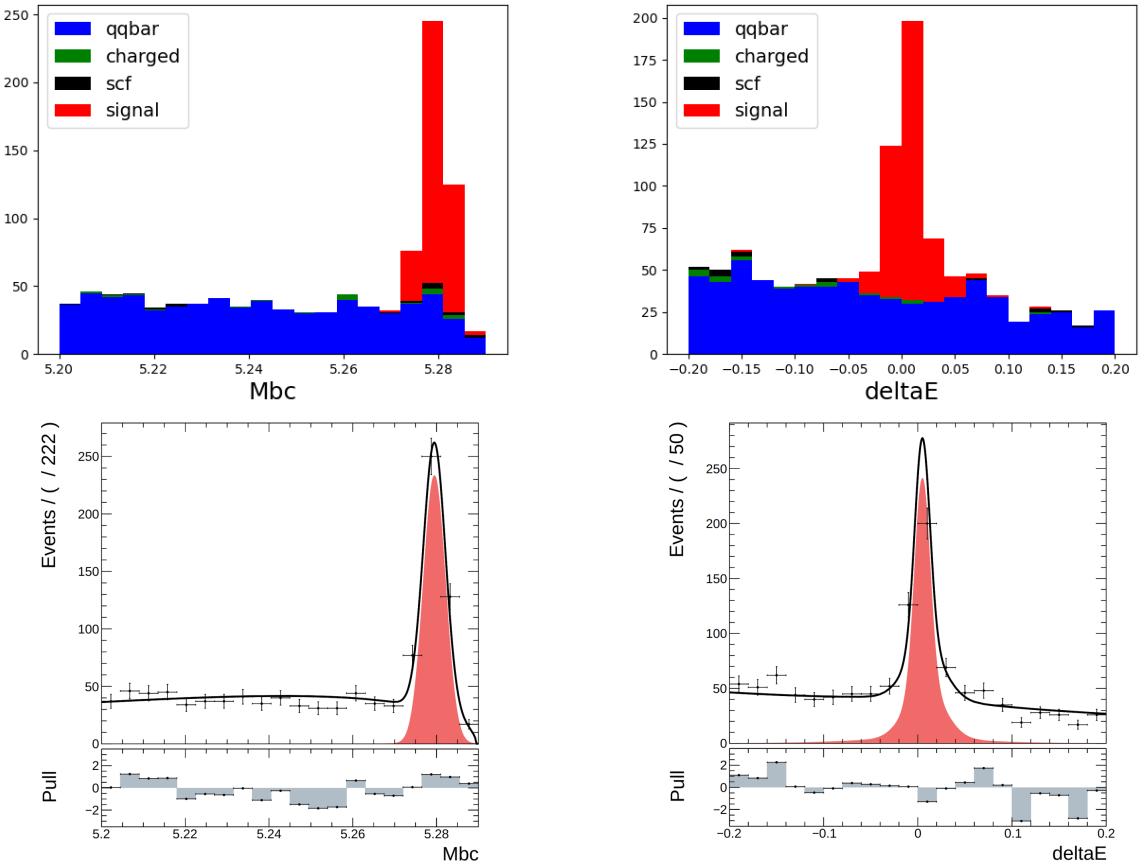


FIG. 17: Top is the stacked plots for generic MC of  $M_{bc}$  and  $\Delta E$ , where each background components are stacked with signal. The bottom is the 2D fit on generic MC13, the red is signal PDF. The integral luminosity is  $1 \text{ ab}^{-1}$

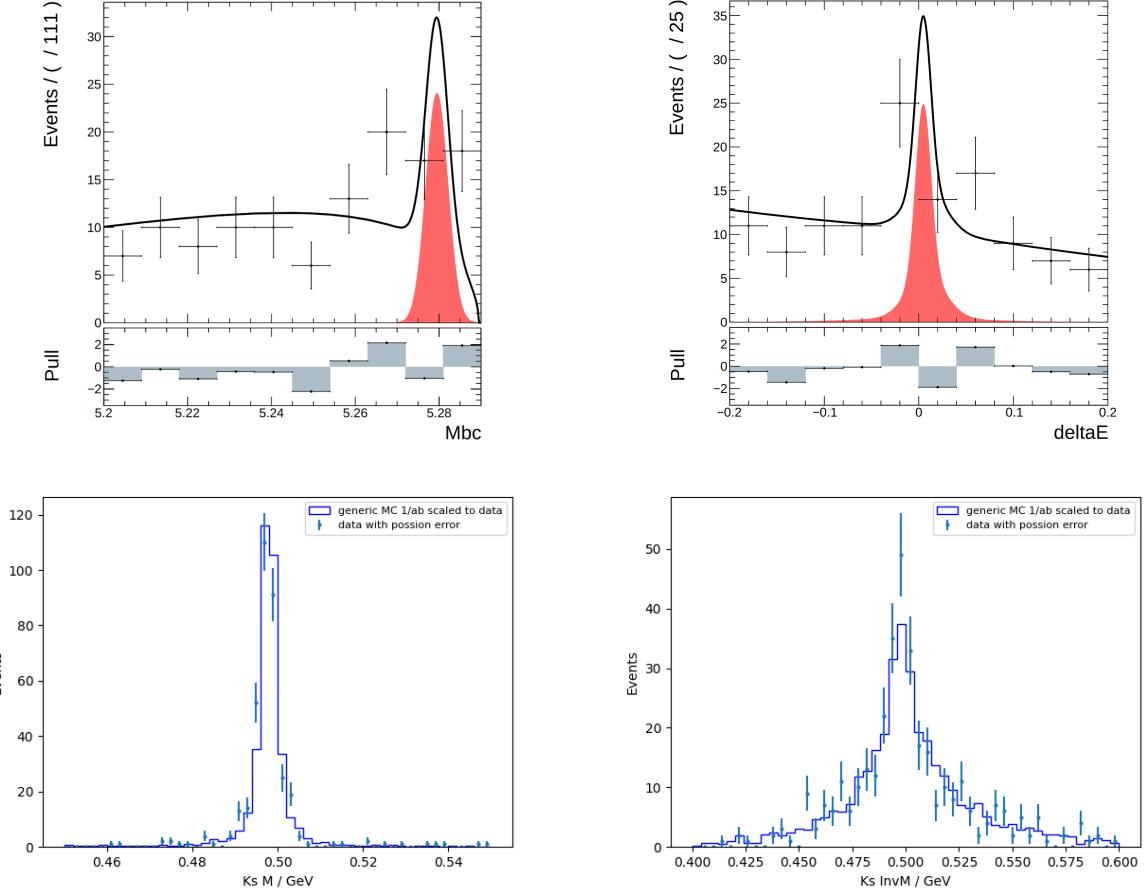


FIG. 18: Top:  $M_{bc}$  and  $\Delta E$  2D fit using  $62.8 \text{ fb}^{-1}$  data, the red is signal PDF. Bottom: “M” and “InvM” from data and generic MC which is scaled to data luminosity. The distributions is in a reasonable consistency.

226      The  $\chi_2/N$  distribution and  $M_{bc}/\Delta E$  in 2D plot from data are shown in FIG.19.

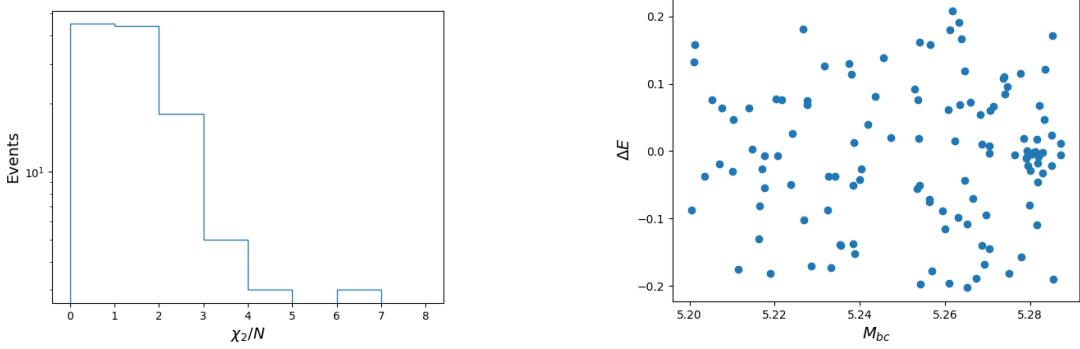


FIG. 19: The left: the distribution of  $\chi_2/N$ , the right: scattered plot of  $M_{bc}$  and  $\Delta E$ . The results are from data.

227      The number of signal events is extracted by integral of fit function Eq(4) over the signal  
228      region which is defined as  $5.27 < M_{bc} < 5.29 \text{ GeV}$  and  $-0.1 < \Delta E < 0.1 \text{ GeV}$ . In  $1 \text{ ab}^{-1}$

generic MC13, the expected signal events with 35% efficiency in this analysis is calculated as:  $7.7 \times 10^8 \times 6 \times 10^{-6} (BR : B^0 \rightarrow 3K_S^0) \times 21\% (BR : K_S^0 \rightarrow \pi^+ \pi^-) \times 35\% \simeq 339$ . The fit result from  $M_{bc}/\Delta E$  yields  $341 \pm 20$  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  $772 \times 10^6 B\bar{B}$  pairs used, signal from data yields  $327 \pm 19$ , which is also consistent. In  $62.8 \text{ fb}^{-1}$  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.

To check linearity of the event number fitted from the  $M_{bc}$  and  $\Delta E$  in this low statistics case, we extract the fraction of continuum backgrounds from generic MC13 sample rescaled to the experimental data luminosity, which yields about 46.5 background events in  $62.8 \text{ fb}^{-1}$ . 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, see FIG.20.  $M_{bc}$  and  $\Delta E$  distribution in each injection test are shown in Appendix D.

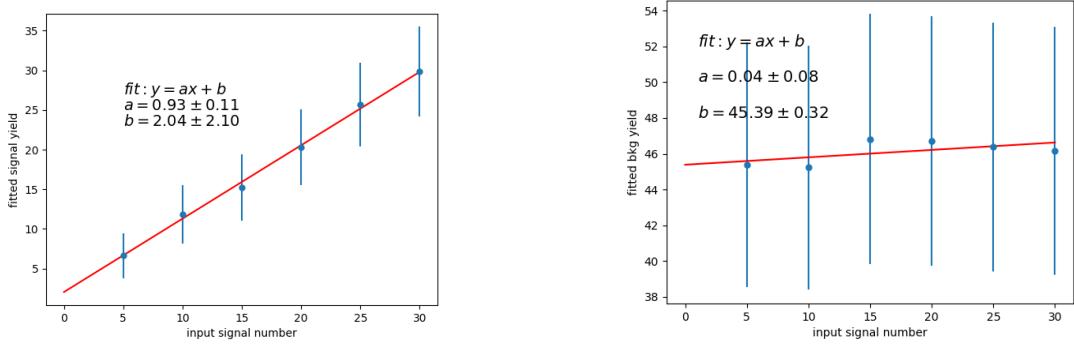


FIG. 20: Injection test for signal extraction. The linearity is clear between input and output signal events number.

### 3. FLAVOR TAGGING

The  $B^0$  flavor  $q$  is determined by associated side B meson using flavor tagging package. The details of flavor tagger package can be referenced from here[9]. The flavor tagging is called for the reconstructed  $B^0$  and  $\mathcal{F} = (q \cdot r)_{FBDT}$  is assign to each tagged events, to evaluate the wrong tagging fraction  $w$  and  $q$ .

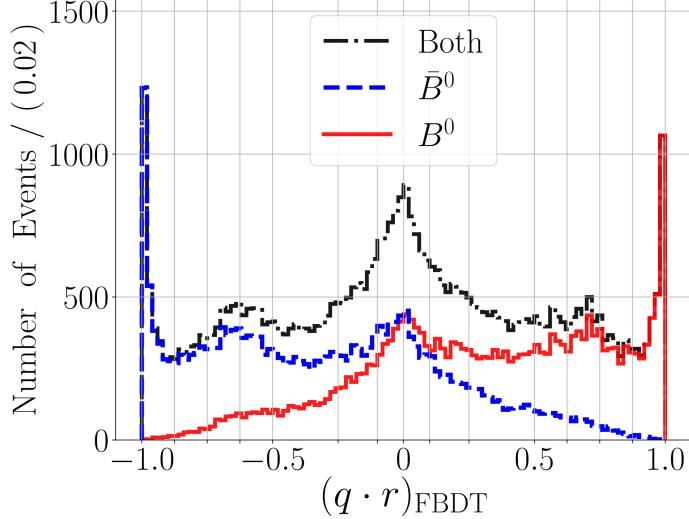


FIG. 21: The distribution of flavor tagger output  $\mathcal{F} = (q \cdot r)_{FBDT}$  for both tag side being  $B^0$  and  $\bar{B}^0$ .

251     The r-bins are intervals that splits the absolute value of  $\mathcal{F}$ . Belle experience on defining  
 252     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  
 253     events that have been successful tagged, they are projected into histogram of  $|\mathcal{F}|$ ,  $w$  can be  
 254     calculated in each bin in the way dividing the number of events that its  $\mathcal{F}$  is opposite to  
 255     its MC flavor to all the events in the same bin. The binned tagging efficiency is calculated  
 256     through dividing the number of successfully tagged events by all events in the same bin.  
 257     Considering  $w = (w_{B^0} + w_{\bar{B}^0})/2$  and  $\Delta w = w_{B^0} - w_{\bar{B}^0}$ , tagging efficiency is  $\epsilon$  and  $\mu$  is  
 258     efficiency difference in  $B^0$  and  $\bar{B}^0$ , see FIG.22. The observed distribution in Eq(1) turns  
 259     into:

$$\mathcal{P}_{sig}^{obs}(\Delta t, q, \epsilon, w) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \epsilon \left\{ 1 - q \cdot \Delta w + q(1 - 2w) \cdot [\mathcal{S} \sin(\Delta M_d \Delta t) + \mathcal{A} \cos(\Delta M_d \Delta t)] \right\} \quad (5)$$

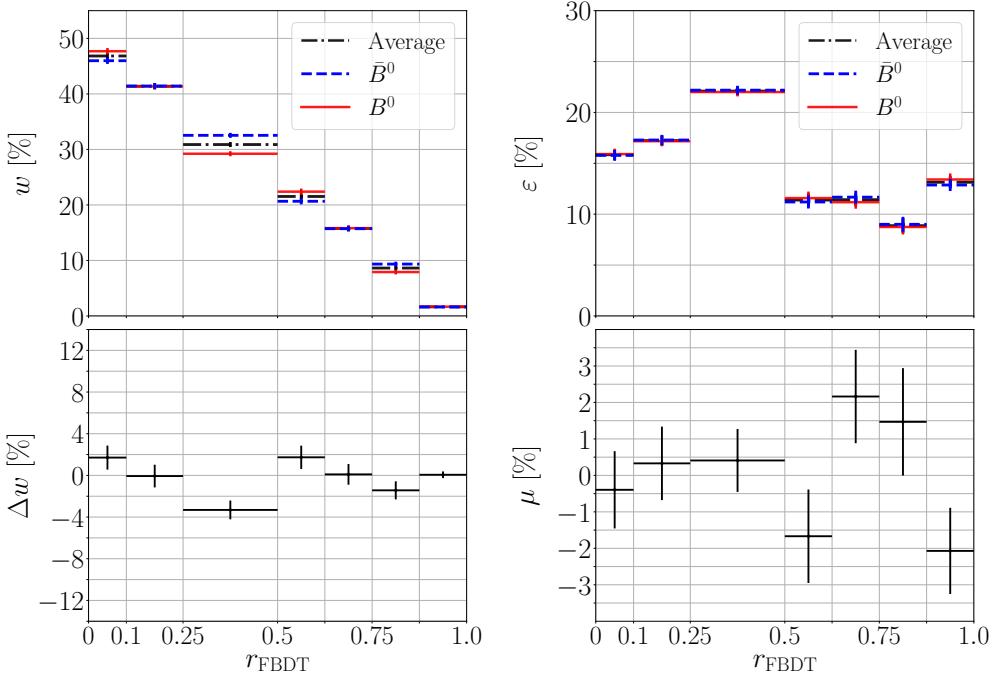


FIG. 22: (left)  $w$  and  $\Delta w$  in r-bins; (right)  $\epsilon$  and  $\mu$  in r-bins, for  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$ .

260 **4. VERTEX RECONSTRUCTION AND  $\Delta t$  RESOLUTION MODEL**

261 **4.1. Vertex Reconstruction**

262 In the vertex reconstruction of a event, we use different options for  $CP$ -side and tag-  
 263 side vertex fit. For  $CP$ -side, we use the recommended fitter option TreeFit[10] with no IP  
 264 constraint used, and for tag-side, the KFit with no IP constraint is chosen[11]. In order  
 265 to better describe the resolution of reconstructed vertices, we check the distribution of  
 266 residual distance for  $CP$ -side with the slice of  $\chi_2/N$  (redChi2). The residual has positive  
 267 dependence with  $\chi_2/N$  so the  $\chi_2/N$  is taken into account for scaling the resolution functions.  
 268 The resolution function in  $CP$ -side is presented and fitted with signal MC sample. It shows  
 269 a small discrepancy when the  $\chi_2/N$  goes to large value, see below:

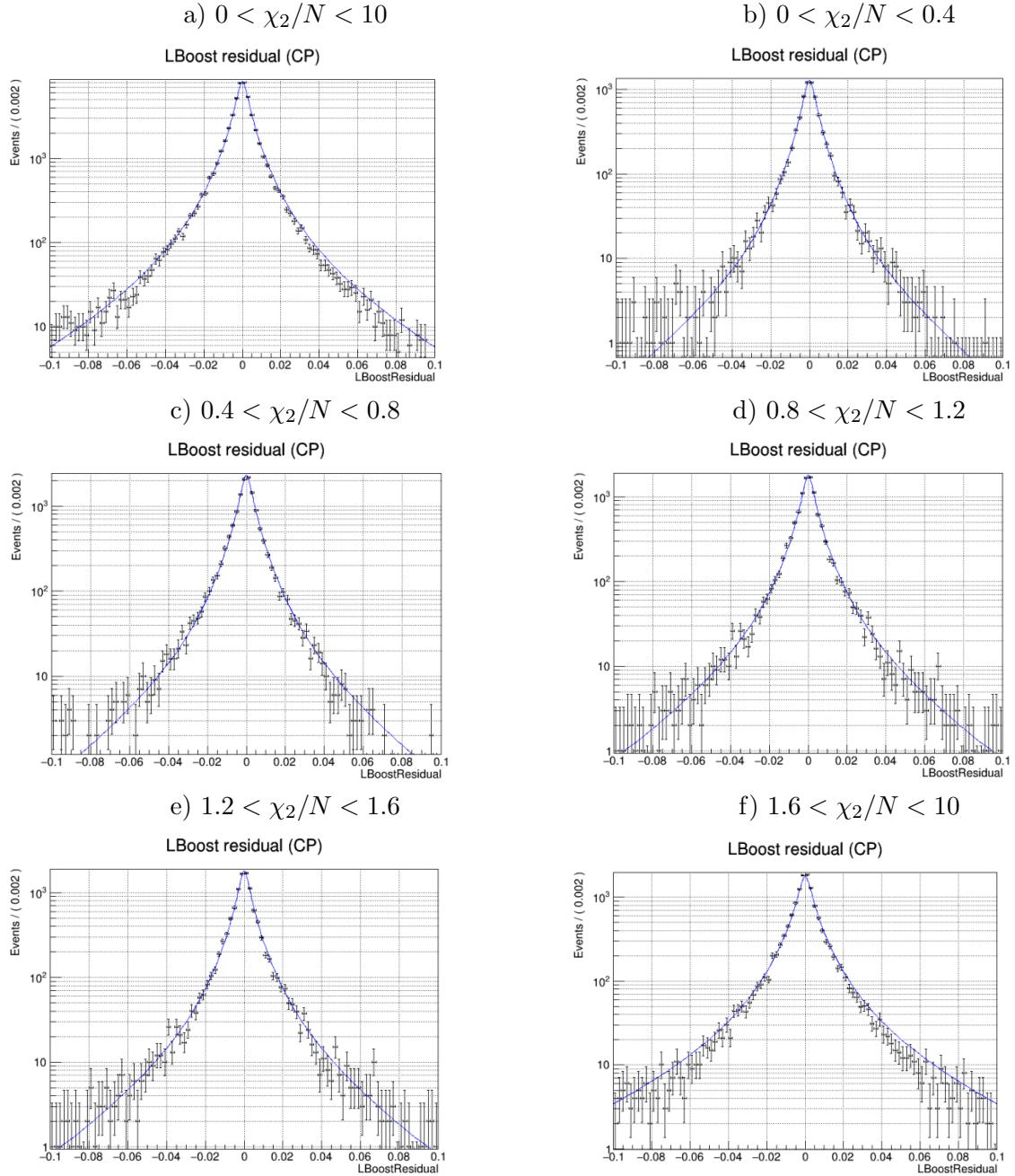


FIG. 23: Resolution functions on *CP*-side, which shows dependence on the  $\chi_2/N$

270     Restrictively speaking, the *CP*-side resolution for  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is slight different from  
 271 other channels such as  $B^0 \rightarrow J/\psi K_S^0$ , due to the lack of the direct charged tracks from the  
 272  $B^0$  vertex. The way of scaling the resolution will be further studied along with more data  
 273 available. Given the current low statistics, the above model works well as an approximation.

274     For tag-side, we use the Belle style artificial model to describe the resolution, which  
 275 receives the contribution from detector resolution, non-primary tracks effects and kinematic  
 276 approximation of  $\Delta t$ . The detail is described here[12].

## 4.2. $\Delta t$ resolution Model

The maximized likelihood fit is used from event-by-event which  $i$  is the index of events. The fit function used to extract  $CP$  parameters can be presented in Eq(6), where the resolution functions in both  $CP$  and tag side as well as the background and outlier are considered. In this study, we don't include the outlier shapes for early analysis to have a more realistic estimation of  $\Delta t$  distribution. The signal fraction  $f_i^{sig}$  is calculated from the 2D distribution of  $M_{bc}$  and  $\Delta E$  event-by-event. The resolution functions are scaled by the  $\chi^2/N$  in both side as conditional variables that can also be accessed event-by-event during the fitting. The resolution function of signal events in  $CP$  and tag side are demonstrated in Eq(8) to Eq(13).

$$\begin{aligned} \mathcal{P}(\Delta t_i, q_i, f_i^{sig}, \mathcal{S}, \mathcal{A}) = & (1 - f_{ol})[f_{sig}\mathcal{P}_{sig}(\Delta t_i) \otimes R_{sig}(\Delta t_i) \\ & + (1 - f_{sig})\mathcal{P}_{bkg}(\Delta t_i) \otimes R_{bkg}(\Delta t_i)] \\ & + f_{ol}\mathcal{P}_{ol}(\Delta t_i) \otimes R_{ol}(\Delta t_i) \end{aligned} \quad (6)$$

$$R_{sig} = R_{cp}(z_{cp} - z'_{cp}) \otimes R_{tag}(z_{tag} - z'_{tag}) \quad (7)$$

for  $CP$ -side,

$$R_{cp}(\delta z_{cp}^{det}) = (1 - f_{cp}^{tail})G(0, s_{cp}^{main} \cdot \sigma_{z_{cp}}) + f_{cp}^{tail}G(0, s_{cp}^{tail} \cdot \sigma_{z_{cp}}) \quad (8)$$

$$\begin{aligned} s_{cp}^{main} &= s_0^{main} + s_1^{main} \cdot \chi_{cp}^2/N \\ s_{cp}^{tail} &= s_0^{tail} + s_1^{tail} \cdot \chi_{cp}^2/N \end{aligned} \quad (9)$$

for tag-side,

$$R_{tag}(z_{tag} - z'_{tag}) = R_{det}^{tag}(\delta z_{tag}^{det}) \otimes R_{np}^{tag}(\delta z_{tag}^{np}) \quad (10)$$

$$R_{det}^{tag}(\delta z_{tag}^{det}) = (1 - f_{tag}^{tail})G(0, s_{tag}^{main} \cdot \sigma_{z_{tag}}) + f_{tag}^{tail}G(0, s_{tag}^{tail} \cdot \sigma_{z_{tag}}) \quad (11)$$

$$s_{tag}^{main/tail} = s_0^{main/tail} + s_1^{main/tail} \cdot \chi_{tag}^2/N \quad (12)$$

$$R_{tag}^{np}(\delta z_{tag}^{np}) = f_\delta \delta(\delta z_{tag}^{np}) + (1 - f_\delta)[f_p E_p(\delta z_{tag}^{np}, \tau_p \cdot \sigma_{z_{tag}}) + (1 - f_p)E_n(\delta z_{tag}^{np}, \tau_n \cdot \sigma_{z_{tag}})] \quad (13)$$

We determine the  $CP$ -side resolution parameters by fitting to signal MC13 as FIG.23(a) shows. Tag-side parameters which is almost mode-independent are adapted from control sample study[11]. The parameters for each components are shown in TABLE VII to IX.

TABLE VII: Parameters in  $R_{cp}$ .

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

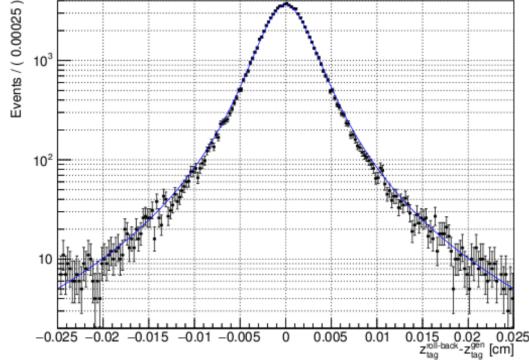


FIG. 24:  $R_{det}^{tag}$  fit

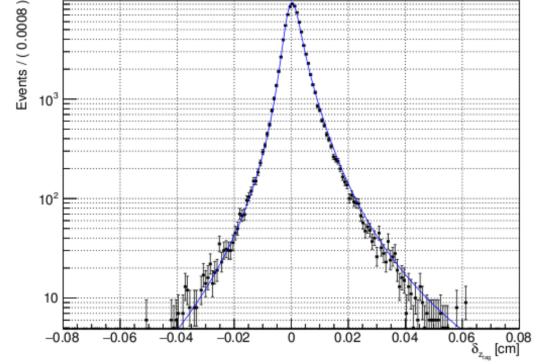


FIG. 25:  $R_{np}^{tag}$  fit

TABLE VIII: Parameters in  $R_{det}^{tag}$

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

TABLE IX: Parameters in  $R_{np}^{tag}$

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

282 For background resolution, we use double-gaussian model which the standard deviation  
283 is scaled by the vertex errors from both  $CP$  and tag-side.

284 The  $P_{bkg}$  and  $R_{bkg}$  are model based on Eq(14) and Eq(15), of which the fit is performed  
285 on sideband events in data cut at  $M_{bc} < 5.26$  GeV. We assume the only background in data  
286 is continuum and perform the fit using this definition of sideband on data where the number  
287 of events are 60.  $\Delta E$  cut is not used in order to increase the number of events that'll be  
288 used. The distribution of  $\Delta t$  for backgrounds is shown FIG.26.

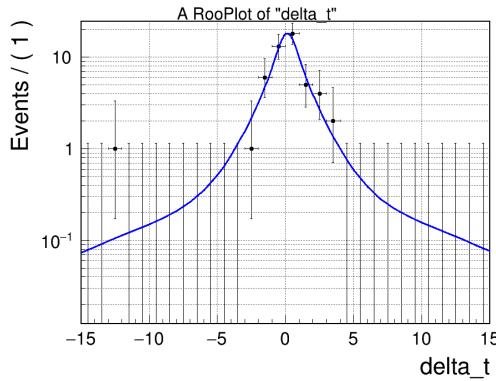


FIG. 26: Background fit on data sideband with  $M_{bc} < 5.26$  GeV

$$\mathcal{P}_{bkg} = f_\delta^{bkg} \delta(\Delta t - \mu_\delta^{bkg}) + (1 - f_\delta^{bkg}) \frac{1}{2\tau_{bkg}} e^{-|\Delta t - \mu_\delta^{bkg}|/\tau_{bkg}} \quad (14)$$

$$R_{bkg} = (1 - f_{tail}^{bkg})G(\Delta t_i, \sigma_{main}^{bkg}\sqrt{\delta_{cp}^2 + \delta_{tag}^2}) + f_{tail}^{bkg}G(\Delta t_i, \sigma_{tail}^{bkg}\sqrt{\delta_{cp}^2 + \delta_{tag}^2}) \quad (15)$$

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

TABLE X: Parameters in Background  $\Delta t$  distribution.

289 **5. BLIND FIT**

290 **5.1.  $CP$  fit on MC samples**

291 Based on the model in Eq(6), we perform the  $CP$  fit on events in signal MC and generic  
 292 MC. The signal and generic MC are generated with phase-space model which contains zero  
 293  $CP$  violation ( $\mathcal{S} = \mathcal{A} = 0$ ) in  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  channel. The event with following cuts are  
 294 selected for  $CP$  fit.

- 295 •  $-70 < \Delta t < 70$  ps
- 296 •  $0 < (\chi_2/N)_{cp} < 8$
- 297 •  $0 < (\chi_2/N)_{tag} < 20$
- 298 • error on tag-side vertex  $< 0.1$  cm
- 299 •  $5.27 < M_{bc} < 5.29$  GeV and  $|\Delta E| < 0.1$  GeV

300 We have 10000 (8873 passing selection) events from signal sample and 415 (373 passing  
 301 selection) events from  $1\text{ab}^{-1}$  generic sample to fit  $CP$  parameters. To mimic the events  
 302 number expected in data sample, we randomly take 30 events from generic sample and  
 303 perform the fit as well. The fit results are shown in FIG.27 to FIG.29 and Eq(16) to Eq(18).

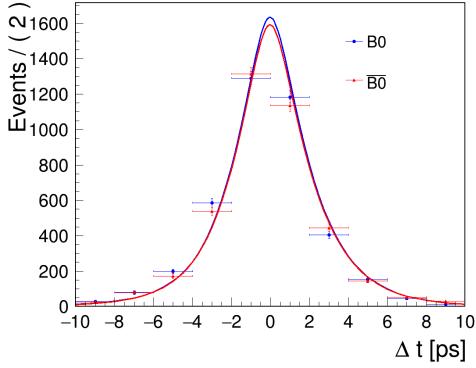


FIG. 27:  $CP$  fit on 8873 signal MC.

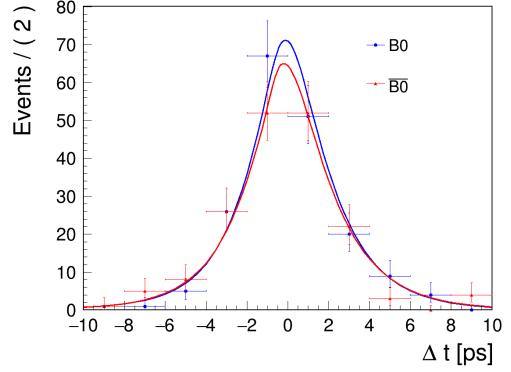


FIG. 28:  $CP$  fit on 373 generic MC.

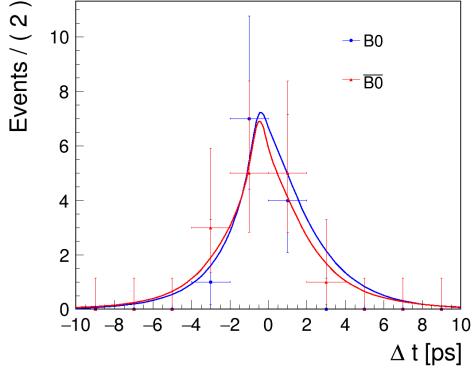


FIG. 29:  $CP$  fit on 30 generic MC.

The fit result for 8873 signal MC is:

$$\begin{aligned} \sin(2\phi_1) &= 0.00 \pm 0.04 \\ \mathcal{A} &= -0.01 \pm 0.02 \end{aligned} \quad (16)$$

the fit result for 373 generic MC is:

$$\begin{aligned} \sin(2\phi_1) &= 0.00 \pm 0.21 \\ \mathcal{A} &= -0.05 \pm 0.07 \end{aligned} \quad (17)$$

the fit result for 30 generic MC is:

$$\begin{aligned} \sin(2\phi_1) &= 0.20 \pm 0.85 \\ \mathcal{A} &= -0.06 \pm 0.30 \end{aligned} \quad (18)$$

304 The fit results are consistent with expectation in non- $CP$  violation from MC input, and  
 305 the statistical uncertainties has the tendency  $\delta \propto 1/\sqrt{N}$  as poission distribution. To test fit  
 306 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  
 307 events selection as well as fit model determination can be found[11]. The fit result over 10000  
 308 events is shown in FIG.30, which results in  $\sin(2\phi_1) = 0.70 \pm 0.05$  and  $\mathcal{A} = -0.01 \pm 0.02$ .  
 309 The results agree with the input.

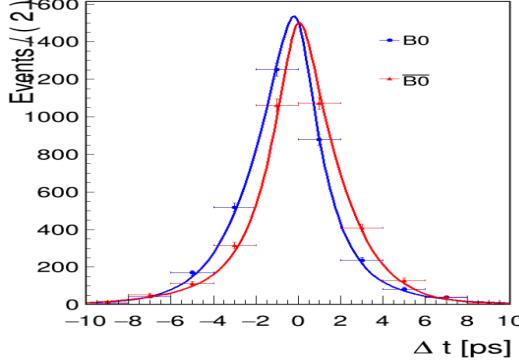


FIG. 30:  $CP$  fit over 10000  $B^0 \rightarrow J/\psi K_S^0$  signal MC.

310      **5.2. Linearity Test**

311      To validate the  $CP$  fit linearity, we generate a series of toy MC sample, which the  $\chi_2$ ,  
 312       $N$  and vertex errors on  $CP$  and tag-side are sampled from distribution of signal MC. The  
 313      resolution functions parameters are kept same as MC fit. The input  $\mathcal{A}$  is set to zero while  
 314      the input value of  $\sin(2\phi_1)$  is running from 0.1 to 0.9. Each dataset contains 10000 events.  
 315      The dependence between input and output are shown in FIG.31. The linearity fit shows a  
 316      good agreement.

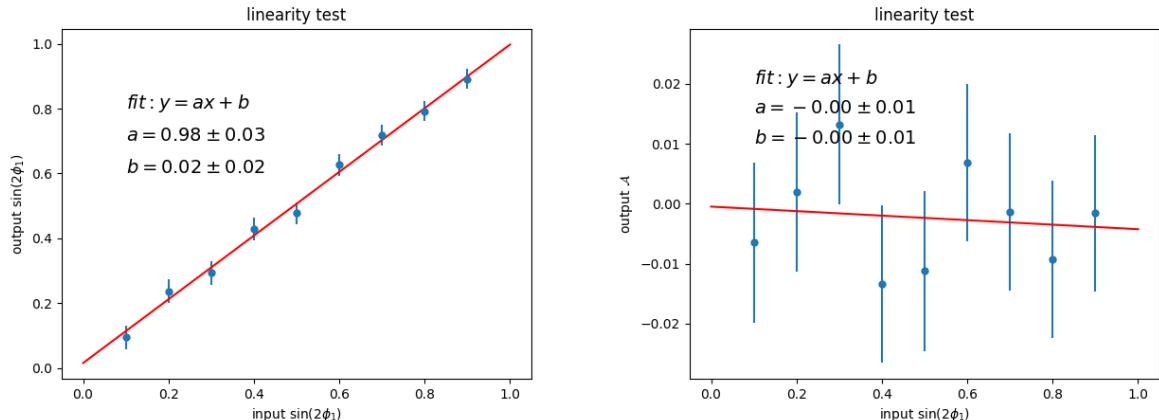


FIG. 31: Linearity test of  $CP$  fit.

317      Also, we fix  $\sin(2\phi_1)$  at zero while floating  $\mathcal{A}$  from 0.1 to 0.9, the dependence between  
 318      input and output are as FIG.32 shows. The linearity fit shows a good agreement.

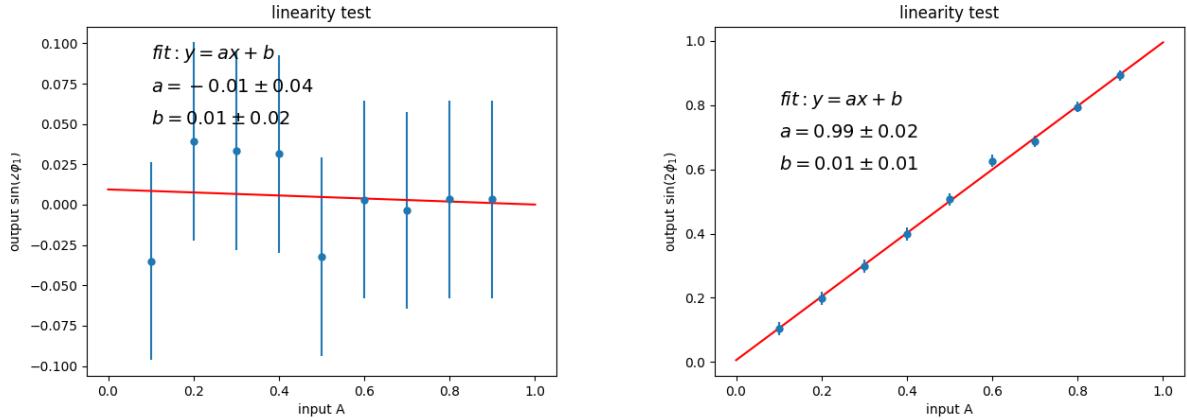


FIG. 32: Linearity test of  $CP$  fit.

### 319 5.3. Toy MC Fit Pull

320 In order to check the fit bias with input-output method, a series of 1000 dataset of 321 toy  
 322 MC has been created containing about 26 events in each. The event number is set based  
 323 on the expected number from signal region in data after the selection.  $\chi_2$ ,  $N$  and vertex  
 324 errors on  $CP$  and tag-side are sampled from distribution of data. The fit to dataset has only  
 325  $\sin(2\phi_1)$  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  $\mathcal{A}$ .

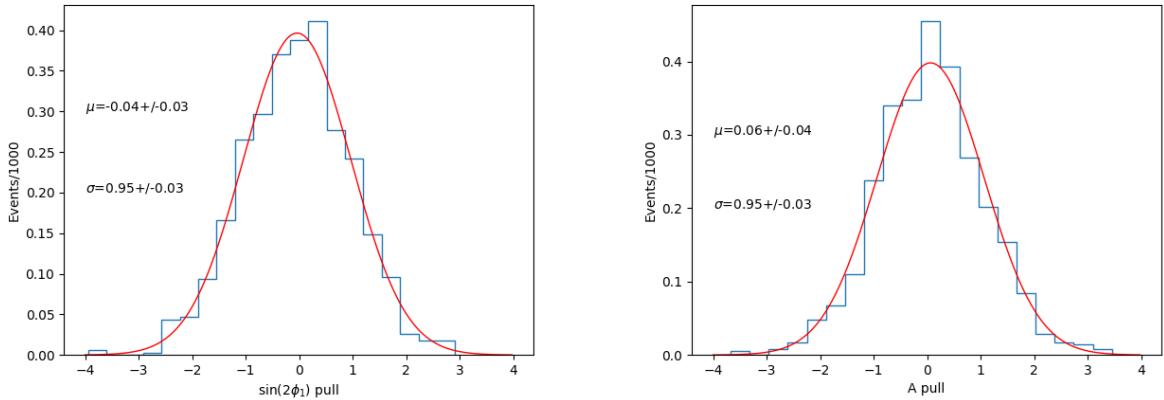


FIG. 33: Pull of  $\sin(2\phi_1)$  and  $\mathcal{A}$  fitted.

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

### 327 5.4. Lifetime and $\Delta m_d$ Fit

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:

$$\mathcal{P}(\Delta t, \tau_{B^0}) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \quad (19)$$

The fit result is  $1.537 \pm 0.024$  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  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  is  $1.431 \pm 0.382$  ps. The result is consistent with PDG value. The distribution of  $\Delta t$  in lifetime fit is shown as FIG.34. The  $B^0$  and  $B^+$  lifetime fit using control sample is also performed and summarized in here[11]. The results are consistent with PDG values.

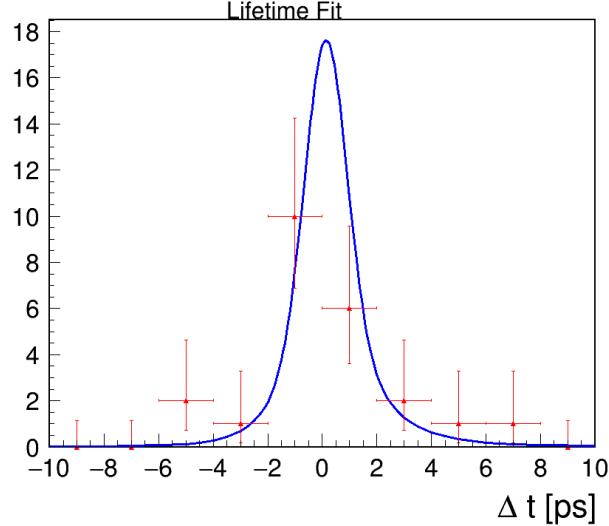


FIG. 34: Lifetime fit on data

To test the fit on physics parameter  $\Delta m_d$ , we generate 200 toy MC sets of  $B^0 \rightarrow K_S^0 K_S^0 K_S^0$  with input  $\Delta m_d = 0.507$  GeV/ $c^2$  where each set contains 26 events as same as data. The fit result is close to normal distribution and the pull of  $\Delta m_d$  is shown in FIG.35.

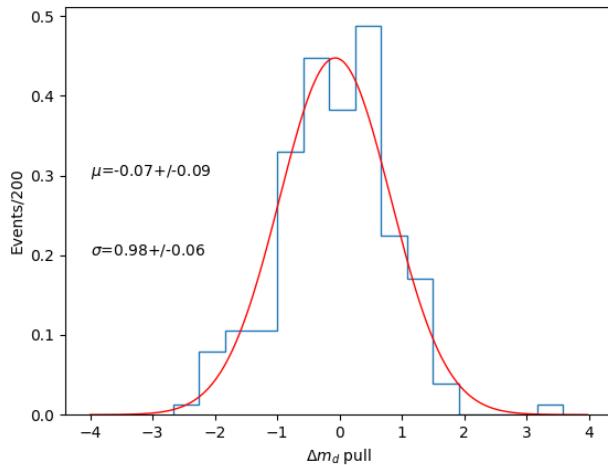


FIG. 35: Pull of  $\Delta m_d$

337        **5.5. Systematics**

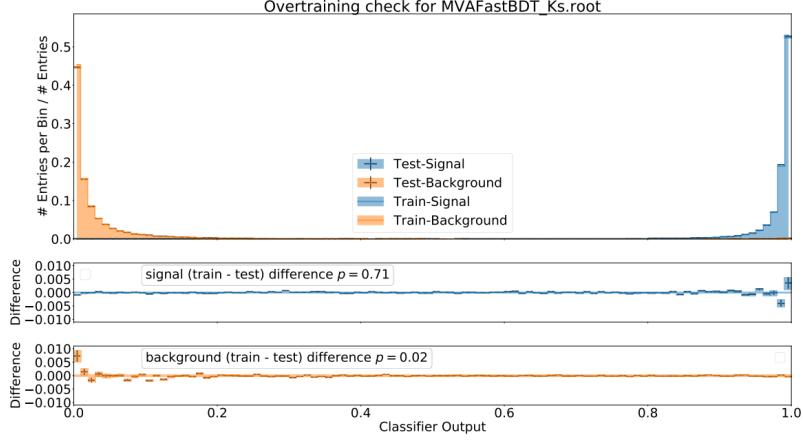
- 338        • resolution functions parameters  
339        • signal fraction  
340        • flavor tagging  
341        • background  $\Delta t$   
342        • fit bias  
343        • physics parameters

344        The program for quickly reproduce the results and systematics errors are prepared. Most  
345        of the items are filled after the data signal box-open is permitted and  $CP$  fitting is fully  
346        validated.

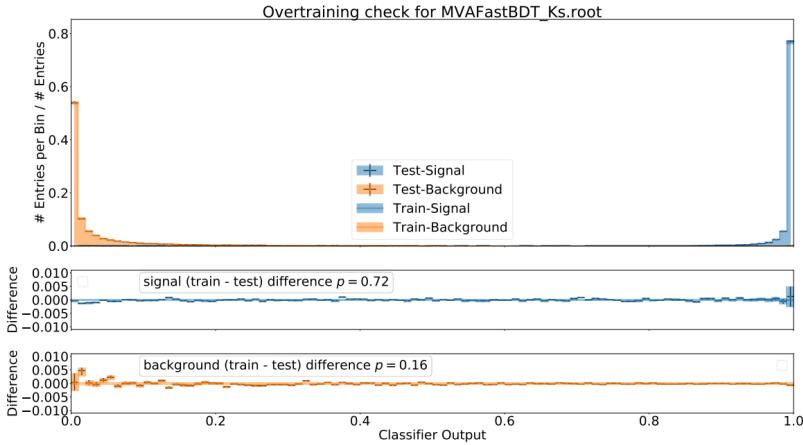
---

- 347 [1] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak*  
348        *interaction*, Progress of theoretical physics **49** (1973) no. 2, 652–657.  
349 [2] K. Kang, H. Park, T. Higuchi, K. Miyabayashi, K. Sumisawa, I. Adachi, J. Ahn, H. Aihara,  
350        S. A. Said, D. Asner, et al., *Measurement of time-dependent CP violation parameters in*  
351        *B0- $\bar{d}$ KsKs decays at Belle*, arXiv preprint arXiv:2011.00793 (2020) .  
352 [3] J. Lees, V. Poireau, V. Tisserand, J. G. Tico, E. Grauges, M. Martinelli, D. Milanes,  
353        A. Palano, M. Pappagallo, G. Eigen, et al., *Amplitude analysis and measurement of the*  
354        *time-dependent C P asymmetry of  $B \rightarrow K S 0 K S 0 K S 0$  decays*, Physical Review D **85**  
355        (2012) no. 5, 054023.  
356 [4] https://stash.desy.de/projects/B2/repos/software/browse/skim/scripts/skim/tcpv.py.  
357 [5] https://software.belle2.org/sphinx/release-04-01-02/\_modules/stdV0s.html#  
358        stdKshorts.  
359 [6] https://confluence.desy.de/display/BI/Software+and+Performance+Workshop+-+V0+  
360        vertex+fitting+working+session.  
361 [7] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic*  
362        *gradient-boosted decision trees for multivariate classification*, CoRR **abs/1609.06119** (2016)  
363        , arXiv:1609.06119. http://arxiv.org/abs/1609.06119.  
364 [8] https://stash.desy.de/projects/B2/repos/software/browse/analysis/examples/  
365        tutorials/B2A702-ContinuumSuppression\_MVATrain.py.  
366 [9] https://docs.belle2.org/record/1905/files/BELLE2-NOTE-PH-2020-013\_v3.2.pdf?  
367        version=1.  
368 [10] F. Abudinén, I. Adachi, R. Adak, K. Adamczyk, P. Ahlburg, J. Ahn, H. Aihara, N. Akopov,  
369        A. Aloisio, F. Ameli, et al., *Measurement of the  $B^0$  lifetime using fully reconstructed*  
370        *hadronic decays in the 2019 Belle II dataset*, arXiv preprint arXiv:2005.07507 (2020) .  
371 [11] https://docs.belle2.org/record/1941/files/BELLE2-NOTE-PH-2020-027.pdf.  
372 [12] https://docs.belle2.org/record/1384/files/belle2-note-resol\_belle.pdf?version=3.

375 A. OVER-FITTING CHECK FOR KSFINDER



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



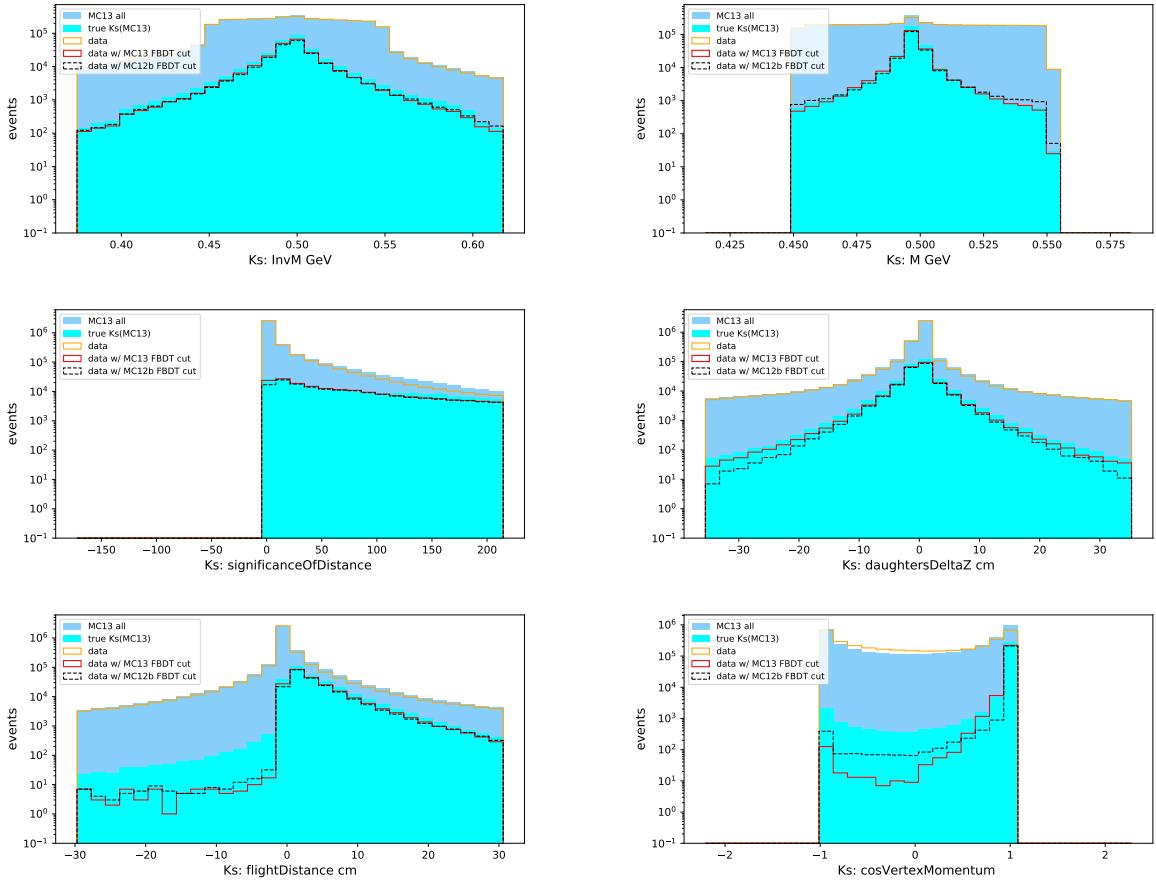
b) Over-fitting check for  $B^0$  generic decay.

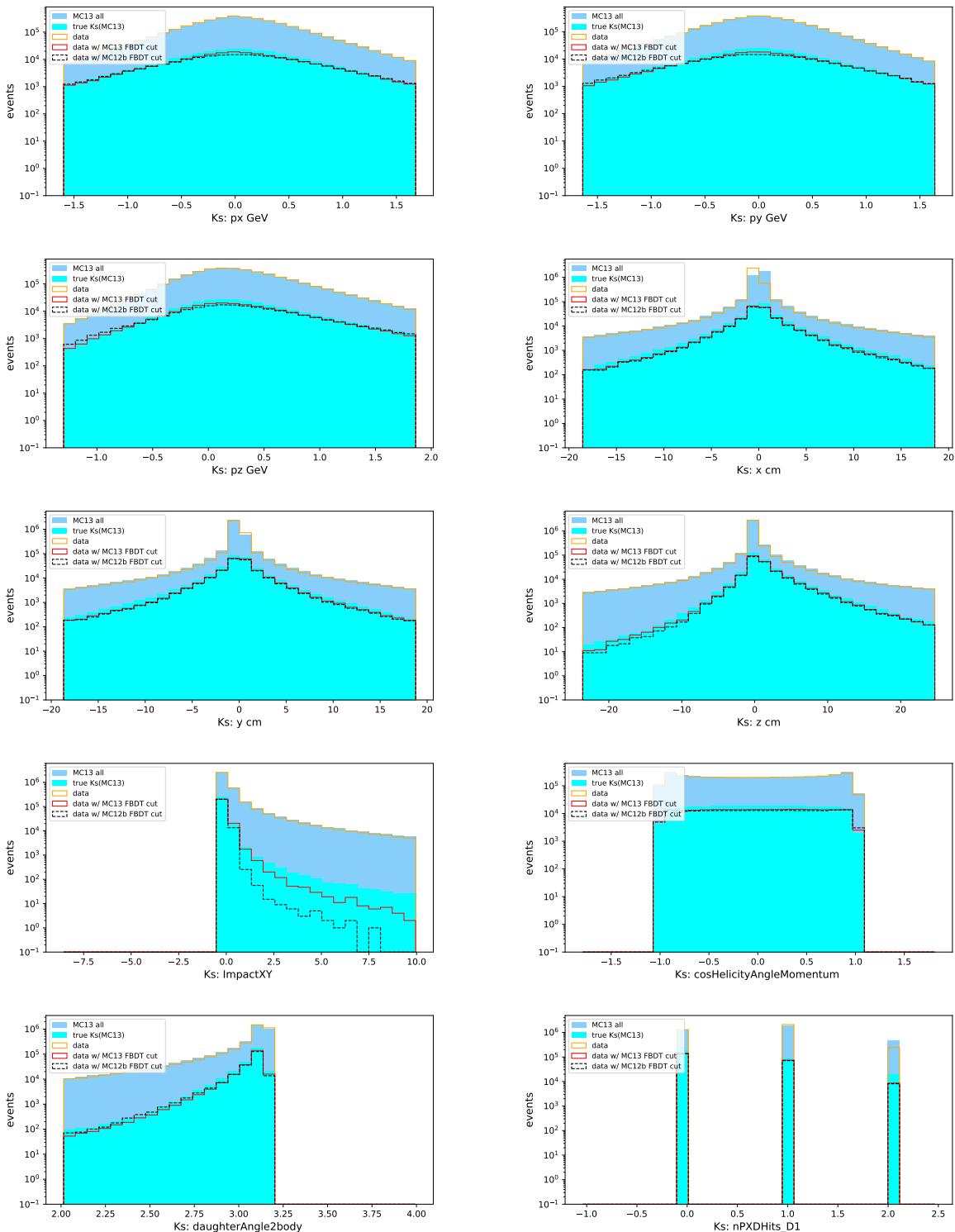
FIG. 36: Over-fitting check for classifiers.

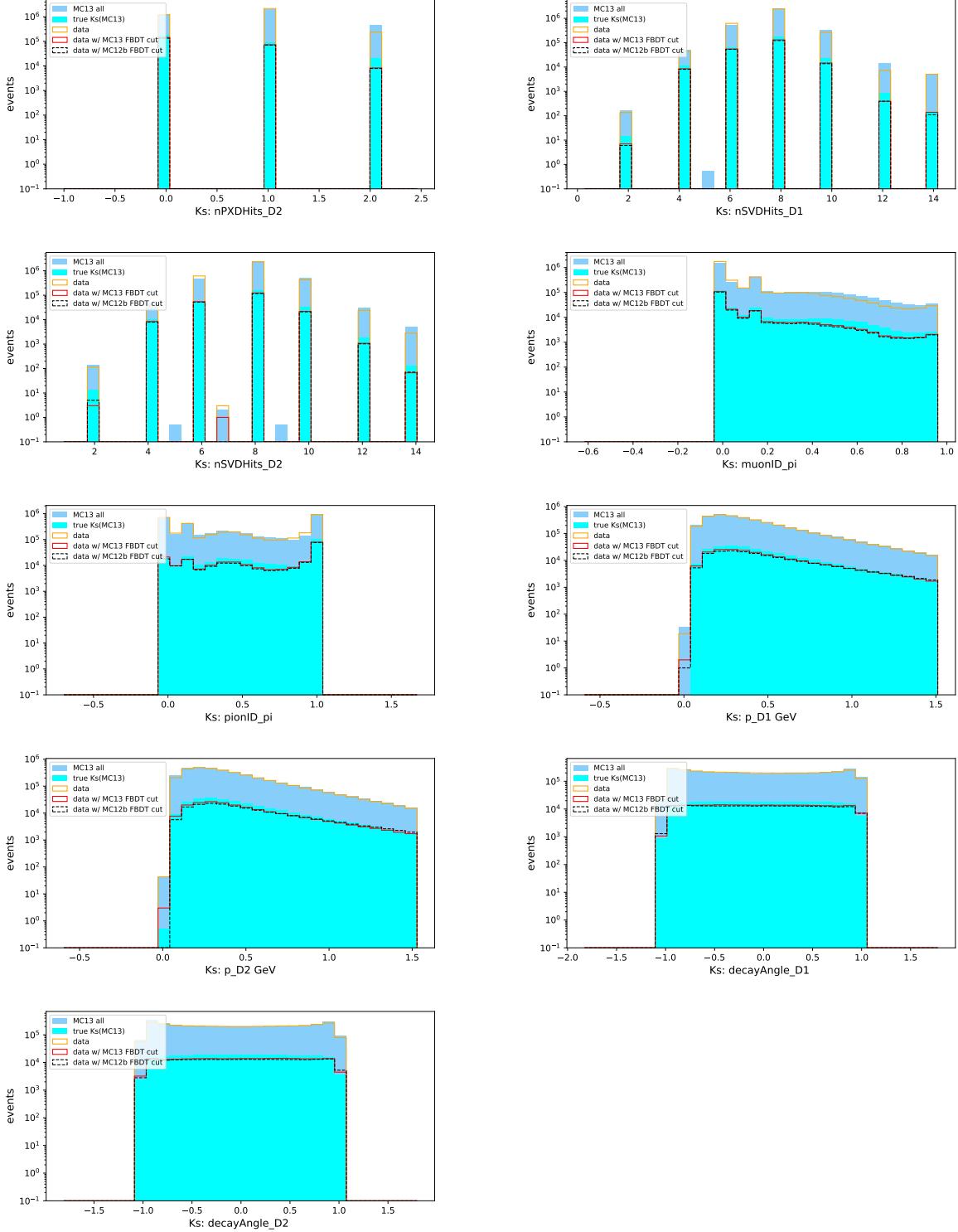
376 **B. KSFINDER AND CS RELATED INFORMATION.**

377 **B.1. KsFinder Variables comparison between data and MC**

FIG. 37: 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.







## B.2. Continuum suppression over-training check

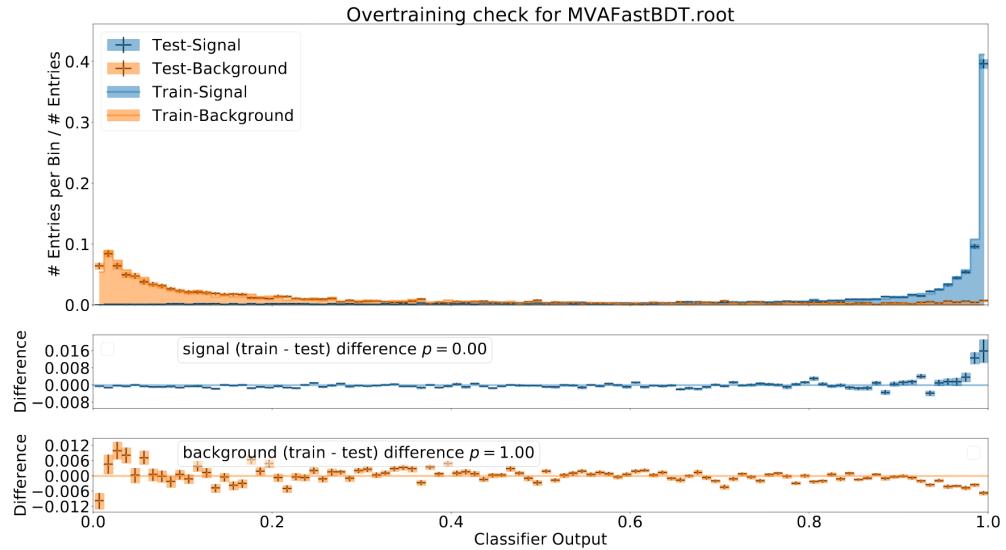


FIG. 39: Over-fitting check of continuum classifier, where no overtrain is found.

## 379 C. RESONANT BACKGROUNDS IN MC

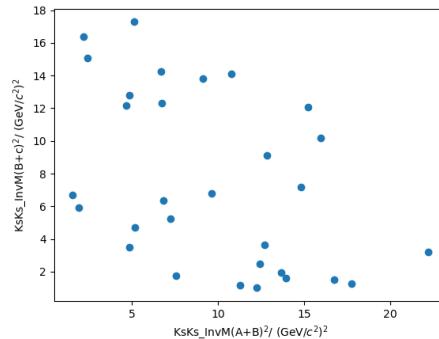
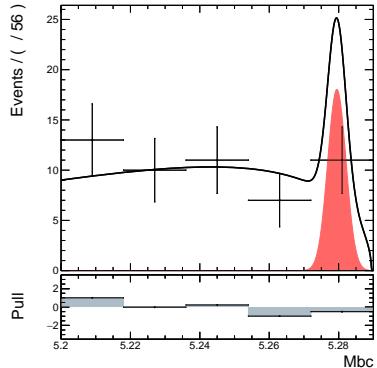


FIG. 40: The Dalitz plot in data signal region, A, B, and C is ordered according the increasing order of the momentum

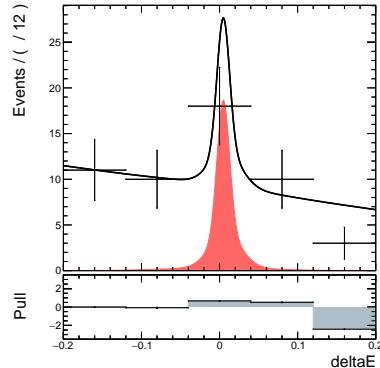
TABLE XI: Expected yield for signal and background resonances  $2.14 \times 10^8 B\bar{B}$  in generic MC. The branching fraction of  $B \rightarrow XK_S$  and  $X \rightarrow 2K_S$  are listed for both PDG value and value in Belle II DEC. file. The number expected from  $CP$ -odd contamination is very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

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

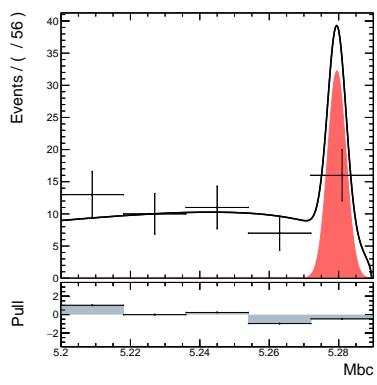
380 D. SIGNAL INJECTION TEST



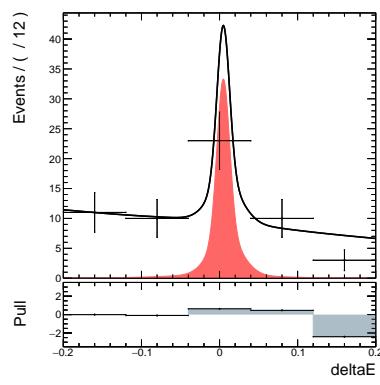
a) signal injected: 5



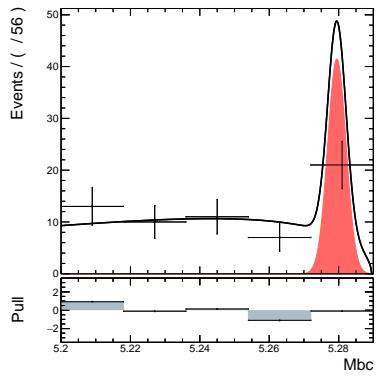
b) signal injected: 5



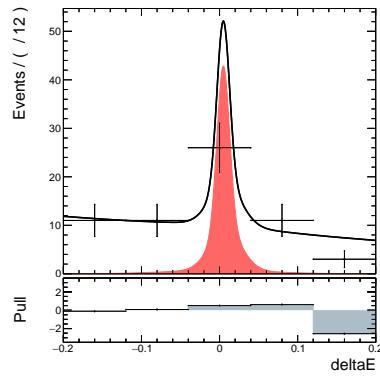
c) signal injected: 10



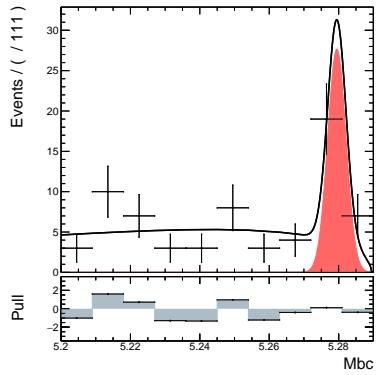
d) signal injected: 10



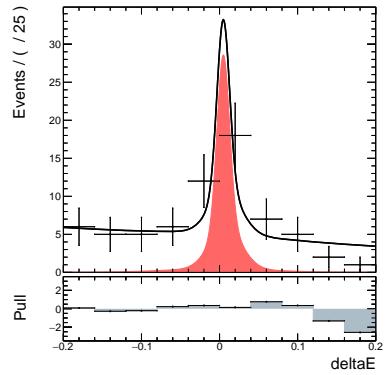
e) signal injected: 15



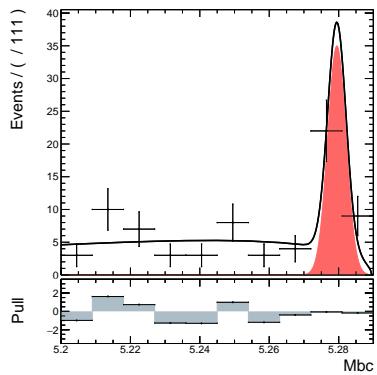
f) signal injected: 15



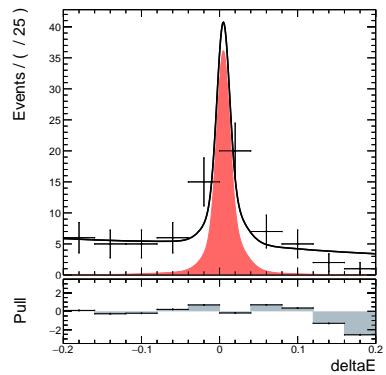
g) signal injected: 20



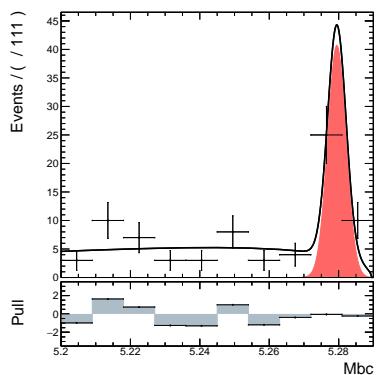
h) signal injected: 20



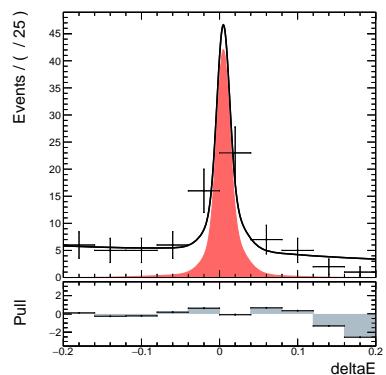
i) signal injected: 25



j) signal injected: 25



k) signal injected: 30



l) signal injected: 30