

---

# Search for the $B^+ \rightarrow K^+ \nu \bar{\nu}$ Decay at Belle II Using a Tagged Analysis Approach

Jiyong Jeong

---



München 2022



---

**Suche nach dem  $B^+ \rightarrow K^+ \nu \bar{\nu}$  Zerfall bei  
Belle II mit einem getaggten  
Analyseansatz**

**Jiyong Jeong**

---

Masterarbeit  
an der Fakultät für Physik  
der Ludwig–Maximilians–Universität  
München

vorgelegt von  
Jiyong Jeong  
aus Korea

München, den 15.03.2022

Supervisor: Prof. Dr. Thomas Kuhr

Proofreader: Dr. Thomas Lück

# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
<b>2</b>	<b>Theory of Flavour Physics</b>	<b>13</b>
2.1	Flavour in the Standard model . . . . .	13
2.2	Cabibbo-Kobayashi-Maskawa(CKM) matrix . . . . .	15
<b>3</b>	<b>Belle II detector</b>	<b>19</b>
3.1	SuperKEKB . . . . .	20
3.2	Vertex detector . . . . .	22
3.3	Central Drift Chamber (CDC) . . . . .	23
3.4	Particle Identification . . . . .	23
3.4.1	Time-Of-Propagation(TOP) - Barrel . . . . .	23
3.4.2	Aerogel Ring-Imaging Cherenkov detector(ARICH) - End-cap . . . . .	24
3.5	Electromagnetic calorimeter (ECL) . . . . .	25
3.6	$K_L^0$ and $\mu$ detection (KLM) . . . . .	26
<b>4</b>	<b>Belle II Analysis Software Framework</b>	<b>27</b>
4.1	Path and Module . . . . .	27
4.2	Reconstruction . . . . .	27
4.3	Full Event Interpretation . . . . .	28
<b>5</b>	<b>Monte Carlo Simulation</b>	<b>31</b>
<b>6</b>	<b>Analysis of <math>B^+ \rightarrow K^+ \nu \bar{\nu}</math> decay</b>	<b>33</b>
6.1	Input variables . . . . .	33
6.1.1	Variables for pre-cut . . . . .	33
6.1.2	Continuum Suppression after pre-cuts . . . . .	38
6.2	Rest of Event . . . . .	45
6.3	Selection cut . . . . .	48
6.3.1	Pre-cut . . . . .	49
6.3.2	Best candidate selection . . . . .	52
6.3.3	fBDT cut . . . . .	53
6.4	Fitting . . . . .	55

<b>7 Machine Learning</b>	<b>60</b>
7.1 fast Boosted Decision Tree(fBDT) . . . . .	60
7.2 Support Vector Machine(SVM) . . . . .	63
7.3 Neural Network . . . . .	69
<b>Summary</b>	<b>73</b>
<b>References</b>	<b>75</b>
<b>A Topology Analysis</b>	<b>78</b>
A.1 Signal( $B^\pm \rightarrow K^\pm \nu \bar{\nu}$ ) . . . . .	79
A.2 Charged B pair . . . . .	81
A.3 ccbar . . . . .	83
<b>Acknowledgement</b>	<b>85</b>

# List of Figures

1.1	Cross sections of upsilon resonances . . . . .	11
1.2	Workflow of the analysis . . . . .	12
2.1	Elementary particles in the Standard Model . . . . .	13
2.2	Meson octet . . . . .	14
2.3	Strength of quark mixing . . . . .	15
2.4	The CKM unitarity triangle . . . . .	18
3.1	Structure of the Belle II detector . . . . .	19
3.2	Structure of SuperKEKB . . . . .	20
3.3	Schematic view of beam collision in the Nano-Beam scheme . . . . .	21
3.4	Structure of the vertex detector . . . . .	22
3.5	Layout of SVD with four layers . . . . .	23
3.6	Structure of quartz bar . . . . .	24
3.7	Detection of Cherenkov photon . . . . .	25
4.1	Structure of path and modules . . . . .	28
4.2	Schematic description of the FEI algorithm . . . . .	29
4.3	The older version FR is compared with FEI . . . . .	30
6.1	Signal and tag side variables for the selection cuts. . . . .	36
6.2	Upsilon variables for the selection cuts . . . . .	37
6.3	Continuum suppression variables of signal and tag B mesons . . . . .	39
6.4	KSFW variables of signal B meson as input variables of fBDT(section 7.1)	41
6.5	KSFW variables of signal B meson as input variables of fBDT(section 7.1)	42
6.6	KSFW variables of signal B meson as input variables of fBDT(section 7.1)	43
6.7	KSFW variables of signal B meson as input variables of fBDT(section 7.1)	44
6.8	The number of ROE particles . . . . .	46
6.9	The number of ECL clusters in the ROE . . . . .	46
6.10	Plot of roeP . . . . .	47
6.11	Plot of roeEextra . . . . .	47
6.12	The number of candidates per event and rank of signal probability in tag side	52
6.14	Plots of fitting for signal and background separately . . . . .	55
6.15	Compsite fit plot . . . . .	57

---

6.16	Residual and pull distribution from composite fitting result . . . . .	58
7.1	Visualization of decision tree . . . . .	61
7.2	ROC curve and score of fBDT . . . . .	62
7.3	The overtraining plot of fBDT . . . . .	62
7.4	Hyperplane of SVM . . . . .	63
7.5	Kernel trick of SVM . . . . .	64
7.6	Correlation matrix of input variables in SVM . . . . .	65
7.7	Overview of hyperparameters and scores in SVM training . . . . .	66
7.8	Permutation feature importances of SVM . . . . .	67
7.9	Schematic diagram of Multi-layer Perceptron (MLP) . . . . .	69
7.10	Permutation feature importances of MLP . . . . .	72

# List of Tables

3.1	Fundamental parameters of SuperKEKB and KEKB . . . . .	21
5.1	Result of MCmatching variables of signal MC data . . . . .	31
6.1	Cross section of physics process of collision at $\sqrt{s} = 10.58$ GeV . . . . .	34
6.2	Test of selection cuts . . . . .	50
6.3	The parameters of the signal fit. . . . .	56
6.4	The parameters of the background fit. . . . .	56
6.5	The parameters of composite fit. . . . .	57
7.1	The hyperparameters of fBDT training . . . . .	61
7.2	The hyperparameters and condition of MLP training . . . . .	71



# Chapter 1

## Introduction

The rare decay  $B^+ \rightarrow K^+\nu\bar{\nu}$  is a flavour-changing neutral current process, which is suppressed in the standard model and thus sensitive to new physics contributions. Monte-Carlo simulated data at Belle II is applied to optimize the separation between signal and background and to estimate the sensitivity of the search. We can search for this decay in upsilon resonance as shown in figure 1.1.

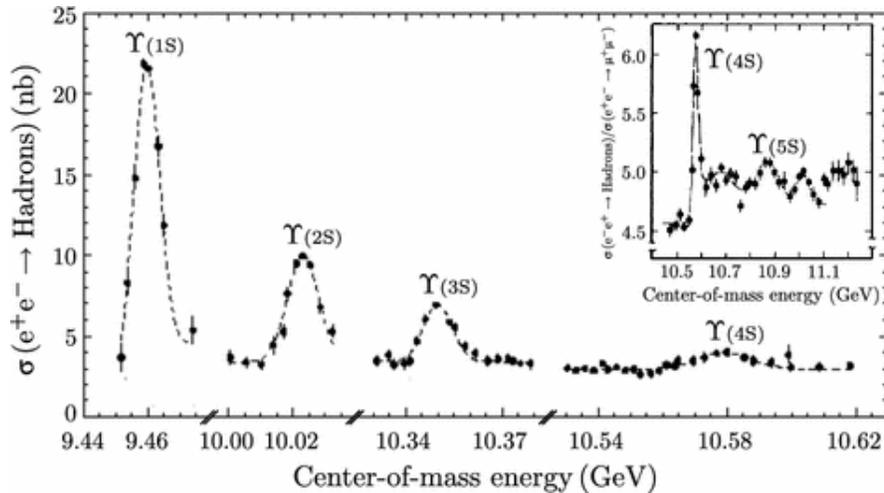


Figure 1.1: Cross sections of upsilon resonances [1]. The center-of-mass energy at upsilon peak is visible.

From the  $\Upsilon(4S)$ , the two B mesons are produced as a pair and one of the B meson can decay to  $K^+\nu\bar{\nu}$ . This process is possible by flavour change of the quark  $b \rightarrow s$ . However, this decay is suppressed in SM prediction. The upper limit of the branching fraction of the decay is found in [3] as

$$Br(B^+ \rightarrow K^+\nu\bar{\nu}) < 1.6 \times 10^{-5} \quad (1.1)$$

As searching for a decay with extremely low branching fraction is strenuous work, Monte Carlo data is produced with higher branching fraction of  $B^+ \rightarrow K^+ \nu \bar{\nu}$  decay. My workflow of the analysis to search for the decay is shown in 1.2.

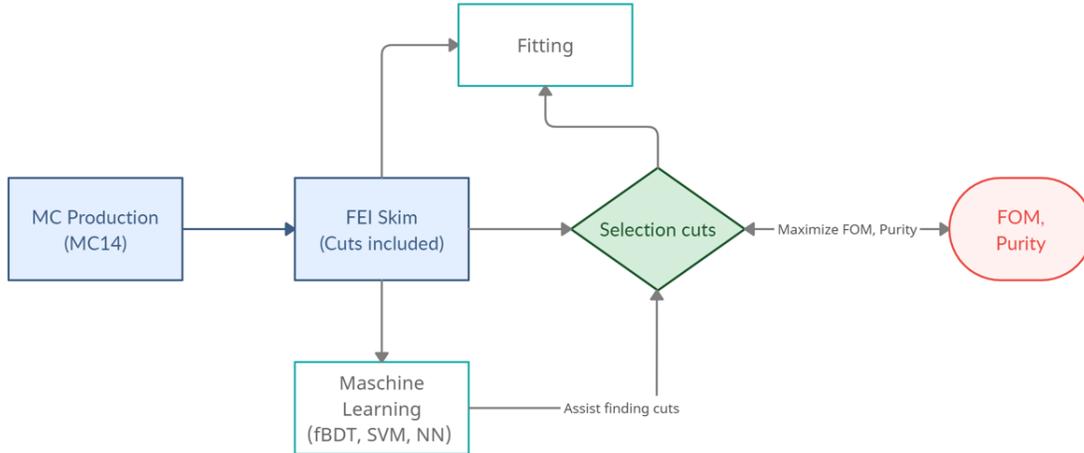


Figure 1.2: Workflow of the analysis

First of all, Monte Carlo datas are produced regularly at Belle II for both signal  $B^+ \rightarrow K^+ \nu \bar{\nu}$  and background from  $\Upsilon(4S)$ . In this work, I do tagged analysis using FEI skimmed data in which the FEI is already applied with selection cut. The FEI(Full Event Interpretation) is the method of Belle II software to reconstruct tag B meson. By the FEI, the information of the tag B meson can be analyzed to search for signal. Then I search for the selection cuts to diminish the amount of background and ideally remove almost all of background at the end. The pre-cut, which is applied in early phase of analysis to reduce background, is searched with help of machine learning(SVM, NN) as shown in section 7.2 and 7.3. The result of the feature importance helps to reveal which variables have good separation between signal and background to make pre-cut between them. The distribution of the variables are displayed in section 6.1. The pre-cut and its test is done in section 6.3.1. Then I do best candidate selection in section 6.3.2 to reduce the number of candidates for signal. After that, I run the fBDT(section 7.1) and find cut on fBDT(section 6.3.3) to reduce remaining background. The fBDT(fast Boosted Decision Tree) is explained in section 7.1. All selection cuts are determined to increase purity and maximize figure of merit(FOM). Finally, the fitting is implemented to measure signal fraction from the data and branching fraction of real data is compared in section 6.4. In section 2, the theoretical knowledge of flavour physics is explained. In section 3, the structure of the Belle II detector is explained with illustration of hardware. In section 4, the Belle II software that I use for analysis is shortly introduced. In the appendix, the part of the topology analysis result is added to show how the exact decays of signal and background appear. In appendix, some parts of TopoAna result is shown, in which the exact decay trees of signal and part of background are displayed.

# Chapter 2

## Theory of Flavour Physics

### 2.1 Flavour in the Standard model

The history of flavour physics started from the discovery of quarks and leptons in the Standard Model, which elucidated mysterious substructure of particles and atoms. So far, the quarks constitute the part of the elementary particles of the Standard Model, which cannot be composed of underlying particles. The quarks, along with leptons, are characterized by the flavour, with which we have six types of quarks in the Standard Model.

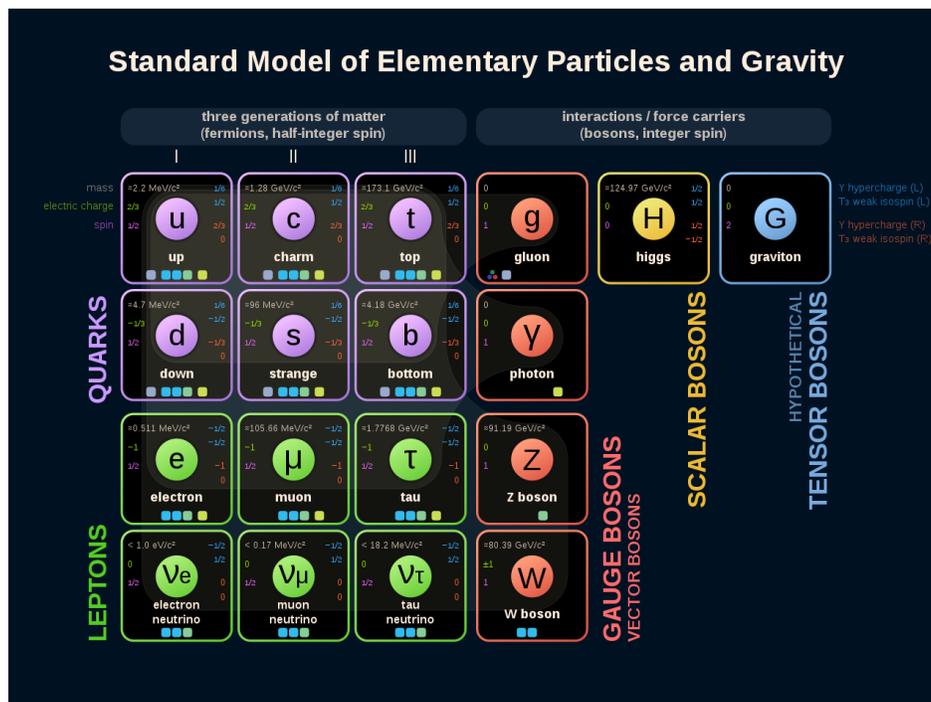


Figure 2.1: Elementary particles in the Standard Model [34]. There are six flavours of quark. The quarks have different physical quantities and quantum number.

As shown in figure 2.1, each quark has identifying quantum numbers, which are isospin  $I_3$  and hypercharge  $Y$ . It is also visible that isospin and hypercharge are identical within up-type quarks, and within down-type quarks likewise. Mass is another distinguishing physical quantity in the determination of the generation. The hypercharge is built by other quantum numbers, which are baryon number  $B$ , lepton number  $L$ , Strangeness  $S$ , Charm  $C$ , Bottomness  $B'$ , Topness  $T$ .

$$Y = B + S + C + B' + T \quad (2.1)$$

These quantum numbers have to satisfy the *Gell-Mann-Nishijima formula* with electric charge  $Q$ .

$$Q = I_3 + Y \quad (2.2)$$

Each quark has its antiparticle, that are called as antiquark and denoted with bar. The antiquarks have almost identical properties, but its electrical charge has opposite sign, while the value of the charge is same. The leptons also have antiparticles, for instance, antineutrino are antiparticle of neutrino. The antiquarks play important role as elementary particle to constitute mesons. For instance, the mesons include B meson, kaon and pion, which participate in rare decay. The mesons are not elementary particle themselves, but composed of quark and antiquark pair. In order to sort the mesons in a organized way, *Eightfold way* is applied and it can make Meson octet as in figure 2.2.

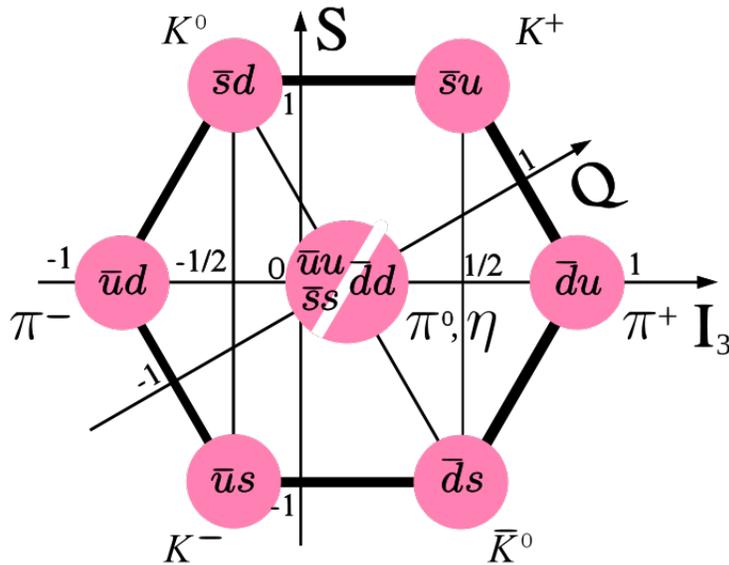


Figure 2.2: A possible example of Meson octet [35].

Such idea was initiated from isospin, in which proton and neutron are regarded as the same particle with different state, because they have almost the same mass and differ in

charge. In meson octet, the isospin, strangeness and charge determine the position of kaons and pions.

In the Standard Model, the Lagrangian is theorized in the way that it is invariant under transformation of gauge symmetry. The symmetry group of the Standard model is  $SU(3) \times SU(2) \times U(1)$ . The Lagrangian of the Standard Model is largely divided in three terms as follows.

$$\mathcal{L}_{SM} = \mathcal{L}_{Kin} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yuk} \quad (2.3)$$

Kinetic term describes the gauge bosons like gluons, W and B bosons. The Higgs term can give the mass to gauge bosons by the Higgs mechanism. Spontaneous symmetry breaking makes gauge group as  $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$ . After the spontaneous symmetry breaking,  $W^\pm, Z^0$  and photon appear with mass in the Standard Model. In order to find the description of the flavour, the Yukawa term of the Lagrangian has to be investigated, from which CKM matrix can be derived.

## 2.2 Cabibbo-Kobayashi-Maskawa(CKM) matrix

The CKM matrix describes the likelihoods of flavour change between the quarks, which is the so-called quark mixing. Following figure 2.3 shows how easily the quark flavour can turn into other flavour.

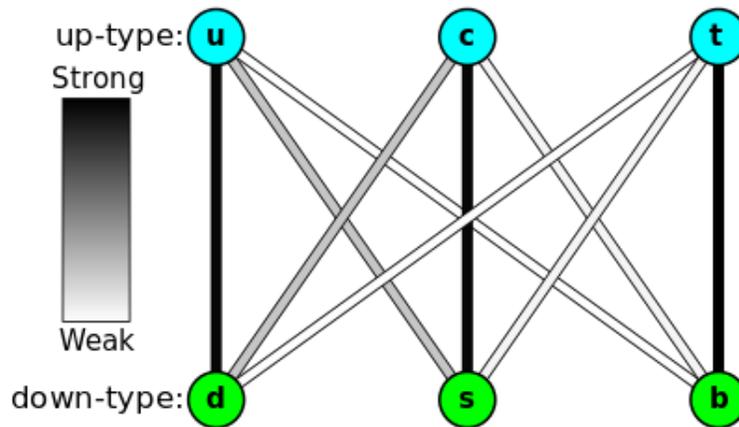


Figure 2.3: Strength of quark mixing(Flavour change). [36]

However, the representation of CKM matrix is required to obtain exact quantitative value of this flavour change. By the fact that weak interaction violates the conservation of quantum number and change the quark flavour, while strong interaction via gluons conserves the quantum number, the Lagrangian of weak interaction has to be applied to

derive CKM matrix. Yukawa term of the Lagrangian [22] is

$$\mathcal{L}_{yuk} = -Y_{ij}^d \bar{Q}^i H d_R^j - Y_{ij}^u \bar{Q}^i \tilde{H} u_R^j + h.c. \quad (2.4)$$

The term  $Q^i$  is doublet pairs of quarks with three generations as follows.

$$Q^i = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \quad (2.5)$$

The right handed quarks have three components in the matrix as  $u_R^i = (u_R, c_R, t_R)$ .

After spontaneous symmetry breaking, the Higgs field produces vacuum expectation value  $v$  and quark mass terms look in matrix form as

$$\mathcal{L}_{yuk} = -\frac{v}{\sqrt{2}} [\bar{d}_L Y_d d_R + \bar{u}_L Y_u u_R] + h.c. \quad (2.6)$$

In this representation, the mass term is described in matrix. Therefore, the mass matrix has to be diagonalized to obtain readable mass values so that the masses are eigenvalues of the mass matrix in this case. For the diagonalization to be realized, we presume that there are diagonal matrices  $M_d, M_u$ . And the unitary matrices  $U_d, U_u$  are applied.  $Y_d Y_d^\dagger, Y_u Y_u^\dagger$  are hermitian and can be diagonalized, while  $Y$  is not hermitian alone. So  $Y Y^\dagger$  can be diagonalized.

$$Y_d Y_d^\dagger = U_d M_d^2 U_d^\dagger, \quad Y_u Y_u^\dagger = U_u M_u^2 U_u^\dagger \quad (2.7)$$

$Y$  alone can have different unitary matrices  $K_d, K_u$  as it doesn't break the diagonalization of (2.7).

$$Y_d = U_d M_d K_d^\dagger, \quad Y_u = U_u M_u K_u^\dagger \quad (2.8)$$

After this consideration, the Lagrangian would look like,

$$\mathcal{L}_{yuk} = -\frac{v}{\sqrt{2}} [\bar{d}_L U_d M_d K_d^\dagger d_R + \bar{u}_L U_u M_u K_u^\dagger u_R] + h.c. \quad (2.9)$$

The unnecessary matrices  $U$  and  $K$  can be cancelled by quark rotation,  $d_R \rightarrow K_d d_R, u_R \rightarrow K_u u_R, u_L \rightarrow U_u u_L, d_L \rightarrow U_d d_L$ . This changes basis of the representation. Now the Lagrangian is in mass basis.

$$\mathcal{L}_{mass} = -m_j^d \bar{d}_L^j d_R^j - m_j^u \bar{u}_L^j u_R^j + h.c. \quad (2.10)$$

The kinetic terms are also affected by this change of basis, since the kinetic terms have the quarks as well.

The kinetic terms are in flavour basis,

$$\begin{aligned} \mathcal{L}_{flavour} \supset (\bar{u}_L \quad \bar{d}_L)^i \left[ i\cancel{\partial} + \gamma_\mu \begin{pmatrix} \frac{g'}{6} B_\mu + \frac{g}{2} W_\mu^3 & \frac{g}{\sqrt{2}} W_\mu^+ \\ \frac{g}{\sqrt{2}} W_\mu^- & \frac{g'}{6} B_\mu - \frac{g}{2} W_\mu^3 \end{pmatrix} \right] \begin{pmatrix} u_L \\ d_L \end{pmatrix}^i \\ + \bar{u}_R^i \left( i\cancel{\partial} + g' \frac{2}{3} \cancel{B} \right) u_R^i + \bar{d}_R^i \left( i\cancel{\partial} - g' \frac{1}{3} \cancel{B} \right) d_R^i \end{aligned} \quad (2.11)$$

The superset symbol  $\supset$  indicates that the Lagrangian actually includes more terms which are not written here. Here we have to apply quark rotation as before to transform it into the mass basis mass basis. After some matrix computations,  $K_u$  and  $K_d$  will be removed by rotations  $d_R \rightarrow K_d d_R$ . What makes interesting for the flavour change is mixing terms of quarks.  $u_R \rightarrow K_u u_R$ .  $B_\mu$  and  $W_\mu^3$  are not involved in quark mixing term. Only  $W^\pm$  are left to be in quark mixing term as,

$$\mathcal{L}_{mass} \supset \frac{e}{\sqrt{2} \sin \theta_w} \left[ W_\mu^+ \bar{u}_L^i \gamma^\mu (V)^{ij} d_L^j + W_\mu^- \bar{d}_L^i \gamma^\mu (V^\dagger)^{ij} u_L^j \right] \quad (2.12)$$

Now the quantities of quark mixing can be summarized in one matrix, which is called as *Cabibbo-Kobayashi-Maskawa (CKM) matrix*.

$$V \equiv U_u^\dagger U_d = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad (2.13)$$

A representation of the CKM matrix is with standard parameters, that are Euler angles  $\theta$  and a phase  $\delta$ . These parameters are from symmetry group and reduced to four by transformation of group.

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{13}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix} \quad (2.14)$$

Here the relations  $c_{ij} \equiv \cos \theta_{ij}$   $s_{ij} \equiv \sin \theta_{ij}$  are applied.

Another representation was realized by Wolfenstein parametrization where  $\lambda \equiv \sin \theta_{12}$  describes, by solely using biggest value  $\theta_{12}$ , since other values are small and negligible.

$$V = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} \quad (2.15)$$

To make geometric representation of CKM matrix, a equation of unitary condition (2.16) is applied.

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad (2.16)$$

This unitary condition has reasonable lengths of all sides to visualized a triangle. The unitary triangle of CKM matrix has a geometric form as in figure 2.4.

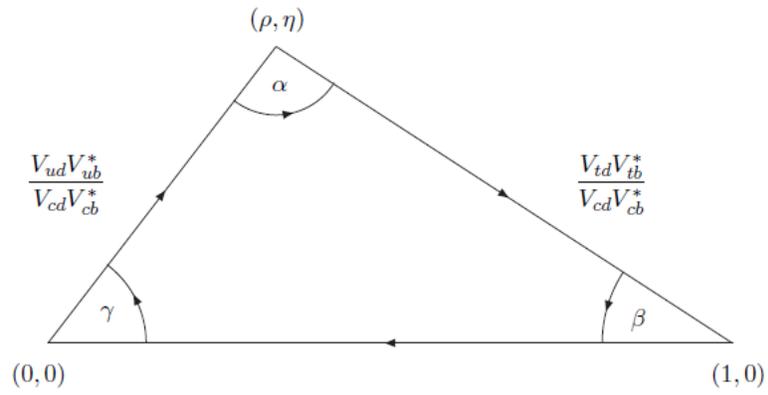


Figure 2.4: The CKM unitarity triangle [24].

# Chapter 3

## Belle II detector

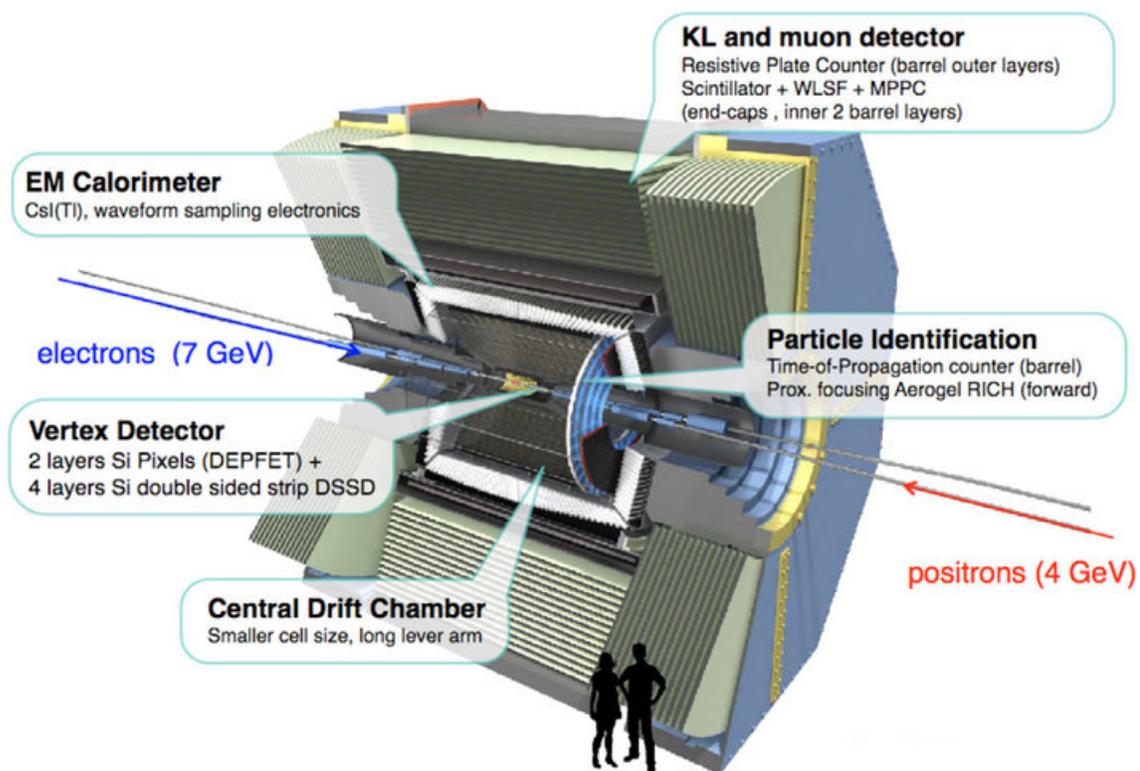


Figure 3.1: Structure of the Belle II detector [37].

### 3.1 SuperKEKB

For the Belle II detector to observe the decay of the  $\Upsilon(4S)$  meson, there has to be an accelerator that can make collision of electron and positron. The SuperKEKB can make it well with high luminosity and its structure is shown in figure 3.2.

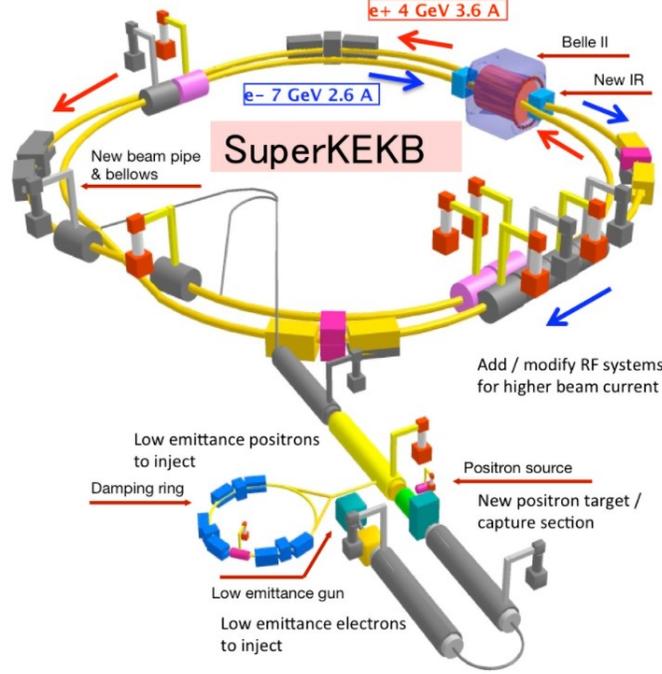


Figure 3.2: Structure of SuperKEKB [37]. Electron and positron are accelerated in the linear accelerator and collide in the Belle II detector.

The SuperKEKB produces two beams of electron and positron in separate rings. One is High Energy Ring(HER) in which electron beam is stored with an energy of 7 GeV. The other one is Low Energy Ring(LER) where positron beam is stored with an energy of 4 GeV. By this collision of electron and positron beams in separate rings, center-of-mass energy  $\sqrt{s}$  of 10.58 GeV can be achieved, which corresponds to the mass of the  $\Upsilon(4S)$  resonance.

The SuperKEKB produces beams with asymmetric energy, 4 GeV and 7 GeV of electron and positron, so that the laboratory system gets Lorentz boost [14] relative to the center of mass frame of  $\Upsilon(4S)$ . The boost has value of  $\beta\gamma$  in SuperKEKB as,

$$\beta\gamma = \frac{P_{e^-} - P_{e^+}}{\sqrt{s}} \simeq \frac{E_{e^-} - E_{e^+}}{\sqrt{4E_{e^-}E_{e^+}}} \simeq 0.28 \quad (3.1)$$

Such boost is used to improve vertex detection of B meson. The B mesons have almost same direction of boost, therefore z direction of B meson can be separated better by using

boosted laboratory system [14]. The exact values of SuperKEKB are shown in the following table with comparison to KEKB.

	KEKB	SuperKEKB
Energy (GeV) (LER/HER)	3.5/8.0	4.0/7.0
$\xi_y$	0.129/0.090	0.090/0.088
$\beta_y^*$	5.9/5.9	0.27/0.41
I(A)	1.64/1.19	3.60/2.262
Luminosity ( $10^{34} \text{cm}^{-2} \text{s}^{-1}$ )	2.11	80

Table 3.1: Fundamental parameters of SuperKEKB and KEKB [9]. Total beam current (I), vertical beam-beam parameter  $\xi_y$  and vertical beta function at the IP  $\beta_y^*$ .

The boost is decreased in SuperKEKB as shown that Energy difference of LER and HER is smaller in SuperKEKB than in KEKB. The value of  $\beta\gamma$  was 0.42 in KEKB. But, the luminosity is greatly improved in SuperKEKB. The luminosity is defined in particle physics as,

$$L = \frac{1}{\sigma} \frac{dN}{dt} \quad (3.2)$$

It includes the cross section( $\sigma$ ) and the number of events(N) with time(t). To obtain high luminosity, the collision of beams has to be more focused and the beam current should be increased in order to obtain the higher number of events. The SuperKEKB accomplished it by development of nano-beam scheme. By nano-beam scheme, the vertical direction of particles, which are related to  $\beta_y^*$  in table 3.1, is reduced. And crossing angle is increased to make area of collision smaller, by that the number of events increases as in figure 3.3.

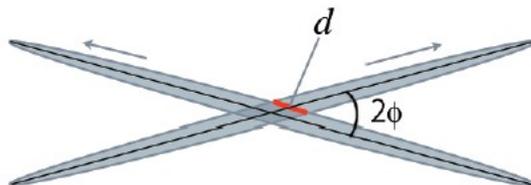


Figure 3.3: Schematic view of beam collision in the Nano-Beam scheme [9]. The larger crossing angle makes smaller area of collision.

## 3.2 Vertex detector

After leaving beam pipe, the particle enter the vertex detector which is composed of a pixel detector(PXD) and a silicon vertex detector(SVD). The structure of the vertex detector is shown in figure 3.4.

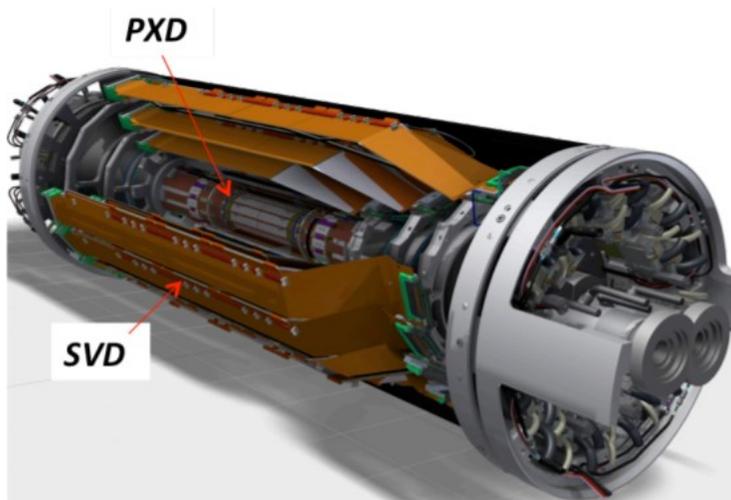


Figure 3.4: Structure of the vertex detector [10]. Inner part is PXN and enclosed by outer SVD. Each part has multiple layers.

PXN is, once more, made of two layers of sensors with radii of 14 mm and 22 mm. To achieve high precision of measurement of vertex in innermost detector, PXN was built by pixel sensors, instead of strips. DEPFET (DEPLETED Field Effect Transistor) pixels are used to detect the position of the tracks, which can make it feasible that pixels can be very thin down to 50 micron. Such thin pixels don't require extra air cooling, because it's only active during the readout, when the particle is detected in sensor.

SVD is surrounding PXN and made of four layers which are double-sided silicon strip detector (DSSDs) as shown in figure 3.5. By consideration of large data size and the number of channel, it was more reasonable to build outer part SVD with strip sensors rather than pixel sensors. The DSSDs construct ladder structure which is labeled with L3 to L6. It measures two dimensional position of particles by horizontal and vertical direction of the strips on the sensor. SVD can detect the particles in range  $17^\circ < \theta < 150^\circ$ . Furthermore, SVD can reconstruct the particle with low transverse momentum, down to a few tens of MeV/c.

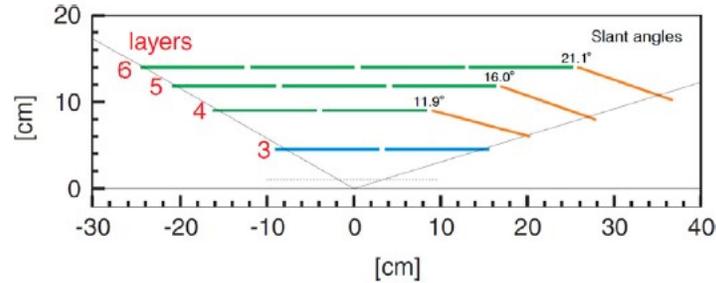


Figure 3.5: Layout of SVD with four layers [10].

### 3.3 Central Drift Chamber (CDC)

The Central Drift Chamber(CDC) is filled with 50% Helium and 50% ethane gas, which is optimal ratio for the performance so far since Belle I. And 14336 sense wires with an inner radius of 160 mm and an outer radius of 1130mm are installed to detect charged particles. When the charged particles enter the CDC, they induce ionization of the filled gas by electric charge of the particles. Then the ionization makes signal in nearest sense wire and the particles can continuously make further ionization as the particles continue to move, which can be observed as a track in CDC. In this work, for example,  $K^+$  can be reconstructed in CDC with this technical method. Moreover CDC can measure energy loss of particle making particle identification of low momentum particle within CDC, without reaching particle identification device.

### 3.4 Particle Identification

The particles that were not identified in CDC have to be identified after they leave the CDC. The particle identification also takes information from other detectors and combines the information for identification of particle. To identify the particles, the particle identification was divided into two parts and it was developed to measure momentum and velocity of particles to infer the mass of particle.

#### 3.4.1 Time-Of-Propagation(TOP) - Barrel

First part of the particle identification is Time-Of-Propagation(TOP), where the principle of Cherenkov radiation plays central role. The speed of light can be slower than  $c$ , when the light is in a material. It means that the particle, which pass the material, can move faster than the speed of light  $c$ . In this case, the Cherenkov photon can be produced. In the TOP, Cherenkov photon is captured in quartz, where it is reflected with Cherenkov angle. Fast particles would make short path to travel to reach photon detectors. The path of the

particles is reconstructed by 2-dimensional space( $x, y$ ) and time  $t$ . With the information of velocity from travel time in quartz, the TOP can deduce the mass of the particles. The structure of quartz bar is shown in figure 3.6.

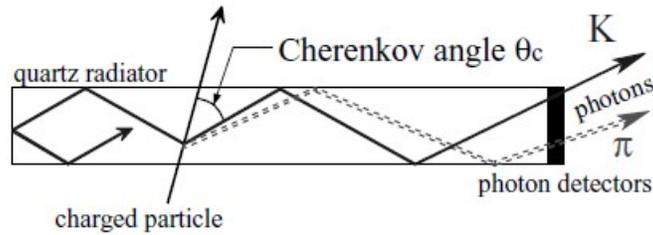


Figure 3.6: The Cherenkov photon passes through quartz and the Cherenkov angle is measured. By dependence of angle on velocity of the particles will make different path and time to reach photon detector. The different information of mass can make feasible to distinguish Kaon and Pion. [9].

### 3.4.2 Aerogel Ring-Imaging Cherenkov detector (ARICH) - End-cap

In ARICH, analog principle from TOP was applied to identify the particles. The particles also produce Cherenkov photons as in TOP, when it passes through aerogel. In this detector as in TOP, the Cherenkov photons have distinguishable angles by different mass of the particles. The Cherenkov photons are emitted and reach the photon detector which is located away. And the photons are detected in ring form. Even though the positions ( $x, y$ ) are different, the same particle will make same radius of ring by the angle. One can observe different rings by Kaon and Pion. Figure 3.7 depicts how the photons were emitted and arrive at photon detector.

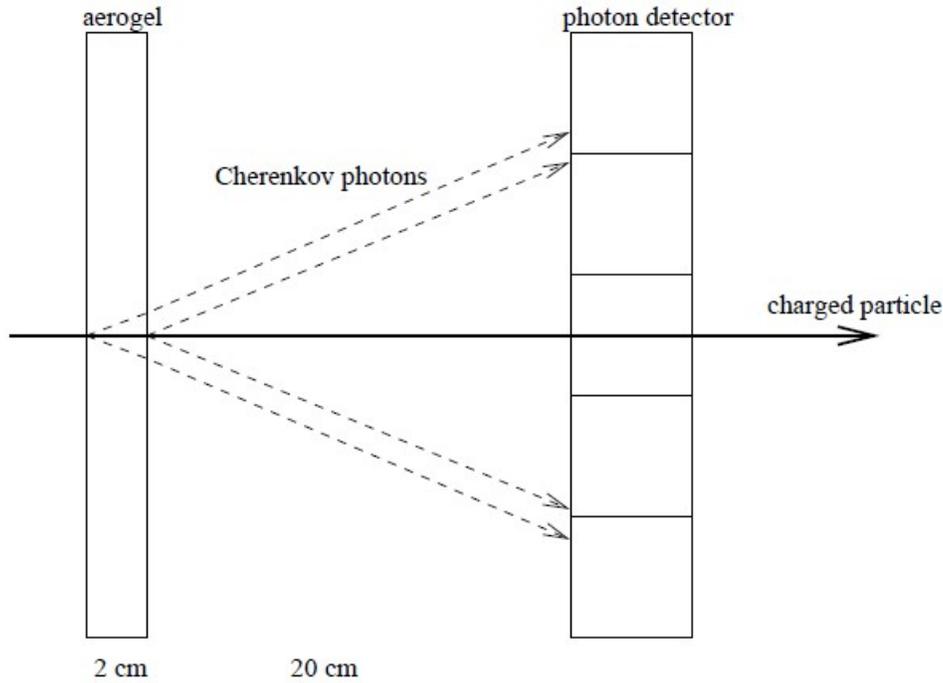


Figure 3.7: Aerogel with 2cm thickness produce the Cherenkov photons with different angle. The photons reach with photon detector as making ring form [9].

### 3.5 Electromagnetic calorimeter (ECL)

The Electromagnetic calorimeter(ECL) measures the deposited energy of photon and electron that reached in ECL after particle identification. High resolution of energy measurement was provided by total 8736 Caesium Iodide crystals doted with Thallium as scintillation material. The barrel part contains 6624 CsI crystals and the two endcaps contain 2112 CSI crystals. The scintillator crystals cover a polar angle region of  $12.4^\circ < \theta < 155.1^\circ$ . Two Hamamatsu Photonics S2744-08 photodiodes were used for readout of scintillation light. The photons that reached ECL create electron positron pairs that make again, as a iterative process, bremsstrahlung photons. This process continues until the energy of the created photons is too low for further processes [13]. Then the energies of all photons is measured together.

Intrinsic energy resolution of the calorimeter was expressed as by following approximate formula [9].

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt{E}}\right)^2 + (1.34\%)^2} \quad (3.3)$$

Superconducting solenoid is located around ECL to generate a magnetic field of 1.5T which

causes curvature of charged tracks. This superconducting solenoid has diameter of 3.4m and length of 4.4m. The muon goes through the ECL without making feasible measurement and can be detected in next part of Belle II detector(KLM).

### 3.6 $K_L^0$ and $\mu$ detection (KLM)

The  $K_L^0$  and muon detector(KLM) is outermost part of the Belle II detector where the  $K_L^0$  and muon are detected. The KLM is composed of the iron plates and detector elements. The iron plates have thickness of 4.7cm and are located outer part of the solenoid. These iron plates have a role to make the magnetic flux return of solenoid. The other role is creating the hadronic showers of the particles in iron plates, which is the detection of the  $K_L^0$ . The detector elements are divided into barrel and endcap part. The barrel part cover the range of angle from  $45^\circ < \theta < 125^\circ$  and this barrel and endcap together cover wider range of angle  $20^\circ < \theta < 155^\circ$  as it is located outer part of Belle II detector. In barrel part, there are 15 detector layers and 14 iron plates, while there are 14 detector layers and 14 iron plates in the endcaps. Muons and non-showering charged hadrons with momentum above 0.6 GeV/c traverse KLM. The glass electrode RPCs detects charged particles in KLM.

# Chapter 4

## Belle II Analysis Software Framework

The Belle II detector as hardware alone cannot investigate the particles. In order to analyze the data gained from the detector, the Belle II software is indispensable. The Belle II Analysis Software Framework(basf2) has numerous functionalities to reconstruct the decay from the raw data.

To operate basf2, the *externals* and *tools* are developed [25]. The externals contains third-party frameworks and libraries, on which basf2 is based and runs. In the externals, for instance, there are python3, git and HEP software like ROOT, Geant4, and EvtGen. The tools are shells and python scripts for the installation and setup of the externals and basf2. And basf2 contains modules, which can selectively loaded by the users.

### 4.1 Path and Module

A *module* is a part of basf2, which executes a certain functionality of basf2. Because the basf2 is split into many modules, it is entailed that many modules have to be run along with in the process of analysis. But it would be fairly cumbersome to write many scripts with code for running each module. Plus, the modules are dependent on each other, in the case of wrong order of the modules, making some module dysfunctional. The more efficient way became feasible by combining the modules with *path* as in figure 4.1.

By using path, one can write many modules in a so-called *steering file* and run at once. In basf2, one can import `modularAnalysis`, which is a python package and contain many functions to run the modules. It is possible that the users of basf2 add or remove certain module according to their need.

### 4.2 Reconstruction

The core part of decay analysis, in terms of software, is reconstruction. The reconstruction enables us to figure out whole decay process in an organized way, which means that one can connect a particle with its daughter particle. To do that, a python function

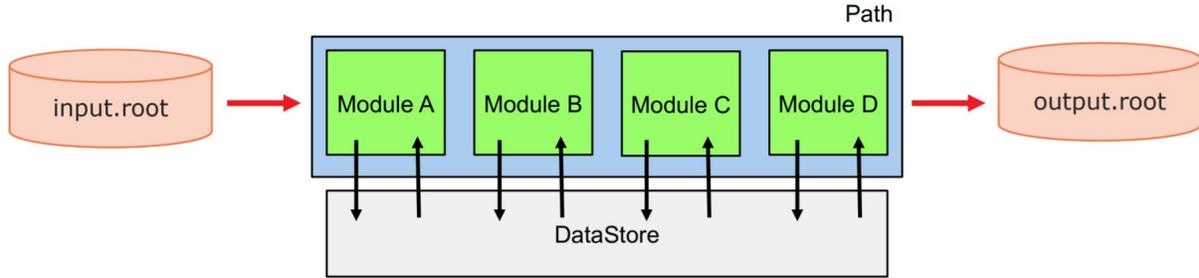


Figure 4.1: Structure of path and modules [26].

`reconstructDecay` of `basf2` is applied with `decayString`. In this work, one `decayString` as example looks like, "`B+:signal -> K+:charged ?nu`". It reconstructs the decay of  $B^+$  and question mark is added on the `nu` to ignore neutrino, because the neutrino doesn't leave any signal in the detector. For the reconstruction of  $\Upsilon(4S)$ , the `decayString` looks like, "`Upsilon(4S):generic -> B+:signal B-:generic`".

An additional functionality of `reconstructDecay` is applying selection cut on the reconstructed particle. The candidates of reconstructed particle, which cannot pass the cut, are removed in the output root file. This cut is classified as *pre-cut*, because it is applied in early phase, before producing output file, while the selection cut during offline analysis is called as *post-cut*. This pre-cut reduces the size of the output file so that it saves storage and memory usage in offline analysis, which is especially advantageous, when high statistics is used for analysis.

### 4.3 Full Event Interpretation

The B mesons pair arises from  $\Upsilon(4S)$  resonance and one B meson is called tag B, which doesn't produce signal  $B^+ \rightarrow K^+ \nu \bar{\nu}$  we are looking for. In some cases, the reconstruction of signal B is not sufficient. For instances, the decay of this work that are including neutrinos cannot have full information about decay only with signal B, because neutrinos have no interaction with detector. The reconstruction of tag B can give information even about signal side like missing energy and momentum. The reconstruction of tag B easily enables full reconstruction of  $\Upsilon(4S)$ , as we can have both reconstructed B mesons from  $\Upsilon(4S)$ . In `basf2`, such reconstruction of tag B is accomplish by *Full Event Interpretation*(FEI) [18], [20].

The FEI algorithm is schematically described in figure 4.2. The FEI get started from top where final-state particles are reconstructed from tracks and ECL clusters. After that, final-state particles are combined to reconstruct intermediate particles. At the end, the tag B meson is finally obtained. In figure 4.3, the *Full Reconstruction*(FR), which is a tagging algorithm used in Belle I experiment, is compared with FEI. FEI requires less computing

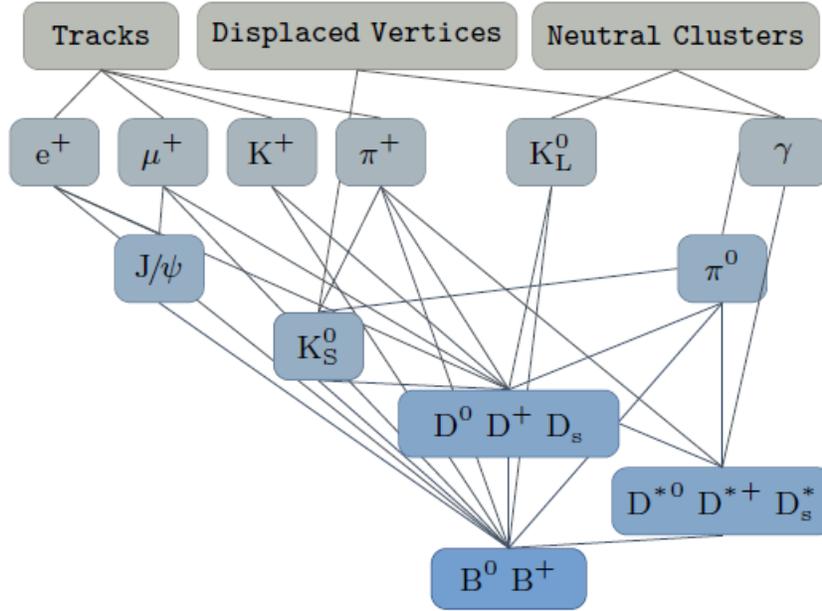


Figure 4.2: Schematic description of the FEI algorithm. The reconstruction is carried out from top to bottom. [20].

time and has higher tag-side efficiency than FR [20].

In the tagging process of FEI, the users can select channel of FEI. Hadronic and semileptonic tagging methods are available in FEI. In Hadronic tagging, the B meson is reconstructed in a hadronic decay mode, then momentum of tag B is known and the sample is purer. But it has demerit having low efficiency and branching fraction. In Semileptonic tagging, the tag B is reconstructed from the decay containing lepton and neutrino. While semileptonic tagging has high efficiency and branching fraction, it lacks full kinematic information, because of the missing neutrino. In this analysis, hadronic tagging is selected from FEI skim.

Main reason of using FEI in this work is for applying cut and doing fBDT training of section 7.1 with the tag variables. Unlike the case of dispensing FEI and tag side, using FEI we can acquire the variables of tag side, which have good properties to reject background. It makes feasible that one can have more chance to find out good selection cut to distinguish the signal from the background. In the following section, the selection cut for both signal and tag side will be described. In this analysis, the FEI skimmed MC data was used, because applying FEI costs longer running time. In FEI skimmed data, FEI is already applied by specialists at Belle II. The users should be aware of the FEI channel and selection cut which were applied in skim.

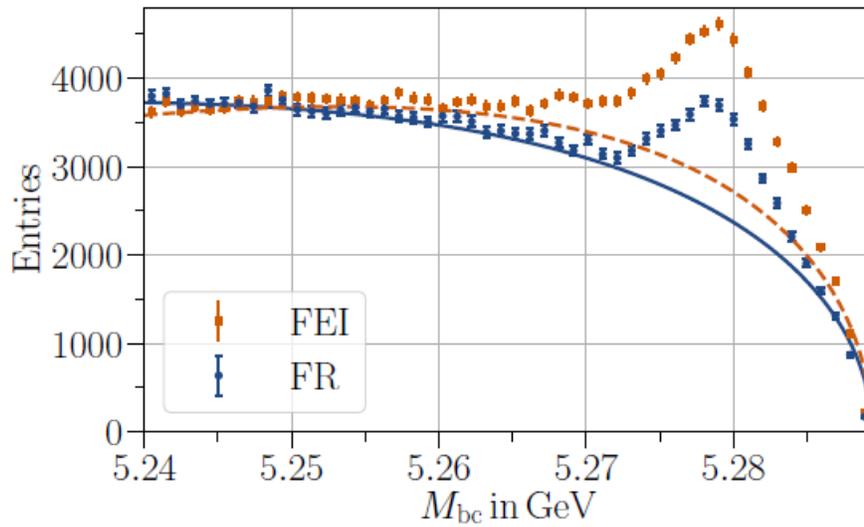


Figure 4.3: The older version FR is compared with the FEI [20]. The FEI has better efficiency to reconstruct a tag B correctly, which has correct beam-constrained mass.

# Chapter 5

## Monte Carlo Simulation

In this analysis, MC datas, that contain signal, B meson pair and continuum, is used and can be compared with the result of real data analysis from the references ??, ??. The large MC data is produced at Belle II, but the analysts can also produce MC data themselves using a python function `generators.add_evtgen_generator` at Belle II software, which imports `EvtGen` generator. The `EvtGen` is a software that can generates events of B and D mesons. The MC data at Belle II is divided into signal MC data( $B^+ \rightarrow K^+ \nu \bar{\nu}$ ) and generic MC data for background with FEI skim, that I used as input data of analysis. Such MC data of Belle II has a valuable feature `MCMatching` that real data doesn't have. The `MCMatching` relates reconstructed particles to MC particles to examine the quality of reconstruction. In this process, reconstructed particles are compared to MC particles. This examination is necessary, because many particles can be wrongly reconstructed and identified. This comparison is implemented by a python function `modularAnalysis.matchMCTruth`, which imports `McMatching`. In the case of B+ signal, the function is written as `modularAnalysis.matchMCTruth(list_name="B+:signal", path=main)`. There are various variables that can examine in `MCMatching`. One of simplest variable is `isSignal`. It is 1.0, if `MCMatching` decides correct reconstruction, and 0.0 otherwise. There are also similar type of `MCMatching` variables and the result of their test is shown in figure 5.1.

	1.0	0.0
<code>isSignal</code>	4663	119
<code>isSignalAcceptMissing</code>	4672	110
<code>isSignalAcceptMissingNeutrino</code>	4663	119
<code>isSignalAcceptMissingGamma</code>	4663	119

Table 5.1: The simple test is made with 265k events of signal MC data to compare `MCMatching` variables. The variables are from signal side.

The other variables in table 5.1 tolerate and accept losing some particles as the name itself explains. As it is shown in section 4.2, the neutrinos are already ignored

during reconstruction. Therefore, `isSignal` and `isSignalAcceptMissingNeutrino` have same value and it didn't matter which one is selected for the calculation of efficiency. `isSignalAcceptMissing` has slightly higher value, which implies that fake signals are most likely to be there.

The `MCMatching` variables are not chosen for making selection cut, because the `MC-Matching` variables don't exist for real data. But, these can be useful to label the data in supervised learning. For example, `isSignal` of signal side is always 0.0 for background, because no background can be reconstructed as signal. These value 1.0 and 0.0 of `isSignal` can be label of machine learning.

# Chapter 6

## Analysis of $B^+ \rightarrow K^+ \nu \bar{\nu}$ decay

### 6.1 Input variables

#### 6.1.1 Variables for pre-cut

In this section, the plots of variables for pre-cut are introduced. No cuts are applied except  $\text{kaonID} > 0.5$  (kaon identification probability). The variables for cut is divided into three types, which are signal side, tag side and epsilon. These variables were searched by using feature importance of machine learning (SVM and neural network). The variables with high feature importance are selected to make pre-cut. Continuum suppression variables don't belong to this section, because these are input variables of fBDT training.

For the plotting of data `WG1Template`, which supports data visualization as a template at Belle II, is imported and applied. Because of the high number of background candidates in the absence of cut, the signal sample is scaled up by 50 times. Then the boundary between signal and background becomes clearly visible and one can easily determine position of pre-cut without losing many signal candidates. Because this is pre-cut, the cuts don't have to be strict and one can tighten the cuts or apply fBDT cut additionally. Choosing highest figure of merit and purity are not considered so that the pre-cuts don't become strict. Running steering file without cut makes runtime, memory usage and storage extremely high, if full statistics applied. Therefore it was run with lowered statistics and I increase statistics after testing and applying all pre-cuts in section 6.3.2. But, reasonable amount of statistics is applied to see separation between signal and background clearly. Here 4834 signal candidates and 1837537 background candidates are produced from 260k events for signal and 8 million events. When only part of entire statistics of MC14 is used, the cross section of each background has to be correctly adjusted. The number of each background candidates are adjusted as in the cross section in [7] by removing excessive amount of

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

Table 6.1: Cross section of physics process of collision at  $\sqrt{s} = 10.58$  GeV [7].

certain background in dataframe. The cross sections of relevant processes are shown in table 6.1.  $\Upsilon(4S)$  decays into either mixed or charged B meson pairs and quark pairs are called as continuum. The charged B mesons are  $B^+$ ,  $B^-$  and mixed B mesons are neutral B. In the plots of this section, the fraction of each background adjusted by this cross section.

In signal side, `missingEnergyOfEventCMS`, which is calculated by the definition,

$$\text{missingEnergyOfEventCMS} = E_{CMS} - \sum_i E_i \quad (6.1)$$

$E_{CMS}$  is the energy of  $\Upsilon(4S)$  and the sum runs over the energies of all reconstructed particles  $E_i$ . The `missingEnergyOfEventCMS` has good separation between signal and background, because signal produces two neutrinos and has more missing energy from neutrinos as a result. In the `Mbc`(beam constrained mass,  $\text{GeV}/c^2$ ) of tag side, there was already FEI skim cut `Mbc`  $> 5.24$   $\text{GeV}/c^2$ , so the plot is empty in the region `Mbc`  $< 5.24$   $\text{GeV}/c^2$ . The signal has peak near B meson mass in `Mbc` plot as signal comes from B meson obviously. For all Rest of Event(ROE) variables, which are explained in section 6.2, signal has a peak at smaller value than background. In upsilon side, ROE has to be zero ideally, because signal can only be produced from upsilon and rest of upsilon doesn't exist. In the plot, the signal peaks are located near zero.

The distribution of the signal and tag side variables is shown in figure 6.1. And the distribution of upsilon variables is shown in figure 6.2. The Rest of Event and `cleanMask`(ROE mask) are explained in section 6.2. ROE variables measure the physical quantity like energy, momentum in the Rest of Event, which is not from the associated given particle in applying ROE. For instance, the `roeP` is momentum of unused tracks and clusters in Rest of Event. The selection cuts on this plots are shown in section 6.3.1. The name of the variables includes either `Bsig` or `Btag`, and either `d0` for a daughter particle(signal B) of upsilon or `d1` for another daughter particle(tag B) of upsilon. If there isn't any `d0` or `d1`, the variable is from upsilon. The definition of the remaining variables are given at Belle II as follows.

`roeE`: Energy of unused tracks and clusters in ROE.[GeV]

`Mbc`: Beam constrained mass.[ $\text{GeV}/c^2$ ]

`roeExtra(maskName)`: Extra energy from ECLClusters in the calorimeter that is not associated to the given Particle.[GeV]

`roeNeextra(maskName)`: Extra energy from neutral ECLClusters in the calorimeter that is not associated to the given Particle.[GeV]

`nROE_Tracks(maskName)`: number of tracks in the related RestOfEvent object that pass the selection criteria.

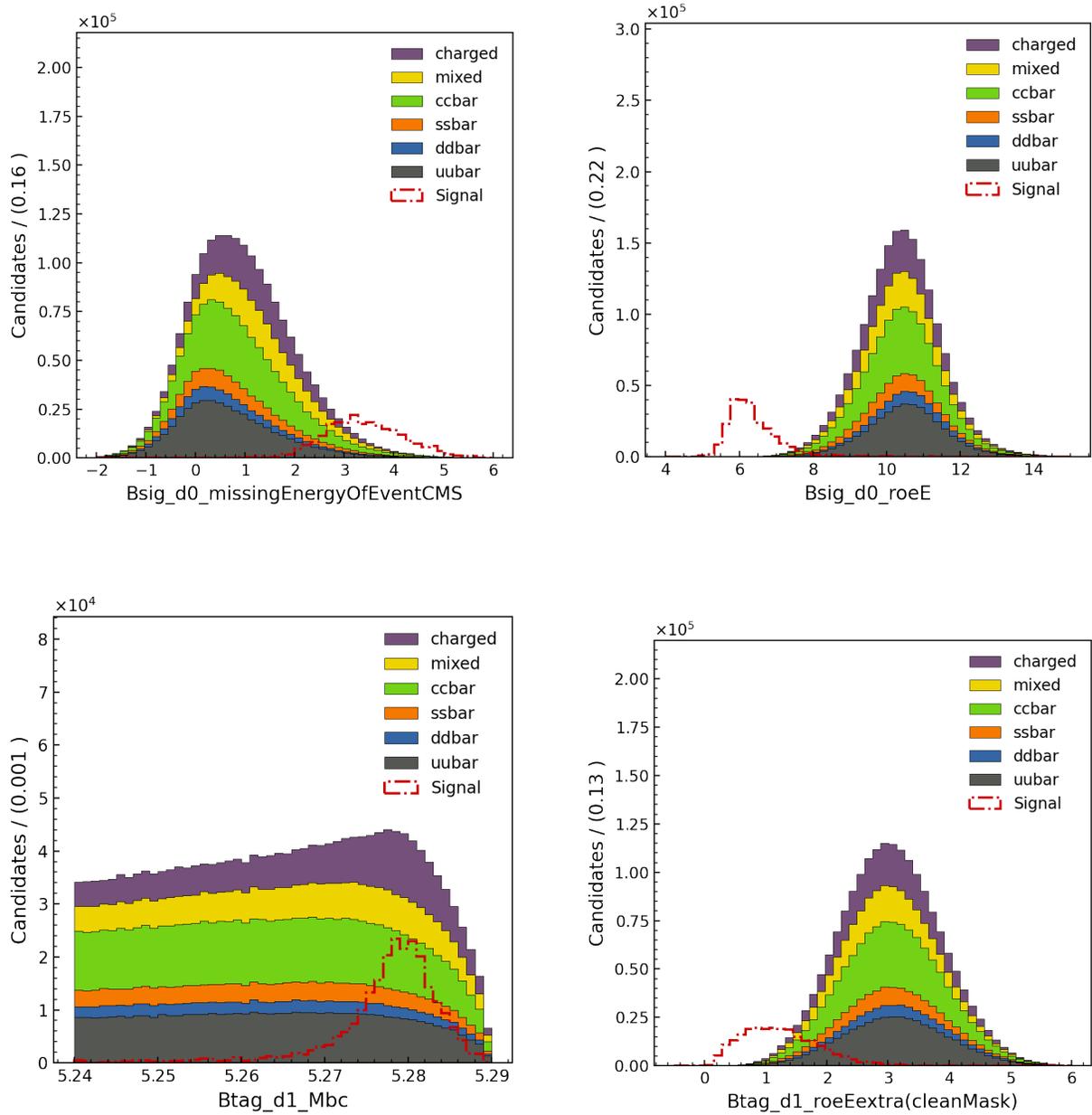


Figure 6.1: Signal and tag side variables for the selection cuts. The selection cuts are shown in section 6.3.1. GeV is used as unit.

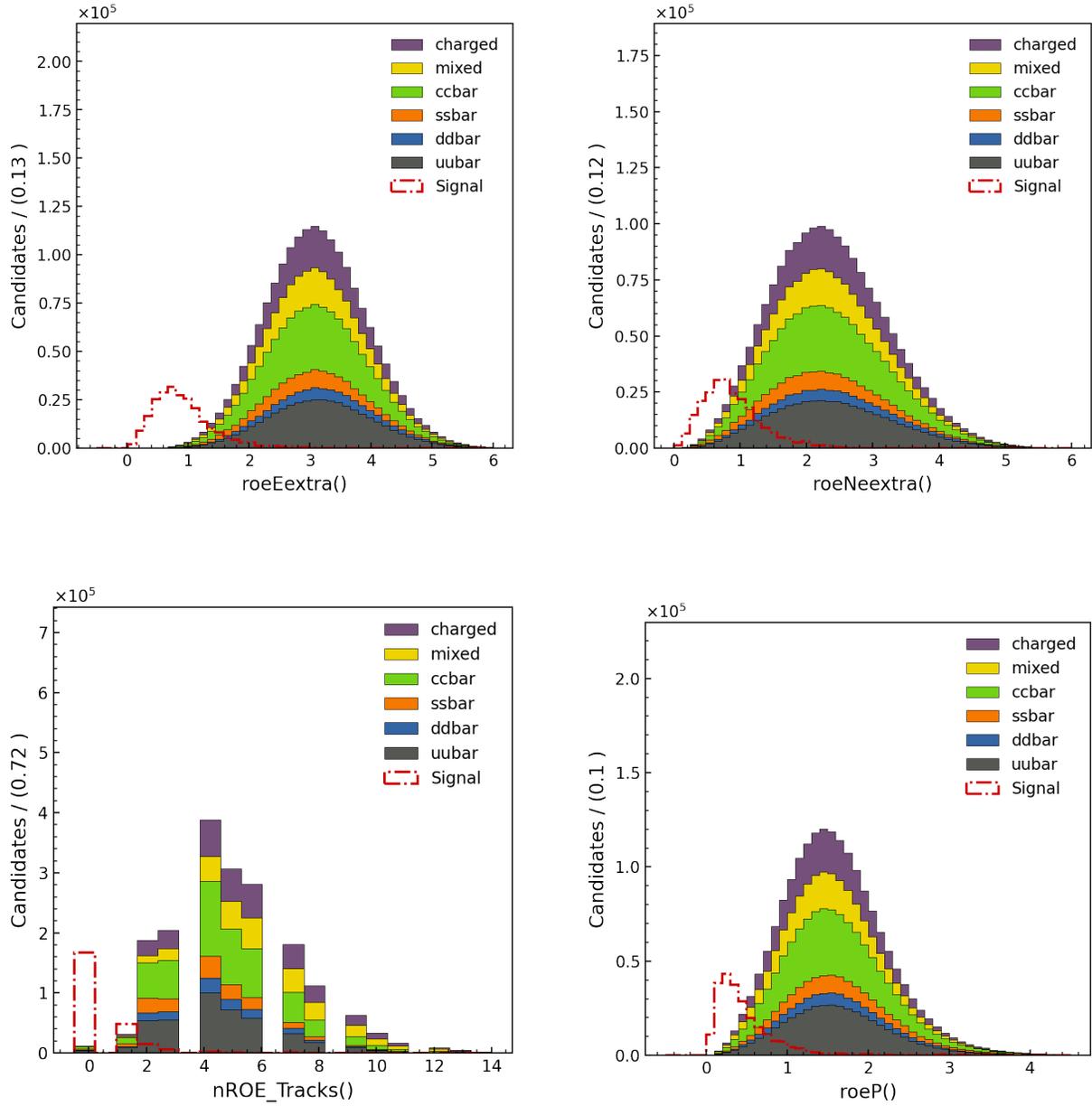


Figure 6.2: Upsilon variables for the selection cuts. The selection cuts are shown in section 6.3.1. GeV is used as unit.

### 6.1.2 Continuum Suppression after pre-cuts

The large fraction of background is contributed by continuum which do not produce B meson pair. The continuum is made of quark pair and have higher momentum than B meson pair. This high momentum makes distinction in thrust and direction of particles. The continuum can be separated from signal with following variables in this section. All of the variables in this section are input variables of fBDT training. All variables are plotted after applying all pre-cuts and making best candidate selection as in the section as in section 6.3.1 and 6.3.2. Because of the effect of the pre-cuts, shape of the plots can differ from original one. The shape of KSFW moment plots with different condition can be found in [5]. There aren't dramatic difference by pre-cut, because continuum suppression variables are not used in pre-cut. But, the most background candidates were removed greatly, thus, height of background peak is lower and comparable with signal. In the plots of this section, signal is not scaled up and stays with right fraction as given. If original plot without pre-cut is shown, then signal needs to be scaled up as in [5], because there are too many background candidates and tiny signal fraction is not visible.

One of the variables that has good discriminating power and feature importance for continuum suppression is R2. To understand how R2 is formulated, the Fox-Wolfram moments has to be defined in advance [7].

$$H_l = \sum_{i,j}^N |p_i| |p_j| P_l(\cos \theta_{i,j}) \quad (6.2)$$

$\theta_{i,j}$  is the angle between the i-th momentum  $\vec{p}_i$  and j-th  $\vec{p}_j$ , and  $P_l$  is the lth-order Legendre polynomial.  $R_l$  is defined with Fox-Wolfram moments as 6.3.

$$R_l = \frac{H_l}{H_0} \quad (6.3)$$

R2 is used more often as special case of  $R_l$ . Thrust is generally defined as 6.4.

$$T = \frac{\sum_{i=1}^N |\vec{T} \cdot \vec{p}_i|}{\sum_{i=1}^N |\vec{p}_i|} \quad (6.4)$$

$\vec{T}$  is the unit vector along which their total projection is maximal. and  $\vec{p}_i$  is i-th momentum. Many variables are defined with this thrust, because the thrust of continuum has good discrimination from B meson pair. **Thrust0m** is the magnitude of the ROE thrust axis. **CosTBT0** is cosine of angle between thrust axis of the signal B and thrust axis of ROE in center-of-mass frame. **CosTBT0** has high discrimination between the signal and the continuum, because thrust axis of the signal B has spherical form of direction and distributed well without peak, while the cosine angle between the two thrust axes of ROE of the continuum has peak at 1.0. **costBz** is cosine of angle between thrust axis of the signal B and z-axis in center-of-mass frame. These variables can also be used with ROE

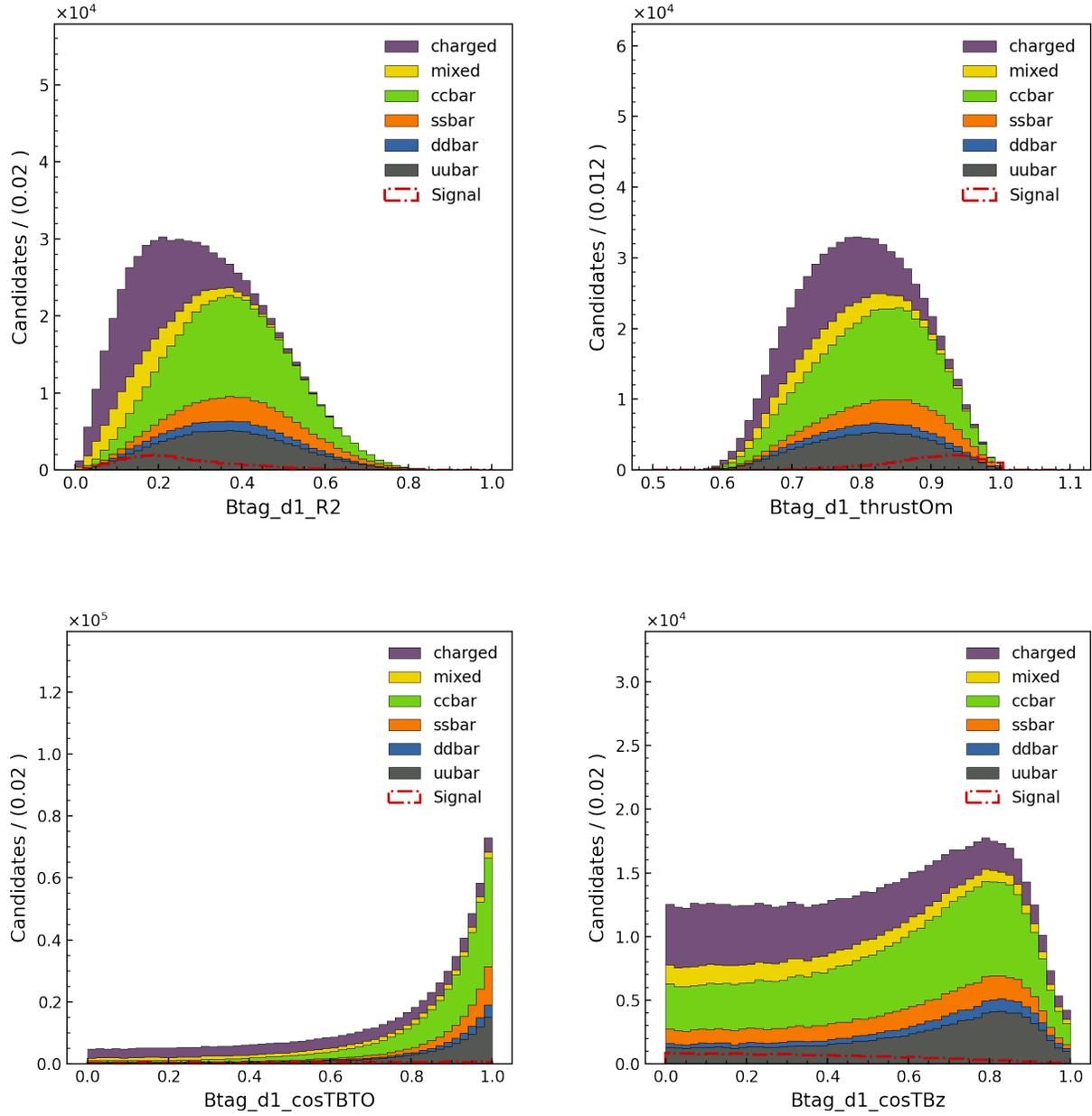


Figure 6.3: Continuum suppression variables of signal and tag B mesons as input variables of fBDT(section 7.1)

mask, but here all variables are applied without ROE mask. The distributions of these variables are shown in figure 6.3.

There are 16 Kakuno-Super-Fox-Wolfram (KSFW) moments which were developed to distinguish B meson pairs from continuum. The defining formulas of KSFW moments are given in [5]. The linear KSFW moments are defined as

$$h_{so\_il} = \sum_b C_{lb} \frac{|p_b| P_l(\cos \theta_{K^+b})}{2(2E_{beam} - E_{K^+})} \quad (6.5)$$

The index  $i$  is defined differently to make sum run with index  $b$  over charged ( $i = 0$ ) and neutral ( $i = 1$ ) ROE particles. The case of  $i = 2$  only calculates missing momentum vector [5]. The angle  $\theta_{K^+b}$  is measured by direction between particle of index  $b$  and signal  $K^+$ .  $C_{lb} = 1$  for even moments  $l$ .  $C_{lb} = 0$  is given for the neutral particles and equals to the product of charges of the particle  $b$  and  $K^+$  for charged particles.  $p_b$  is the momentum of index  $b$  and  $P_l$  is the  $l$ th-order Legendre polynomial. The quadratic KSFW moments are defined as

$$h_{oo\_l} = \sum_a \sum_b C_{lab} \frac{|p_a| |p_b| P_l(\cos \theta_{ab})}{(2(2E_{beam} - E_{K^+}))^2} \quad (6.6)$$

The index  $a$  and  $b$  run over exclusively ROE particles.  $C_{lab} = 1$  is given for even moments  $l$ . For odd moments,  $C_{lab} = 0$  is given, if  $a$  or  $b$  is neutral, or equals to the product of the charges for charged particles. The distributions of KSFW variables are shown in figures 6.4, 6.5, 6.6, 6.7.

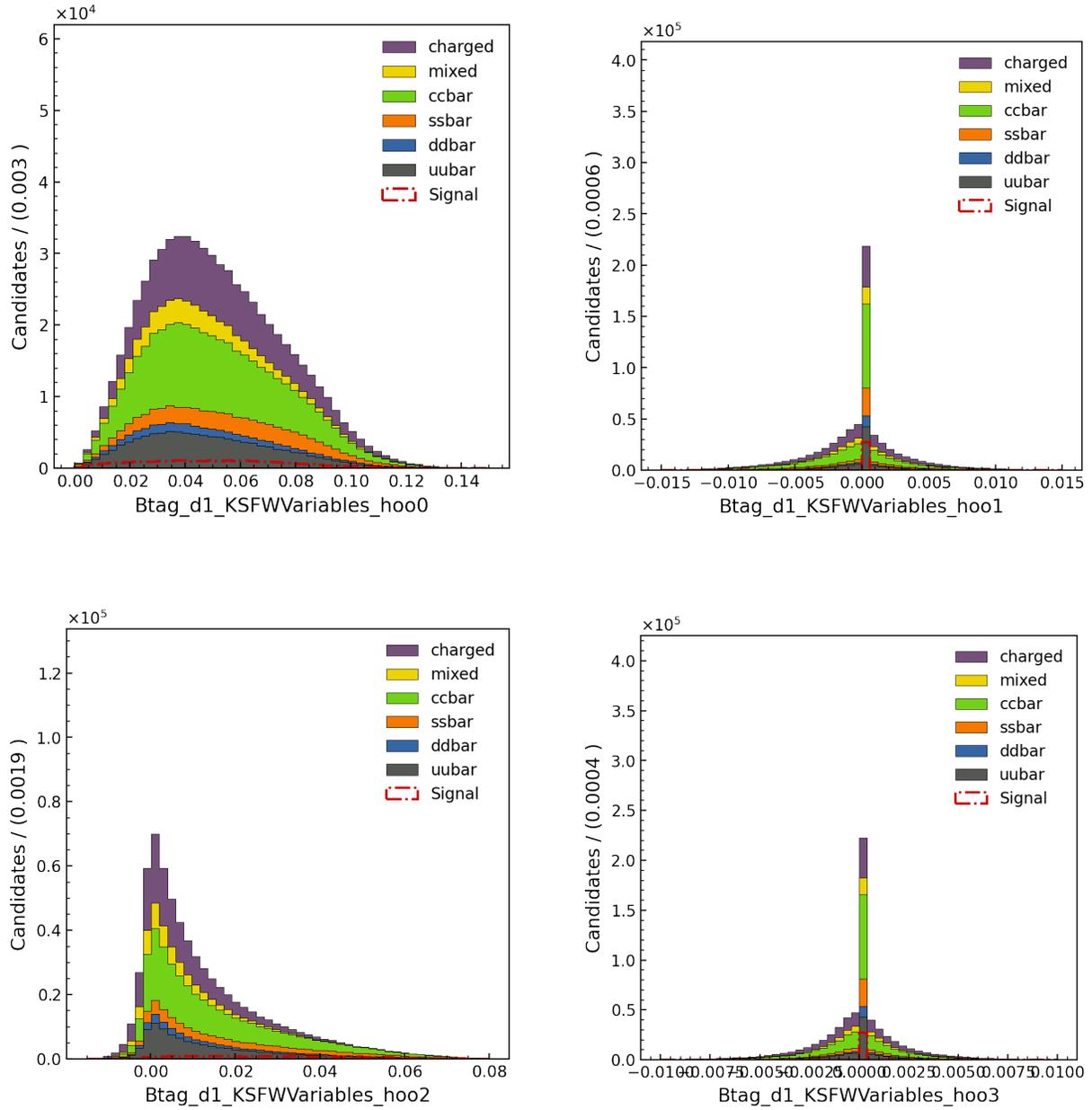


Figure 6.4: KSW variables of signal B meson as input variables of fBDT(section 7.1)

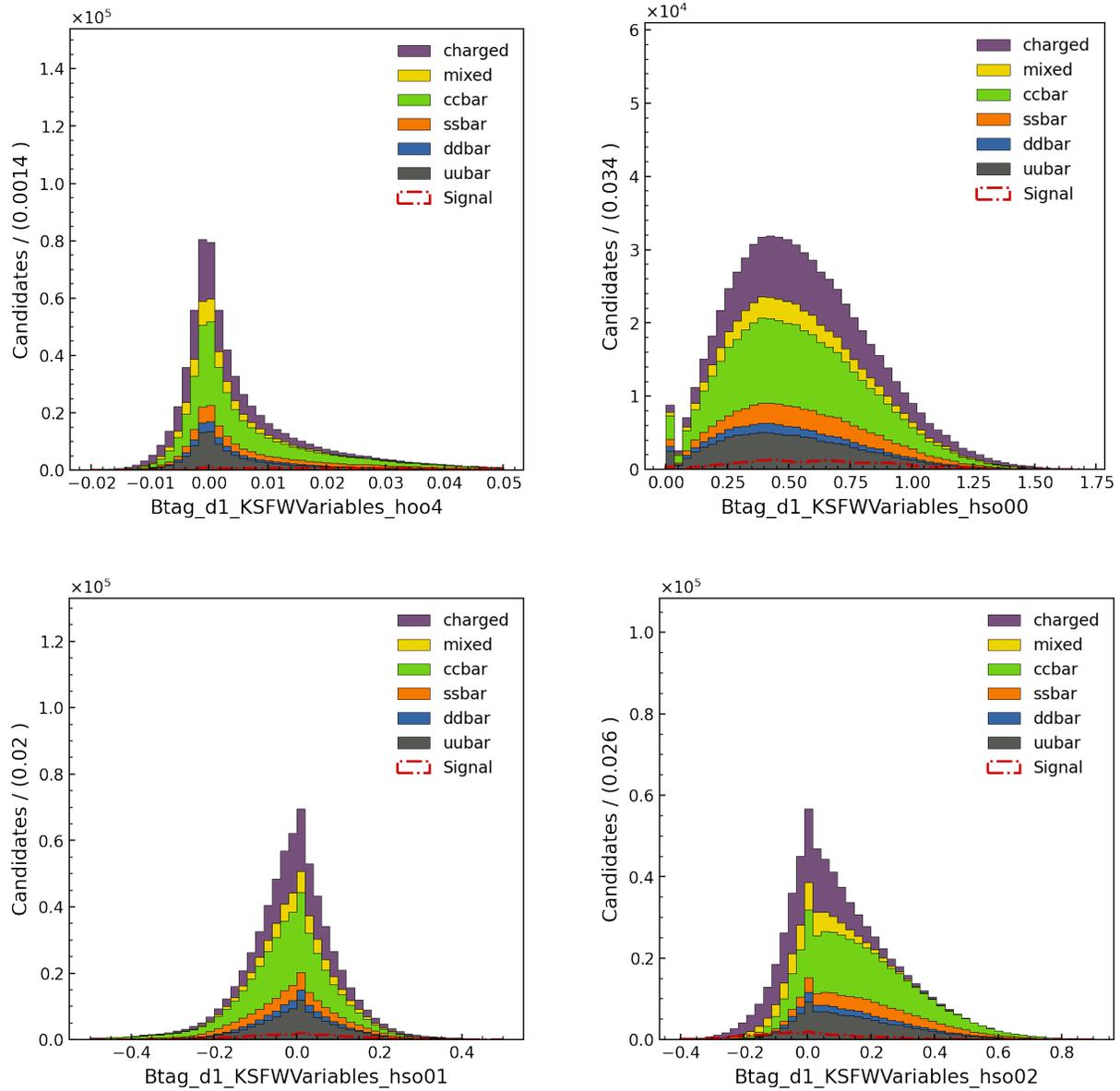


Figure 6.5: KSW variables of signal B meson as input variables of fBDT(section 7.1)

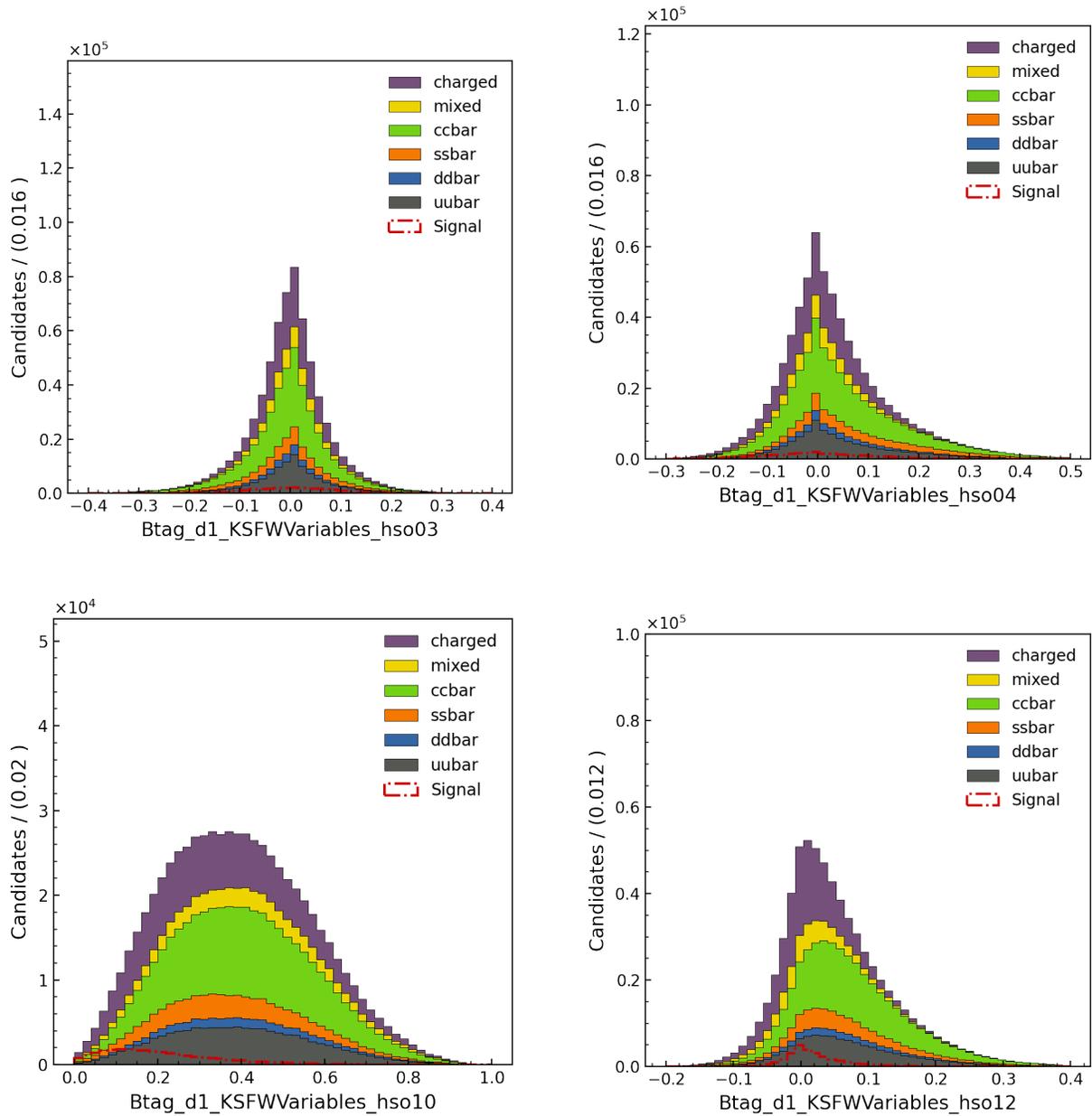


Figure 6.6: KSW variables of signal B meson as input variables of fBDT(section 7.1)

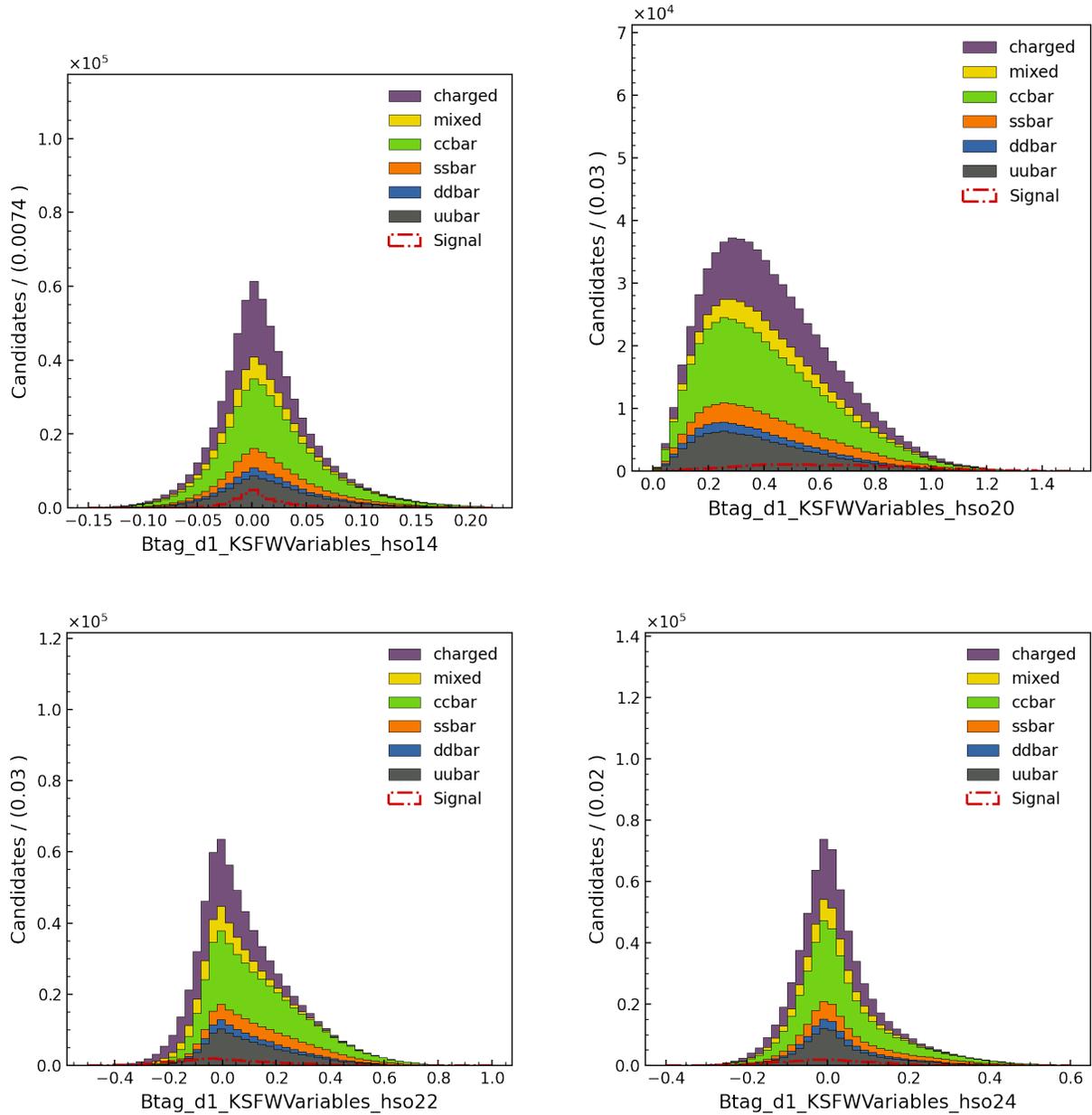


Figure 6.7: KSFV variables of signal B meson as input variables of fBDT(section 7.1)

## 6.2 Rest of Event

Independent of actual cut on the next section, Rest of Event is another useful method to clean up background. Rest of Event enables looking into variables without including signal particle. In other words, Rest of Event shows the distribution of background, which should be cleaned up by applying ROE mask(selection cut). The ROE mask is the set of selection cuts which removes candidates in Rest of Event. The ROE mask of the analysis is defined as follows.

### ROE mask

**Track:**  $nCDCHits > 0$ ,  $\theta_{InCDCAcceptance} == 1$ ,  $pt > 0.1$ ,  $abs(dr) < 1$ ,  $abs(dz) < 3$

**ECL cluster:**  $p \leq 0.05$ ,  $useCMSFrame(p) \leq 3.2$

First of all, the definition of the variables have to be introduced. `nCDCHits` is the number of CDC hits associated to the track. `thetaInCDCAcceptance` returns true, if particle is within CDC angular acceptance  $17^\circ < \theta < 150^\circ$ . `p` is momentum magnitude(GeV/c) and `useCMSFrame(p)` is the momentum(GeV/c) in CMS frame. And `dr` is transverse distance(cm) in respect to IP for a vertex and `dy` is vertex or POCA in case of tracks `z` in respect to IP. `nCDCHits > 0` is implemented, because CDC has better momentum resolution than vector detector, which is necessary for the analysis of the decay. Additionally, `thetaInCDCAcceptance` is made true so that only the angle range is analyzed. `abs(dr) < 1`, `abs(dz) < 3` rule out the long-lived particles and background tracks, which are not coming from the IP. The cuts of ECL cluster remove the background from photons.

In this analysis, the FEI reconstructed tag B meson, therefore, Rest of Event of Upsilon ideally should be empty. This result is shown in first figure 6.8. In all plots, the ROE mask is compared with the result without ROE mask.

In both plots, ROE mask make the distribution of number of tracks move to zero. In signal MC data, ideally, there should be no particle as ROE of upsilon by definition. For the background, all continuums and BB pair datas were combined in single dataframe and used to make plot. In ROE of background, there are still other particles as background. The next figure 6.9 shows the number of ECL clusters in the ROE. The ROE particles were reduced by ROE mask by same principle, but here the second ECL cluster cuts  $p \leq 0.05$ ,  $useCMSFrame(p) \leq 3.2$  of ROE mask removed ROE particles.

The next two figures 6.10, 6.11 show the momentum and extra energy of ROE particles. In these plots, the ROE mask doesn't make remarkable difference, even though the number of ROE particles decreased clearly. However, the ROE variables like these can be used for selection cut in the following sections, regardless of presence of ROE mask, because the distribution of these variables still can produce good separation between signal and background.

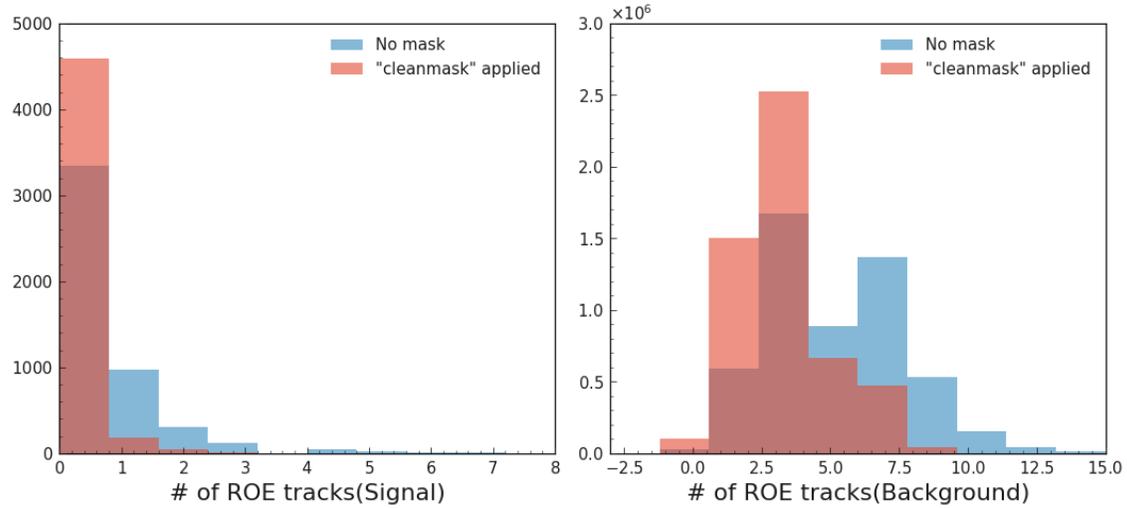


Figure 6.8: The number of ROE particles is shown. Left plot is signal MC and right one is generic MC(background).

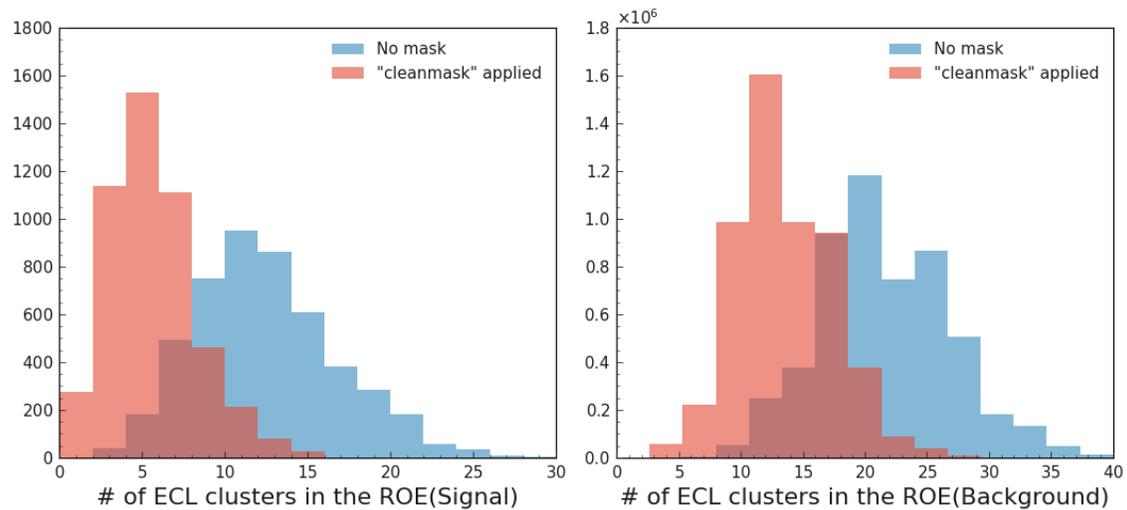


Figure 6.9: The number of ECL clusters in the ROE is shown. Left plot is signal MC and right one is generic MC(all of background).

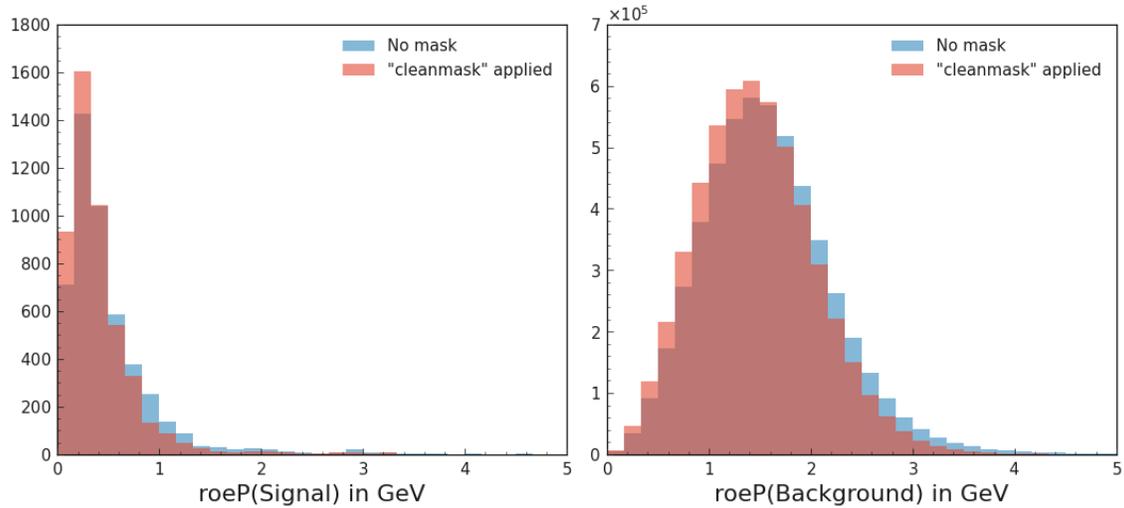


Figure 6.10: The variable  $\text{roeP}$  is shown. It is momentum of unused tracks and clusters in ROE. Left plot is signal MC and right one is generic MC(background).

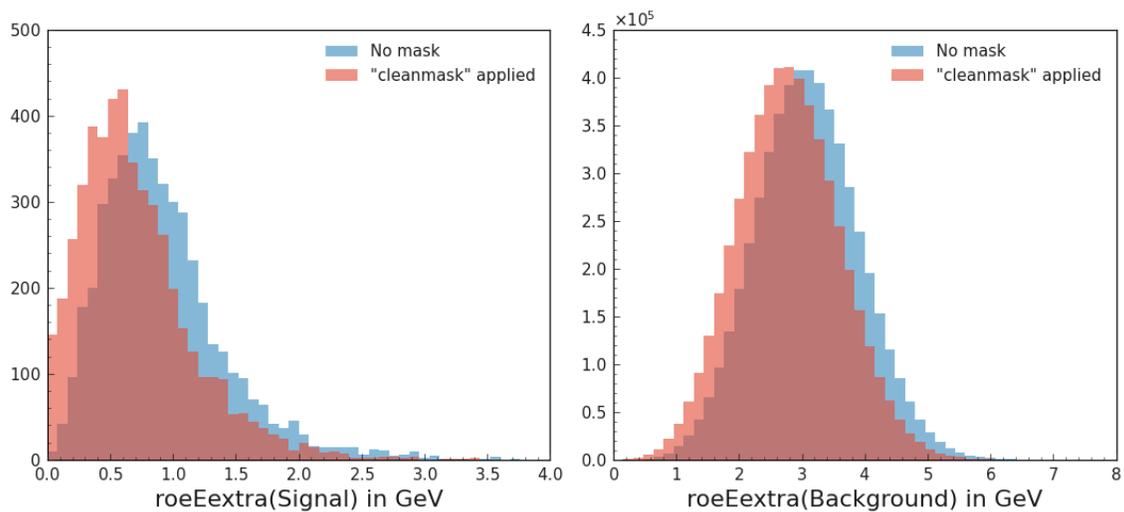


Figure 6.11: The variable  $\text{roeEExtra}$  is shown. The extra energy coming from the particles which are not associated to the given Particle(Upsilon). Left plot is signal MC and right one is generic MC(background).

### 6.3 Selection cut

Searching for selection cut is main part of analysis, because the goal is removing background and gain pure signal sample. The types of selection cuts are organized in this category.

1. FEI skim cuts
2. pre-cuts from feature importances
3. Best candidates selection(sigProb cut)
4. fBDT cut

First thing to be mentioned is that FEI skim, because the FEI skimmed dataset is used to save computation time. As FEI is already applied by Belle II researcher, FEI skim includes its own cuts to remove common background.

#### List of the FEI skim cuts

**Event:**  $n\text{CleanedTracks} \geq 3, n\text{CleanedECLClusters} \geq 3, \text{visibleEnergyOfEventCMS} > 4,$   
 $2 < E\_ECL\_FEI < 7$

**B+:**  $Mbc > 5.24, \text{abs}(\text{deltaE}) < 0.200, \text{sigProb} > 0.001, \text{extraInfo}(\text{dmID})==25$

**Track:**  $\text{abs}(z0) < 2.0, \text{abs}(d0) < 0.5, \text{pt} > 0.1$

**Cluster:**  $E > 0.1, 0.296706 < \text{theta} < 2.61799$

These cuts were applied, before starting analysis. FEI skim should have loose cuts to remove background and retain true B mesons. The definitions of the variables for the selection cut are given at Belle II as follows.

**nCleanedTracks:** The number of clean Tracks in the event Clean tracks are defined by the tracks which pass the given cut assuming a pion hypothesis.

**nCleanedECLClusters:** The number of clean Clusters in the event Clean clusters are defined by the clusters which pass the given cut assuming a photon hypothesis.

**visibleEnergyOfEventCMS:** The visible energy in center-of-mass frame.[GeV]

**E\_ECL\_FEI:** Cleaned tracks and clusters in ECL.

**Mbc:** Beam constrained mass of the B meson.[GeV/ $c^2$ ]

**deltaE:** Difference between upsilon energy and half the center of mass energy.[GeV]

**sigProb:** The probability for correct reconstruction of tag B by FEI.

**extraInfo(dmID):** Extra information stored under the dmID.

**z0:** The tracking parameter, the z-coordinate of the point-of-closest-approach.

**d0:** The tracking parameter, the signed distance to the point-of-closest-approach in the  $r - \pi$  plane.[cm]

**pt:** Transverse momentum.[GeV/c]

**E:** Energy.[GeV]

**theta:** Polar angle.[rad]

After that, the variables with good discrimination were searched. Machine learning of scikit-learn was applied to produce feature importances of variables and the variables with high importance were selected to make selection cut. And these cuts moved in the steering file as pre-cut to reduce size of output root file, and statistics(the number of events) is increased up to full dataset. After running steering file again with full statistics, best candidate selection was made. Then, finally, fBDT was run to classify the signal and background for higher purity and figure of merit.

### 6.3.1 Pre-cut

#### List of the selection cuts

**KaonID:**  $\text{kaonID} > 0.5$  applied in advance.

**Signal:**  $\text{missingEnergyOfEventCMS} > 1.85$  and  $\text{roeE} < 8.0$

**Tag:**  $\text{roeEextra}(\text{cleanMask}) < 2.5$  and  $\text{Mbc} > 5.265$

**Upsilon:**  $\text{roeEextra}() < 2.0$  and  $\text{nROE\_Tracks}() < 3.0$  and  $\text{roeP}() < 1.5$  and  $\text{roeNeextra}() < 2.0$

The variables and cuts were determined to make pre-cut sufficiently loose. It means that only little amount of signal candidates should be lost by pre-cut.

First of all, kaonID cut was applied, before applying other cuts. The kaonID cut removed roughly half of kaon candidates which are very likely to be wrongly reconstructed. If the kaonID is not present, the signal candidates are twice more, but the half of them have  $\text{isSignal}=0.0$ . The purity was sacrificed by kaonID cut as the half of signal candidates is removed, but the better candidates with  $\text{isSignal}=1.0$  for are obtained. The other cuts were determined as above. The ROE mask doesn't play important role to make cut, because ROE mask doesn't make better separation between signal and background, but it moves both in the direction of zero. Therefore, it is quite irrelevant, whether one chooses ROE variable with mask or without mask, if we only apply the selection cut. These pre-cuts were tested, before I apply in steering file. It is fairly infeasible to run steering file with full statistics of MC14, if pre-cut is not applied, because memory and storage limit in work environment. I applied these cuts with fraction of full statistics, which approximately amounts to 260k events for signal and 8 million events for background. But, a lot of candidates were already removed by FEI skim cuts. In the table 6.2, it is shown how these cuts affect the number of candidates, figure of merit and purity that are shown in (6.8) and (6.11).

In this table, the cuts are additive from left to right. At the end, where upsilon cut was added, all of cuts are applied and highest FOM and purity are achieved. Efficiency for

	kaonID	Signal cut added	Tag cut added	Upsilon cut added
Sig cand	4834	4609	4294	4077
Bkg cand	5307608	72886	33676	14390
FOM	$2.2938 \cdot 10^{-6}$	$1.8948 \cdot 10^{-5}$	$2.7169 \cdot 10^{-5}$	$3.9638 \cdot 10^{-5}$
Purity	0.091%	5.95 %	11.31 %	22.08 %

Table 6.2: Selection cuts remove the background candidates. After all cuts, 99.73% of background candidates are removed. FOM and purity increases as more cuts are added.

FOM is calculated for correct reconstruction of both signal and tag side. After all cuts are applied, 99.73% of background candidates are removed. But, remaining background still lowers purity. For the next step of removing background, all cuts are moved in steering file to be pre-cut. Since pre-cuts reduce size of output root file, the statistics can be now increased to maximum. Then new output file is used in best candidate selection and fBDT for higher purity to be accomplished.

Usually, one try to maximize the FOM by removing background in order to accomplish best sensitivity. Conventional formulations of FOM [28] look simplified as (6.7).

$$a) \frac{S}{\sqrt{B}} \quad \text{or} \quad b) \frac{S}{\sqrt{B+S}} \quad (6.7)$$

S is the number of signal candidates and B is the number of background candidates. However, this formulation offers limited merit to obtain useful sensitivity. a) is simplest formulation, but it can cause divergence problem(infinite value), as  $B \rightarrow 0$ . It overestimates sensitivity at low background [28]. This could be often problematic, because analyst endeavors to remove the background candidates as much as possible. This divergence problem is shunned in b) by adding S. Then this formulation looks a bit similar to purity and has same demerit as purity. It is not clear how much the value b) should be maximized. One can intentionally maximize it up to 99% or even more as loosing massive signal candidates and reconstruction efficiency. Another problem which occurs, if the actual signal fraction of real data is not known. In MC data, signal fraction is set arbitrarily high to make analysis easier. Thus, the value b) and purity are simply influenced by the way how much signal sample is produced in MC data. The cross section of real data is not concerned.

In this analysis, a different useful formulation of FOM (6.8) was exclusively used from [28] to have more advantages of analysis.

$$FOM = \frac{\epsilon(t)}{a/2 + \sqrt{B(t)}} \quad (6.8)$$

$\epsilon(t)$  is reconstruction efficiency and  $a$  is the number of sigma.  $a = 3$  is used to calculate FOM.  $B(t)$  is the number of background candidates after all cuts.  $t$  is defined as set of all cuts which are applied before calculating FOM. The equation (6.8) is quite different from

usual FOM, as it was calculated by using efficiency. This formulation can prevent from losing efficiency massively, when the FOM is being maximized. Background  $B(t)$  should be removed as much as possible, but efficiency  $\epsilon(t)$  should remain as moderate value at the same time.

The formulation of purity is explained in [17]. The signal purity is given as

$$p = \frac{N_S(1 - \alpha)}{N_S(1 - \alpha) + N_B\beta} \quad (6.9)$$

The  $N_S$  and  $N_B$  are the number of signal and background candidates respectively without applying any cut.  $\alpha$  is the fraction of signal, which is wrongly classified as background, and  $\beta$  is the fraction of background, which is wrongly classified as signal [17].  $\alpha$  and  $\beta$  can be reformulated as

$$1 - \alpha = \frac{N_S(t)}{N_S} \quad , \quad \beta = \frac{N_B(t)}{N_B} \quad (6.10)$$

$t$  is again the set of all cuts that are applied in signal or background sample. By combining the two equations of 6.9 and 6.10, the signal purity is now simply as

$$p = \frac{N_S(t)}{N_S(t) + N_B(t)} \quad (6.11)$$

### 6.3.2 Best candidate selection

Before doing best candidate selection, the all pre-cuts are implemented and statistics is increased up to entire MC14 dataset, which amounts 50 million events for signal and 5.687 billion events( $1 ab^{-1}$ ) for background. The two informative variables for best candidates selection were chosen for plots 6.12a, 6.12b. In the left plot 6.12a, one can find that one

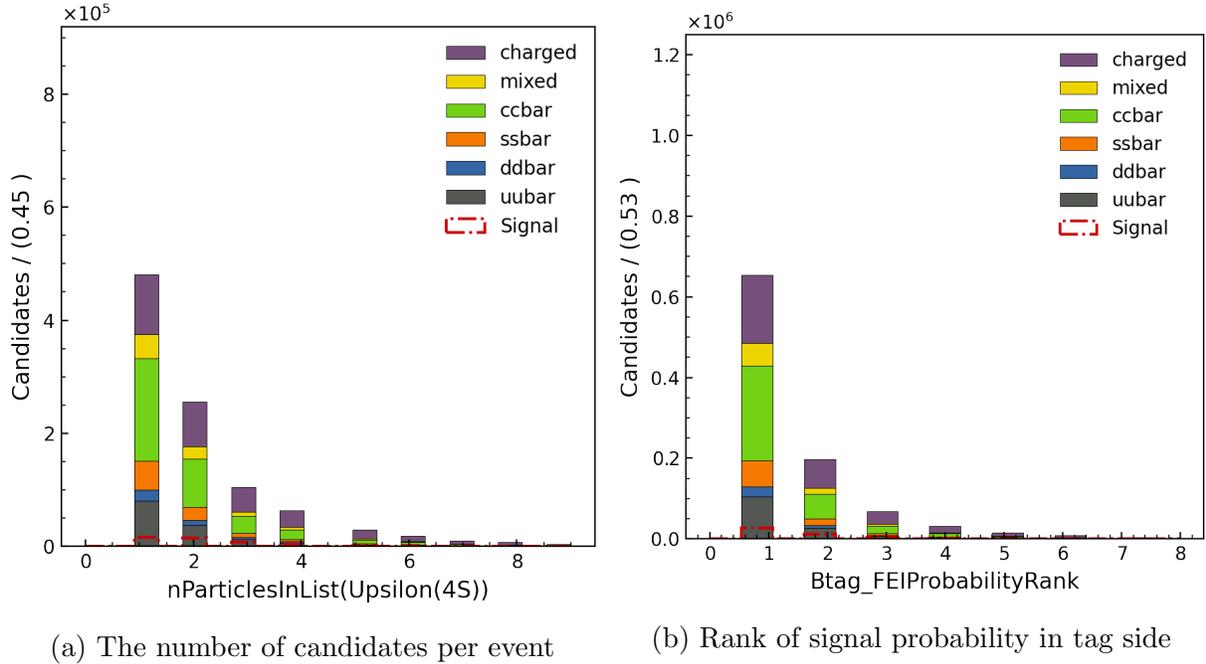


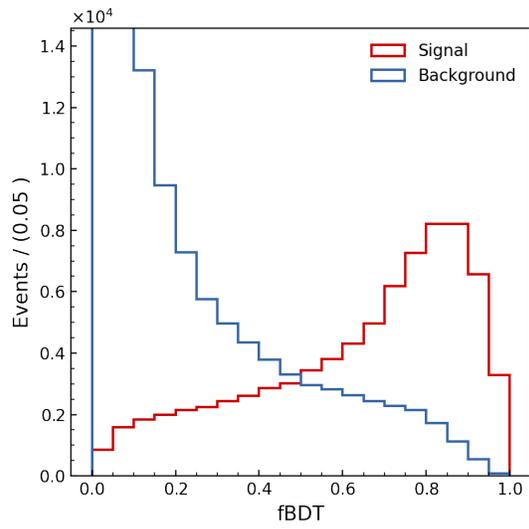
Figure 6.12: There are still multiple candidates per event. Only one candidate per event can be selected by signal probability rank.

event produces many candidates. Because not all candidates can be the particle that we are searching for, only one candidate per event should be selected by some criteria. In right plot 6.12b, the candidates are ranked by signal probability of tag B. The lower value of the plot means that the candidates have higher signal probability. By assessing this rank, one candidate per event with highest signal probability is selected. Before selecting the best candidates, there were 47290 signal candidates and 983334 background candidates. The signal candidates are simply from the signal MC data. After the best candidate selection, there are 27626 signal candidates and 671850 background candidates. However it makes more sense that we have one candidate per event of tag B to search for signal.

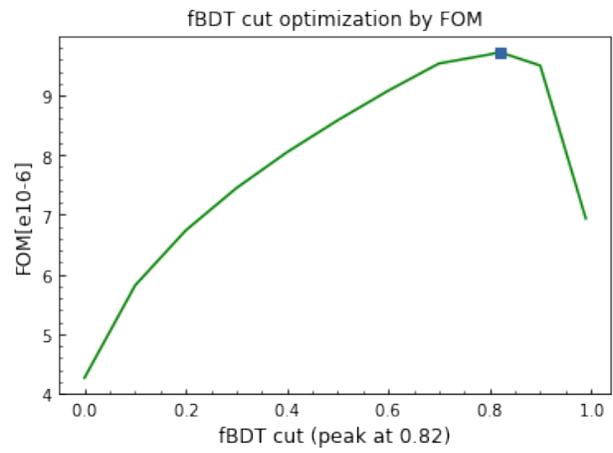
### 6.3.3 fBDT cut

After making pre-cut and best candidate selection, one can make cut on the output of fBDT training. Since the most variables with good discrimination were already applied as pre-cut, the continuum suppression variables become the input variables in fBDT training. The shape of these variables which became input of fBDT are shown in section 6.1.2 and the principle and the result of fBDT is explained in section 7.1. The fBDT is applied to find final cut, which removes more background candidates.

In figure 6.13a, the output distribution of fBDT is shown. In this plot, the signal and background are classified and have separation. By making a cut on this plot 6.13a, we can finally obtain more pure signal sample as many background candidates are purged by cut. Determination of cut requires some criteria, which depend on goal of analysis. If the highest purity is demanded, the cut move as near as possible to 1.0 with sacrificed efficiency. By the cut  $fBDT > 0.991$  the number of background candidates can be zero. However, it would be more desirable to use FOM in equation 6.8 as criteria to determine fBDT cut for saving efficiency. In figure 6.13b, the FOM is plotted with fBDT cut. As the plot shows, the FOM doesn't increase unceasingly, it reaches peak and decreases. This enables determination of cut on the peak where FOM is maximal with certain cut. The cut near 1.0 maximizes the purity, but FOM is not maximum as the efficiency drops. In the result of fBDT training, the FOM reaches peak by the cut  $fBDT > 0.82$ . However, here the target variable for label is `isNotContinuumEvent`. It doesn't distinguish between signal and other B meson pair. Thus, the other B meson pairs are included in signal. To obtain correct signal sample, the fBDT output result is combined in previous input dataframe and redefine signal as `isSignal` of tag side equals 1.0. Then we can also remove wrongly reconstructed signal by tag side. There were 16953 signal candidates and 155539 background candidates before fBDT cut. Then there are finally 7699 signal candidates and 18594 background candidates after the cut  $fBDT > 0.82$ . This number of candidates makes 29.28% purity. This count is calculated only by the test dataset of fBDT, which has smaller size of data, because test dataset is not used in training and is suitable for further analysis. The efficiency is 0.01584% after cut  $fBDT > 0.82$ . By this fBDT cut, the continuum is mostly removed, because the input variables are continuum suppression variables. The B meson pairs are not effectively removed by this fBDT cut. The B meson pairs should have been distinguished and removed by pre-cut, for instance, `missingEnergyOfEventCMS` cut. Because it can distinguish signal with two neutrinos from other B meson pairs, which don't produce the neutrino.



(a) fBDT output distribution



(b) fBDT cut is chosen with maximum FOM.

## 6.4 Fitting

The input data for fitting has the same number of events as in section 6.3(260k events for signal and 8 million events for background).The same pre-cut are applied on the input data, but only one cut of `missingEnergyOfEventCMS` is removed for fitting. This is reason why there are a bit more signal candidates and its shape of the distribution is different. I used `Roofit`[33] as a framework for an unbinned maximum likelihood fit. At first, I fit the signal and background sample separately with different probability density functions. The plot of fit looks as in figure 6.14.

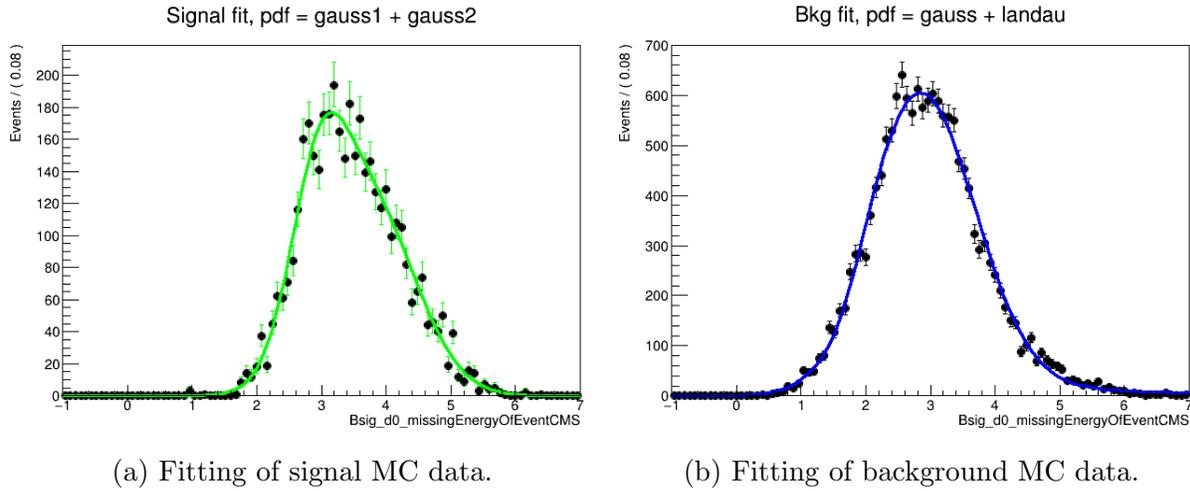


Figure 6.14: Fitting for signal and background by roofit.

Since applying only one probability density function cannot make precise fit, two functions are applied for each signal and background. Then the curve of fit can become more flexible to adjust to input data. Two Gaussian functions are summed for signal, and one Gaussian and one landau function are summed for background. The Gaussian function is basic probability density function defined as

$$g(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \quad (6.12)$$

The  $\mu$  is mean value and  $\sigma^2$  is variance. These two parameters have to be optimized during fit process. The landau function looks as

$$p(x; \mu, c) = \frac{1}{\pi c} \int_0^\infty e^{-t} \cos\left(t \left(\frac{x - \mu}{c}\right) + \frac{2t}{\pi} \log\left(\frac{t}{c}\right)\right) dt \quad (6.13)$$

The mean and variance are not defined as Gaussian function, but the analogous parameters are used. The  $\mu$  is the location parameter and  $c$  is the scale parameter. Both parameters have certain ranges as  $c \in (0, \infty)$  and  $\mu \in (-\infty, \infty)$ . These parameters are optimized

well by fit, and now the parameters are fixed before making composite fit of signal and background. Exceptionally, sigma and scale of background functions are not fixed and can float. Fixing the parameters averts screw-up of searching for optimal parameters during composite fitting of signal and background. Additionally, the proportions of each probability density functions in composite fit is also fixed. In tables 6.3 and 6.4, the results of the parameters of signal and background fit are shown.

	Value	Error
Proportion	$6.896e^{-1}$	$5.36261e^{-2}$
Mean of Gauss1	2.947	$3.82018e^{-2}$
Mean of Gauss2	3.698	$5.00640e^{-2}$
Sigma of Gauss1	$4.343e^{-1}$	$3.19250e^{-2}$
Sigma of Gauss2	$7.351e^{-1}$	$1.61256e^{-2}$

Table 6.3: The parameters of the fit result of signal. The proportion is the fraction of second Gaussian function.

	Value	Error
Proportion	$1.638e^{-1}$	$1.37085e^{-2}$
Mean of Gauss	2.914	$1.44904e^{-2}$
Location of Landau	2.558	$8.13646e^{-2}$
Sigma of Gauss	$8.125e^{-1}$	$9.03393e^{-2}$
Scale of Landau	$4.216e^{-1}$	$3.97969e^{-2}$

Table 6.4: The parameters of the fit result of background. The proportion is the fraction of Landau function.

The proportion means that the portion of first probability density function, when the two probability density functions are summed. The functions are summed in a way as

$$S = fg_1(x) + (1 - f)g_2(x) \quad (6.14)$$

$g_1$  is the first probability density function and  $g_2$  is the second probability density function.  $S$  is the summed probability function of  $g_1$  and  $g_2$ .  $f$  determines the proportion(fraction) of the probability density function. This principle is same in all cases that I summed two probability density functions. Now the fit has to be executed jointly by both signal and background such as real data. Then I combine again the two probability density functions of signal and background to make final fit. The plot of composite fit is shown in figure 6.15. The resulting parameters of composite fit are shown in table 6.5.

From the table 6.5, one can see the estimated signal and background yield from fit. Actual signal and background yields from input data is 4099 for signal and 15805 for background. The fit result deviates slightly from the correct value of input data, but it is acceptable approximately within 1 sigma. Plus, the both sigma of Gauss and Scale of

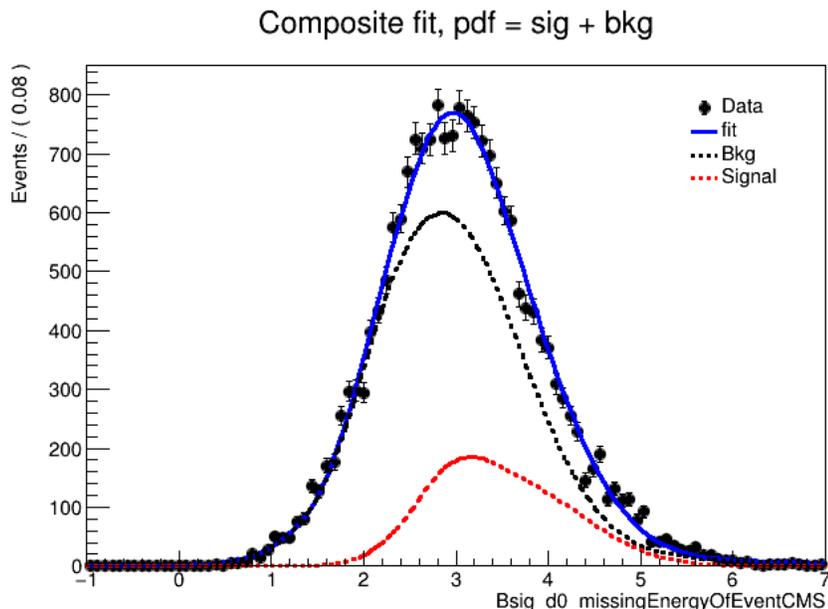


Figure 6.15: Composite fitting of signal and background for missing energy(GeV) in center-of-mass frame.

	Value	Error
Signal yield	$4.3042e^{+3}$	$2.21206e^{+2}$
Background yield	$1.55937e^{+4}$	$2.45162e^{+2}$
Sigma of Gauss(background)	$8.108e^{-1}$	$8.59253e^{-3}$
Scale of Landau(background)	$4.086e^{-1}$	$3.29244e^{-2}$

Table 6.5: The parameters in composite fit result. Signal and background yields are extracted the fraction of each data from fit.

Landau decreased as they float during the composite fit. The shape of the fit could be slightly different, if these two parameter didn't float.

It is necessary to examine the quality and validation of the fit. One way is calculating residual and pull, then one can make plot with the residual and pull values in RooFit.

The residual is simply the difference between actual value and fit value for each bin. Pull is also defined for each bin as

$$Pull = \frac{N_{fit} - N_{data}}{\sigma} \quad (6.15)$$

$N_{fit}$  is value of fit and  $N_{data}$  is actual value of input data for that bin. And  $\sigma$  is statistical uncertainty of each bin. In ideal case of the fit, approximately 68% of all residual are within 1 sigma around 0 and rest is not. By this residual distribution, the 68% of the all pull is in the range  $[-1, 1]$ .

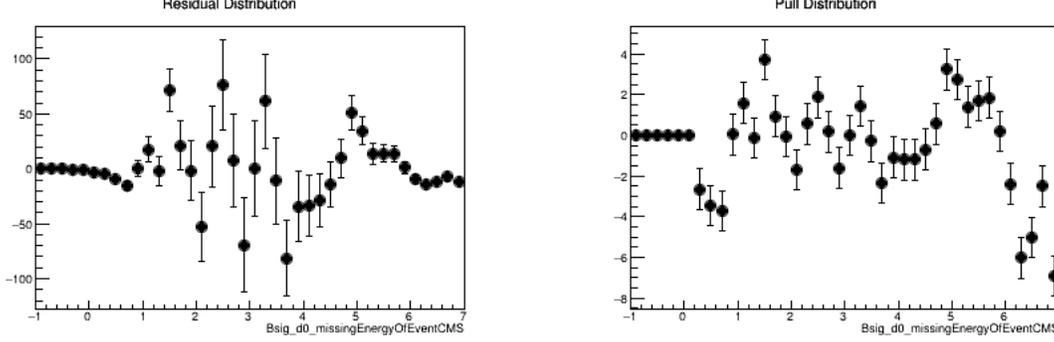


Figure 6.16: Residual and pull distribution from composite fitting result.

Another way for examining validity of fitting is calculating value of  $\chi^2$ . Its definition and resulting value from fit are shown in equation 6.16.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = 7.234 \quad (6.16)$$

$O_i$  is fit value and  $E_i$  is actual value of input data. There are four floating parameters of fit and 40 bins are given. This value of  $\chi^2$  is divided by the number of degrees of freedom. Thus, the ideal value of this  $\chi^2$  calculation should be close to 1 as a good fit.

It is also feasible to calculate branching fraction with the signal yield of fit. The formula of the branching fraction looks as

$$Br = \frac{N_{fit}}{2f_{+-}N_{BB}\epsilon} \quad (6.17)$$

$N_{fit}$  is the value of the signal yield from fitting and  $N_{BB}$  is the number of events of B meson pairs without cut.  $\epsilon$  is the reconstruction efficiency, and  $f_{+-}$  is multiplied, because the decay is now considered only from charged B meson pair. The number of events of neutral B meson pair should not be in calculation.  $f_{+-}$  is 0.5 here for MC data analysis. In MC data, one can only take the number of events from charged B meson pair in calculation. In many previous result of experiment, the branching fraction limit is known, for example, a strict limit was found in [3].

$$Br(B^+ \rightarrow K^+\nu\bar{\nu}) < 1.6 \times 10^{-5} \quad (6.18)$$

From this branching fraction limit, we can calculate limit of signal yield. I insert the efficiency value from the input data of the fitting, in which the efficiency for correct reconstruction of both signal and tag side is 0.0048409. The  $N_{BB}$  is  $471 \times 10^6$  in [3] and I use  $f_{+-} = 0.5$ . Then the limit of  $N_{fit}$  can be calculated as

$$N_{fit} < Br \cdot 2f_{+-} \cdot N_{BB} \approx 41 \quad (6.19)$$

---

This value gives the limit of the signal yield from the real data. From this result, we can realize that the signal fraction of  $B^+ \rightarrow K^+ \nu \bar{\nu}$  of real data is much lower than MC data, because we obtained  $N_{fit} = 4304$  from the fitting of MC data, even though the real data has much more the number of events than the input MC data. It indicates that analysis and fitting of real data would be much more challenging than MC data, because of extreme rareness of the signal sample.

# Chapter 7

## Machine Learning

In the analysis of rare decays, it is fairly arduous to search for variable one by one, because one cannot know beforehand, which variable has good discriminating power between signal and background. To save the energy and time of analyst, *Multivariate Analysis*(MVA) is contained in basf2, in which the multiple variables are processed and used to distinguish signal from background by exploiting the correlation between variables. In this analysis, I used fBDT, SVM and neural network as shown in this section. The fBDT is already core part of basf2 and widely used in Belle II research. I have run fBDT to make final cut at the end of analysis. I used SVM to produce feature importances to search variables for cuts. I have also run neural network(Multilayer Perceptron) along with SVM. The neural network is more modern method, therefore its application and utility should be researched further at Belle II.

### 7.1 fast Boosted Decision Tree(fBDT)

FastBDT is stochastic gradient-boosted decision trees with speed optimization [19]. Its training and evaluation are well developed within basf2. The principle of decision trees for classification is seemingly akin to making selection cut. But, it differs from the fact that decision tree works with multiple variables and cuts are applied many times within one variables(feature). A schematic overview of decision tree is shown in figure 7.1.

The input variables of fBDT are continuum suppression variables including KSFV variables, which are shown in section 6.1.2. These could come from signal side, tag side and `isSignal`. After testing fBDT training of all variables, tag side variables have highest score and discrimination for classification than signal and `isSignal` variables. As a result of that, tag side variables were chosen for fBDT training.

Since the process of fBDT is supervised learning, the data has to be labeled. Here `isNotContinuumEvent` is used as target variable to label, instead of `isSignal`, because continuum suppression variable cannot distinguish well between signal and background

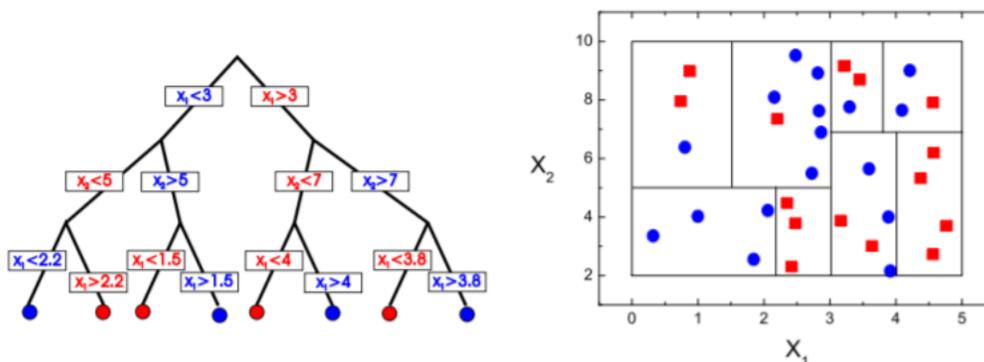


Figure 7.1: The left image is visualization of decision trees with two features  $x_1$  and  $x_2$ . The right image shows applied decision trees on two dimensional space. Each line on the space corresponds to a cut of tree. [29].

fBDT parameters	value
The number of tree	200
The number of level	3
Shrinkage	0.1
The number of cuts	4

Table 7.1: The hyperparameters of fBDT training.

from B meson pair. The hyperparameters of fBDT training is used as shown in table 7.1.

The higher number of trees and level can make classification more precise, but it has to be adjusted moderately, because high value can cause overfitting. Shrinkage makes some parameter closer to 0 during training so that the training model becomes less complex. Shrinkage(learning rate) should be lowered sufficiently to get precise classification, but lower value requires higher training time. The training time was short with the low shrinkage due to speed optimization of fBDT. The result of fBDT training is shown below. At first, ROC curve of fBDT training is shown in figure 7.2.

The test dataset is assigned as 30% of whole dataset. The test dataset has to be used for further analysis, because the training dataset can be biased after the training. The test and training dataset are shown separately, but both results have almost identical curves. It means that the overfitting was avoided successfully. In figure 7.3, the difference between test and train dataset is plotted for both signal and background. The difference between test and train dataset is quite small as it also confirms that the result is reasonable to be used for making cut. Signal is distributed near 1.0, as target variable `isNotContinuumEvent` is given to signal. The process and result of making cut on fBDT distribution is elaborated in section 6.3.3.

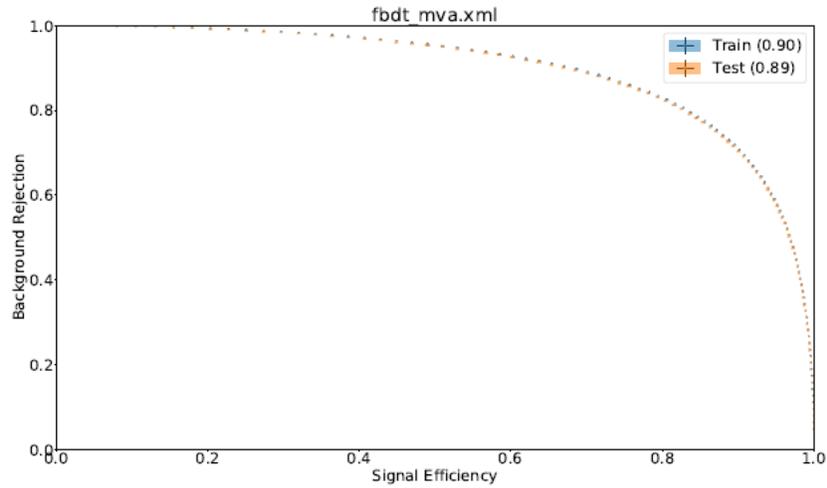


Figure 7.2: ROC curve and score of test and training dataset. The higher AUC (Area under the ROC Curve) implies better performance in classification.

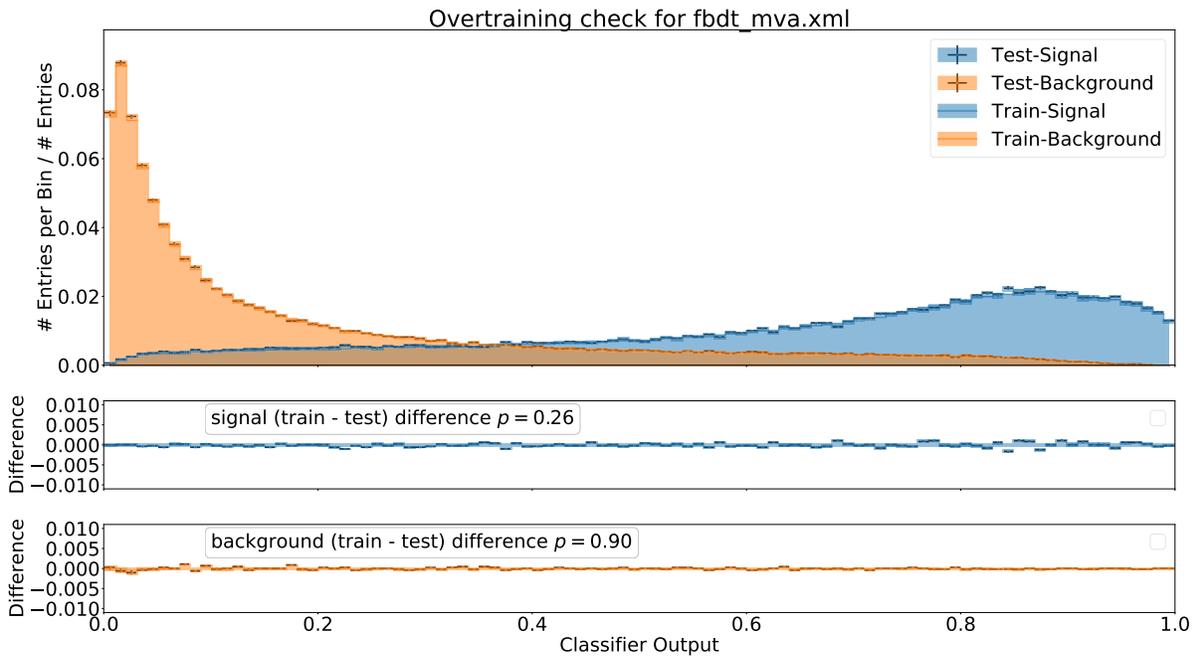


Figure 7.3: The overtraining plot with signal and background distribution.

## 7.2 Support Vector Machine(SVM)

Support vector machine(SVM) is a classical method of machine learning for classification. It works especially well for binary classification, which is often the case of particle physics(signal and background).

The classification works by splitting data with hyperplane in multidimensional space as shown in figure 7.4. In particle physics, one side of hyperplane would be classified as signal and other side becomes background. The support vectors adjust the position of hyperplane so that the distance between two different types of data becomes maximized, which leads to better quality of classification. In 2D space, hyperplane is simply line, which is one dimensional object. And in 3D space, hyperplane is 2 dimensional plane. As this way generalizes, in  $n$  dimension, hyperplane is  $n-1$  dimensional object and it cannot be visualized anymore.

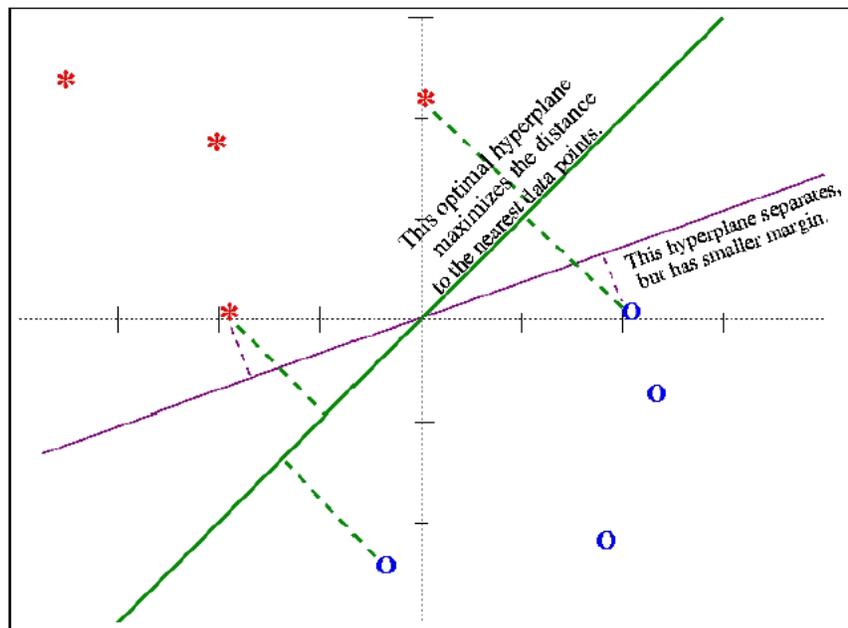


Figure 7.4: Hyperplane of SVM [30].

In many cases of classification in multidimensional space, the hyperplane cannot easily split data, if the separation in dataset is not large. To solve this problem, *kernel* method was developed. The kernel maps the data in different space as shown in figure 7.5. In the new space, it is possible that the data has better separation to be classified. There are many types of kernels and the dimension of new space is determined by type of kernel. In this analysis, Radial basis function(RBF) kernel is used, which maps the data in infinite dimensional space. Equation 7.1 shows the form of RBF kernel.  $\gamma$  is hyperparameter of

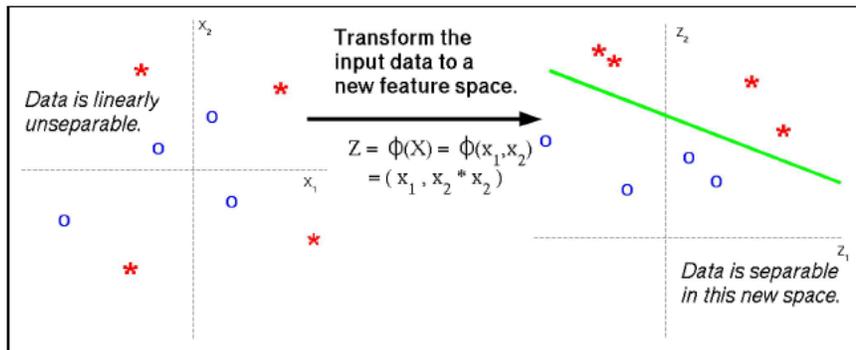


Figure 7.5: Kernel trick of SVM [30].

SVM and  $\|\mathbf{x} - \mathbf{x}'\|^2$  is distance between two feature vectors.

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2) \quad (7.1)$$

How the RBF kernel maps data in infinite dimensional space can be shown intuitively by making expansion of RBF kernel as follows.

$$\begin{aligned} \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|^2\right) &= \exp\left(\frac{2}{2}\mathbf{x}^\top \mathbf{x}' - \frac{1}{2}\|\mathbf{x}\|^2 - \frac{1}{2}\|\mathbf{x}'\|^2\right) \\ &= \exp(\mathbf{x}^\top \mathbf{x}') \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \exp\left(-\frac{1}{2}\|\mathbf{x}'\|^2\right) \\ &= \sum_{j=0}^{\infty} \frac{(\mathbf{x}^\top \mathbf{x}')^j}{j!} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \exp\left(-\frac{1}{2}\|\mathbf{x}'\|^2\right) \end{aligned} \quad (7.2)$$

By expansion of RBF kernel, the sum runs over the infinite number of inner product with index  $j$ . Because the each inner product term is polynomial kernel, the RBF kernel is sum of infinite number of the polynomial kernels. Each polynomial kernel has finite dimension and all polynomial together contribute to infinite dimension of RBF kernel by sum. More strict proofs of theoretical foundation have to be found in the reference of mathematics. From now on, the result with MC14 input data will be shown. I labeled the data as signal, if it is signal MC data. And the rest of generic MC data is labeled as background. First of all, correlation matrix of input features is made in figure 7.6. It displays how the input features are similar with each other. Strongly correlated features contribute to classification with similar amount. Typically, ROE variables without ROE mask and same ROE variables with ROE mask have very high correlation, because only position of peak is shifted by ROE mask. And `roeE`, `roeP` and `roeM` have maximum correlation. Therefore, one of them could be removed to avoid redundant classification and high memory and storage usage.

Another heat map in figure 7.7 shows an overview of hyperparameters of SVM.  $C$  determine how much we forbid wrong classification. If the hyperplane split the datas, it is inevitable that some fraction of datas are classified as wrong class. Low  $C$  makes the

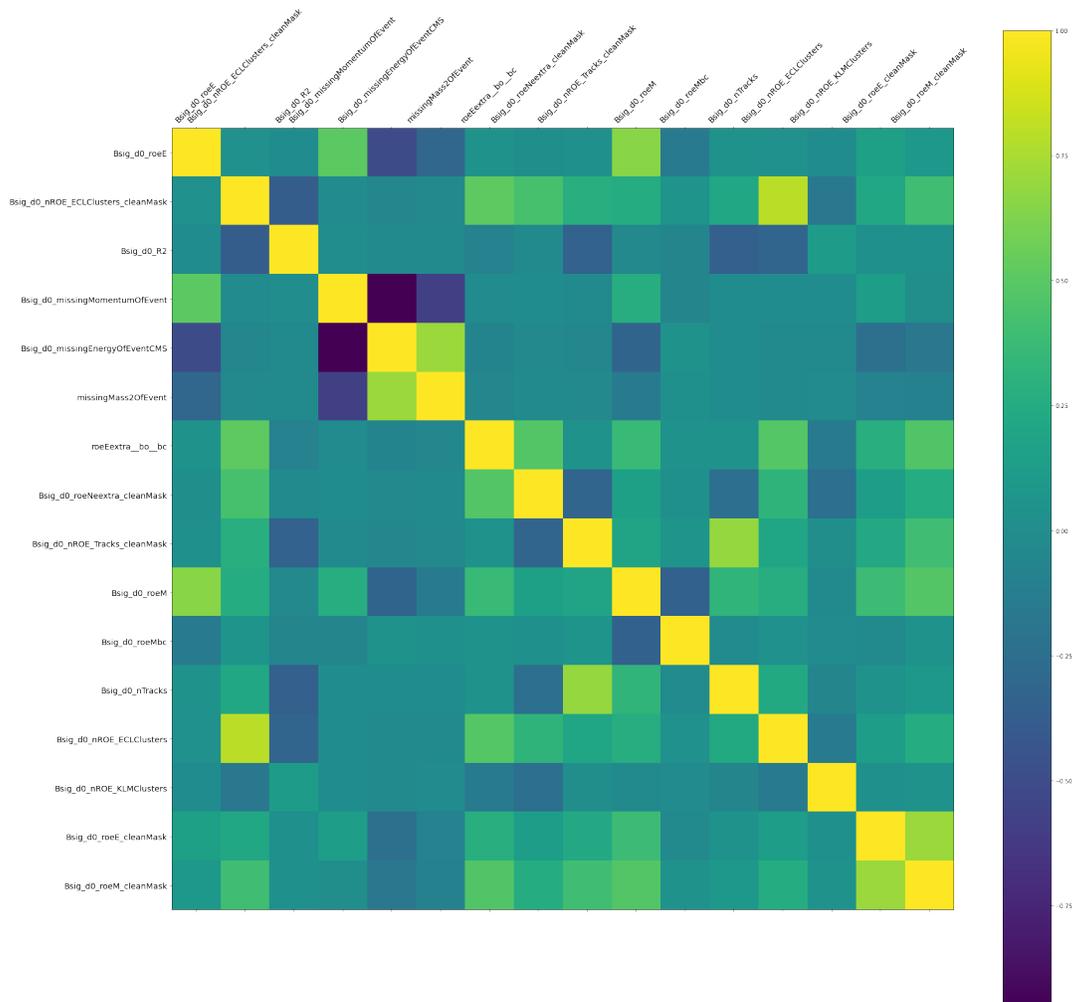


Figure 7.6: Correlation matrix of input variables. The variables which have maximum correlation(1.0) with each other were removed.

training more tolerant with such wrong points of datas. And gamma makes the hyperplane more flexible and closer to datas. The higher score is feasible with high gamma, but high value of gamma can cause overfitting more easily. This gamma can be interpreted as analogy of the number of tree and level in fBDT, even though it has geometrically different shape.

In scikit-learn, permutation importance is defined as follows. For each feature  $j$  of the dataset, the dataset is shuffled randomly and one obtains corrupted(shuffled) dataset. The shuffle repeats with certain number( $K$ ). One can choose the number of repetition. The shuffled dataset would have low score( $s_{k,j}$ ) as it contributes worse in machine learning training. Then the difference of scores( $s, s_{k,j}$ ) by a feature between original dataset and

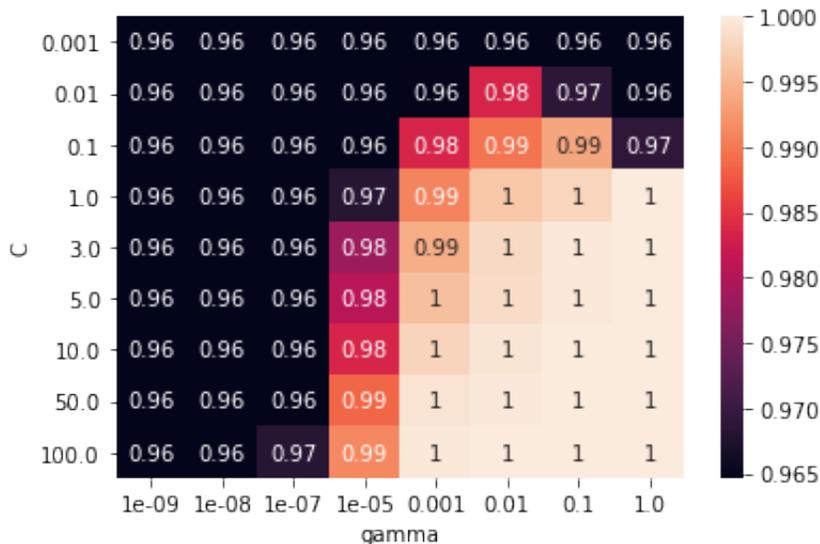


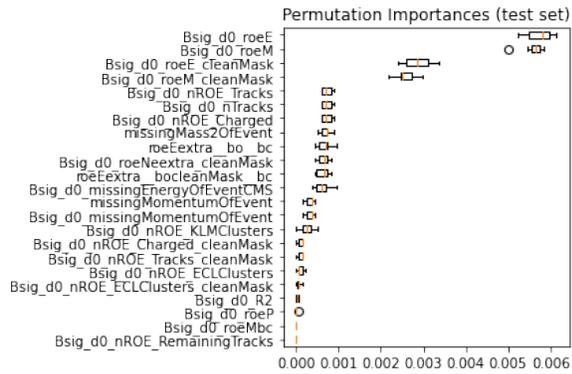
Figure 7.7: Overview of hyperparameters and scores with some data for test. The statistics was reduced to calculate many times with short run time. The higher C and gamma tend to make higher scores.

shuffled dataset is defined as importance.

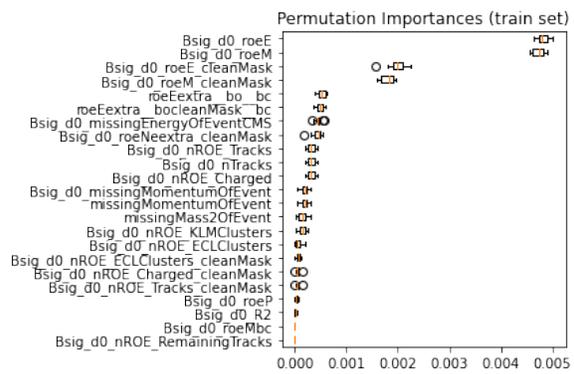
$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j} \quad (7.3)$$

If there was bad separation between signal and background, the difference of scores(importance) would be small, because original feature has already low score( $s$ ). Therefore, the separation between signal and background is very likely to be proportional to permutation importance. To make reasonable analysis, the number of shuffles( $K$ ) is 10 in this training. The high number of shuffles increases runtime and CPU usage very greatly, because it has to shuffle all input features of training repeatedly. Because of this reason, the 260k events for signal and 8 million events for background are used as input data, which are not full statistics of MC14 dataset. Nevertheless, the output result of feature importance in figure 7.8 gives reasonable information about which variables have good discriminating power. Because it is supervised learning, the data has to be labeled. All type of background datas were combined as one dataframe and labeled as 0. And signal MC data was labeled as 1. The target variable is not used to label the data.

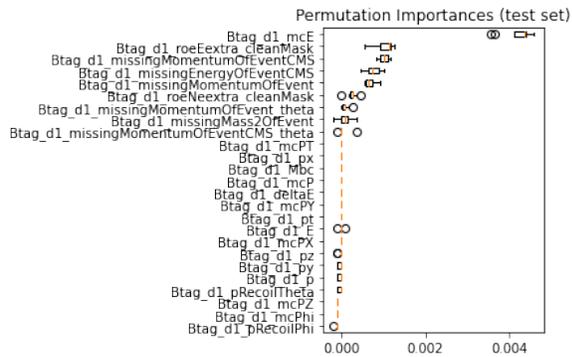
By using reduced statistics(260k events for signal and 8 million events for background), it is more convenient to run many times in short time until we find the variables with good discrimination between signal and background. SVM and Neural network that I applied don't have speed optimization yet unlike fBDT. However, the feature importance is dependent on the machine learning method that was used to calculate, it cannot perfectly be proportional to actual separation of variables. The importances can give different values, when hyperparameter or machine learning method itself is changed. To solve this problem



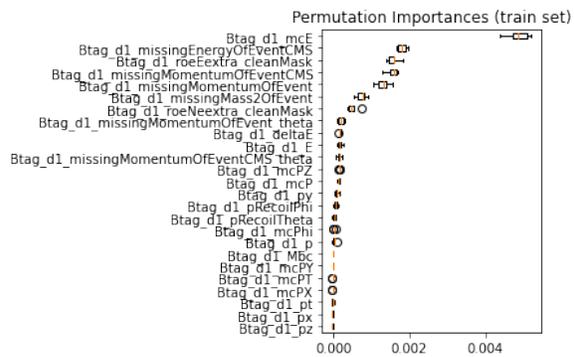
(a) Signal side features of test set



(b) Signal side features of train set



(c) Tag side variables of test set



(d) Tag side variables of train set

Figure 7.8: Permutation feature importances of SVM

and make the result more objective independent of certain machine learning training, I have run neural network with same condition and produced the feature importances.

The result of this SVM training as follows. Score and AUC score are always 0.99 for both signal and tag side. Test dataset is assigned as 25% of whole dataset. The test and train dataset have same score(0.99). And the feature importance result is very similar in test and train set. It means that overfitting is avoided successfully. The hyperparameter  $C = 1.0$  is set for all and gamma is 'scale'. This gamma in the case of 'scale' is defined in scikit-learn as equation 7.4.

$$\text{gamma}(= \text{scale}) = \frac{1}{n\_features \cdot \text{Var}(X)} \quad (7.4)$$

$n\_features$  is simply the number of features and  $\text{Var}(X)$  is variance.

## 7.3 Neural Network

Neural network is implemented with same input data of SVM and this can give comparable result. The 260k events for signal and 8 million events for background are used as input data, which is same in SVM training. The Multi-layer Perceptron(MLP) of scikit-learn is applied for classification. The diagram that illustrates the mechanism of MLP is shown in 7.9.

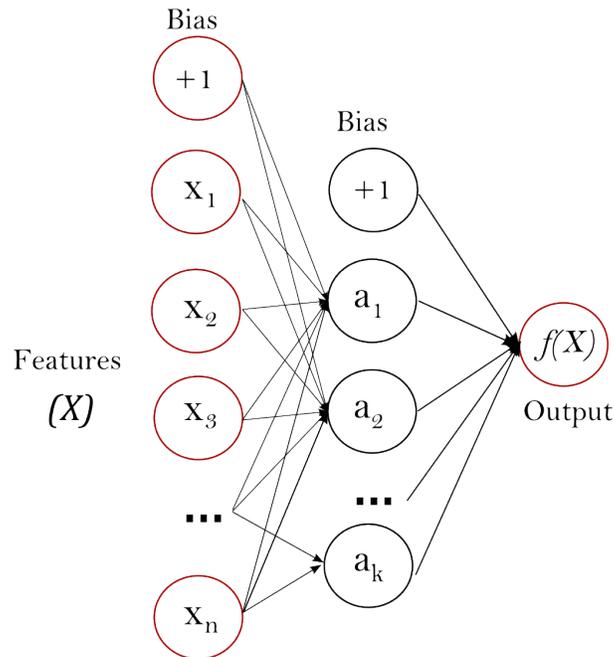


Figure 7.9: A schematic diagram of Multi-layer Perceptron (MLP) with one hidden layer from scikit-learn library [38].

The perceptron is connected to other perceptrons of previous layers with a value which is called *weight*. The sum of all weights in one perceptron becomes input value in an activation function. The activation function produces output value and transfer it to the next layer. After it repeats in all hidden layer, the output is calculated. The dimension of  $\mathbf{X}$  in the final output function  $f(\mathbf{X})$  can differ from the dimension of the input features. The final output function has to pass logistic function to be value between 0 and 1 in binary classification, which squeezes the value between 0 and 1. In more general case, where many classes exist, the output function has to pass *softmax function* that are general version of the logistic function as

$$\text{softmax}(z)_i = \frac{\exp z_i}{\sum_{l=1}^k \exp z_l} \quad (7.5)$$

$l$  corresponds to class and  $k$  is the total number of classes. In neural network, the weight and hyperparameters are adjusted by minimizing the loss function. The loss function

makes output for measurement of how much the prediction of neural network training deviates from actual value. The lower value of loss function output means more precise prediction. There are several different ways to make this process. In Stochastic Gradient Descent(SGD), the neural network updates weights minimizing the Cross-Entropy loss function. The Cross-Entropy loss function looks as follows in binary classification.

$$Loss = -(y \log(p) + (1 - y) \log(1 - p)) \quad (7.6)$$

$p$  is predicted output of neural network and  $y$  is actual binary indicator(label). In Limited-memory BFGS(L-BFGS), the Hessian matrix is approximated to minimize the loss function and optimize the hyperparameters. The first-order partial-derivative of multivariate function is simply gradient of the function. In second-order partial-derivative of the function, there are a lot of components and all possible values can be written as the components of the *Hessian matrix*, which is shown in following equation.

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \quad (7.7)$$

However, calculating the all components of this Hessian matrix demands challenging computing power. Therefore, the L-BFGS uses the method to approximate the Hessian matrix and find descending direction for the optimization. The detailed mechanism of L-BFGS is elaborated in [32]. The epoch is the number determining how many times the training data passes the neural network. The higher epoch can produce better quality of classification, but it should be moderate value to avoid overfitting. The resulting hyperparameters and condition of the MLP training is shown in table 7.2.

First of all, ReLU function is chosen as activation function, which always produces zero as output value for negative input value. The parameter alpha is analogous to gamma of SVM, which can be adjusted to avoid overfitting. The low alpha is more prone to incur overfitting. The hidden layer sizes as (15,) implies that MLP has one hidden layer with 15 units. The solver L-BFGS is chosen more often in smaller dataset. For larger data, SGD(stochastic gradient descent) is often chosen as the solver. The tol stands for *tolerance for optimization*. The tol determines convergence in which the minimal loss or maximal score is reached. The convergence is defined as a point where the loss or score is not being improved for the `n_iter_no_change(10)` iteration. `max_iter` is defined as the number so that the solver iterates until convergence or the `max_iter`. `max_fun` is the maximum number of loss function calls, with which the solver cannot iterate more than this number.

activation	ReLU: $f(x) = \max(0, x)$
alpha	1
hidden layer sizes	(15,)
solver	L-BFGS
tol	0.0001
max_fun	15000
max_iter	200
n_iter_no_change	10

Table 7.2: The hyperparameters and condition of MLP training for both signal and tag side.

Early stopping, which breaks the training before the maximum number of the iteration, is not executed in this training. And many hyperparameters which are only relevant for SGD or Adam solver are not included in the table, because the L-BFGS is applied. The result of the feature importance of MLP training is shown in figure 7.10. The result shows the similar order as SVM shows, but `nTracks`(the number of tracks based on event) has clearly higher importance in MLP than SVM. It has quite good separation between signal and background, but it was not used in pre-cut, because pre-cut is made as loose cut and applying cut on `nTracks` without losing signal candidates was difficult. The MLP training also has score(0.99) and the overfitting is avoided, because the test and train set produced similar result.

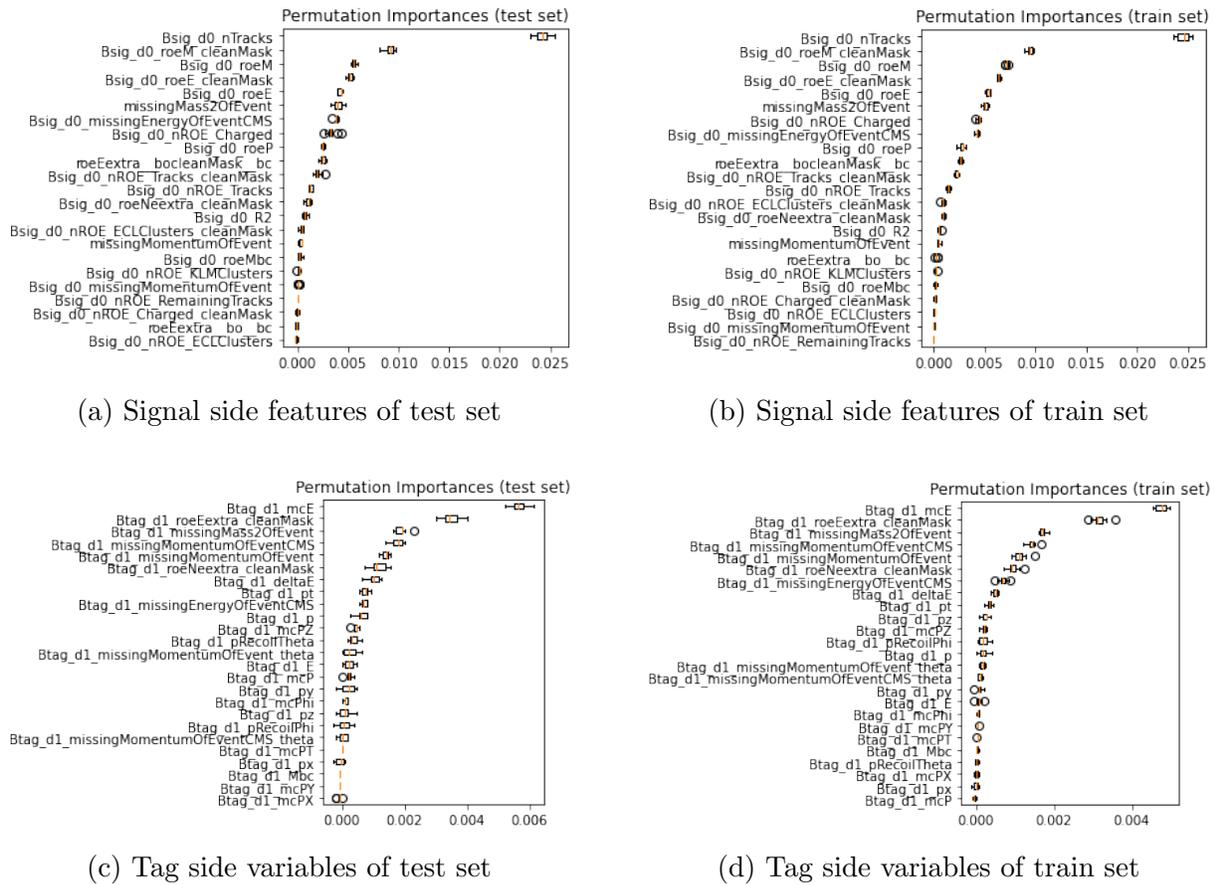


Figure 7.10: Permutation feature importances of MLP

# Summary

In this analysis, the rare decay  $B^+ \rightarrow K^+ \nu \bar{\nu}$  is searched. The crucial method that supports achieving this goal was Full Event Interpretation. As I used FEI skimmed data, many selection cuts were previously applied and diminished size of background, by which the runtime and memory usage could also be reduced during the analyzing the data. Plus, the reconstruction of B meson offered more variables of tag B meson to find selection cut and also to make input data in fBDT.

In general, the continuum could be easily distinguished from signal and ruled out to obtain pure signal sample, because the continuum has distinct physical characteristics and Belle II research already organized the recommended continuum suppression variables very well. But, the background of B meson pairs is harder, because the signal is also produced from the B meson. To distinguish the background of B meson, I had to apply `modularAnalysis.buildEventKinematics`, by which the 'missing' variables can be imported. Using such variables with 'missing' is reasonable, because the signal decay  $B^+ \rightarrow K^+ \nu \bar{\nu}$  produces two neutrinos that cannot be detected in Belle II detector and can be only found with missing physical quantities. But, some variables of event kinematics also don't have distinguishing signal distribution from B meson pair. It is worth that the more event kinematics variables are searched to remove more background candidates of B meson as it implements distinguishing physical quantities of neutrino. The selection cuts including pre-cuts could be compared by examining the purity and FOM. However, as the result of the fitting is shown, the purity cannot be absolute value to measure quality of the analysis, because the MC data has much higher signal sample than real data and this fraction of signal is easily manipulated in MC production process, which doesn't have specific meaning as the fraction of the real data. But, relative value of purity for the comparison of selection can be a intellectual way to examine the quality of the selection cut. The pre-cut couldn't be easily made for high purity, because the pre-cut is intended to be loose for further analysis process and should preserve the signal candidates.

In machine learning, SVM and MLP of scikit-learn were used at first with low statistics to repeat run many time so that the good candidate variable for the selection cut can be found. Unlike fBDT, the SVM and MLP don't have speed optimization for analyzing Belle II data. This was also why the input data of SVM and MLP had lower statistics.

In fBDT, the large dataset( $1 \text{ ab}^{-1}$ ) can be input data in training without causing high runtime. When it comes to the score of ML, the SVM and NN always had higher score than fBDT. Presumably, the different input variable made better classification and higher score with SVM and NN. As shown in the section 6.1, some input variables for the SVM and NN, which were also used as pre-cut, have certainly stronger discriminating power than continuum suppression variables for the fBDT. This is also visible in the distributions of input variables in section 6.1. Therefore, one cannot easily conclude that fBDT made worse performance than other methods. One of limit of the fBDT training in this analysis was that the fBDT classified only to distinguish the continuum from the signal. The background of B meson pair should also be distinguished. It would be instructive to search for different input variables, in addition to `missingEnergyOfEventCMS`, to classify the signal better from other B meson. Then determination of the new target variable and hyperparameter optimization in fBDT would be needed. The search for such variables can demand a measurement of separation between signal and background. For example, this measurement is Jensen-Shannon distance in [5]. In this analysis, I tried to measure the separation with the permutation feature importance of ML. It is desirable that more methods for the measurement of separation are searched.

Lastly, the fitting was performed to estimate signal yield and compare it with the result from the branching fraction limit of the real data. The primary issue is that the real data has extremely tiny signal fraction( $\lesssim 41$ ) compared to MC data. The analyzing such a tiny signal sample appears to be almost infeasible with the current technique. One can import smaller signal sample from MC data for fit to organize a similar condition of real data. The many new technique and methods still have to be searched to analyze in more challenging circumstance.

# Bibliography

- [1] D. Besson and T. Skwarnicki *UPSILON SPECTROSCOPY: Transitions in the Bottomonium System*. Ann. Rev. Nucl. Part. Sci. 43, 333, 1993
- [2] O. Lutz, S. Neubauer, M. Heck, T. Kuhr, A. Zupanc,... *Search for  $B \rightarrow h^{(*)}\nu\bar{\nu}$  with the full Belle  $\Upsilon(4S)$  data sample*. arXiv:1303.3719v3 [hep-ex], 2013.
- [3] J. P. Lees, V. Poireau, V. Tisserand, ... *Search for  $B \rightarrow K^{(*)}\nu\bar{\nu}$  and invisible quarkonium decays*. arXiv:1303.7465v2 [hep-ex], 2013
- [4] J. Grygier, P. Goldenzweig, M. Heck, ... *Search for  $B \rightarrow h\nu\bar{\nu}$  decays with semileptonic tagging at Belle*. arXiv:1702.03224v3 [hep-ex], 2017
- [5] Filippo Dattola, Sasha Glazov, Simon Kurz, Cyrille Praz and Slavomira Stefkova, *Studies of  $B^\pm \rightarrow K^\pm\nu\bar{\nu}$  decay using inclusive tagging method*. BELLE2-NOTE-PH-2020-057, 2021
- [6] C. Cecchi, G. De Nardoy, E. Manoniz, M. Merolax, *Study of  $R_2$  distribution and  $B$  counting in Early Phase 3 Data*. BELLE2-NOTE-PH-2019-025, 2020
- [7] E. Kou, P. Urquijo1, W. Altmannshofer,... *The Belle II Physics Book*. arXiv:1808.10567 [hep-ex], 2019
- [8] J. Grygier, P. Goldenzweig, M. Heck, ... *The Physics of the B Factories*. Springer, 2014
- [9] Z. Dolezal and S. Uno, ... *Belle II Technical Design Report*. 2010.
- [10] L. Andricek, J. Baudot, F. Forti and G. Rizzo, ... *Initial requirements for an upgraded VXD in Belle II*. 2019.
- [11] James Meier Samuel Kahn, *Hadronic Tag Sensitivity Study of  $B \rightarrow K^{(*)}\nu\bar{\nu}$  Selective Background Monte Carlo Simulation at Belle II*. 2019.
- [12] Jo-Frederik Krohn,  *$K_L^0$  identification studies for Belle II*. 2016.
- [13] Oskar Tittel, *SM Compatibility of  $B \rightarrow K^{(*)}\pi$  Measurements and Analysis on  $B^+ \rightarrow K^{(*)}(892)^+\pi^0$  in Belle II*. 2021.

- [14] Alberto Martini, *Measurement of the time dependent asymmetry in the decay  $B^0 \rightarrow K_S^0 \pi^0 \gamma$  using the Silicon Vertex Detector of the Belle II experiment.* 2016.
- [15] Nadia Toutounji, *Reconstruction Methods for Semi-leptonic Decays of B-mesons with the Belle II Experiment.* 2018.
- [16] Markus Röhrken, *The Belle Experiment.* SpringerLink. 2013.
- [17] Thomas Keck, *The Full Event Interpretation for Belle II.* 2014.
- [18] Thomas Keck, *Machine learning algorithms for the Belle II experiment and their validation on Belle data.* 2017.
- [19] Thomas Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment.* Comput Softw Big Sci 1:2. 2017.
- [20] T. Keck, F. Abudinén, F.U. Bernlochner, ... *The Full Event Interpretation An exclusive tagging algorithm for the Belle II experiment.* arXiv:1807.08680v4 [hep-ex], 2019.
- [21] James Kahn, Thomas Kuhr, and Martin Ritter, *Selective background Monte Carlo simulation at Belle II.* J. Phys.: Conf. Ser. 1525 012089, 2020
- [22] Matthew D. Schwartz, *Quantum Field Theory and the Standard Model.* Cambridge University Press, 2013
- [23] Y. Grossman, *Introduction to flavour physics.* Cornell University, 2010
- [24] Jure Zupan, *Introduction to flavour physics.* arXiv:1903.05062v2 [hep-ph], 2019
- [25] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, N. Braun, *The Belle II Core Software.* 2018
- [26] Yo Sato, *Introduction to the analysis package.* Belle II Starter Kit workshop, 2020
- [27] Xingyu Zhou, Shuxian Du, Gang Li, Chengping Shen *TopoAna: A generic tool for the event type analysis of inclusive Monte-Carlo samples in high energy physics experiments.* Computer Physics Communications, 2021
- [28] Giovanni Punzi, *Sensitivity of searches for new signals and its optimization.* PHYS-TAT2003, SLAC Sep. 8-11, 2003
- [29] Gerhard Bohm, Günter Zech *Einführung in Statistik und Messwertanalyse für Physiker.* 2006
- [30] Dustin Boswell, *Introduction to Support Vector Machines.* 2002

- 
- [31] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, volume 12, 2011
- [32] Yunhai Xiao, Zengxin Wei, Zhiguo Wang, *A limited memory BFGS-type method for large-scale unconstrained optimization*. Computers & Mathematics with Applications, Volume 56, Issue 4, 2008
- [33] W. Verkerke, D. Kirkby, *The RooFit toolkit for data modeling*. arXiv:physics/0306116 [physics.data-an], 2003
- [34] *Standard Model*, URL: [https://en.wikipedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg#/media/File:Standard\\_Model\\_of\\_Elementary\\_Particles\\_and\\_Gravity.svg](https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg#/media/File:Standard_Model_of_Elementary_Particles_and_Gravity.svg), Accessed: 2021-07-07
- [35] *Meson-octet*, URL: <https://de.wikipedia.org/wiki/Datei:Meson-octet-small.svg>, Accessed: 2021-07-07
- [36] *Quark weak interactions*, URL: [https://commons.wikimedia.org/wiki/File:Quark\\_weak\\_interactions.svg](https://commons.wikimedia.org/wiki/File:Quark_weak_interactions.svg), Accessed: 2021-07-07
- [37] *SuperKEKB and Belle II*, URL: [https://www.belle2.org/project/super\\_kekb\\_and\\_belle\\_ii](https://www.belle2.org/project/super_kekb_and_belle_ii), Accessed: 2021-07-21
- [38] *MLP at scikit-learn*, URL: [https://scikit-learn.org/stable/modules/neural\\_networks\\_supervised.html](https://scikit-learn.org/stable/modules/neural_networks_supervised.html), Accessed: 2022-03-11

# Appendix A

## Topology Analysis

TopoAna is a software for topology analysis that is available for Belle II research [27]. It offers an opportunity to look into whole decay trees of signal and background in MC data. By looking at result of the topology analysis, one can strategically come up with the selection cut or criteria to rule out certain dominant decay in background. The result of TopoAna is ranked by nEtr(the number of entries). The decay trees with high nEtr are shown in this appendix and rest of them is not included.

A.1 Signal( $B^\pm \rightarrow K^\pm \nu \bar{\nu}$ )

Table 1: Decay trees and their respective initial-final states.

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCEtr
1	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	302	66	66
2	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^- D^0, \rho^- \rightarrow \pi^0 \pi^-, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	338	57	123
3	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \pi^- D^0, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^- K^+ K^-$ )	1	35	158
4	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^- D^{*0}, \rho^- \rightarrow \pi^0 \pi^-, D^{*0} \rightarrow \pi^0 D^0, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	345	29	187
5	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^0 \pi^+ \pi^+ D^{*-}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, D^{*-} \rightarrow \pi^- \bar{D}^0, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- \pi^- K^+ K^-$ )	250	27	214
6	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^{*0}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	262	27	241
7	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \pi^0 \pi^- \pi^- D^{*+}, D^{*+} \rightarrow \pi^+ D^0, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- \pi^- K^+ K^-$ )	25	26	267
8	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	740	26	293
9	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^0 \rightarrow \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^- K^+ K^-$ )	1097	24	317
10	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \pi^- D^{*0}, D^{*0} \rightarrow \pi^0 D^0, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	79	23	340
11	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \pi^+ \pi^- \pi^- D^0, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	224	21	361
12	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^+ \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^- K^+ K^-$ )	263	19	380
13	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow D^{*0} a_1^-, D^{*0} \rightarrow D^0 \gamma, a_1^- \rightarrow \rho^0 \pi^-, D^0 \rightarrow \pi^0 \pi^+ K^-, \rho^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^- \gamma$ )	140	18	398
14	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^- D^0, \rho^- \rightarrow \pi^0 \pi^-, D^0 \rightarrow \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^- K^+ K^-$ )	299	18	416
15	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^{*0}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^{*0} \rightarrow \bar{D}^0 \gamma, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^- \gamma$ )	329	18	434
16	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \bar{D}^{*0} a_1^+, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, a_1^+ \rightarrow \rho^0 \pi^+, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+, \rho^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	81	17	451

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCEtr
17	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^- D^0, \rho^- \rightarrow \pi^0 \pi^-, D^0 \rightarrow K^- a_1^+, a_1^+ \rightarrow \rho^0 \pi^+, \rho^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	725	17	468
18	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^{*0}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, \bar{D}^0 \rightarrow \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	257	16	484
19	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^- D^{*0}, \rho^- \rightarrow \pi^0 \pi^-, D^{*0} \rightarrow D^0 \gamma, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^- \gamma$ )	277	16	500
20	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \bar{D}^{*0} a_1^+, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, a_1^+ \rightarrow \rho^0 \pi^+, \bar{D}^0 \rightarrow \pi^- K^+, \rho^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	308	16	516
21	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \pi^0 \pi^+ \pi^- \pi^- D^{*0}, D^{*0} \rightarrow D^0 \gamma, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^- \gamma$ )	36	15	531
22	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^0 \pi^- D^0, \rho^0 \rightarrow \pi^+ \pi^-, D^0 \rightarrow \pi^0 \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	143	15	546
23	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^0 \pi^+ \pi^+ \pi^- \bar{D}^{*0}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	251	15	561
24	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^+ \pi^+ \pi^- \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^0 \rightarrow \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	413	15	576
25	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^+ \bar{D}^{*0}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, \bar{D}^0 \rightarrow \pi^0 \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- K^+ K^-$ )	672	15	591
26	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^0 \pi^+ \pi^+ D^{*-}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, D^{*-} \rightarrow \pi^0 D^-, D^- \rightarrow \pi^- \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	306	14	605
27	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^0 \rightarrow \pi^+ \pi^- K_S^0, K_S^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^-$ )	141	13	618
28	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \rho^- D^0, \rho^- \rightarrow \pi^0 \pi^-, D^0 \rightarrow \pi^+ \pi^- \bar{K}^*, \bar{K}^* \rightarrow \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^- \pi^- K^+ K^-$ )	191	13	631
29	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \rho^+ \bar{D}^0, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \rho^+ \rightarrow \pi^0 \pi^+, \bar{D}^0 \rightarrow K^+ a_1^-, a_1^- \rightarrow \rho^0 \pi^-, \rho^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	259	13	644
30	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow D^{*0} a_1^-, D^{*0} \rightarrow \pi^0 D^0, a_1^- \rightarrow \rho^0 \pi^-, D^0 \rightarrow \pi^0 \pi^+ K^-, \rho^0 \rightarrow \pi^+ \pi^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	1424	13	657
31	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \pi^0 \pi^+ \pi^+ \pi^- \bar{D}^{*0}, B^- \rightarrow \nu_e \bar{\nu}_e K^-, \bar{D}^{*0} \rightarrow \pi^0 \bar{D}^0, \bar{D}^0 \rightarrow \pi^- K^+$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	502	12	669
32	$\Upsilon(4S) \rightarrow B^+ B^-, B^+ \rightarrow \nu_e \bar{\nu}_e K^+, B^- \rightarrow \pi^0 \pi^+ \pi^- \pi^- D^{*0}, D^{*0} \rightarrow \pi^0 D^0, D^0 \rightarrow \pi^+ K^-$ ( $\Upsilon(4S) \rightarrow \nu_e \bar{\nu}_e \pi^0 \pi^0 \pi^+ \pi^+ \pi^- \pi^- K^+ K^-$ )	1187	12	681

## A.2 Charged B pair

Table 1: Decay trees and their respective initial-final states.

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCEtr
1	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^0\pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ $(\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	286	80	80
2	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow \mu^-\bar{\nu}_\mu D^{*0}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^{*0} \rightarrow \pi^0D^0,$ $D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow \mu^-\bar{\nu}_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	16168	76	156
3	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \pi^0\pi^+\pi^+D^{*-}, B^- \rightarrow \mu^-\bar{\nu}_\mu D^{*0}, D^{*-} \rightarrow \pi^-\bar{D}^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+,$ $D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow \mu^-\bar{\nu}_\mu\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-)$	41735	51	207
4	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow \mu^-\bar{\nu}_\mu D^0, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow \mu^-\bar{\nu}_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	7482	50	257
5	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow \mu^-\bar{\nu}_\mu D^{*0}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^{*0} \rightarrow D^0\gamma,$ $D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow \mu^-\bar{\nu}_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	47956	46	303
6	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \pi^0\pi^-\pi^-D^{*+}, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^{*+} \rightarrow \pi^+D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+,$ $D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	127168	39	342
7	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow e^+\nu_e\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^0\pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ $(\Upsilon(4S) \rightarrow e^+\nu_e\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	34282	38	380
8	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \bar{D}^0\gamma, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^0\pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ $(\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	132795	38	418
9	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow e^-\bar{\nu}_e D^0, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow e^-\bar{\nu}_e\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	223564	36	454
10	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow e^-\bar{\nu}_e D^{*0}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^{*0} \rightarrow \pi^0D^0,$ $D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow e^-\bar{\nu}_e\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	184765	35	489
11	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^{*0} \rightarrow \pi^0D^0,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	16945	34	523
12	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow e^-\bar{\nu}_e D^{*0}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^-K^+, D^{*0} \rightarrow \pi^0D^0,$ $D^0 \rightarrow \pi^0\pi^+K^-$ $(\Upsilon(4S) \rightarrow e^-\bar{\nu}_e\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-)$	61188	34	557

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCEtr
13	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \pi^0\pi^+\pi^-\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^0\pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $\Upsilon(4S) \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	58871	33	590
14	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow \pi^0\pi^+\pi^-\bar{D}^{*0}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^{*0} \rightarrow \pi^0D^0,$ $D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	32585	32	622
15	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^0, B^- \rightarrow \rho^-D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	59324	32	654
16	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \pi^-D^{*0}, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+,$ $D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-$ )	72516	32	686
17	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \pi^+\pi^-\pi^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-$ )	71631	29	715
18	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \pi^0\pi^+\pi^+D^{*-}, B^- \rightarrow e^-\bar{\nu}_eD^{*0}, D^{*-} \rightarrow \pi^-\bar{D}^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+,$ $D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow e^-\bar{\nu}_e\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	23209	28	743
19	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow e^-\bar{\nu}_eD^{*0}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^{*0} \rightarrow D^0\gamma,$ $D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow e^-\bar{\nu}_e\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-\gamma$ )	22992	27	770
20	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \pi^0\pi^+\pi^+D^{*-}, B^- \rightarrow \rho^-D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^0\pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $\Upsilon(4S) \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	47494	27	797
21	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow e^+\nu_e\bar{D}^{*0}, B^- \rightarrow \pi^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $\Upsilon(4S) \rightarrow e^+\nu_e\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-$ )	94916	27	824
22	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \rho^-D^{*0}, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^{*0} \rightarrow D^0\gamma,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-\gamma$ )	376688	27	851
23	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \bar{D}^0\gamma, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-\gamma$ )	74370	26	877
24	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \mu^+\nu_\mu\bar{D}^{*0}, B^- \rightarrow \rho^-D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, \rho^- \rightarrow \pi^0\pi^-, D^0 \rightarrow \pi^+K^-,$ $\bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $\Upsilon(4S) \rightarrow \mu^+\nu_\mu\pi^0\pi^0\pi^+\pi^-\pi^-K^+K^-$ )	113562	26	903
25	$\Upsilon(4S) \rightarrow B^+B^-, B^+ \rightarrow \rho^+\bar{D}^0, B^- \rightarrow \pi^0\pi^-\pi^-D^{*+}, \rho^+ \rightarrow \pi^0\pi^+, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^{*+} \rightarrow \pi^+D^0,$ $D^0 \rightarrow \pi^0\pi^+K^-$ ( $\Upsilon(4S) \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	156471	26	929

## A.3 ccbar

Table 1: Decay trees and their respective initial-final states.

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCEtr
1	$e^+e^- \rightarrow \pi^0\pi^+\pi^-D^{*+}D^{*-}\gamma^I, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-\pi^-K^+K^-\gamma^I$ )	100319	84	84
2	$e^+e^- \rightarrow \pi^0\pi^+\pi^-D^{*+}D^{*-}, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-\pi^-K^+K^-$ )	343224	65	149
3	$e^+e^- \rightarrow \pi^+\pi^-D^{*+}D^{*-}\gamma^I, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^+\pi^+\pi^-\pi^-\pi^-K^+K^-\gamma^I$ )	94207	53	202
4	$e^+e^- \rightarrow \pi^+\pi^-\rho^-D^{*+}\bar{D}^{*0}, \rho^- \rightarrow \pi^0\pi^-, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-\pi^-K^+K^-$ )	9250	51	253
5	$e^+e^- \rightarrow \pi^0\pi^+\pi^-D^{*0}\bar{D}^{*0}\gamma^I, D^{*0} \rightarrow \pi^0D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	33609	51	304
6	$e^+e^- \rightarrow \pi^0\pi^+\pi^-D^{*+}\bar{D}^{*0}, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	106145	49	353
7	$e^+e^- \rightarrow \pi^+\pi^-D^{*+}D^{*-}, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	541814	48	401
8	$e^+e^- \rightarrow \pi^+\pi^+\pi^-D^{*+}D^{*0}\gamma^I, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	18162	47	448
9	$e^+e^- \rightarrow \pi^+\pi^-\rho^+D^{*-}D^{*0}, \rho^+ \rightarrow \pi^0\pi^+, D^{*-} \rightarrow \pi^-\bar{D}^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	58868	46	494
10	$e^+e^- \rightarrow \pi^0\pi^+\pi^-D^{*+}D^{*0}\gamma^I, \rho^- \rightarrow \pi^0\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	167638	45	539
11	$e^+e^- \rightarrow \pi^0\pi^+D^{*+}D^{*0}\gamma^I, D^{*+} \rightarrow \pi^+D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	90363	44	583
12	$e^+e^- \rightarrow \pi^+\pi^-\rho^+D^{*+}D^{*0}\gamma^I, \rho^+ \rightarrow \pi^0\pi^+, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	264880	44	627
13	$e^+e^- \rightarrow \pi^+\pi^-\omega D^{*+}D^{*0}, \omega \rightarrow \pi^0\pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	120373	43	670
14	$e^+e^- \rightarrow \pi^+\pi^-\rho^-D^{*+}\bar{D}^{*0}\gamma^I, \rho^- \rightarrow \pi^0\pi^-, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	200357	43	713
15	$e^+e^- \rightarrow \pi^+\pi^-\pi^-D^{*+}\bar{D}^{*0}, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	470853	43	756
16	$e^+e^- \rightarrow \rho^0\pi^+\pi^-D^{*+}D^{*0}, \rho^0 \rightarrow \pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	509230	43	799
17	$e^+e^- \rightarrow \pi^+\pi^-\pi^-D^{*+}\bar{D}^{*0}\gamma^I, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	80530	39	838

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCEtr
18	$e^+e^- \rightarrow \pi^0\pi^0\pi^+\pi^-D^{*+}D^{*-}\gamma^I, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	129107	39	877
19	$e^+e^- \rightarrow \pi^0\pi^+\pi^-D^{*0}\bar{D}^{*0}, D^{*0} \rightarrow \pi^0D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	196476	39	916
20	$e^+e^- \rightarrow \pi^0\pi^+\pi^+\pi^-D^{*+}D^{*-}, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	202149	39	955
21	$e^+e^- \rightarrow \pi^+\pi^-\omega D^{*+}D^{*-}\gamma^I, \omega \rightarrow \pi^0\pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	5995	38	993
22	$e^+e^- \rightarrow \pi^0\pi^-D^{*+}\bar{D}^{*0}, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	19476	38	1031
23	$e^+e^- \rightarrow \pi^0\pi^+\pi^-\rho^+D^{*-}D^{*0}, \rho^+ \rightarrow \pi^0\pi^+, D^{*-} \rightarrow \pi^-\bar{D}^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	165338	38	1069
24	$e^+e^- \rightarrow \pi^0\pi^+\pi^-\rho^+D^{*-}D^{*0}\gamma^I, \rho^+ \rightarrow \pi^0\pi^+, D^{*-} \rightarrow \pi^-\bar{D}^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	10567	37	1106
25	$e^+e^- \rightarrow \pi^0\pi^-D^{*+}\bar{D}^{*0}\gamma^I, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	44343	37	1143
26	$e^+e^- \rightarrow \pi^0\rho^0\pi^-D^{*+}\bar{D}^{*0}\gamma^I, \rho^0 \rightarrow \pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	67347	37	1180
27	$e^+e^- \rightarrow \pi^0\pi^+\pi^-\omega D^{*+}D^{*-}, \omega \rightarrow \pi^0\pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	137375	37	1217
28	$e^+e^- \rightarrow \pi^0\pi^-\omega D^{*+}\bar{D}^{*0}, \omega \rightarrow \pi^0\pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	230976	37	1254
29	$e^+e^- \rightarrow \pi^0\pi^-D^{*+}\bar{D}^{*0}\gamma^I, \omega \rightarrow \pi^0\pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	255940	37	1291
30	$e^+e^- \rightarrow \pi^0\rho^0\pi^+\pi^-D^{*+}D^{*-}, \rho^0 \rightarrow \pi^+\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	304006	37	1328
31	$e^+e^- \rightarrow \pi^0\pi^+\pi^-\rho^-D^{*+}\bar{D}^{*0}\gamma^I, \rho^- \rightarrow \pi^0\pi^-, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	338000	37	1365
32	$e^+e^- \rightarrow \pi^0\pi^+\pi^-\pi^-D^{*+}\bar{D}^{*0}\gamma^I, D^{*+} \rightarrow \pi^+D^0, \bar{D}^{*0} \rightarrow \pi^0\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	464869	37	1402
33	$e^+e^- \rightarrow \pi^0\pi^+\rho^-D^{*+}D^{*-}, \rho^- \rightarrow \pi^0\pi^-, D^{*+} \rightarrow \pi^+D^0, D^{*-} \rightarrow \pi^-\bar{D}^0, D^0 \rightarrow \pi^0\pi^+K^-, \bar{D}^0 \rightarrow \pi^0\pi^-K^+$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	59801	36	1438
34	$e^+e^- \rightarrow \pi^0\pi^+\pi^+\pi^-D^{*+}D^{*0}\gamma^I, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-\gamma^I$ )	141592	36	1474
35	$e^+e^- \rightarrow \pi^0\pi^+D^{*+}D^{*0}, D^{*+} \rightarrow \pi^+D^0, D^{*0} \rightarrow \pi^0D^0, \bar{D}^0 \rightarrow \pi^0\pi^-K^+, D^0 \rightarrow \pi^0\pi^+K^-$ ( $e^+e^- \rightarrow \pi^0\pi^0\pi^0\pi^0\pi^+\pi^+\pi^-\pi^-K^+K^-$ )	277876	35	1509

# Acknowledgement

First of all, I want to thank Prof. Dr. Thomas Kuhr for allowing me to work with this topic and always guide me to find better direction of research.

And I also thank Dr. Thomas Lück for making many meetings to discuss about both code and physics analysis in detail. It was very nice teaching and collaboration.

I also thank Nathalie Eberlein for her enthusiasm teaching me and many students in group.

And I thank Dr. Sviatoslav Bilokin for helps, especially, regarding fBDT.

I cannot enumerate all advices and supports of the group members in this limited page.

Lastly, I want to mention that I enjoyed talking with all group members about any topic.

# Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

Ort, Datum, Unterschrift: