國立臺灣大學理學院物理學系

碩士論文

Department of Physics

College of Science

National Taiwan University

Master Thesis

在 Belle II 實驗中對 B^+ 介子衰變至 $\mu^+\nu$ 之敏感度分析

Sensitivity Study of $B^+ \to \mu^+\nu$ in the Belle II Experiment

金彥廷

Yen-Ting Chin

指導教授: 張寶棣 博士

Advisor: Pao-Ti Chang, Ph.D.

中華民國 110 年 2 月

February, 2021

國立臺灣大學碩士學位論文
口試委員會審定書

在 Belle II 實驗中對B^+介子衰變至$\mu^+\nu$之敏感度分析
Sensitivity Study of $B^+ \to \mu^+\nu$ in the Belle II Experiment

本論文係金彥廷君（R07222022）在國立臺灣大學物理學系研究所完成之碩士學位論文，於民國 109 年 1 月 25 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

張寶棣 （簽名）

（指導教授）

王名儒

張敏娟

徐靜苇

Acknowledgements

I would like to thank my advisor, Professor Pao-Ti Chang, who pointed out a research direction for me and gave me countless precious suggestion whenever I presented my research updates at the weekly NTU Belle analysis meeting. He also patiently gave me instructions and corrected my misconceptions about technical issues of my research through innumerable times of discussion with me. Without his help, I would certainly not have been able to finish this thesis.

I would also like to thank the members of my master thesis defense committee, Professor Ming-Chuan Chang, Doctor Jing-Ge Shiu, and Professor Min-Zu Wang, for providing me with new perspectives about my research topic. A thank also goes to Professor Kai-Feng Chen and Professor Yuan-chin Chang for their consultation in solving statistical problems. I would also like to thank the members of Belle II Semileptonic Working Group, Markus Prim, Florian Bernlochner, Racha Cheaib, William Sutcliffe, Mario Merola, and Steven Robertson. Without their generous help and suggestions, I could not have succeeded in solving many problems in my study.

I would also like to express my appreciation to my friends and colleagues in NTUHEP, Chuan-Yu Chang, Janice Chen, Wen-Yi Chen, Yu-Tan Chen, Pin-Chun Chou, Kaining Chu, Yu-Chieh Ku, Ching-Hua Li, You-Hsuan Liang, Cheng-Wei Lin, Jie-Cheng Lin, Yuan-Ru Lin, Tzu-An Sheng, Han-Sheng Wang and those that are not mentioned above,

doi:10.6342/NTU202100639

for giving me support both academically and mentally. During my stay in this lab, I have learned a lot from you including how to improve my presentation skills.

Finally, I would like to thank my mother for educating me to lead a balanced life and supporting me to pursue my interest in physics.
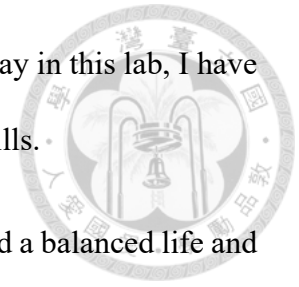
摘要

本研究旨在預測「假設標準模型對於 $B^+ \to \mu^+\nu$ 之分支分數預測正確,則在不同 Belle II 資料量下,$B^+ \to \mu^+\nu$ 之統計顯著性分別為何」。本研究使用蒙地卡羅方法產生大量資料樣本以模擬 e^+e^- 在 SuperKEKB 加速器中以 10.58 GeV 質心能量對撞所產生之各種物理事件。這些物理事件包含了 $e^+e^- \to q\bar{q}$、$e^+e^- \to \tau^+\tau^-$ 以及 $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0/B^+B^-$ 之過程。我們也模擬了 Belle II 探測器對這些物理事件之反應。本研究使用多變量分析以區隔 $B^+ \to \mu^+\nu$ 事件及背景事件。我們為不同物理過程建立機率密度函數並產生艾西莫夫資料,並使用其剖面概似函數預測不同 Belle II 資料量下 $B^+ \to \mu^+\nu$ 之統計顯著性。根據預測,我們需要多於 $7\,\mathrm{ab}^{-1}$ 之資料量以超越 5σ 之顯著性以宣稱發現 $B^+ \to \mu^+\nu$ 衰變。

關鍵字: B 介子、稀有 B 衰變、Belle II 實驗、SuperKEKB、渺子

Abstract

In this study, we intend to estimate the statistical significance of $B^+ \to \mu^+\nu$ when various amounts of Belle II data are available in the future, assuming that the Standard Model estimation of the branching fraction of $B^+ \to \mu^+\nu$ is correct. The Monte-Carlo method is applied to generate a large amount of data samples to simulate the physics events produced by e^+e^- collision on the center-of-mass energy of 10.58 GeV in the SuperKEKB accelerator. These physics events include the $e^+e^- \to q\bar{q}$, the $e^+e^- \to \tau^+\tau^-$, and the $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0/B^+B^-$ processes. The Belle II detector responses to these physics events are also simulated. In this study, we use multivariate analysis to distinguish the $B^+ \to \mu^+\nu$ events from the background events. We establish probability density functions for different physics processes and generate Asimov data sets. We then use the profile likelihood function of the Asimov data sets to estimate the statistical significance of $B^+ \to \mu^+\nu$ when different amounts of Belle II data are available. According to the estimation, we will need to collect more than 7 ab^{-1} of data in order to surpass the significance of 5σ in order to claim the discovery of $B^+ \to \mu^+\nu$.
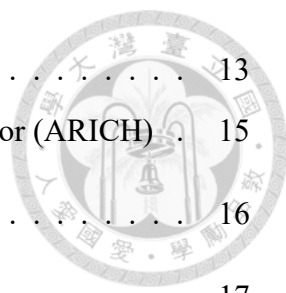
Contents

List of Figures

List of Tables

1 Introduction

1.1 Standard Model

The Standard Model (SM) is a theory describing the properties of all known elementary particles and the interactions among them. Particles in the SM are classified into 3 categories: fermions, gauge bosons, and the Higgs bosons.

Gauge bosons are spin-1 particles that mediate the fundamental interactions. There are four kinds of gauge bosons, including photons, eight types of gluons, $W^{\pm}$ bosons, and Z bosons. Photons serve as the mediators for electromagnetic interactions, $W^{\pm}$s and Zs carry the weak interactions, and gluons mediate the strong interactions. The spin-0 Higgs bosons play an important role in SM via the "Higgs mechanism" to generate mass of all the elementary particles except for photons and gluons. Fermions are spin-1/2 particles and can be further classified into two types, quarks and leptons, depending on whether the particles can interact via the strong interaction. Both quarks and leptons consist of six types of particles and can be classified into three generations. All the particles can interact via the weak interaction. Only quarks and gluons, which carry color charges, can interact via the strong interaction. Particles carrying electric charges can interact via the electromagnetic interaction.

The properties of all the elementary particles are summarized in Fig. 1.1

Standard Model of Elementary Particles
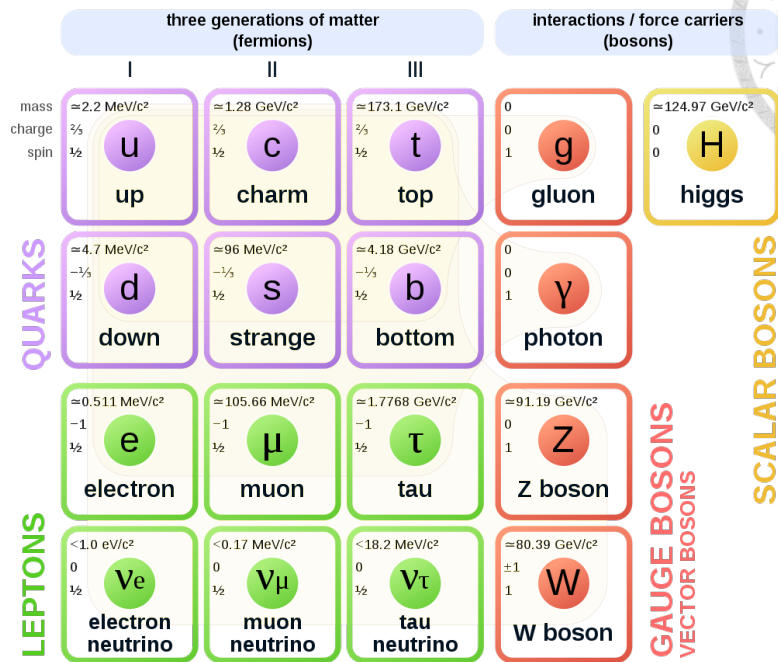
Figure 1.1: SM particles[1]

1.2 B physics

B mesons refer to two-quark states that consist of a b quark and another light quark. These mesons are denoted as B^0, B^+, B_s^0 and B_c^+ for $d\bar{b}$, $u\bar{b}$, $s\bar{b}$ and $c\bar{b}$ states respectively. The B mesons were first observed by the CLEO collaboration in 1983 at the Cornell Electron Storage Ring[2]. The B mesons were reconstructed from $B^- \to D^0\pi^-$, $\overline{B}^0 \to D^0\pi^-\pi^+$, $\overline{B}^0 \to D^{*+}\pi^-$, and $B^- \to D^{*+}\pi^-\pi^-$ decay channels. (It should be noted in the first place that, throughout this thesis, when a certain particle is referred to, its anti-particle is also taken into consideration even if it is not specified in the text. For example, when the decay channel $B^- \to D^0\pi^-$ is mentioned, it automatically implies that the conjugated channel $B^+ \to \overline{D}^0\pi^+$ is also accounted for.) The reconstructed B signal is shown in Fig. 1.2.

The B physics refers to the studies related to the properties of B mesons, including

2
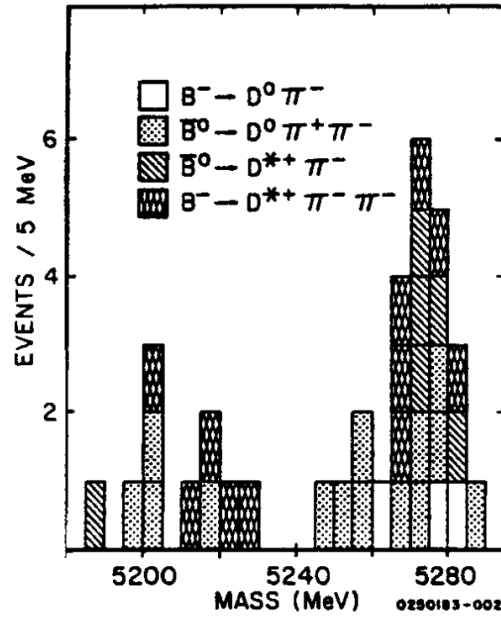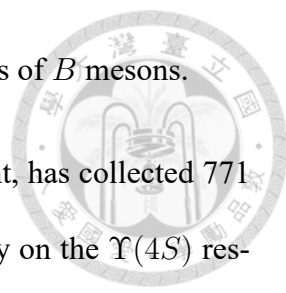
Figure 1.2: Reconstruction of B meson by CLEO collaboration, 1983[2]. A clear peak at the nominal $\overline{B}^0$ and B^- mass at ≈ 5.28 GeV is visible.

the branching fractions of B decays, CP violations in B decays, the B meson life time, and the $B^0 - \overline{B}^0$ mixing parameters. These studies are essential for determining the Cabibbo-Kobayashi-Maskawa (CKM) quark flavor mixing matrix parameters, which explains the phenomenon of CP violation. In addition, the measurements of rare B decays can serve as powerful probes for lots of different new physics (NP) models beyond SM.

1.3 Introduction to the Belle II Experiment

The Belle II experiment aims to study the properties of B mesons, which are produced by SuperKEKB, an e^+e^- accelerator, located in the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. The Belle II detector is at the collision point of SuperKEKB, where the $\Upsilon(4S)$ mesons are produced by colliding e^+e^- pairs at the energy of the $\Upsilon(4S)$ resonance (10579.4 ± 1.2 MeV). Each $\Upsilon(4S)$ meson will further decay to either a B^+B^- or a $B^0\overline{B}^0$ pair, each with approximately 50% of branching fraction. The

Belle II detector is used to detect the final states of the decay products of B mesons.
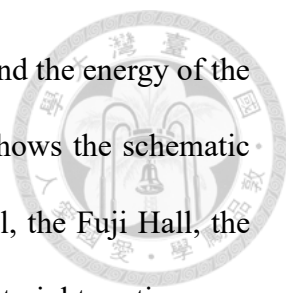
The predecessor of the Belle II experiment, the Belle experiment, has collected 771 million $B\bar{B}$ pairs, corresponding to 711 fb^{-1} of integrated luminosity on the $\Upsilon(4S)$ resonance. It has reached the instantaneous luminosity of 2.11×10^{34} cm^{-2}s^{-1}, which was a world record. The Belle II Experiment, with an upgraded accelerator, is expected to reach 40 times higher instantaneous luminosity and collect 50 times more data than the Belle experiment. As a result, new physics can be probed by studying the decay channels of B mesons with a higher sensitivity. The above mentioned possibilities of new physics include but are not limited to possible new CP sources in $b \to sq\bar{q}$ transition (including $B \to \phi K_S^0$, $B \to \eta' K_S^0$ and $B \to K_S^0 K_S^0 K_S^0$ decays), charged Higgs in 2 Higgs doublet models (2HDM) mediating the purely leptonic $B \to \mu\nu$ and $B \to \tau\nu$ decays, and new physics effects occurring in the loop diagram of $b \to s\gamma$ and $b \to d\gamma$ ($B \to X_s\gamma$ and $B \to X_d\gamma$) processes[12]. The Belle II experiment can also crosscheck some of the B decay anomalies emerged from the experimental results of other B-factories such as BaBar and LHCb, for example, the measurement of R(D) and R(D*) which deviated from the SM value. Being able to trigger high energy single photon in the barrel region of its electromagnetic calorimeter, the Belle II experiment is also dedicated to dark sector search where the event topology is expected to contain a single high energy photon, for example, $e^+e^- \to \gamma A'$[13].

1.4   SuberKEKB Accelerator

The content of this section is based on Ref. [14].

SuperKEKB is a circular e^+e^- double-ring accelerator in KEK, Tsukuba. The en-

ergy of the electron beam in the High Energy Ring (HER) is 7 GeV and the energy of the positron beam in the Low Energy Ring (LER) is 4 GeV. Fig. 1.3 shows the schematic view of SuperKEKB. Four experimental buildings, the Tsukuba Hall, the Fuji Hall, the Oho Hall, and the Nikko Hall, are located at the middle of the four straight sections, respectively. The total circumference is 3016 m. The Belle II detector is located at the one and only e^+e^- interaction point of SuperKEKB in the underground fourth level of the Tsukuba Hall.

SuperKEKB is upgraded from the old KEKB accelerator. The SuperKEKB accelerator is designed to reach an instantaneous luminosity of 8×10^{35} $\mathrm{cm^{-2}s^{-1}}$, 40 times higher than the luminosity of KEKB. In order to increase the luminosity, a "nano beam scheme" is adopted, in which a large crossing angle (83 mrad) is used, and the final-focusing magnets are placed very close to the Interaction Point (IP) so that the beam size is squeezed to nano-meter scale. Compared to the KEKB accelerator, the beam current is also increased (2.6 A to 3.6 A for LER and 1.1 A to 2.6 A for HER). The first collision of SuperKEKB is produced on April 26, 2018. On June 15, 2020, SuperKEKB has reached an instantaneous luminosity of 2.22×10^{34} $\mathrm{cm^{-2}s^{-1}}$, setting a new world record.
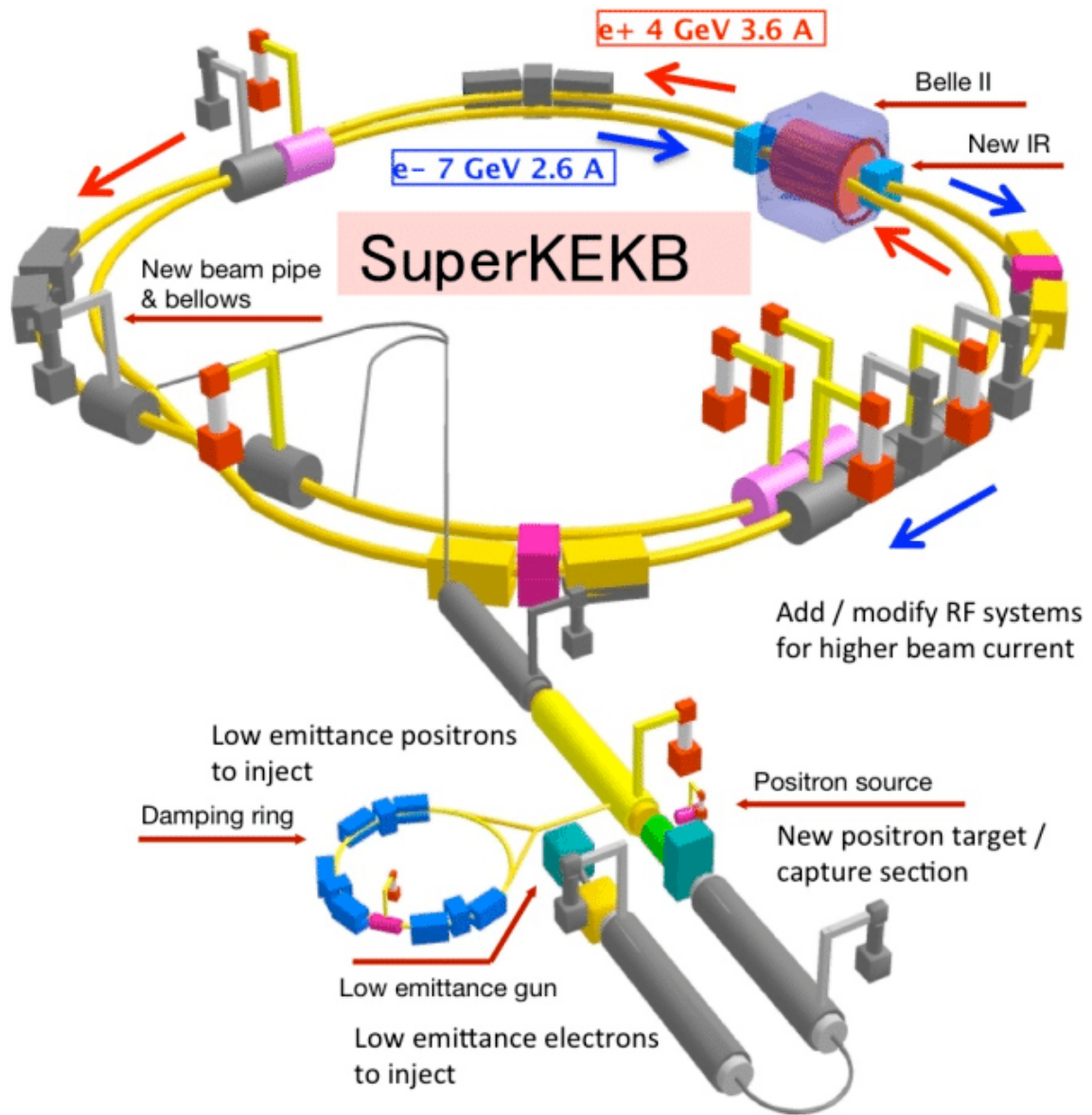
Figure 1.3: Schematic view of SuperKEKB[3]

Ring		LER	HER	Unit
Beam energy	E	4	7.007	GeV
Circumference	C	3016.315		m
Half crossing angle	θ_x	41.5		mrad
Piwinski angle	ϕ_{Piw}	24.6	19.3	rad
Horizontal emittance	ϵ_x	3.2(1.9)	4.6(4.4)	nm
Vertical emittance	ϵ_y	8.64	11.5	pm
Coupling		0.27	0.28	%
Beta function at IP	β_x^*/β_y^*	32/0.27	25/0.30	mm
Horizontal beam size	σ_x^*	10.1	10.7	μm
Vertical beam size	σ_y^*	48	62	nm
Horizontal betatron tune	v_x	44.530	45.530	
Vertical betatron tune	v_y	46.570	43.570	
Momentum compaction	α_p	3.25	4.55	10^{-4}
Energy Spread	σ_ϵ	8.14(7.96)	6.49(6.34)	10^{-4}
Beam current	I	3.60	2.60	A
Number of bunches	n_b	2500		
Particles/bunch	N	9.04	6.53	10^{10}
Energy loss/turn	U_0	1.87	2.45	MeV
Long damping time	τ_z	21.6	29.0	ms
RF frequency	f_{RF}	508.9		MHz
Total cavity voltage	V_c	9.4	15.0	MV
Total veam power	P_b	8.33	7.5	MW
Synchrotron tune	v_s	-0.0427	-0.0280	
Bunch length	σ_z	6.0(5.0)	5.0(4.9)	mm
Beam-beam parameter	ξ_x/ξ_y	0.0028/0.088	0.0012/0.081	

Table 1.1: Machine parameters of SuperKEKB. Inside the parentheses are values without intra-beam scattering[8]
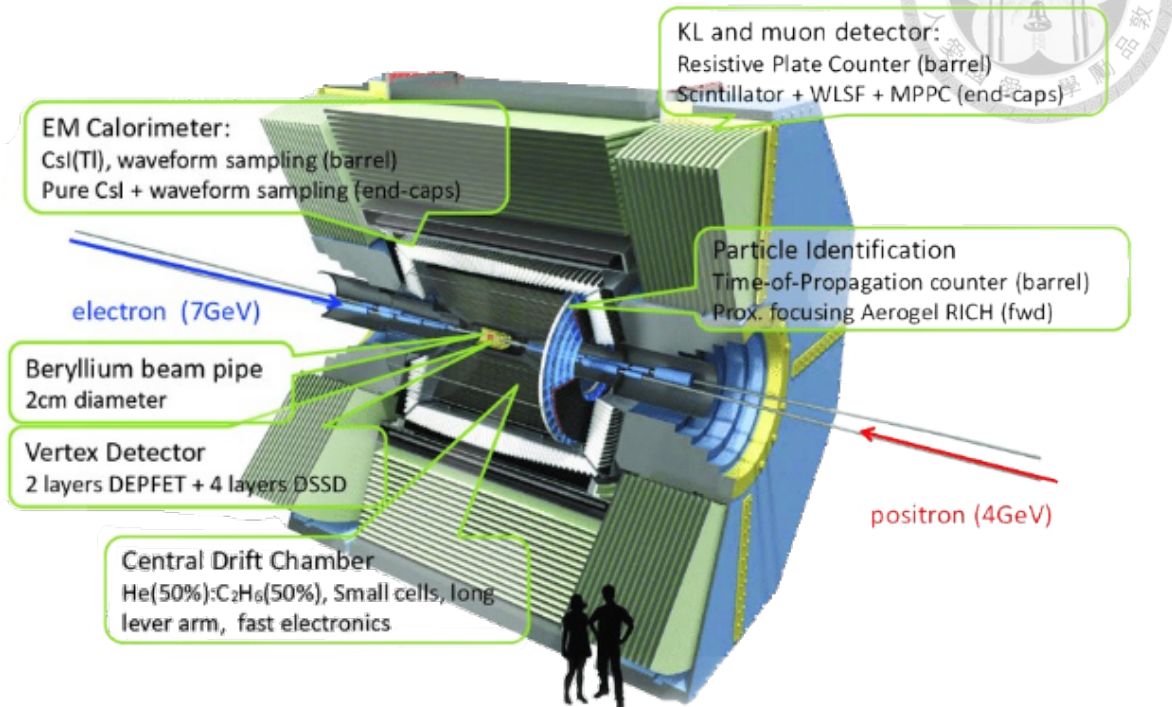
Figure 1.4: Schematic view of the Belle II detector[4]

1.5 Belle II Detector

The materials, the tables, and the plots presented in this section, if not specifically noted, are all from Ref. [15] and Ref. [5].

Fig. 1.4 shows the schematic view of the Belle II detector. The Belle II detector consists of several sub-detectors. These sub-detectors are responsible for particle identification, tracking, and energy measurement. It also consists of a trigger system to veto backgrounds and to decide which events should be recorded.
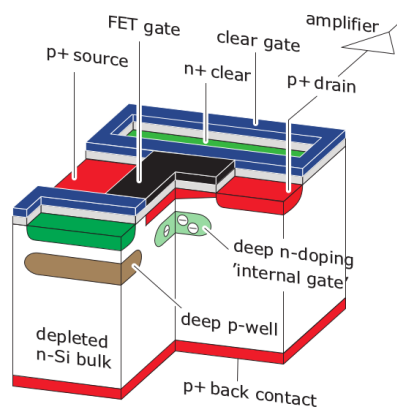
Figure 1.5: A DEPFET sensor[5]
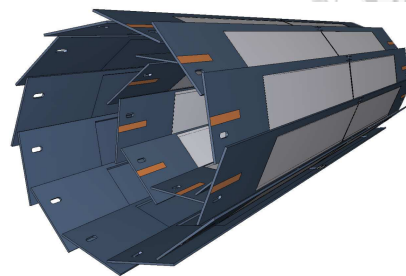
Figure 1.6: Schematic view of PXD[5]

1.5.1 Tracking System

The tracking system is used to detect the charged particles and determine the vertex positions of decay processes. It includes three sub-detectors: PXD, SVD, and CDC. The charged particles leave hits in these sub-detectors due to various physics processes. By connecting these hits, the tracks of the charged particles can be reconstructed. A magnetic field of around 1.5 T parallel to the beam pipe is applied to the sub-detectors, bending the trajectories of the charged particles by Lorentz force. By knowing the helix parameters of the curved trajectories, we can calculate the four-momentum of the particles.

1.5.1.1 Pixel Detector (PXD)

PXD is a two-layer cylindrical detector with the radii being 14 mm and 22 mm. It consists of 8 planar silicon substrates for the inner layer and 12 for the outer layer. The DEPleted Field EffecT (DEPFET) pixels array (see Fig. 1.6) mounted on the detector is responsible for detecting the charged particles. Each pixel consists of a FET mounted on the fully-depleted silicon substrate. An electric potential minimum is formed in the silicon substrate due to "sideward depletion". The potential minimum is called an "Internal

Gate". When a charged particle passes by, electron-hole pairs are generated in the silicon substrate. By flowing to the potential minimum, the electrons modulate the current that goes through the FET, resulting in a hit on the detector.

1.5.1.2 Silicon Vertex Detector (SVD)

SVD consists of four layers of Double-Sided Silicon Detectors (DSSD)[16]. These rectangular shaped silicon sensors are arranged in a windmill structure (Fig. 1.7) surrounding the beam pipe and the inner PXD detector.

To increase the acceptance in the forward region to account for the boost of the center-of-mass (CM) frame, the sensors in the forward region are slanted (see Fig. 1.8).

The sensors are "double-sided": facing toward the beam pipe is the p-side of the sensors with long strips parallel to the z axis, while the n-side of the sensors with short strips along the ϕ direction faces toward outside. When a charged particle passes by, electron-hole pairs are generated in the silicon sensors. The high voltage and the magnetic field applied on the sensors will determine the direction of the drifting carriers: the holes will drift to the p-side and the electrons will drift to the n-side, resulting in a hit. The drifting electrons and holes will both spread along the ϕ direction because of the Lorentz force. Due to higher mobility of the electrons, there will be a larger spread for the drifting electrons than the holes. Thus, the sensor plane is tilted to reduce the difference in the extent of spread. (see Fig. 1.9).
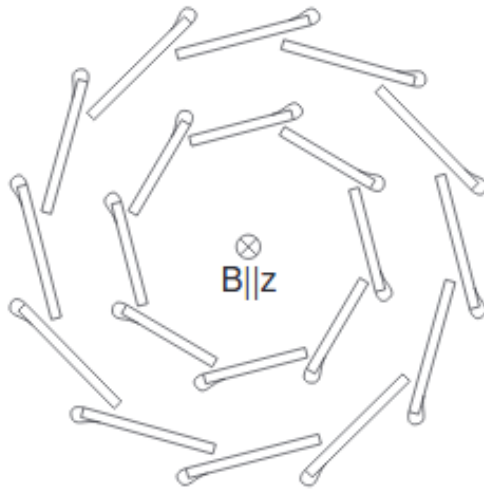
Figure 1.7: Windmill structure of the SVD[5]
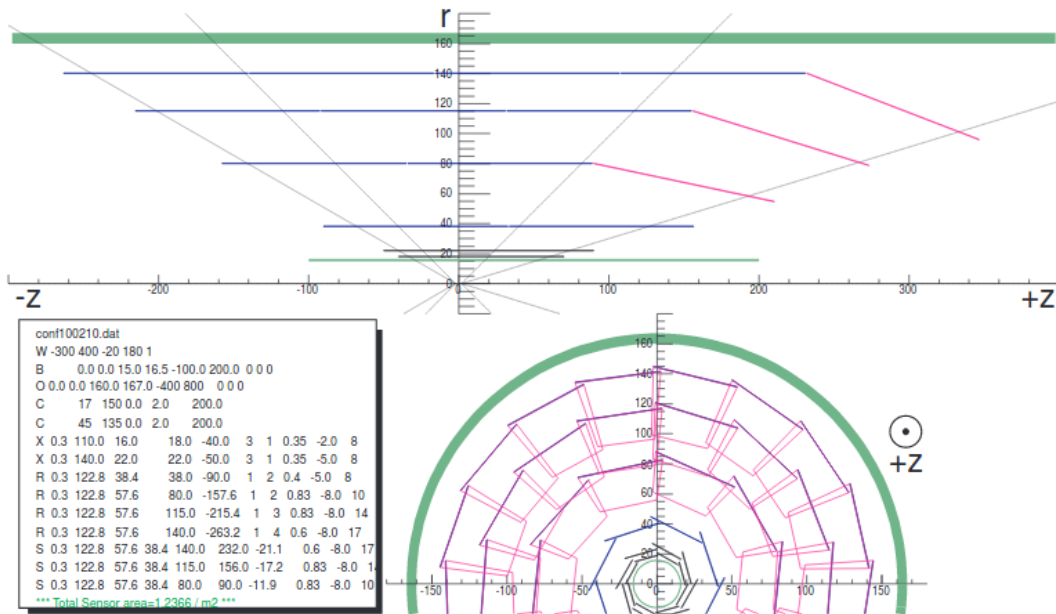
Figure 1.8: Configuration of the 4 layers of SVD and the 2 layers of PXD. All dimensions are in mm.[5]
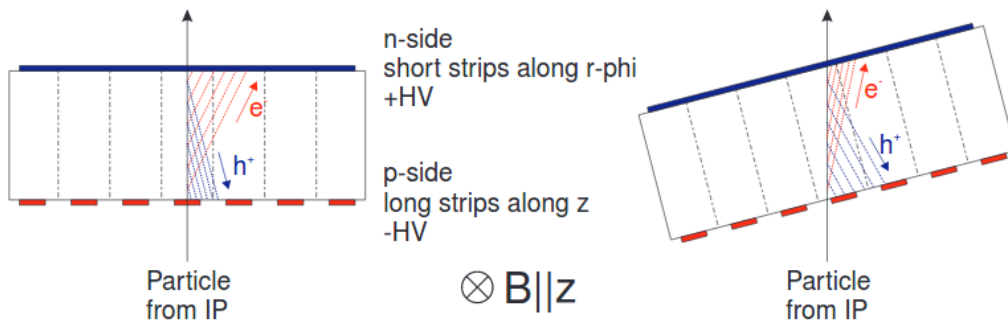
Figure 1.9: With the existence of magnetic field, the drifting carriers are deflected. On the left-hand-side, the spread for the electrons is larger than the holes due to its higher mobility. On the right-hand-side, the sensor is tilted to minimize the overall charge spread.[5]

11
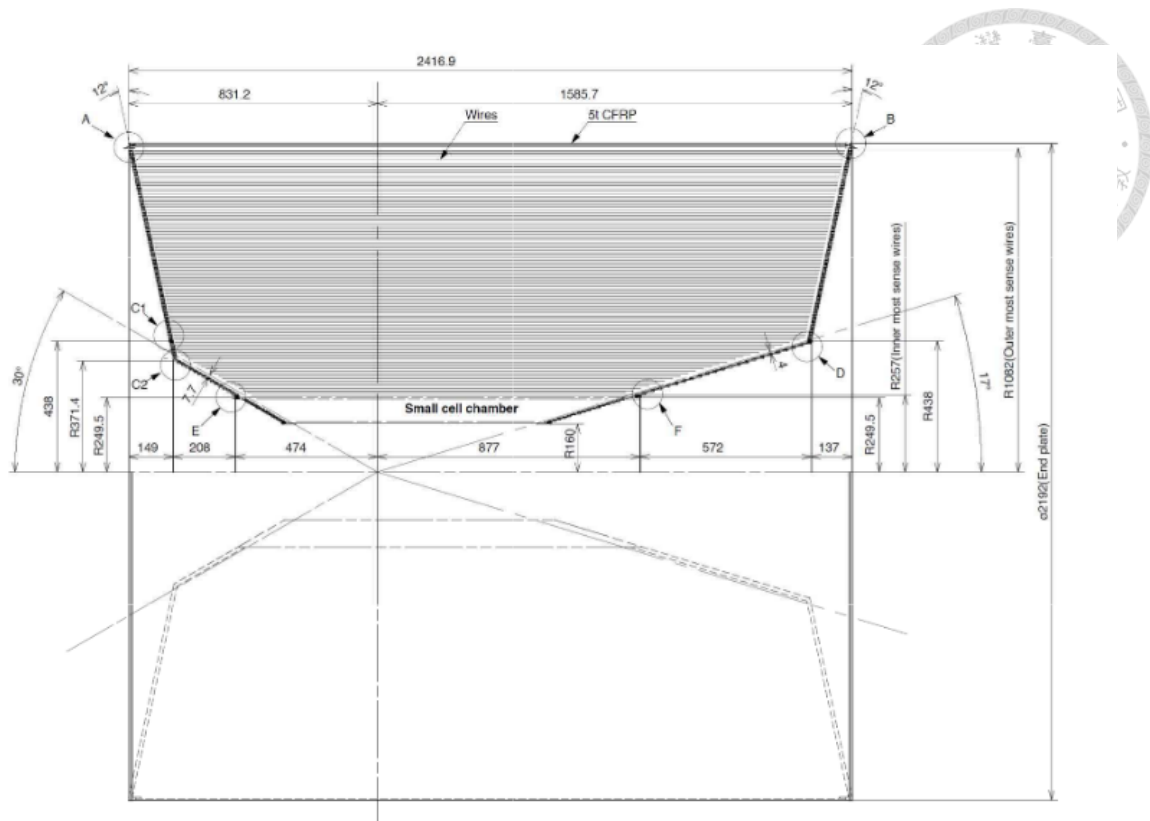
Figure 6.2: Main structure of the CDC.

Figure 1.10: Overview of the structure of CDC[5]

1.5.1.3   Central Drift Chamber (CDC)

CDC is a cylindrical detector filled with He-C_2H_6 gas. Two kinds of wires, the sense wires and the field wires, are alternatively placed in the detector, forming a wire array. The He-C_2H_6 gas is ionized when high energy charged particles pass by. The electrons from ionization are driven by the electric field between the nearby ground-level field wire and the high-voltage sense wire. They are then collected by the sense wire, which creates hits. In CDC, there are 32 layers of axial wires and 24 layers of stereo wires. The axial wires are parallel to the beam pipe when the stereo wires are tilted in order to measure the z position of the charged tracks.
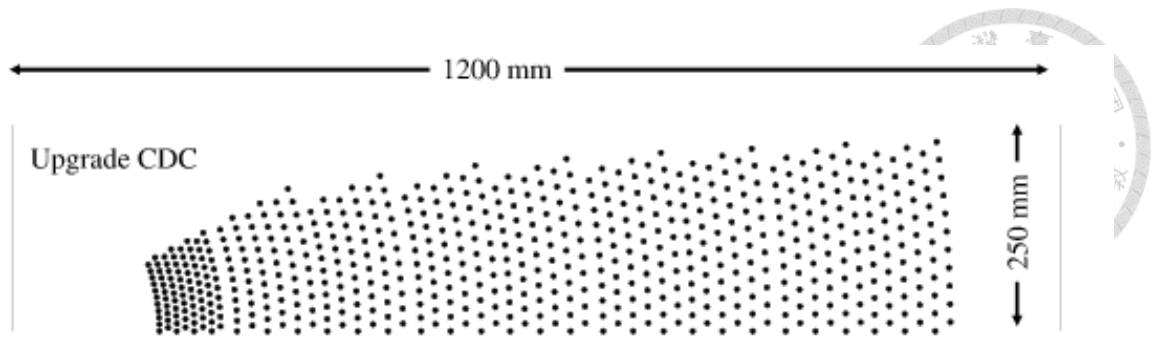
Figure 1.11: Wire configuration of CDC[5]

1.5.2 Particle Identification (PID) System

The sub-detectors in the PID system are mainly responsible for distinguishing K^+ from π^+. These two particles behave similarly (both of them are charged particle and has a high possibility to decay into $\mu^+\nu$) but have different masses.

1.5.2.1 Time of Propagation Detector (TOP))

The TOP detector accounts for the particle identification in the barrel region of the Belle II detector. It consists of sixteen modules surrounding the beam pipe cylindrically. Each module is composed by two quartz (silica) bars each sized at $125 \times 45 \times 2$ cm, a focusing mirror at the forward end of the bars, and an array of Micro-Channel-Plate PhotoMultiplier Tubes (MCP-PMT) glued to a prism at the backward end[6]. The quartz bars serve as Cherenkov light radiators while charged particles pass through the module. The Cherenkov light generated in the bars will experience total inflection. The total inflected light will eventually reach the end of the bar and then be reflected by the focusing mirror. The reflected light will then be focused on MCP-PMT. The sketch of the TOP module and the operational principle of the module are shown in Fig. 1.12. By measuring the Cherenkov angle θ_c, we can determine the speed of the particle by the following formula:
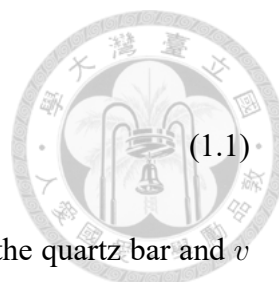
$$cos(\theta_c) = c/nv \qquad (1.1)$$

, where c is the speed of light in vacuum, n is the refractive index in the quartz bar and v is the speed of the particle. Then, by the relation between the relativistic momentum, the speed, and the mass,

$$p = \gamma m v \qquad (1.2)$$

, we can determine the type of the particle.
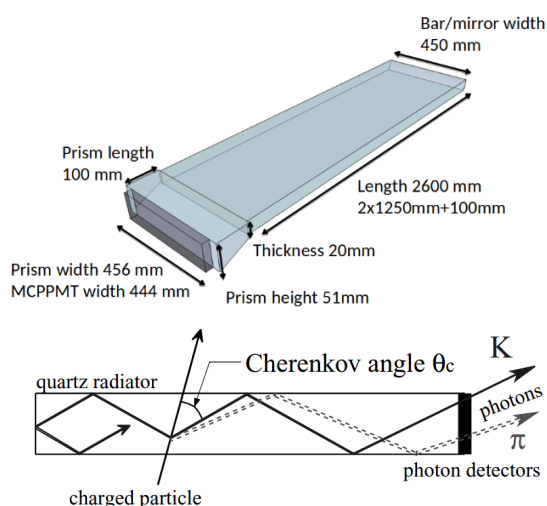
Figure 1.12: Up: sketch of the TOP module[6], down: Operational principle of the module[5]
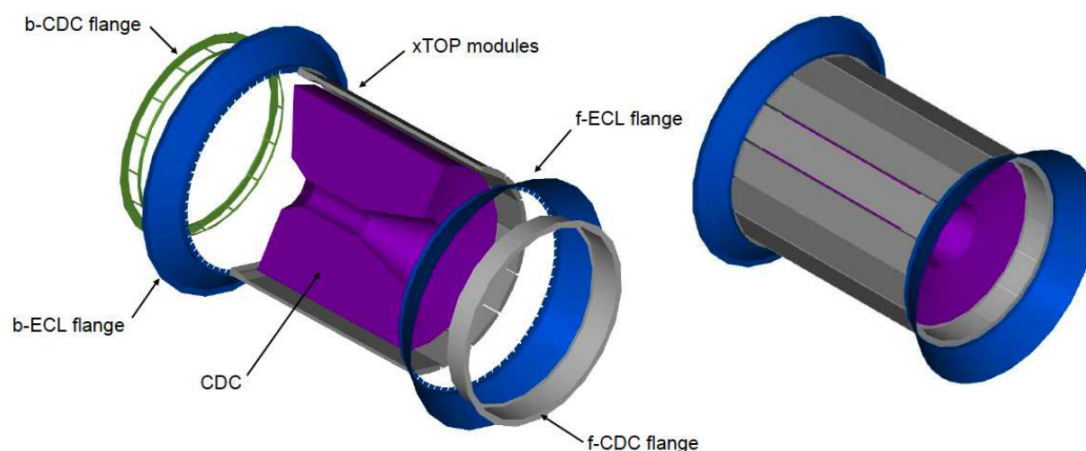
Figure 1.13: Structure of TOP mounted on CDC[5]

1.5.2.2 Aerogel Ring-Imaging Cherenkov Detector (ARICH)

The ARICH detector is responsible for the particle identification at the forward end-cap of the Belle II detector. The detector consists of two layers of aerogel tiles where the Cherenkov light is produced by the passing charged particles, an expansion volume where the Cherenkov ring is formed, and an array of Hybrid Avalanche Photo-Detectors (HAPD) which can detect the position of the Cherenkov photon and reconstruct the ring image. The type of the passing particle will be determined by the size of the ring image.
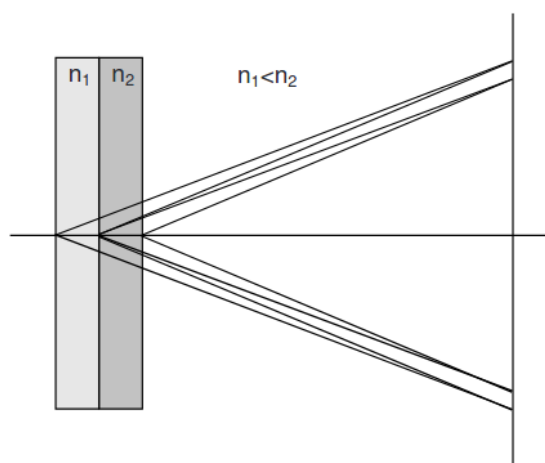
Figure 1.14: Conceptual view of the two layers of aerogel and the photon detectors[5]
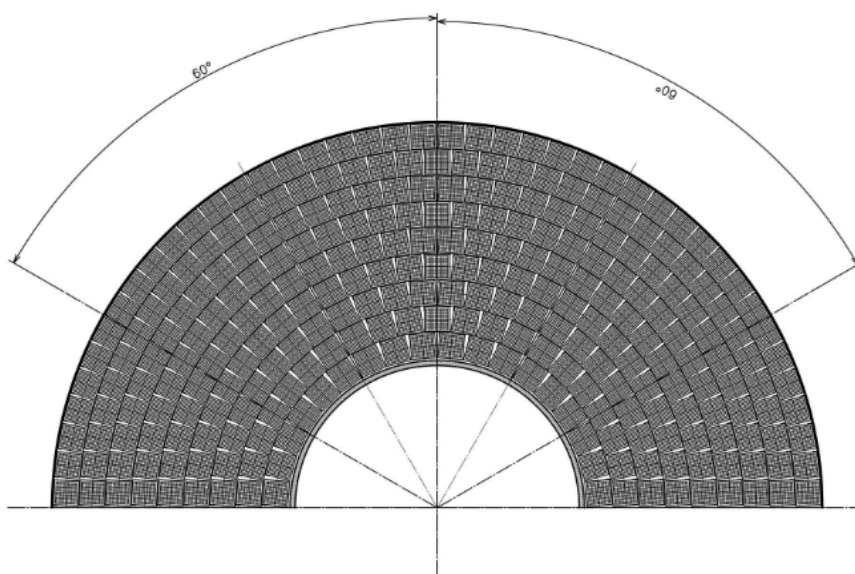
Figure 1.15: The structure of ARICH[5]
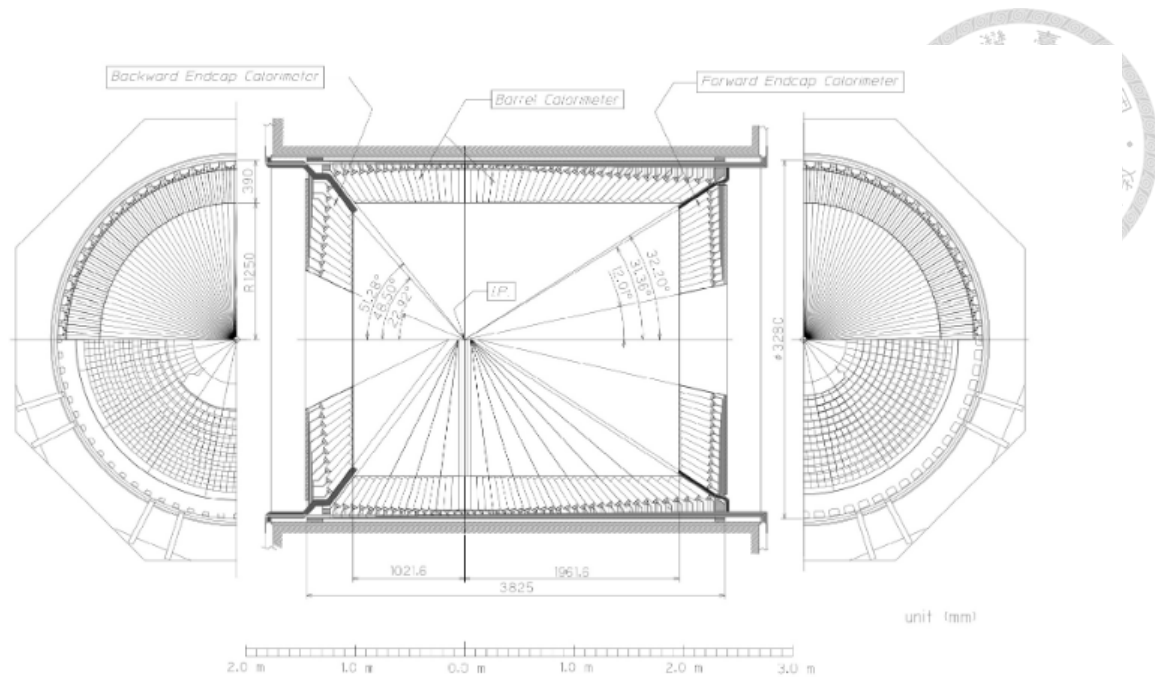
Figure 1.16: The structure of ECL

1.5.3 Electromagnetic Calorimeter (ECL)

Measurement of photon energy and identification of electrons are conducted by ECL. The detector contains 6624 CSI(TI) scintillator crystals in the barrel region and 2112 pure CSI crystals in the endcap region. At an energy scale above 10 MeV, the electron interaction within materials is dominated by bremsstrahlung, where an electron splits into a photon and a lower energy electron. On the other hand, at such an energy scale, the interaction between photons and materials is dominated by pair production, where an e^+e^- pair is produced. Within the CSI crystals that are high Z and low in radiation length, the electrons and photons passing by will shower into multiple low energy particles. Once the energy of the showered electrons are below the critical energy, 10 MeV, where ionization effect dominates, the ionization of the material will cause the scintillator crystals to illuminate. The photons produced by the luminescence of the scintillator crystals will be recorded by the two Hamamatsu S2744- 08 photodiodes mounted at the rear of each crystal, and then the deposit energy of the shower will be measured.

16

1.5.4 K_L^0 and μ^+ Detector (KLM)

The KLM detector is responsible for detecting μ^+s and neutral hadrons (for example, K_L^0s). It consists of alternating active detectors and iron plates layers. The iron plates have two functions: they provide 3.9 or more interaction lengths that allow hadrons to shower in it and at the same time serve to return the magnetic field back to the magnet. In the barrel region (BKLM), the glass-electrode Resistive Plate Chambers (RPC) are used for active detection. In the design of RPC, a high voltage is applied between two parallel glass-sheet electrodes. When the gas between the glass electrodes is ionized by charged particles passing by (either μ^+s or products of hadronic showers), the high voltage accelerates the electrons and ions produced by the ionization, causing an avalanche. The signal of the avalanche is then read out by two strips located on each side of RPC.

In the endcap region (EKLM), due to the high background rates, scintillator based detectors are used instead of RPC. Readout is done with silicon photomultiplier (SiPM) working in the Geiger mode.
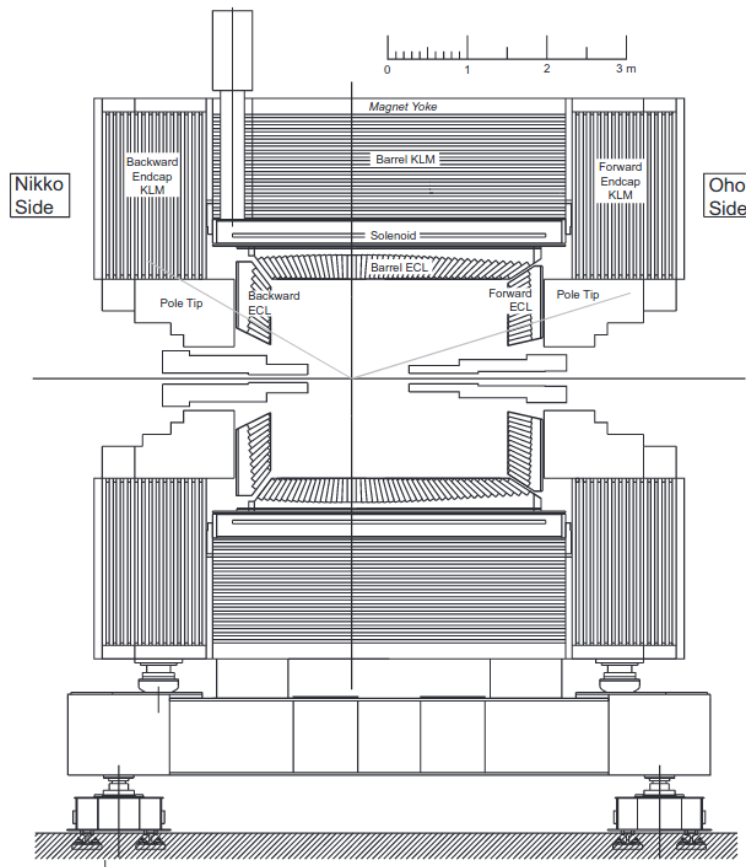
Figure 1.17: The side view of KLM
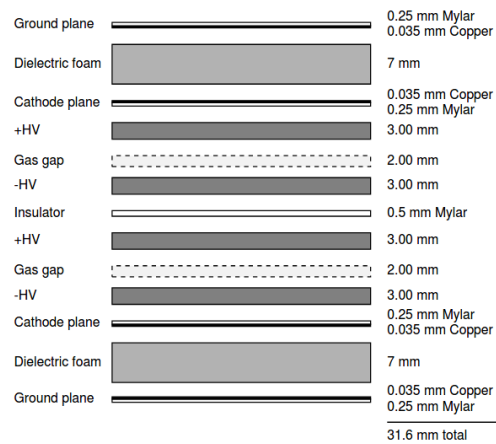
Figure 1.18: A superlayer of RPC

18
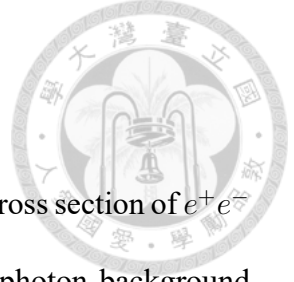
1.5.5   Trigger

Due to the high level of beam induced background and the large cross section of e^+e^- interaction that does not produce $\Upsilon(4S)$, such as Bhabha and two-photon background events, a well-performed trigger system is needed in order to veto these background events with low multiplicity (with fewer tracks and clusters than the $\Upsilon(4S)$/continuum events) and to select the events that we are interested in. We list several processes with their trigger rates and cross sections at the Belle II nominal luminosity of 8×10^{35} cm^{-2}s^{-1} in Table 1.2. Unlike the $B\overline{B}$ events, the Bhabha and the two-photon events are not physics processes that we are interested in. However, they are essential for the measurement of integrated luminosity. Instead of being directly vetoed, the trigger rate of these events are pre-scaled down by a factor of more than 100 to reduce the data size.

The trigger system consists of 4 sub-detector triggers and one Global Decision Logic (GDL). The four sub-detectors that participate in the trigger system are: CDC, TOP, ECL and KLM. CDC provides track finding results, ECL provides information of deposit energy and clusters, TOP provides timing information, and KLM provides information of muons. The information of the sub-detector triggers will be passed to GDL, and then

Physics process	Cross section (nb)	Rate (Hz)
$\Upsilon(4S) \to B\overline{B}$	1.2	960
Hadron production from continuum	2.8	2200
$\mu^+\mu^-$	0.8	640
$\tau^+\tau^-$	0.8	640
Bhabha ($\theta_{\mathrm{lab}} \geq 17°$)	44	350[a]
$\gamma\gamma$ ($\theta_{\mathrm{lab}} \geq 17°$)	2.4	19[a]
2γ processes ($\theta_{\mathrm{lab}} \geq 17°, p_t \geq 0.1$ GeV)	~ 80	~ 15000
Total	~ 130	~ 20000

[a] rate is pre-scaled by a factor of 1/100

Table 1.2: The total cross sections and trigger rates of various physics processes at $L = 8 \times 10^{35}$ cm^{-2}s^{-1} at the energy of $\Upsilon(4S)$ resonance[5]

GDL will make the final decision whether the event is triggered or not. For example, an event must consist of more than 3 tracks in CDC and 4 isolated ECL clusters and must have deposit energy larger than 1.0 GeV in ECL to be triggered by GDL as an $\Upsilon(4S)$/ continuum-like event.
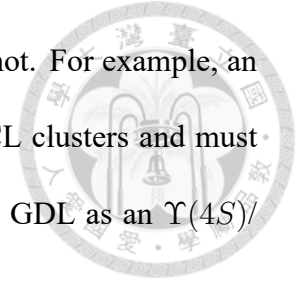
1.6 Final State Particles Reconstruction and Identification in Belle II

In this section, the procedure to reconstruct and identify individual particles by the information from the readout of the sub-detectors is described.

1.6.1 Charged Particle Reconstruction

The content of this sub-section is based on Ref. [17].

The four momentum of a charged particle is reconstructed by performing a track fit combining the hits and the clusters information in PXD, SVD and CDC. Tracks that come from the vicinity of IP are identified by the CDC global track finding algorithm, while the short tracks and tracks that are distant from IP are identified by the CDC local track finding algorithm. Low momentum tracks with high curvature are reconstructed by the standalone SVD track finder. The CDC track finding results are combined with the information of PXD and SVD. They are then fitted with K^+, π^+, and p^+ hypotheses in the final stage.

1.6.2 Charged Particle Identification

Identification of charged particles is done by combining the information from the sub-detectors. These information includes:

- **dE/dx:** dE/dx is defined as the energy loss per distance. According to the Bethe formula, the mean energy loss per distance in the tracking system depends on the particle velocity β. Different particles with the same momentum would have different velocity and therefore different dE/dx. The value of dE/dx is useful especially for identifying low energy particles which are not motive enough to leave the tracking system and to leave hits or clusters in farther sub-detectors.

- **Cherenkov light in TOP and ARICH:** In order to separate K^+s from π^+s that are hard to be distinguished by the tracking system, the ECL, or the KLM, the Cherenkov light angles of these charged particles measured in TOP and ARICH are used for particle identification. The angles follow Eq. 1.1, which is useful to identify particles with known momentums and different masses.

- **Shower pattern in ECL:** All charged particles shower and decelerate in ECL. However, only the electrons lose all of their energy and are stopped in ECL. This is because the bremsstrahlung effect, dominating the process of shower, is inversely proportional to the mass of the particle (see page 18 of Ref. [18]). The mass of an electron is much smaller than the mass of other charged final-state particles decaying from B. Therefore, electrons are the only particles of which the energy loss is dominated by bremsstrahlung and that can be stopped by ECL. The difference of the shower pattern between electrons and other charged particles can be used for electron identification.

- **Hits pattern in KLM:** The μ^+s and the long lived charged hadrons have different scattering pattern in the KLM detector. Thus these particles can be distinguish using the information of KLM.

1.6.2.1 Matching Between ECL (KLM) and CDC

In the ECL (KLM), charged and neutral particles can both produce cluster signals. To know whether the clusters are produced by charged or neutral particles, trajectories of charged particles are extrapolated to ECL (KLM) to see if any of the clusters in the ECL (KLM) can be matched with these tracks.

1.6.3 Global PID of Charged Particles

From the information extracted from the sub-detectors, each charged track is given six particle likelihood values for the six types of charged final states:

$$L(\mathrm{i}) = \prod_{\mathrm{j}=\{\text{subdetectors}\}} L_{\mathrm{j}}(\mathrm{i}) \tag{1.3}$$

, where j runs through all the sub-detectors and i is one of the six mass hypotheses: (e^+, μ^+, π^+, K^+, p^+, d^+). The global PIDs for the six hypotheses are defined as below:

$$\mathrm{i\text{-}ID} = \frac{L(\mathrm{i})}{\sum_{\mathrm{i}'=e^+,\mu^+,\pi^+,K^+,p^+,d^+} L(\mathrm{i}')} \tag{1.4}$$

. The values of the global PIDs of each charged track are used for determining the particle types. For example, if a track has a high π-ID, it means that the track is more likely to be a π^+ than other types of particles.

22

1.6.4 Photon Identification

An ECL cluster is identified to be "neutral" if the cluster is not matched with any track in the tracking system. These ECL neutral clusters are identified as photon candidates. The photon momentum can be calculated from the deposit energy in ECL.

1.6.5 K_L^0 Identification

Neutral KLM clusters which are not matched with any track are identified as K_L^0 candidates. We should note that KLM is not able to measure the energy and momentum of K_L^0s.

1.7 Reconstruction of Intermediate States and the B Meson in Belle II

In a typical measurement of the branching fraction of a certain B decay mode, the four-momentum of the final states is added up in order to reconstruct the intermediate states (like D^0, which can decays to kaons and pions) and the B meson. For the reconstructed B meson, the beam-constrained mass (M_{bc}) and the energy difference (ΔE) can be defined:

$$M_{\text{bc}} = \sqrt{\left(E_{\text{beam}}^*/2\right)^2 - \left(p^*\right)^2} \tag{1.5}$$

$$\Delta E = E^* - \left(E_{\text{beam}}^*/2\right) \tag{1.6}$$

, where E^*_{beam} is the beam energy in the CM frame, p^* is the reconstructed CM-frame momentum of the B meson, and E^* is the reconstructed CM-frame energy of the B meson. According to momentum conservation, the energy of the B meson should be equal to half of the e^+e^- total energy in the CM frame. Thus, for a correctly reconstructed B meson, M_{bc} should equals to the nominal B mass, and ΔE should be zero. M_{bc} and ΔE are the most commonly used variables to identify the signal for B decay analyses that does not consist of missing particles like neutrinos. For decay modes that consist of neutrinos, which cannot be detected by the Belle II detector, other variables like momentum of final state particles or methods like the Full Event Interpretation (FEI)[19] are used for identifying the signal.

2 Motivation

It is an important task to measure the $B^+ \rightarrow l^+\nu$ branching fractions in Belle II since these channels are relatively precise in theoretical predictions and feasible for probing new physics beyond SM.

The SM prediction for the $B^+ \rightarrow l^+\nu_l$ branching fraction is:

$$\mathcal{B}\left(B^+ \rightarrow l^+\nu_l\right) = \frac{G_{\mathrm{F}}^2 m_B m_l^2}{8\pi}\left(1 - \frac{m_l^2}{m_B^2}\right) f_B^2 \left|V_{ub}\right|^2 \tau_B \qquad (2.1)$$

, where G_{F} is the Fermi coupling constant, m_B and m_l stand for the mass of the B meson and the mass of the final-state lepton respectively, f_B is the decay constant, V_{ub} is one of the CKM matrix element, and τ_B is the lifetime of the B meson.

The Feynman diagram of the SM tree-level $B^+ \rightarrow l^+\nu_l$ is illustrated in Fig 2.1:
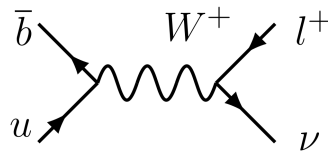
Figure 2.1: Feynman diagram of the tree-level $B^+ \rightarrow l^+\nu_l$

The theoretical branching fractions for the three $B^+ \rightarrow l^+\nu_l$ modes are listed in Table 2.1. The branching fractions of the $\mu^+\nu_\mu$ and $e^+\nu_e$ modes are much smaller than that of the $\tau^+\nu_\tau$ mode due to the helicity suppression. However, the $\tau^+\nu_\tau$ mode suffers

25

mode	$\mathcal{B}\left(B^+ \to l^+ \nu_l\right)$
$B^+ \to \tau^+ \nu_\tau$	$(9.59 \pm 1.28) \times 10^{-5}$
$B^+ \to \mu^+ \nu_\mu$	$(4.31 \pm 0.57) \times 10^{-7}$
$B^+ \to e^+ \nu_e$	$(1.01 \pm 0.13) \times 10^{-11}$
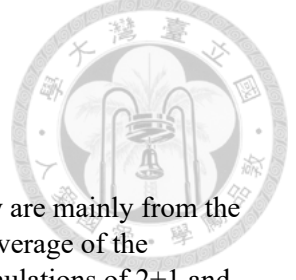
Table 2.1: $\mathcal{B}\left(B^+ \to l^+ \nu_l\right)$ theoretical branching fractions. The uncertainty are mainly from the CKM matrix element V_{ub} and the decay constant f_B. For V_{ub} we used the average of the inclusive and exclusive results, and for f_B we use the average of LQCD simulations of 2+1 and 2+1+1 dynamical quark flavors. The values of the parameters for calculation of the branching fractions are mostly from Ref. [9]

from multi-neutrinos in the final state and has to be identified by tagging the accompanying B meson, hadronically or semileptonically, resulting a low identification efficiency $(O(10^{-3}))$. Compared to the $\tau^+ \nu_\tau$ mode, the $\mu^+ \nu_\mu$ and $e^+ \nu_e$ modes are much "cleaner", with large missing energy and a high-momentum charged track in the final states, enabling signal identification with untagged method. While the expected branching fraction of $B^+ \to e^+ \nu$ is too small to be measured even in Belle II, a measurement with 2.8 sigma significance for $B^+ \to \mu^+ \nu$ has been reported by the Belle collaboration using 772 million $B\overline{B}$ pairs[20]. If what Belle reports is not due to the upper fluctuation, we expect to observe the decay $B^+ \to \mu^+ \nu$ using a sample of 6 ab^{-1}[21].

New physics (NP) may contribute to the decay $B^+ \to l^+ \nu$ and deviate the corresponding branching fractions predicted by SM. It was shown that in the type II of the Two Higgs Doublet Model (2HDM-II), the W^+ boson can be replaced with a charged Higgs H^+[22], and the branching fractions of $B^+ \to \tau^+ \nu$ and $B^+ \to \mu^+ \nu$ can be both enhanced with the ratio between them unchanged. Further more, in a general 2HDM (for example, G2HDM), which allows extra Yukawa Couplings, the ratio between $\mathcal{B}\left(B^+ \to \tau^+ \nu\right)$ and $\mathcal{B}\left(B^+ \to \mu^+ \nu\right)$ might deviate from SM[23]. Other supposed new physics particles, such as the S_1 leptoquark[24], might also change the prediction of $B^+ \to \mu^+ \nu$ and $B^+ \to \tau^+ \nu$.

In this work, we use a large amount of simulated data sample to estimate the amount

of data needed in the future for the discovery of $B^+ \to \mu^+\nu$ in Belle II. We also provide side-band plots to compare the simulated sample and the collected data.

3   Data Samples

3.1   Monte-Carlo Samples

Simulated data samples that are generated by Monte-Carlo (MC) include the signal samples ($B^+ \to \mu^+\nu$) and the background samples (other B decay channels or other e^+e^- interactions which do not produce $\Upsilon(4S)$). The generated samples are listed here:

Signal samples:

- **$B^+ \to \mu^+\nu$:** The samples are generated using the EvtGen[25] package. First, a B^+B^- pair will be generated by EvtGen. Then, one of the B meson, denoted as the "signal B", will decay to $\mu^+\nu$, and the other B, denoted as the "companion B", will decay into some certain final state other than $\mu^+\nu$. The probability for a B meson to decay into a certain final state is based on its branching fraction. Thousands of B decay channels are available in the EvtGen of Belle II. These channels include but not limit to $b \to c$ transition, $b \to u$ transition with hadronic final states, and $B \to X_s\gamma$. The available channels with their branching fractions are listed in the Belle II Decay Table. The decay of the "companion B" in the EvtGen must follow the channels and the branching fractions listed in the Belle II Decay Table. One should note that $B^+ \to \mu^+\nu$ is NOT in the Decay Table.
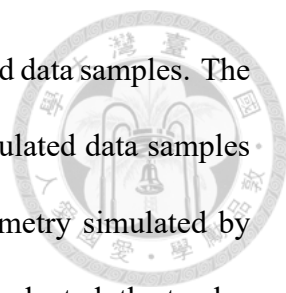
sample	statistics
$B^+ \to \mu^+ \nu$	1230 ab^{-1} (6×10^5 events)
$B^+ \to X_u l \nu$	10.0 ab^{-1}
$B^0 \to X_u l \nu$	11.4 ab^{-1}
Other $B\bar{B}$	1 ab^{-1}
$e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$ and $c\bar{c}$	1 ab^{-1}
$e^+e^- \to \tau^+\tau^-$	1 ab^{-1}

Table 3.1: MC statistics

Background samples:

- **Generic $B\bar{B}$:** The samples are generated using the EvtGen package. In each collision event, either a B^+B^- or a $B^0\bar{B}^0$ pair is generated. The decay of each of the B mesons will follow a certain channel written in the Belle II Decay Table according to its branching fraction.

- **$B \to X_u l \nu$:** The samples are generated using the EvtGen package, with one of the B meson decays to $X_u l \nu$ and the other decays into some final state written in the Belle II Decay Table according to its branching fraction.

- **$e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$ or $c\bar{c}$:** These samples are generated by the KKMC generator[26]. In most of the time, the collision of e^+e^- will not produce $\Upsilon(4S)$ that can decays to $B\bar{B}$. In the $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$ or $c\bar{c}$ case, a pair of quarks is directly produced by the e^+e^- collision. The fragmentation process of the quark is simulated by PYTHIA8 and the decays of the fragmentation products are simulated by EvtGen.

- **$e^+e^- \to \tau^+\tau^-$:** These samples are generated by the KKMC generator. In this case a $\tau^+\tau^-$ pair is produced by the e^+e^- collision. The $e^+e^- \to \tau^+\tau^-$ and the $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$ or $c\bar{c}$ samples are collectively called "continuum background"

The statistics of each MC simulated data sets are listed in Table. 3.1. In each collision

event, simulated beam induced background is overlaid on the simulated data samples. The beam induced background along with the MC particles from the simulated data samples will be together detected by a virtual Belle II detector with its geometry simulated by the Geant4 software. After reconstruction of tracks and clusters is conducted, the tracks, clusters, PID likelihood, and the corresponding generator-level MC particles of the tracks and the clusters will be stored into mini data summary tape (mDST) format.

3.2 $B \to X_u l\nu$ MC Samples and the Hybrid Model

The $B \to X_u l\nu$ (or $b \to u l\nu$ in the quark level) processes, where X_u denotes a $u\bar{u}$ or a $u\bar{d}$ system, is the dominant background of the study of $B^+ \to \mu^+\nu$. Therefore, they need to be treated carefully. The inclusive $B \to X_u l\nu$ and the resonant $B \to \pi, \rho, \eta, \eta', \omega l\nu$ samples are generated separately. The inclusive samples are generated using the BLNP[27] decay model. For resonant samples, π, ρ and $\omega l\nu$ are generated using the Bourrely-Caprini-Lellouch (BCL) model[28], while η and $\eta' l\nu$ are generated using the ISGW2 model.

In order to prevent double counting the resonant part, the inclusive MC samples were reweighted before being combined with the resonant MC samples. All the MC samples are binned on (q^2, E_l^B, m_X) as follows:

$$m_X : [0., 1.4, 1.6, 1.8, 2., 2.5, 3., 3.5],$$

$$E_l^B : [0., 0.5, 1., 1.25, 1.5, 1.75, 2., 2.25, 3.],$$

$$q^2 : [0., 2.5, 5., 7.5, 10., 12.5, 15., 20., 25.]$$

The decision of binning is based on the fact that in each bin the inclusive decay can be well described by the decay model. In each bin, the inclusive MC samples are reweighted

and combined with the resonant MC samples. The weight of the inclusive MC samples in

each bin is calculated by:

$$w_{i} = \frac{I_{i} - R_{i}}{I_{i}} \tag{3.1}$$

, where R_{i} is the expected number of events from the resonant decay and I_{i} is the expected

total number of events predicted by the standalone inclusive decay model in the ith bin.

Finally, the Hybrid Model for $B \rightarrow X_u l \nu$ is created by:

$$\mathbf{H} = \mathbf{wI} + \mathbf{R} \tag{3.2}$$

, where $\mathbf{H}$, $\mathbf{I}$ and $\mathbf{R}$ are the hybrid, inclusive and resonant models, and $\mathbf{w}$ is the weight.

3.3   Real Data Samples

34.58 ab^{-1} of real collision data collected by Belle II during 2019 and the first half

of 2020 is used to compare with the MC samples.
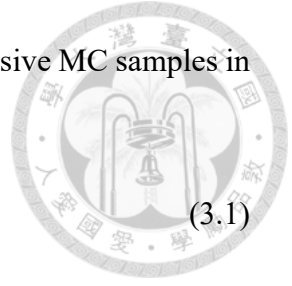
4 Analysis Strategy and Event Selection

4.1 Analysis Strategy

$B^+ \to \mu^+\nu$ is a two-body decay channel, with a high momentum μ^+ and a missing ν. Due to the rareness of high momentum leptons in the $B\overline{B}$ events, a powerful skim can be applied to suppress most of the generic $B\overline{B}$ background (described in Sec. 4.2.1). The μ^+ momentum of $B^+ \to \mu^+\nu$ in the B^+ rest frame, denoted as p_μ^B, can be calculated from momentum conservation,

$$p_\mu^B = \frac{m_B^2 - m_\mu^2}{2m_B} = 2.639 \text{ GeV} \tag{4.1}$$

, where m_B and m_μ are the masses of B^+ and μ^+ respectively. However, p_μ^B cannot be directly measured by the tracking system. What can be directly measured is the CM-frame momentum p_μ^*, which is a flat distribution between 2.45 and 2.85 GeV (see Fig. 4.1a). In this analysis, the CM-frame momentum is boosted back to the B^+ rest frame by reconstructing the companion B^+ from the Rest Of Event (ROE). After the boosting procedure, the shape of the resulting p_μ^B, peaking at 2.639 GeV, can be employed to discriminate the signal and the background and to extract the signal yield (see Fig. 4.1b).
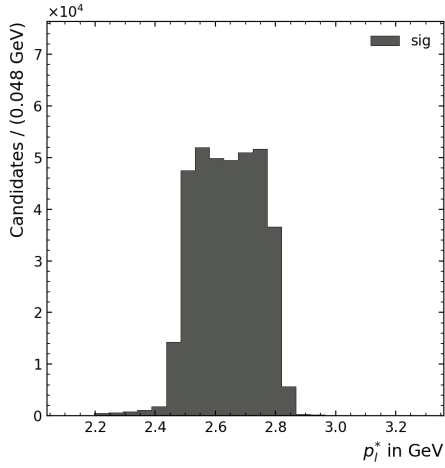
(a) Signal μ^+ momentum in the CM frame
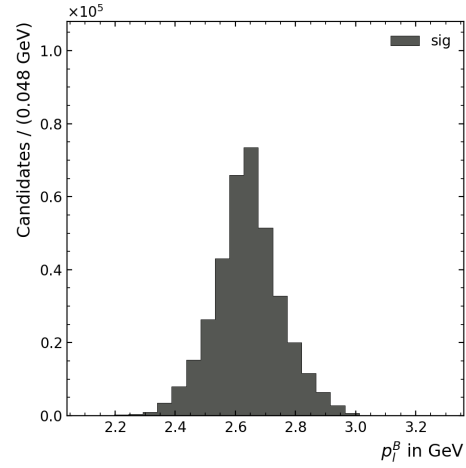
(b) Signal μ^+ momentum in the B rest frame

4.2   Event Selection

4.2.1   Skim

Due to the unique kinematic feature of $B^+ \to \mu^+ \nu$ mentioned in Sec. 4.1, a `LeptonicUntagged` skim is applied to select the high momentum lepton candidates, resulting in a background reduction rate of around $97\% \sim 98\%$. In the `LeptonicUntagged` skim, we require that in each event at least one charged particle pass the following criteria:

- μ-ID > 0.5 or e-ID > 0.5

- $p_l^* > 2.0$ GeV

The charged particles that pass the above criteria are selected as the signal lepton candidates of $B^+ \to \mu^+ \nu$.

To suppress the low multiplicity and the Bhabha events, we also require the number of tracks in each event to be more than two.

particle type	cut
charged particles	$p^* < 3.2$ GeV
	$N_{\text{CDCHits}} > 0$
	$d_0 < 10$ cm and $|z_0| < 10$ cm
ECL clusters	$p > 0.05$ GeV and $p^* < 3.2$ GeV
KLM clusters	No selection

Table 4.1: ROE selection

4.2.2 Rest of Event

For each signal lepton candidate, we exclude the candidate itself and define a rest of event (ROE) object as the remaining charged tracks, ECL clusters and KLM clusters in an event that pass the selection criteria in Table 4.1. The ECL and KLM clusters in the ROE are respectively identified as photons and K_L^0s in default. The charged particles in the ROE are identified and classified using the following criteria:

if μ-ID > 0.5 and `isFromIP` and `isGoodTrack`:

muon;

else if e-ID > 0.5 and `isFromIP` and `isGoodTrack`:

electron;

else if p-ID > 0.5 and `isFromIP` and `isGoodTrack`:

proton;

else if K-ID > 0.5 and `isFromIP` and `isGoodTrack`:

kaon;

else:

pion;

The definition of `isFromIP` and `isGoodTrack` is summarized in Table 4.2.

34

variable	cut
isFromIP	$d_0 < 0.5$ cm and $|z_0| < 2$ cm
isGoodTrack	$N_{\text{CDCHits}} > 20$ and isInCDCAcceptance

Table 4.2: Definition of isFromIP and isGoodTrack

4.2.2.1 Derivation of p_μ^B

In each event, if the lepton candidate we selected is really the signal μ^+ from $B^+ \rightarrow \mu^+\nu$, then, ideally, we can assume that all the tracks and the clusters in the ROE form the companion B meson. According to momentum conservation, The direction of the CM-frame momentum of the companion B should be opposite to the direction of the CM-frame momentum of the signal B. That is to say, we can know the direction of the signal B momentum by the formula

$$\hat{p}^*_{\text{ROE}} = \hat{p}^*_{B,\text{companion}} = -\hat{p}^*_{B,\text{signal}} \tag{4.2}$$

. The magnitude of the signal B^+ momentum is set to $\left|\vec{p}^*_{B,\text{signal}}\right| = 0.332$ GeV, which is calculated by employing the momentum conservation condition,

$$|\vec{p}^*_B| = \sqrt{(\frac{E^*_{\text{beam}}}{2})^2 - m_B^2} = 0.332 \text{ GeV} \tag{4.3}$$

, where E^*_{beam} is the CM-frame beam energy and m_B is the nominal mass of B^+. Once the direction and the magnitude of the signal B^+ momentum is known, we can boost p_μ^* back to p_μ^B by Lorentz Transformation (see Fig. 4.1a and Fig. 4.1b).

4.2.3 Pre-Selection

After the skim, more stringent selection criteria are applied on the $B^+ \rightarrow \mu^+\nu$ candidates:

- **μ-ID> 0.95:**

  A stringent μ-ID cut is set to reject fake π^+s that are mostly from the continuum background.

- **$d_0 < 0.5$ cm and $|z_0| < 2$ cm:**

  d_0 and z_0 are the transverse distance and the z coordinate of the point of closest approach of the tracks with respect to IP. The selections on d_0 and z_0 ensure that the tracks come from IP.

- **$2.2 < p_\mu^* < 3.3$ GeV:**

  The CM-frame μ^+ momentum is limited in the region between 2.2 and 3.3 GeV.

- **$N_{\text{Tracks,highP}} == 1$:**

  We require that there is only one high momentum track within each event. The high momentum tracks are defined to be tracks with $p^* > 2.2$ GeV, $d_0 < 10$ cm and $|z_0| < 10$ cm.

- **$M_{\text{bc}}^{\text{ROE}} > 5.1$ GeV:**

  The $M_{\text{bc}}^{\text{ROE}}$ is defined by $\sqrt{\left(E_{\text{beam}}^*/2\right)^2 - \left(p^{*\text{ROE}}\right)^2}$, where E_{beam}^* is the CM-frame beam energy and $p^{*\text{ROE}}$ is the total CM-frame momentum of the ROE.

- **$-2 < \Delta E^{\text{ROE}} < 2$ GeV:**

  The ΔE^{ROE} is defined by $E^{*\text{ROE}} - (E_{\text{beam}^*}/2)$, where $E^{*\text{ROE}}$ is the total CM-frame energy of the ROE.

4.2.4 Continuum Suppression

On the one hand, because the B mesons are spin-0 particles having low momentum (≈ 0.3 GeV), the event shape of $B\bar{B}$ events is spherical. On the other hand, because the

36

momentum of the q and the τ^+ from the $e^+e^- \to q\overline{q}, \tau^+\tau^-$ interactions are high, the event

shape is more jet-like (see Fig. 4.2). Because of the difference of the event shape between
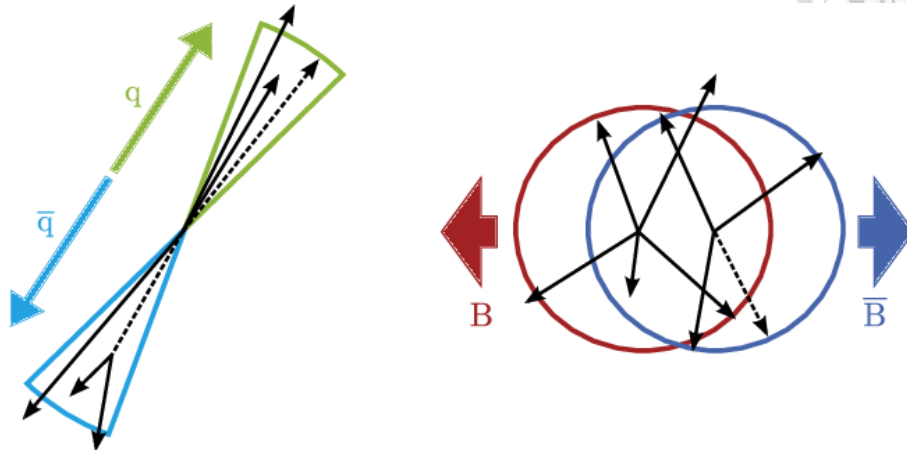
Figure 4.2: Spherical-like $B\overline{B}$ events and jet-like continuum events[7]

the $B\overline{B}$ events and the continuum events, we can use some "shape variables", including

KSFW, R2, Thrust and etc., to distinguish the $B\overline{B}$ events from the continuum events.

In addition, the distributions of the dilution factor of the flavor tagging result (described

below) differ between the $B\overline{B}$ events and the continuum events. Therefore, the variable

can also be used to distinguish continuum from $B\overline{B}$. We use the default MVA package

of Belle II, the FastBDT[29], to train the signal against the continuum background. The

variables used in the FastBDT are:

- **Cleo Cones:**

  The Cleo Cones variables are the momentum sums in every 10 degrees of the polar

  angle, where the momentum direction of the signal μ^+ is defined as the zero degree

  of the polar angle. The μ^+ momentum is not included in the momentum sums.

  Otherwise, the momentum sum in the cone containing the signal μ^+ would have

  large correlation with p^*_μ.

- **KSFW moments:**

  The KSFW moments are several variables describing the event shape. See Ch. 9 of

Ref. [30] for details.

- **R2:**

  R2 is defined by $\frac{\sum_{i,j} |p_i^*| |p_j^*| P_2\left(\cos\left(\theta_{ij}\right)\right)}{\sum_{i,j} |p_i^*| |p_j^*|}$, where P_2 is the second-order Legendre polynomial and i, j iterate over all the particles in both the signal and the ROE side.

- **ThrustO:**

  ThrustO is the magnitude of the thrust axis of ROE. the thrust axis is defined to be the axis $\vec{n}$ that maximize $T(\vec{n}) = \frac{\sum_i |\vec{p_i} \cdot \vec{n}|}{\sum_i |\vec{p_i}|}$, where i iterates over all the particles in ROE.

- **CosTBTO:**

  CosTBTO is the cosine of the angle between the thrust axis of the signal B (in this case it is the momentum direction of the signal μ^+) and the thrust axis of ROE.

- **CosTBz:**

  CosTBz is the cosine of the angle between the thrust axis of the signal B (in this case it is the momentum direction of the signal μ^+) and the beam axis.

- $q_\mu \times qr_{\text{out}}$**:**

  qr_{out} is defined as the flavor tagging result q (1 for B^+ or B^0, and -1 for B^- or $\overline{B}^0$) multiplied by the dilution factor r. The flavor tagger is a MVA based algorithm that is originally used for tagging the flavor type of the companion B meson (distinguishing B^0 from $\overline{B}^0$) in analyses relative to time dependent CP violation of the B meson. Interestingly, it was found that the dilution factor of the flavor tagging result is a useful variable for continuum suppression. For continuum events, the dilution factor r should be close to zero. As for signal events, it should be close to 1. On the one hand, for a perfectly-tagged signal event, the charge of the signal muon, q_μ,

multiplied by qr_{out} should equals to -1. On the other hand, for a continuum event, $q_\mu \times qr_{\text{out}}$ should be close to zero.

The distribution of the continuum suppression variables are shown in Appendix C.1

4.2.5 $B \to X_u l \nu$ **Suppression**

The $B \to X_u l \nu$ decay channels are the second largest background in this analysis, preceded only by the continuum background. In order to suppress the $B \to X_u l \nu$ background, we inclusively reconstruct several $B \to X_u l \nu$ channels and veto the events with a high "$b \to u l \nu$ probability". The detailed procedure of the suppression is described below.

4.2.5.1   Reconstruction

The reconstructed $B \to X_u l \nu$ channels are listed in Table 4.3. We apply the same algorithm used in the Full Event Interpretation (FEI) [19] to reconstruct and select π^+s and π^0s, except that we do not reconstruct photon conversion ($\gamma \to e^+ e^-$) in the procedure of π^0 reconstruction. The same μ^+ candidates for $B^+ \to \mu^+ \nu$ are selected as the lepton candidates for $B \to X_u l \nu$. At this stage, we allow multiple $B \to X_u l \nu$ candidates to coexist in each event.

4.2.5.2 $B \to X_u l \nu$ **MVA**

For each of the 11 reconstructed $B \to X_u l \nu$ channels, we train a FastBDT to distinguish the MC true $B \to X_u l \nu$ from the fake ones. We used 1.00 ab^{-1} of $B^+ \to X_u l \nu$ and 1.14 ab^{-1} of $B^0 \to X_u l \nu$ MC samples in each BDT. Before training the BDTs, we

doi:10.6342/NTU202100639

reconstructed channels
$B^0 \to \pi^- \mu^+ \nu$
$B^+ \to \pi^+ \pi^- \mu^+ \nu$
$B^0 \to \pi^- \pi^+ \pi^- \mu^+ \nu$
$B^+ \to \pi^+ \pi^- \pi^+ \pi^- \mu^+ \nu$
$B^+ \to \pi^0 \mu^+ \nu$
$B^0 \to \pi^0 \pi^- \mu^+ \nu$
$B^+ \to \pi^0 \pi^+ \pi^- \mu^+ \nu$
$B^0 \to \pi^0 \pi^- \pi^+ \pi^- \mu^+ \nu$
$B^+ \to \pi^0 \pi^0 \mu^+ \nu$
$B^0 \to \pi^0 \pi^0 \pi^- \mu^+ \nu$
$B^+ \to \pi^0 \pi^0 \pi^+ \pi^- \mu^+ \nu$

Table 4.3: 11 reconstructed $B \to X_u l \nu$ channels

classify the reconstructed $B \to X_u l \nu$ candidates into three categories:

- **Well-reconstructed**: in this category, each μ^+, π^+ and π^0 is well reconstructed and matched with MC

- **Self-cross-feed**: in this category, some of the πs are well-reconstructed and matched with MC, while the other πs are not matched with MC.

- **Background**: in this category, none of the πs is matched with MC.

The event topology of "self-cross-feed" samples can be close to either that of "well-reconstructed" samples or that of "background" samples, depending on the number of πs that are matched with MC. As a result, "self-cross-feed" samples are not used in the BDTs. In order to identify the $B \to X_u l \nu$ decays, for each of the 11 modes, the "well-reconstructed" samples are trained against the "background" samples. The training variables are listed below:

- $M_{\mathbf{bc},Y}$:

  $M_{\mathrm{bc},Y}$ is defined by $\sqrt{(E^*_{\mathrm{beam}}/2)^2 - (p^*_Y)^2}$, where p^*_Y is the reconstructed CM-frame momentum of the $X_u l$ system.

- **ΔE_Y:**

  ΔE_Y is defined by $E_Y^* - (E_{\text{beam}}^*/2)$, where E_Y^* is the reconstructed CM-frame energy of the $X_u l$ system.

- **$M_{\text{bc}}^{\text{ROE}}$:**

  See Sec. 4.2.4 for the definition.

- **ΔE^{ROE}:**

  See Sec. 4.2.4 for the definition.

- **$\frac{2E_B^* E_Y^* - M_B^2 - m_Y^2}{2p_B^* p_Y^*}$:**

  This variable is usually used for background suppression in semi-leptonic B decay analyses, where E_B^*, p_B^*, and M_B are the nominal CM-frame energy, the nominal momentum, and the nominal mass of a B meson, respectively; E_Y^*, p_Y^*, and m_Y are the reconstructed CM-frame energy, the reconstructed momentum, and the reconstructed mass of the $X_u l$ system, respectively. For a "well-reconstructed" $B \to X_u l \nu$ decay, this variable should be equal to the cosine of the angle between the signal B meson and the $X_u l$ system; and, thus, the value of the variable should be within the range of [-1,1].

- **p(i):**

  p(i) is the momentum of the ith π daughter.

- **sigProb(i):**

  sigProb(i) is the sigProb of the ith π daughter, where the sigProb is the signal probability derived from the π selection MVA described in Ref. [19].

- **daughterProductOfSigProb:**

  daughterProductOfSigProb is the product of the sigProbs of all the π daughters.
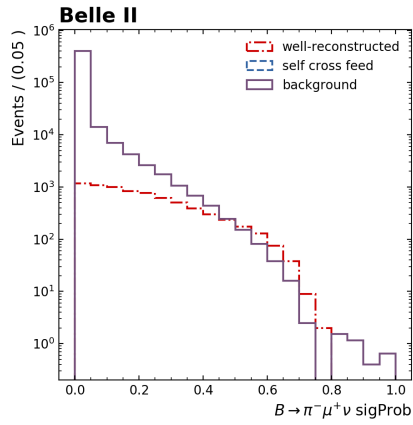
- **$\cos(\theta_{Y,i})$:**

  $\cos(\theta_{Y,i})$ is the cosine angle between the $X_u l$ system and the i$^{\text{th}}$ π, where i iterates over all the π daughters..
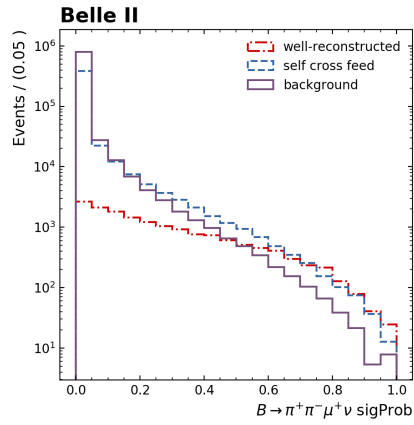
- **$\cos(\theta_{i,j})$:**

  $\cos(\theta_{i,j})$ is the cosine angle between the i$^{\text{th}}$ π and the j$^{\text{th}}$ π daughters, where i and j iterate over all the π daughters.
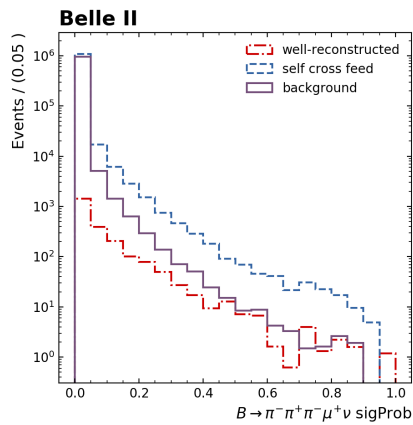
The distributions of the training variables for one of the 11 reconstructed decay channels, the $B^0 \to \pi^0 \pi^- \mu^+ \nu$ channel, are shown in Appendix C.2. After training all the 11 classifiers, there will be a FastBDT output for each candidate in each event. The candidate must belongs to one of the 11 reconstructed modes. The FastBDT outputs of the three categories for the 11 modes are shown in Fig. 4.3. We select the candidate with the highest FastBDT output in each event and define the value of the output to be the "$b \to u l \nu$ probability" of the event. The higher the "$b \to u l \nu$ probability" of an event is, the more the event topology is similar to a $B \to X_u l \nu$ decay. In Fig. 4.3l we compare the distribution of "$b \to u l \nu$ probability" of $B^+ \to \mu^+ \nu$ and $B \to X_u l \nu$ to show that the variable can be utilized to distinguish these two components.

(a) $B \to \pi^- \mu^+ \nu$
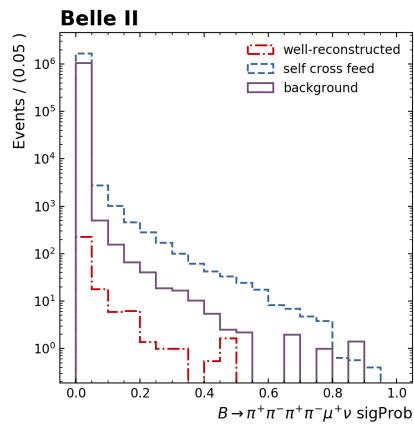
(b) $B \to \pi^+ \pi^- \mu^+ \nu$

(c) $B \to \pi^- \pi^+ \pi^- \mu^+ \nu$

(d) $B \to \pi^+ \pi^- \pi^+ \pi^- \mu^+ \nu$
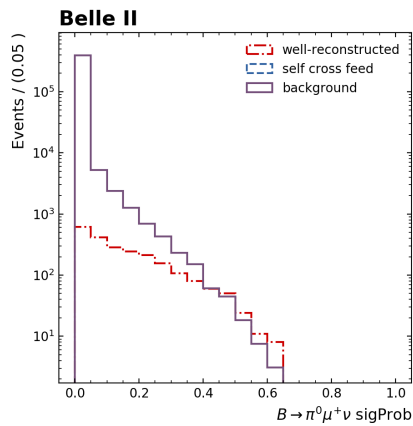
(e) $B \to \pi^0 \mu^+ \nu$
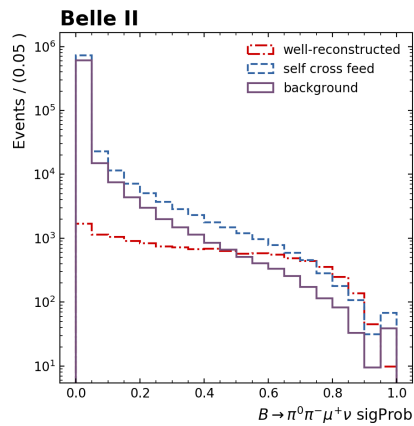
(f) $B \to \pi^0 \pi^- \mu^+ \nu$
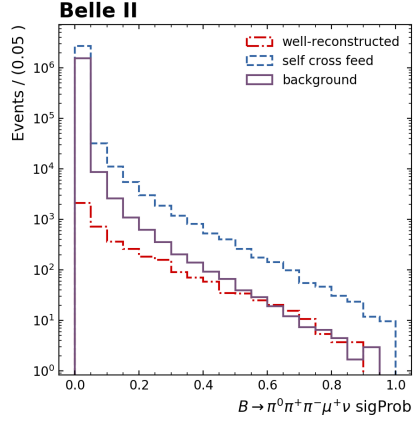
(g) $B \to \pi^0 \pi^- \pi^+ \mu^+ \nu$
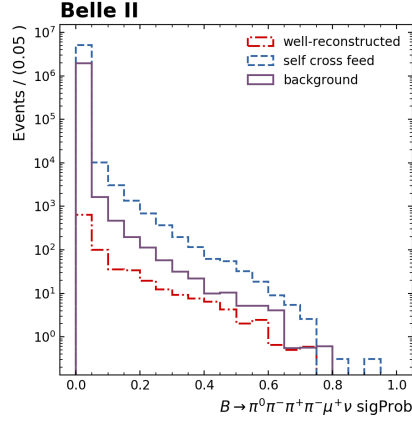
(h) $B \to \pi^0 \pi^- \pi^+ \pi^- \mu^+ \nu$

(i) $B \to \pi^0 \pi^0 \mu^+ \nu$

(j) $B \to \pi^0 \pi^0 \pi^- \mu^+ \nu$
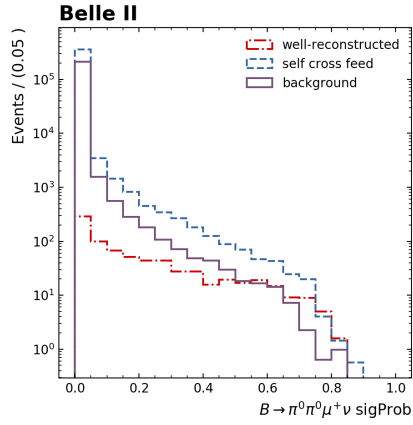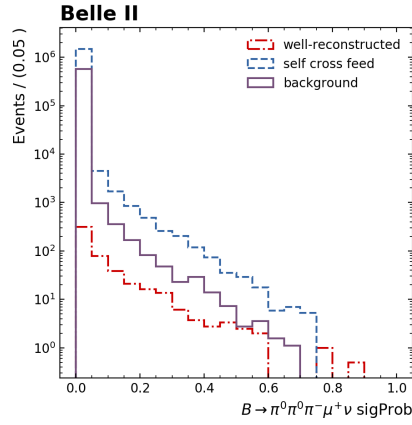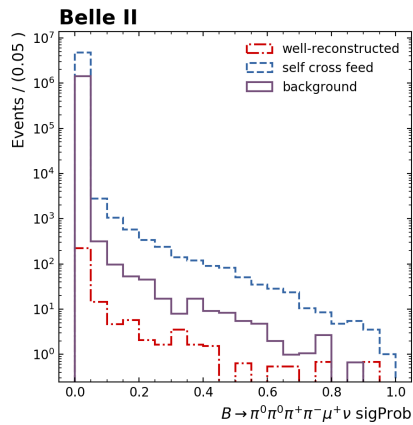
(k) $B \to \pi^0 \pi^0 \pi^+ \pi^- \mu^+ \nu$
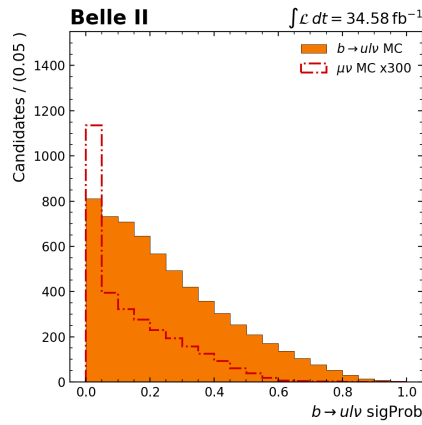
(l) $B^+ \to \mu^+ \nu$ v.s. $B \to X_u l \nu$

Figure 4.3: (a)-(k): BDT output for the 11 reconstructed modes, (l): $b \to ul\nu$ probability of $B^+ \to \mu^+ \nu$ and $B \to X_u l \nu$
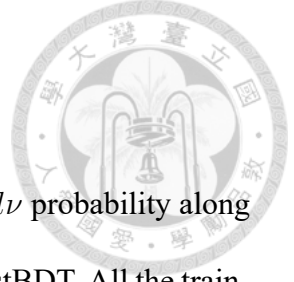
4.2.6 Combined Background Suppression MVA

Utilizing the continuum suppression BDT output and the $b \to ul\nu$ probability along with other variables, we train a combined background suppression FastBDT. All the training variables are listed as follows.

- **$b \to ul\nu$ probability**

- **continuum suppression output**

- **$E_{\mathrm{ECL}}^{\mathbf{ROE}}$**

- **$\Delta E^{\mathbf{ROE}}$**

- **$M_{\mathbf{bc}}^{\mathbf{ROE}}$**

The $E_{\mathrm{ECL}}^{\mathrm{ROE}}$ is defined as the total ECL deposit energy in the ROE.

We define a transformation formula for the BDT output.

$$BDT_{\mathrm{trans}} = \log\left(\frac{BDT_{\mathrm{out}}}{1 - BDT_{\mathrm{out}}}\right) \tag{4.4}$$

We cut on $BDT_{\mathrm{trans}} > 3.8$ to suppress background. The cut is optimized with Figure Of Merit (FOM), defined by

$$\frac{N_{\mathrm{signal}}}{\sqrt{N_{\mathrm{signal}} + N_{\mathrm{background}}}} \tag{4.5}$$

The BDT output distributions before/after the transformation are shown in Fig 4.5a/Fig 4.5b.
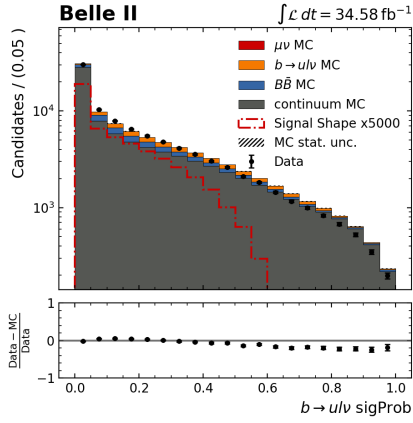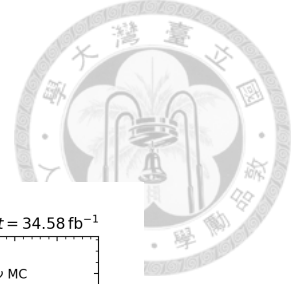
(a) $b \to u l \nu$ probability
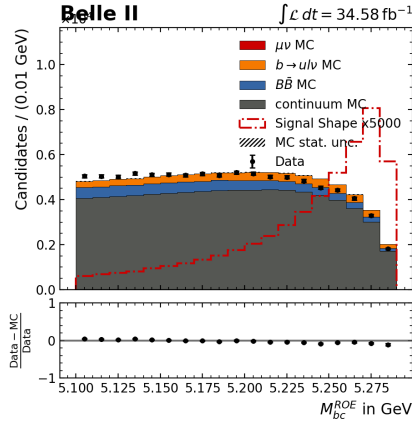
(b) $M_{\mathrm{bc}}^{\mathrm{ROE}}$
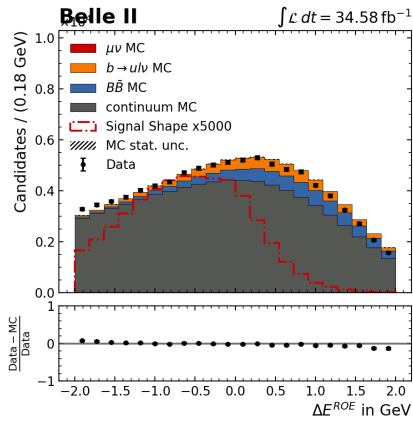
(c) ΔE^{ROE}
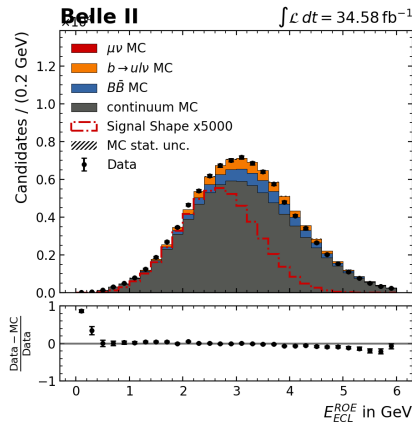
(d) $E_{\mathrm{ECL}}^{\mathrm{ROE}}$
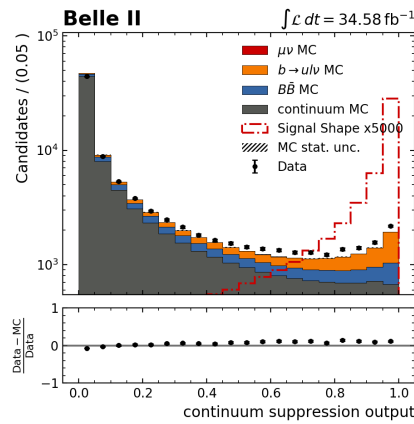
(e) Continuum suppression output

Figure 4.4: Training variables for the combined background suppression MVA
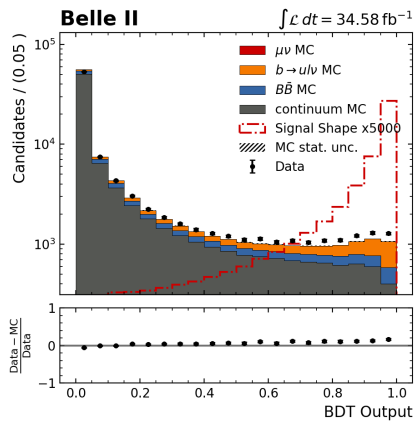
(a) Combined MVA output
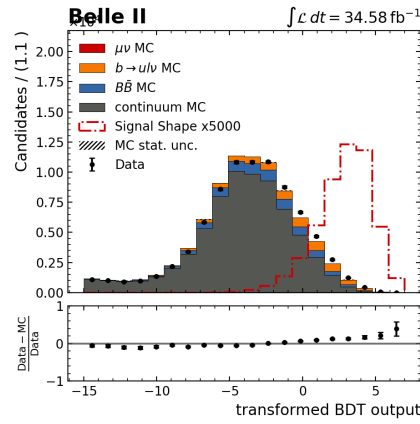
(b) Transformed combined MVA output

Figure 4.5: Output of the combined background suppression MVA

4.3 Calibration of Muon Identification Efficiency

The μ^+ selection efficiency of the μ-ID cut in the MC samples could be different from that in the real data. The fake rate, which is defined as the retention rate of the wrongly-identified μ^+ candidates after the μ-ID selection, could differ between the real data and the MC samples, too. Calibration tables[31] are provided by the Belle II Performance Group in order to calibrate the data-MC discrepancy of the efficiency and that of the fake rate. There are three calibration tables, one is for calibrating the efficiency of correctly identified μ^+s, the other two are used to correct the fake rate of wrongly-identified μ^+s that are actually K^+s or π^+s. To calibrate μ^+ efficiency, correctly identified μ^+s in the MC samples are binned on (p,θ), where p is the lab frame momentum and θ is the polar angle. Then the μ^+s in each bin are weighted by

$$C_i^\mu = \frac{N_{\text{data},i}^\mu}{N_{\text{MC},i}^\mu} \tag{4.6}$$

, where i stands for the bin index and C_i^μ is the calibration factor of bin i in the μ^+ efficiency calibration table. For K^+ or π^+ that are misidentified as μ^+, the K^+/π^+ fake rate table is used and the particles are weighted in the same manner:

$$C_i^\pi = \frac{N_{\text{data,i}}^\pi}{N_{\text{MC,i}}^\pi}, \tag{4.7}$$

$$C_i^K = \frac{N_{\text{data,i}}^K}{N_{\text{MC,i}}^K} \tag{4.8}$$

In Belle II, the measurement of lepton ID calibration factors is conducted by studying $ee \to eell$, $ee \to ee$, $ee \to ll\gamma$, and $J/\psi \to ll$ channels for lepton efficiency calibration, and $K_S^0 \to \pi\pi$, $D^* \to D(K\pi)\pi$, and $ee \to \tau(1\pi)\tau(3\pi)$ for $\pi/K \to$ lepton fake rates. These studies are conducted using a tag-and-probe method described in Ref. [31].

48

4.4 Side-Band Plots

After applying the combined MVA selection on all the MC samples and the real data samples, we compare the p_μ^B distribution of the MC samples with the 34.58 fb^{-1} real data in the side-band region in Fig. 4.6. The definition of the side-band region is:

$$2.2 < p_\mu^B < 2.45 \text{ GeV},$$

$$2.85 < p_\mu^B < 3.2 \text{ GeV}$$

Here, we show two side-band plots on the p_μ^B dimension in the region of $BDT_{\text{trans}} < 4.6$ and $BDT_{\text{trans}} > 4.6$ respectively. We do not define side-band region on the BDT_{trans} dimension due to the broad signal distribution on BDT_{trans}.
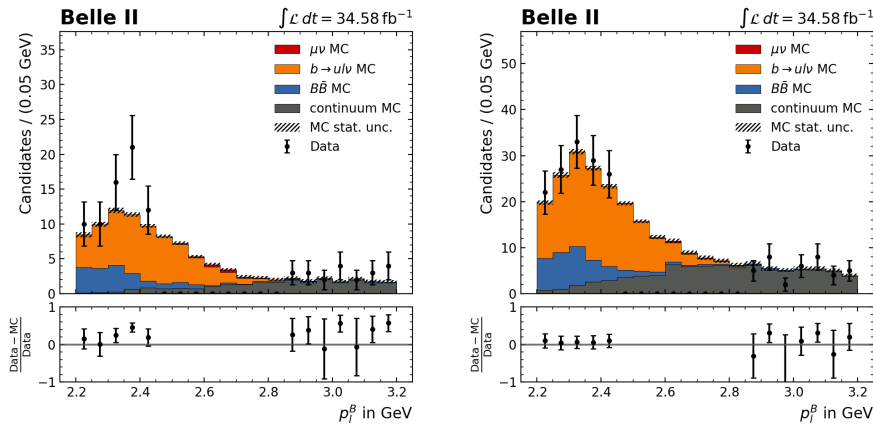
(a) Side-band plot for $BDT_{\text{trans}} > 4.6$ (b) Side-band plot for $BDT_{\text{trans}} < 4.6$

Figure 4.6: Side-band plots for the 2 categories

5 Signal Extraction

In this chapter, the procedure of extracting the signal yield and calculating the branching fraction is elaborated.

We construct a two-dimensional histogram Probability Density Function (PDF) for each of the signal, the $B\bar{B}$, the $B \to X_u l\nu$ and the continuum processes using the MC samples that pass the final selection. The two dimensions of the PDFs are p_μ^B and BDT_{trans}. All the PDFs are constructed in the region of $2.2 < p_\mu^B < 3.2$ and $3.8 < BDT_{\text{trans}} < 12$. On the p_μ^B dimension, the histogram is equally divided into 20 bins. On the BDT_{trans} dimension, the $5.8 < BDT_{\text{trans}} < 12$ region is treated as one bin because of the low statistics in the region, and the $3.8 < BDT_{\text{trans}} < 5.8$ region is equally divided into 10 bins. Thus, there are totally $20 \times (10+1)$ bins for each 2D histogram PDF. The constructed PDFs are shown in Fig. 5.1.
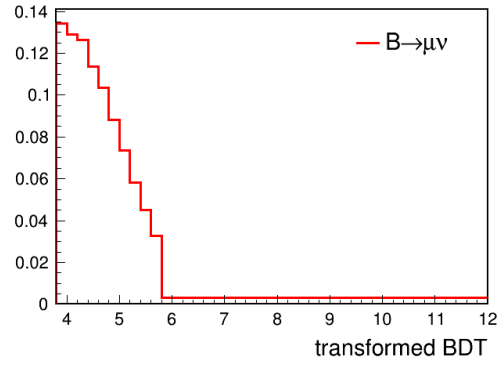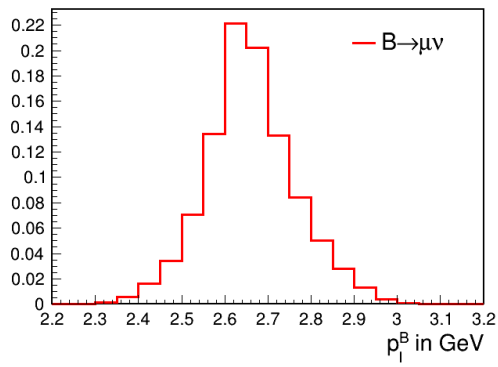
(a) $B^+ \to \mu^+\nu$ PDF model, projected to the p_μ^B and the BDT_{trans} dimensions
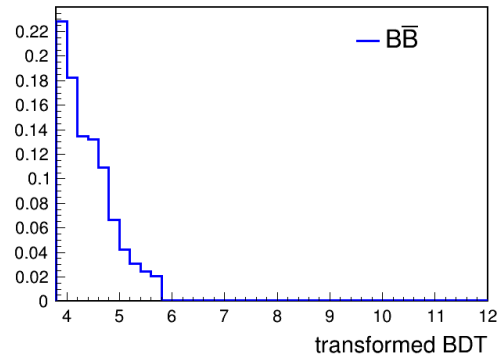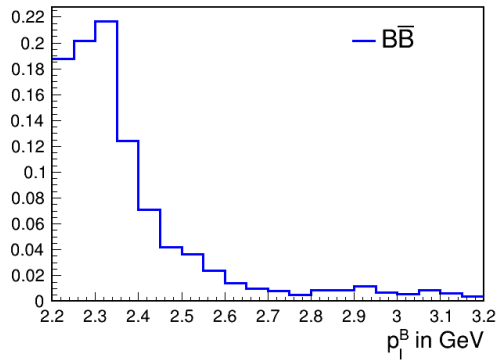
(b) $B\bar{B}$ PDF model, projected to the p_μ^B and the BDT_{trans} dimensions
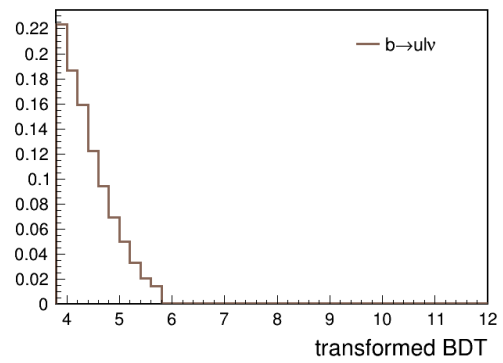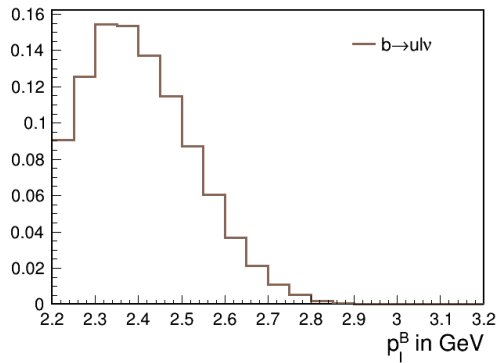
(c) $B \to X_u l\nu$ PDF model, projected to the p_μ^B and the BDT_{trans} dimensions
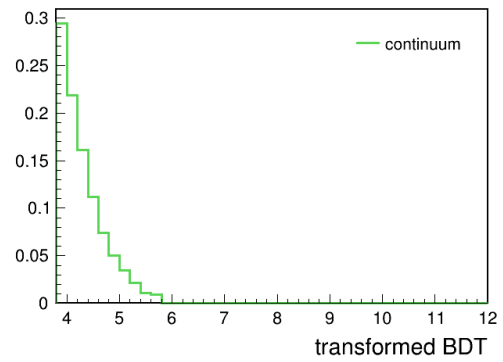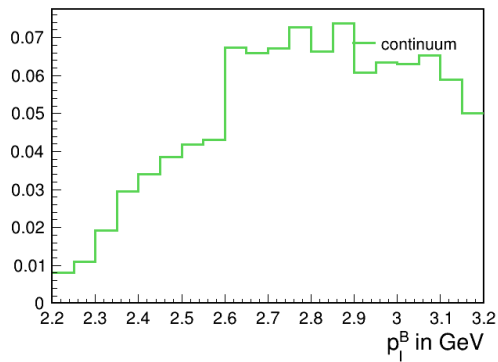
(d) continuum PDF model, projected to the p_μ^B and the BDT_{trans} dimensions

Figure 5.1: PDF models

51

A 2D binned maximum likelihood fitter is constructed based on the four PDFs (signal, $B\bar{B}$, $B \to X_u l\nu$ and continuum) to extract the signal yield. The likelihood function is constructed as

$$L = \prod_{i=1}^{N} \mathcal{P}(n_i | \sum_{k=1}^{4} p_{ik}\theta_k) \qquad (5.1)$$

, where

$k = 1, 2, 3, 4$ labels the signal, the $B\bar{B}$, the $B \to X_u l\nu$ and the continuum processes respectively,

i is the bin index,

n_i is the total number of data in the ith bin,

θ_k is the parameter of interest (number of events of the kth physics process),

p_{ik} is the probability for an event of the kth process to be in the ith bin according to the kth PDF,

$\mathcal{P}$ is the Poisson distributions,

$N = 20 \times 11 = 220$ is the total number of bins.

The likelihood function will be maximized to extract the number of signal, $B\bar{B}$, $B \to X_u l\nu$ and continuum processes ($\theta_1 \sim \theta_4$).
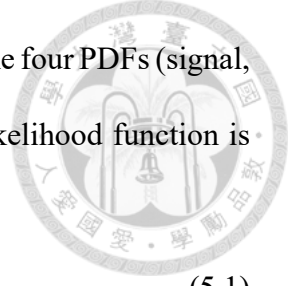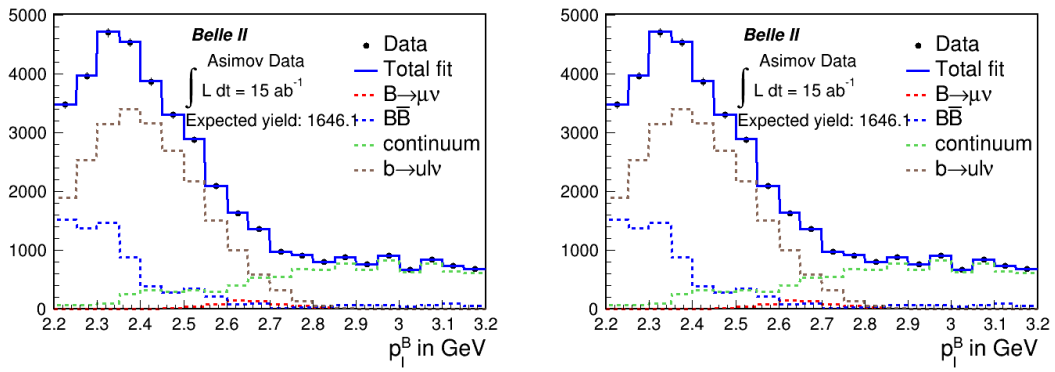
Figure 5.2: Fit to Asimov data

5.1 Fit to Asimov Data

An Asimov data set is defined as a pseudo data set with its distribution exactly the same as the presumed PDF distribution. In this analysis, the four constructed PDFs for the four physics processes are weighted by their expected yield. These weighted PDFs are then superpositioned. Then, 15 Asimov data sets for luminosity of 1 ab^{-1} to 15 ab^{-1} are generated using the resultant PDF. While performing the binned maximum likelihood fit to the Asimov data set, the yields of the signal, the $B\overline{B}$, the $B \rightarrow X_u l \nu$ and the continuum processes should be exactly the same as the values expected by SM. Such feature can be used to check whether the definition of the likelihood function is correct. Furthermore, the likelihood function constructed by the Asimov data sets can be used to estimate the significance of $B^+ \rightarrow \mu^+ \nu$ for each luminosity (detailed explanation is in Sec. 5.3). We only show the result of the fit to the 15 ab^{-1} Asimov data set here in Fig. 5.2. The results of the fits to 1 ab^{-1} to 14 ab^{-1} Asimov data sets are shown in Appendix. C.3. The fitted signal yields and the statistical errors of the Asimov data sets are listed in Table. 5.1.

luminosity	yield and stat. error
1 ab^{-1}	109.7 ± 49.2
2 ab^{-1}	219.5 ± 69.6
3 ab^{-1}	329.2 ± 85.3
4 ab^{-1}	439.0 ± 98.5
5 ab^{-1}	548.7 ± 110.1
6 ab^{-1}	658.5 ± 120.6
7 ab^{-1}	768.2 ± 130.3
8 ab^{-1}	878.0 ± 139.3
9 ab^{-1}	987.7 ± 147.7
10 ab^{-1}	1097.4 ± 155.8
11 ab^{-1}	1207.1 ± 163.4
12 ab^{-1}	1316.9 ± 170.7
13 ab^{-1}	1426.7 ± 177.6
14 ab^{-1}	1536.4 ± 184.3
15 ab^{-1}	1646.1 ± 190.8

Table 5.1: Yields and errors of the Asimov data sets

5.2 Toy MC Test

In order to assure that the fitter is unbiased, toy MC tests for luminosity of 1 ab^{-1} to 15 ab^{-1} are performed. We generate pseudo data sets based on the likelihood function in Eq. 5.1 with $\theta_1 \sim \theta_4$ fixed to the SM prediction. We then fit to these data sets by maximizing the same likelihood function with $\theta_1 \sim \theta_4$ floated. For each luminosity, 1000 data sets are generated and fitted. We define the "pull" of each fitted result as:

$$\frac{\hat{\theta}_{\mathrm{k}} - \theta_{\mathrm{k}}^0}{\sigma_{\mathrm{k}}} \tag{5.2}$$

, where $\hat{\theta}_{\mathrm{k}}$ and θ_{k}^0 are the fitted and expected yield, and σ_{k} is the fitted error of $\hat{\theta}_{\mathrm{k}}$. We only show the signal pull distribution for the 15 ab^{-1} toy data set here in Fig. 5.3. As shown in the plot, for an unbiased fitter, the distribution of pull should be very close to a normal distribution. The signal pull distributions for the 1 ab^{-1} to 14 ab^{-1} data sets are shown in Appendix C.3.
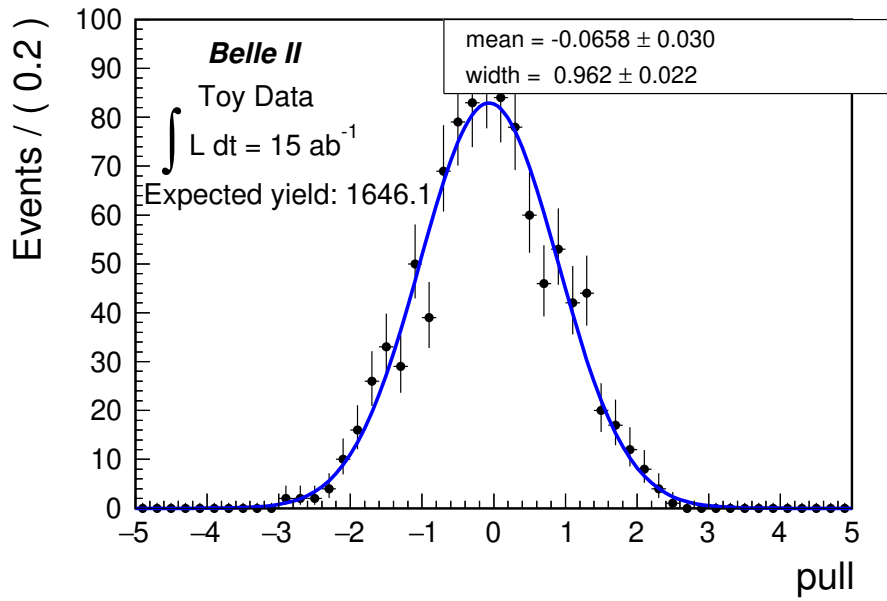
Figure 5.3: Pull distribution

5.3 Significance Estimation

We use the profile likelihood distribution, $L(\theta_1)$, of each Asimov data set to estimate the significance. We let $\theta_2 \sim \theta_4$ float and then maximize the likelihood with different θ_1 values to obtain the distribution. In order to include the systematic uncertainty, the likelihood function is further smeared by a bifurcated Gaussian, $\mathcal{G}_b$. The smeared likelihood function is defined as:

$$L'(\theta_1) = \int_{-\infty}^{\infty} L(\theta_1')\mathcal{G}_b(\theta_1 - \theta_1', \sigma_+, \sigma_-)d\theta_1' \tag{5.3}$$

, where

$$\mathcal{G}_b(\theta_1 - \theta_1', \sigma_+, \sigma_-) = \begin{cases} e^{\frac{-(\theta_1 - \theta_1')^2}{2\sigma_+^2}}, & \text{if } \theta_1 > \theta_1' \\ e^{\frac{-(\theta_1 - \theta_1')^2}{2\sigma_-^2}}, & \text{if } \theta_1 < \theta_1' \end{cases} \tag{5.4}$$

, where $\sigma_{\pm}$ are the total asymmetric systematic uncertainty. Please see Ch. 7 for a detailed discussion for the systematic uncertainties.
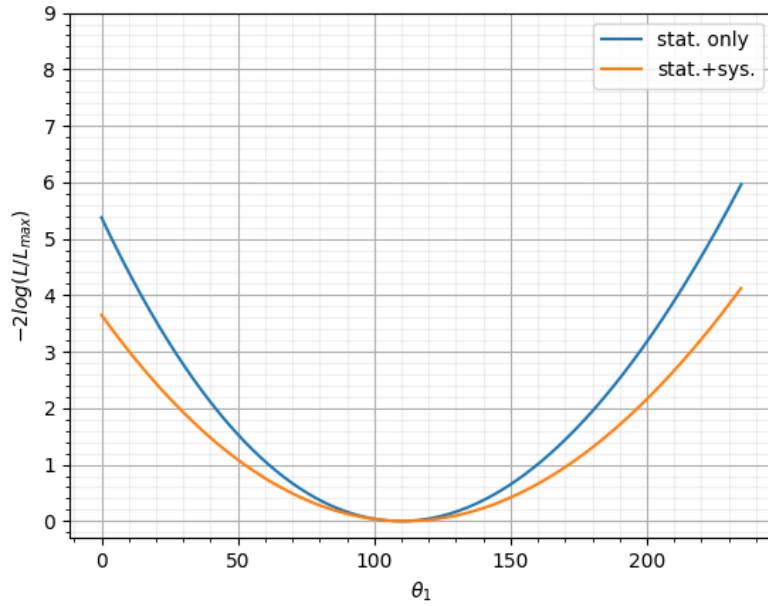
Figure 5.4: The distribution of q as a function of θ_1 of the 1 ab^{-1} Asimov data set

In the next step, a log likelihood ratio is defined:

$$q(\theta_1, \hat{\theta}_1) = -2 \log \frac{L'(\theta_1)}{L'(\hat{\theta}_1)} \tag{5.5}$$

, where $\hat{\theta}_1$ is the fitted signal yield. The distribution of q as a function of θ_1 of the 1 ab^{-1} Asimov data set is shown in Fig. 5.4.

According to Wilk's theorem [32], the likelihood ratio q follows a χ^2 distribution with 1 degree of freedom, denoted as $\chi^2(q)$. Thus, for the null hypothesis where $\theta_1 = 0$, the corresponding p-value of $q(0, \hat{\theta}_1)$ is:

$$p = \int_{q(0,\hat{\theta}_1)}^{\infty} \chi^2(q) dq \tag{5.6}$$

The p-value is then transformed to a Z score by the following relation:

$$p = 1 - \int_{-Z}^{+Z} \mathcal{N}(x) dx \tag{5.7}$$

, where $\mathcal{N}$ is a normal distribution. For each luminosity, we quote "$Z\sigma$" as the estimated significance. The estimation of significance for different luminosity using the Asimov data sets will be shown in Ch. 8
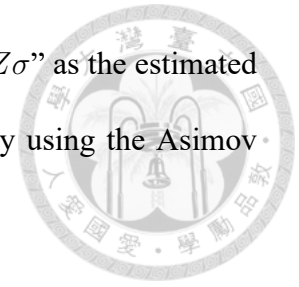
6　Control Channel Study

In the analysis of $B^+ \to \mu^+\nu$, the decay topology of the B^+ on the signal side has been fully understood. However, the knowledge on the companion B, which decays generically, is limited. As a result, the distribution of the variables that are calculated from the four-momentum of the companion B or its daughter, including $M_{\mathrm{bc}}^{\mathrm{ROE}}$, ΔE^{ROE}, and the shape variables in the continuum suppression, might not be well-simulated by the Monte-Carlo. Thus, there might be a discrepancy between the distribution of the signal Monte-Carlo samples and the signal in real data. Such discrepancy would be inherited by the output of the combined background suppression MVA, which is trained using the variables mentioned above.

In Sec. 4.2.6, a selection on the combined background suppression MVA, $BDT_{\mathrm{trans}} > 3.8$, has been set. If there is a discrepancy between the BDT_{trans} distribution of the signal in real data and that of the signal MC samples, the selection efficiency of the signal in real data ($\epsilon_{\mathrm{data}}^{\mathrm{sig}}$) and the selection efficiency of the signal MC samples ($\epsilon_{\mathrm{MC}}^{\mathrm{sig}}$) will be different. Thus, we have to derive a calibration factor $C = \frac{\epsilon_{\mathrm{data}}^{\mathrm{sig}}}{\epsilon_{\mathrm{MC}}^{\mathrm{sig}}}$ to correct the efficiency of the MC samples back to the efficiency of the real data. Because it is not legal to look at the real data in the signal box of $B^+ \to \mu^+\nu$ without the permission from the Belle II collaboration, if we want to check the discrepancy between the real data and the MC samples and derive the calibration factor, we'll have to study a decay mode which has a similar decay topology

to $B^+ \to \mu^+\nu$ as a "control channel". In this analysis, we study $B^+ \to \overline{D}^0(K^+\pi^-)\pi^+$ as our control channel. The $B^+ \to \overline{D}^0(K^+\pi^-)\pi^+$ decay mode has a high energy π^+ which mimics the μ^+, and a $\overline{D}^0$ which is ignored in all the calculation of the training variables in the MVAs to mimic the behavior of the missing neutrino. In the following sub-sections, the procedure to study $B^+ \to \overline{D}^0(K^+\pi^-)\pi^+$ and the derivation of the calibration factor will be explained.

6.1 MC Samples

The MC samples used for the control channel study is listed in Table 6.1.

sample	statistics
$B^+ \to \overline{D}^0(K^+\pi^-)\pi^+$	3.11 ab^{-1} (6×10^5 events)
$B\overline{B}$	1 ab^{-1}
$e^+e^- \to u\overline{u}, d\overline{d}, s\overline{s}$ and $c\overline{c}$	1 ab^{-1}
$e^+e^- \to \tau^+\tau^-$	1 ab^{-1}

Table 6.1: MC statistics

6.2 Pre-Selection on Signal Side

The selection on the signal-side of $B^+ \to \overline{D}^0\pi^+$ is listed as follows:

- $M_{\mathrm{bc}} > 5.27$ GeV

- $-0.1 < \Delta E < 0.1$ GeV

- $1.84 < M_{D^0} < 1.89$ GeV

- $\mathcal{L}_{K/\pi} > 0.6$ for K^+, where $\mathcal{L}_{K/\pi}$ is the PID likelihood ratio of the K^+ and π^+ hypotheses

- π-ID > 0.1 for the π^- that comes from the $\overline{D}^0$

- $|z_0| < 2$ cm and $d_0 < 0.5$ cm for all the tracks in the signal side

For each event, we only select the best candidate with its M_{D^0} being the closest to its nominal value (1.865 GeV).

6.3 Pre-Selection on the ROE Side and the Event N_{Tracks} Selection

The definition of the ROE is the same as what was listed in Table. 4.1. The selection on the ROE side and the event N_{Tracks} selection are listed as follows:

- $M_{\text{bc}}^{\text{ROE}} > 5.1$ GeV

- $-2 < \Delta E^{\text{ROE}} < 2$ GeV

- $N_{\text{Tracks}} > 3$

- $N_{\text{Tracks,highP}} == 1$

6.4 MVA Calibration

We do not train a new MVA classifier to separate the signal $B^+ \to \overline{D}^0 \pi^+$ from the background. Instead, in order to calibrate the MVA selection of $B^+ \to \mu^+ \nu$, the same weight file from the combined background suppression MVA of $B^+ \to \mu^+ \nu$ is applied on the control channel to derive the BDT_{out} and the BDT_{trans} distribution of $B^+ \to \overline{D}^0 \pi^+$. To mimic the $B^+ \to \mu^+ \nu$ decay topology, when we calculate the training variables in the continuum suppression MVA, the $B \to X_u l \nu$ MVAs, and the combined background suppression MVA, all of the $\overline{D}^0$s from the signal B^+ are treated as missing particles and the momentum of π^+s from the B^+ are scaled by 1.144 to mimic the signal μ^+s from $B^+ \to \mu^+ \nu$. The number of 1.144 is the ratio between the nominal μ^+ momentum in the B frame of the $B^+ \to \mu^+ \nu$ decay and the nominal π^+ momentum in the B frame of the $B^+ \to \overline{D}^0 \pi^+$ decay, The unscaled π^+ momentum in the B rest frame, the output of the combined background suppression MVA, and a subset of training variables in the continuum suppression MVA and the combined background suppression MVA are shown in Fig. 6.1.
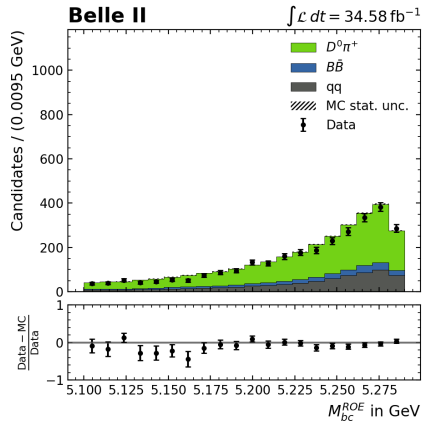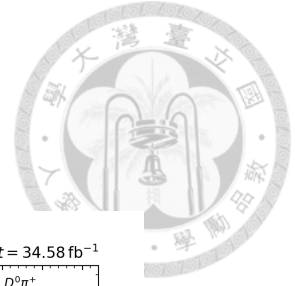
(a) $M_{\mathrm{bc}}^{\mathrm{ROE}}$
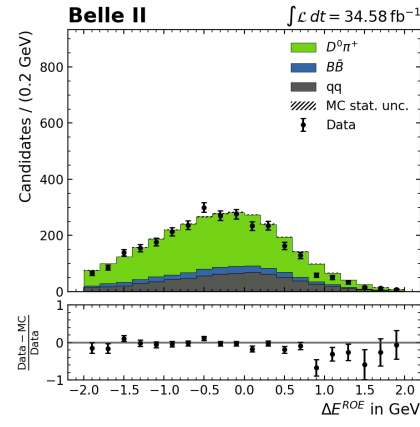
(b) ΔE^{ROE}
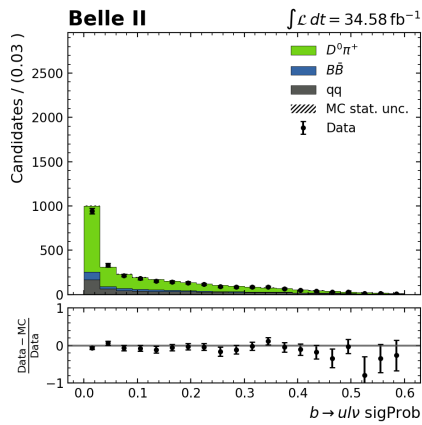
(c) $b \to u l \nu$ probability
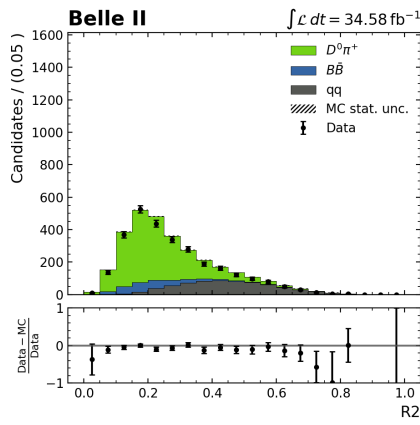
(d) R2
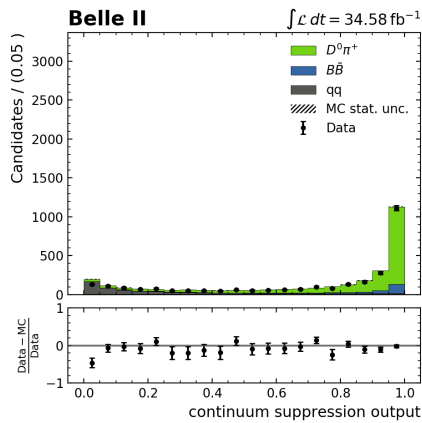
(e) continuum suppression output
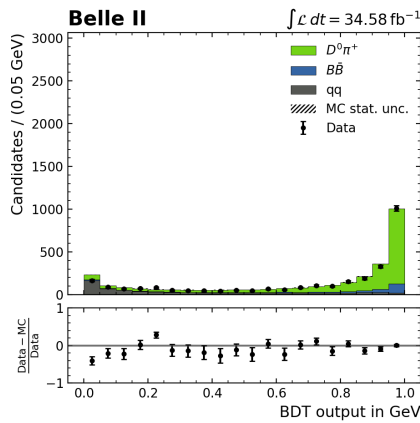
(f) combined MVA output
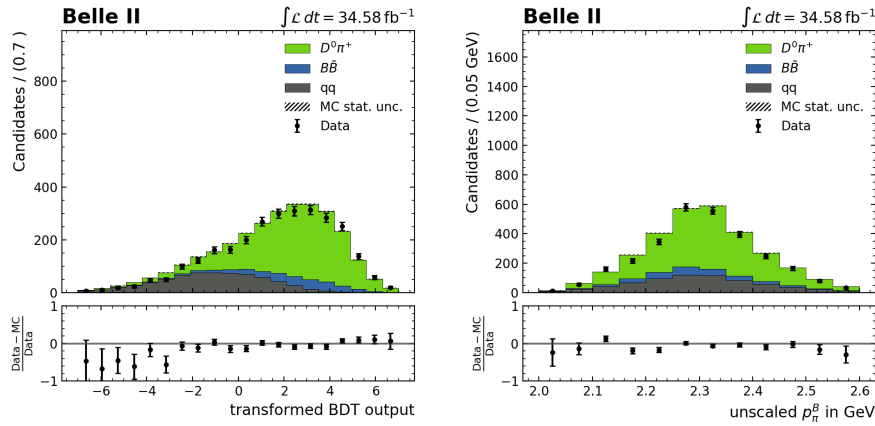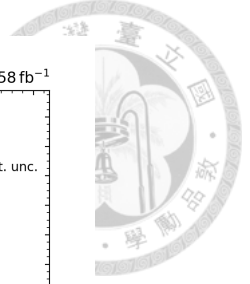
(g) transformed combined MVA output   (h) unscaled p_π^B

Figure 6.1: MVA variables, MVA output, and the π^+ momentum of $B^+ \rightarrow \overline{D}^0\pi^+$

The calibration factor to account for the difference between the MVA selection efficiency of the signal in real data (the signal yield) and that of the signal MC samples is defined as follow:

$$C_{\mathrm{MVA}} = \frac{\epsilon_{\mathrm{data}}^{\mathrm{sig}}}{\epsilon_{\mathrm{MC}}^{\mathrm{sig}}} = \frac{\dfrac{N_{\mathrm{yield,pass}}^{\mathrm{sig}}}{N_{\mathrm{yield,tot}}^{\mathrm{sig}}}}{\dfrac{N_{\mathrm{MC,pass}}^{\mathrm{sig}}}{N_{\mathrm{MC,tot}}^{\mathrm{sig}}}} \tag{6.1}$$

, where $N_{\mathrm{MC,pass}}^{\mathrm{sig}}$ and $N_{\mathrm{yield,pass}}^{\mathrm{sig}}$ are the amount of signal MC samples and the amount of signal in real data that pass the MVA selection, $N_{\mathrm{MC,tot}}^{\mathrm{sig}}$ and $N_{\mathrm{yield,tot}}^{\mathrm{sig}}$ are the total amount of signal MC samples and the total amount of signal in real data. While the value of $\epsilon_{\mathrm{MC}}^{\mathrm{sig}}$ can be well determined by MC simulation, to determine $\epsilon_{\mathrm{data}}^{\mathrm{sig}}$, some steps are needed to be taken. First, we separate the MC samples and the real data samples into four categories: $S_{\mathrm{data,pass}}$, $S_{\mathrm{data,fail}}$, $S_{\mathrm{MC,pass}}$ and $S_{\mathrm{MC,fail}}$, denoting the MC/real data samples that pass/fail the MVA selection respectively. Next, we construct four one-dimensional unbinned PDFs on the ΔE dimension. Two of them are for the $\overline{D}^0\pi^+$ signal that pass/fail the MVA selection and the other two are for the $B\overline{B}$+continuum background that pass/fail the MVA selection. To construct the four PDFs, we use several kinds of functions that are listed in Table. 6.2 to describe the ΔE distribution of the signal and the background samples. These PDFs are shown in Fig. 6.2. Finally, a simultaneous unbinned maximum likelihood fit is performed

	$\overline{D}^0\pi^+$, for both $S_{\text{MC,pass}}$ and $S_{\text{MC,fail}}$	$B\overline{B}$+continuum, for both $S_{\text{MC,pass}}$ and $S_{\text{MC,fail}}$
functions	two Gaussians and one Crystal Ball	one polynomial and one Gaussian
notes	The widths of the three functions are floated when fitting to the data, but the ratios between the widths are fixed.	The Gaussian accounts for the $\overline{D}^0K^+$ in the $B\overline{B}$, and the polynomial accounts for the other $B\overline{B}$+continuum background. The ratio between the polynomial and the Gaussian is floated when fitting to the data.

Table 6.2: The functions used in the PDF models

on the ΔE distribution of $S_{\text{data,pass}}$ and $S_{\text{data,fail}}$ to derive $\epsilon_{\text{data}}^{\text{sig}}$ and the corresponding uncertainty. Detailed explanation about the simultaneous fit is described in Appendix A.1. The fitting result is shown in Fig. 6.3. The $\epsilon_{\text{MC}}^{\text{sig}}$, $\epsilon_{\text{data}}^{\text{sig}}$ and the calibration factor for the MVA selection efficiency derived from the fit is listed in Table. 6.3:

	ϵ^{sig}	C_{MVA}
real data	$0.2779^{+0.0115}_{-0.0112}$	$1.085^{+0.045}_{-0.043}$
MC samples	0.2562	

Table 6.3: Efficiencies and the calibration factor

(a) $B^+ \to \overline{D}^0 \pi^+$, $S_{\text{MC,pass}}$

(b) $B^+ \to \overline{D}^0 \pi^+$, $S_{\text{MC,fail}}$

(c) $B\overline{B}$+continuum, $S_{\text{MC,pass}}$

(d) $B\overline{B}$+continuum, $S_{\text{MC,fail}}$
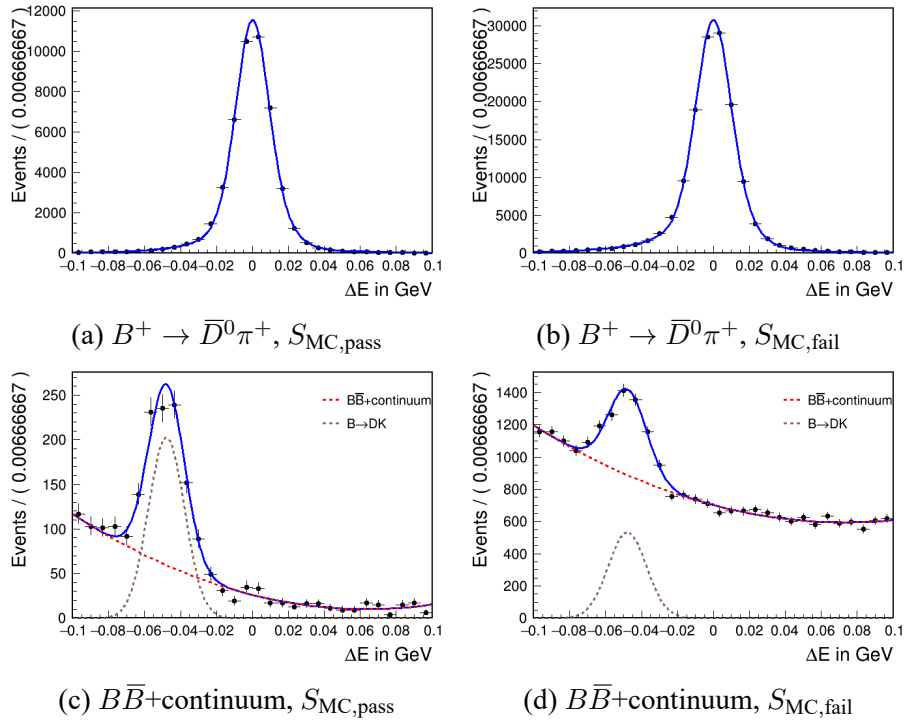
Figure 6.2: PDF models for control channel

(a) Fit to $S_{\text{data,pass}}$

(b) Fit to $S_{\text{data,fail}}$
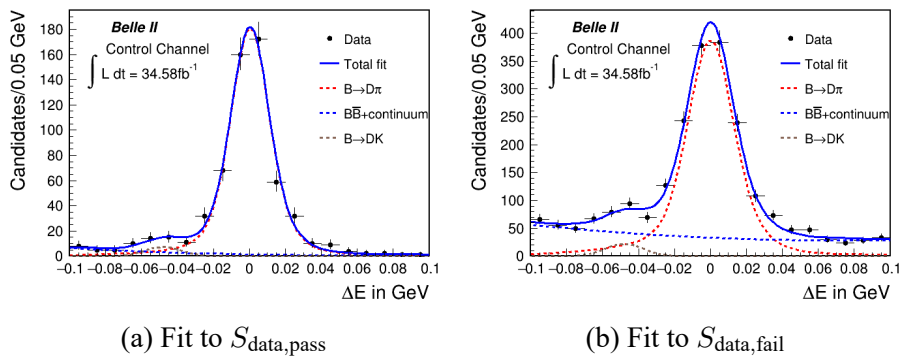
Figure 6.3: Result of simultaneous fit

7 Systematic Uncertainty

7.1 Hybrid Model of $B \to X_u l\nu$

Three sources of systematic uncertainties from the Hybrid Model of $B \to X_u l\nu$ are considered: the uncertainty of the form factors for the resonant modes, the uncertainty of the branching fractions, and the uncertainty caused by using various inclusive decay models to generate $B \to X_u l\nu$.

7.1.1 Form Factors for $B \to \pi l\nu$, $B \to \rho l\nu$ and $B \to \omega l\nu$

There is one form factor, $f_+ (q^2)$, for the $B \to \pi l\nu$ mode, and three form factors, $V (q^2)$, $A_1 (q^2)$ and $A_{12} (q^2)$, for the $B \to \rho l\nu$ and the $B \to \omega l\nu$ modes. These form factors can be expanded using the Bourrely-Caprini-Lellouch (BCL) expansion[28]. For the calculation of $f_+ (q^2)$, the coefficients of the first four terms in the BCL expansion[28], b_0^+, b_1^+, b_2^+, and b_3^+, are included. As for the calculation of $V (q^2)$, $A_1 (q^2)$, and $A_{12} (q^2)$, the coefficients of the first three terms in BCL of each form factor, which are, v^0, v^1, and v^2 for $V (q^2)$, a_1^0, a_1^1, and a_1^2 for $A_1 (q^2)$, and a_{12}^0, a_{12}^1, and a_{12}^2 for $A_{12} (q^2)$, are included. For more details about the form factors and the SM prediction of the $B \to \pi, \rho, \omega l\nu$ modes, please see Apendix B.1 and B.2.

	$B \to \rho$	$B \to \omega$		$B \to \pi$
a_1^0	0.26 ± 0.03	0.24 ± 0.03	b_0^+	0.419 ± 0.013
a_1^1	0.39 ± 0.14	0.34 ± 0.24	b_1^+	-0.495 ± 0.054
a_1^2	0.16 ± 0.41	0.09 ± 0.57	b_2^+	-0.43 ± 0.13
a_{12}^0	0.30 ± 0.03	0.27 ± 0.04	b_3^+	0.22 ± 0.31
a_{12}^1	0.76 ± 0.20	0.66 ± 0.26		
a_{12}^2	0.46 ± 0.76	0.28 ± 0.98		
v^0	0.33 ± 0.03	0.30 ± 0.04		
v^1	-0.86 ± 0.18	-0.83 ± 0.29		
v^2	1.80 ± 0.97	1.72 ± 1.24		

Table 7.1: The BCL expansion coefficients of the $B \to \rho$ and $B \to \omega$ are quoted from the light-cone sum rule (LCSR) prediction[10], and the coefficients of the $B \to \pi$ are quoted from the combined result of LCSR, LQCD, Belle and BaBar[11]

To calculate the total systematic uncertainty, twenty sets of BCL coefficients are generated based on a multidimensional Gaussian function using the covariance matrix of these coefficients as the Gaussian function's multidimensional width. The differential decay rates of the $B \to \pi, \rho, \omega l\nu$ modes are recalculated using each set of the new coefficients and the $B \to X_u l\nu$ MC samples are reweighted accordingly. The 20 sets of the reweighted MC samples are then used to construct 20 new 2D histogram PDFs for the $B \to X_u l\nu$ process. Along with the signal, the $B\bar{B}$ and the continuum PDFs constructed in Ch. 5, these 20 PDFs are fitted to the 1 ab^{-1} Asimov data set to derive 20 $B^+ \to \mu^+ \nu$ signal yields. The distribution of the 20 yields is fitted using a bifurcated Gaussian. The results of the positive and negative errors of the bifurcated Gaussian function divided by the mean are then quoted as the relative asymmetric systematic uncertainties.

7.1.2 Inclusive and Resonant $B \to X_u l\nu$ Branching Fractions

The inclusive and resonant branching fractions of $B \to X_u l\nu$ and their uncertainties are listed in Table 7.2. To calculate the total systematic uncertainty caused by the uncertainties of these branching fractions, 100 sets of branching fractions are generated based on Gaussian functions, with the means of the Gaussian functions being the central values

67

	$X_u l\nu$	$\pi l\nu$	$\rho l\nu$	$\omega l\nu$	$\eta l\nu$	$\eta' l\nu$
charged modes	22.04	1.45	2.94			
($\times 10^{-4}$)	± 3.19	± 0.05	± 0.21			
neutral modes	20.44	0.78	1.58	0.38	0.23	1.19
($\times 10^{-4}$)	± 2.96	± 0.03	± 0.11	± 0.06	± 0.08	± 0.09

Table 7.2: Branching fraction of $B \to X_u l\nu$ modes and their uncertainties

and the widths of the Gaussian functions being the uncertainties of the branching fractions. We then use the similar method described in Sec. 7.1.1 to derive 100 signal yields and calculate the relative asymmetric systematic uncertainties.

7.1.3   Various Inclusive $B \to X_u l\nu$ Decay Models

To calculate the systematic uncertainty, we compare two inclusive models, the BLNP model and the DFN model[33]. We use these two models to generate two inclusive $B \to X_u l\nu$ MC data sets. These two inclusive data sets are reweighted and combined with the resonant data sets as described in Sec. 3.2 to form two Hybrid Models. The two combined data sets both pass the combined MVA selection and are used to generate two histogram PDFs for the $B \to X_u l\nu$ process. Along with the other PDFs constructed in Ch. 5, we fit these two PDFs to the 1 ab^{-1} Asimov data sets. The difference between the two $B^+ \to \mu^+ \nu$ signal yields are divided by the 1 ab^{-1} signal yield listed in Table. 5.1, and the result is quoted as the relative systematic uncertainty.

7.2   Lepton Identification

In Sec. 4.3, to calibrate the efficiency of the μ-ID selection, each signal μ^+ candidate is reweighted by a calibration factor C_i^μ, where i is the bin index of the (p, θ) array. Each C_i^μ has a corresponding uncertainty, and the uncertainties of all the calibration factors are

fully correlated. These uncertainties of the calibration factors can cause a total systematic uncertainty in our analysis. With C_i^μ being the mean values and their corresponding uncertainties being the widths of the Gaussians, 100 sets of C_i^μ are generated. The μ^+ candidates in the signal MC samples are reweighted by these C_i^μ and then summed to derive 100 expected $B^+ \to \mu^+\nu$ yields. The distribution of these expected yields is fitted with a bifurcated Gaussian. The widths divided by the mean of the bifurcated Gaussian are then quoted as the relative asymmetric systematic uncertainties.

7.3 MVA Selection

In Ch. 6, a control channel is studied to derive the calibration factor for the MVA selection on $BDT_{\text{trans}} > 3.8$. The relative systematic uncertainty of this calibration is:

$$\sigma_{\text{rel,MVA}} = \sigma_{\text{MVA}}/C_{\text{MVA}} \tag{7.1}$$

, where C_{MVA} is the calibration factor and σ_{MVA} is the absolute uncertainty.

7.4 Tracking

The systematic uncertainty of the track reconstruction is caused by the difference between the tracking efficiency of real data and that of MC samples in Belle II. Such discrepancy is measured using $e^+e^- \to \tau^+\tau^-$ events, in which one of the τ decays to either a μ^+ or a e^+ ($\tau \to l^\pm \nu_l \bar{\nu}_\tau$, $l = e, \mu$, referred as 1-prong decay) and the other τ decays to 3 charged π^+ ($\tau \to 3\pi^\pm \nu_\tau + n\pi^0$, referred as 3-prong decay).

A tag-and-probe method is used for measuring the tracking efficiency. According to

charge conservation, the presence of three charged tracks in the final state with the total

charge equaling to ± 1 infers the existence of the fourth charged track that makes the total

charge equal to zero. The three charged tracks are tagged to probe for the fourth track.

The tracking efficiency ϵ_{track} is defined by Eq. 1 in Ref. [34]:

$$\epsilon_{\text{track}} \cdot A = \frac{N_4}{N_3 + N_4} \tag{7.2}$$

, where A is the acceptance of the tracking system in Belle II, N_4 is the number of events

where the fourth track is reconstructed, and N_3 is the number of events where the fourth

track is not reconstructed.

According to Eq. 2 and Eq. 3 in Ref. [34], the data-MC discrepancy of the tracking

efficiency is:

$$\delta = 1 - \frac{\epsilon_{\text{track}}^{\text{data}}}{\epsilon_{\text{track}}^{\text{MC}}} \tag{7.3}$$

For more details about how this discrepancy and the systematic uncertainty are derived,

please see Ref. [34].

7.5 Number of $B\overline{B}$

In Belle II, the number of $B\overline{B}$ is measured by subtracting the non-$B\overline{B}$ events from the

on-resonance hadronic events. The on-resonance data is defined as the data produced by

the e^+e^- collision at the $\Upsilon(4S)$ resonance (10.58 GeV). In addition to the on-resonance

data, Belle II also produce some off-resonance data where the e^+e^- pairs collide at an

center-of-mass energy that is 60 MeV below the $\Upsilon(4S)$ resonance. The off-resonance

data does not contain any $B\overline{B}$ process because no $\Upsilon(4S)$ can be produced. Therefore,

the off-resonance data can be used to represent the non-$B\overline{B}$ events in the measurement of

70

$N_{B\bar{B}}$. To be specific,

$$N_{B\bar{B}} = \frac{(N_{\text{had}}^{\text{on-res}} - R_{\text{lumi}} \times N_{\text{had}}^{\text{off-res}} \times k)}{\epsilon_{B\bar{B}}} \qquad (7.4)$$

, where $N_{\text{had}}^{\text{on-res}}$ is the number of on-resonance data that pass the hadronic event selection, $N_{\text{had}}^{\text{off-res}}$ is the number of off-resonance data that pass the hadronic event selection, R_{lumi} is the ratio of luminosity between on-resonance and off-resonance data, $\epsilon_{B\bar{B}}$ is the hadronic selection efficiency of the $B\bar{B}$ events estimated by MC samples, and k accounts for the difference between the hadronic selection efficiencies for various kinds of non-BB processes in the on-resonance data and that in the off-resonance data, being defined as:

$$k = \frac{\sum_i \epsilon_i \sigma_i}{\sum_i \epsilon_i' \sigma_i'} \qquad (7.5)$$

, where ϵ_i and σ_i stand for the hadronic selection efficiency and the cross section, respectively, of the ith kind of process in on-resonance data, and ϵ_i' and σ_i' stand for the efficiency and the cross section, respectively, of the ith kind of process in the off-resonance data. For more details about the measurement of $N_{B\bar{B}}$ and the corresponding systematic uncertainty, pleas see Ref. [35].

	$X_u l\nu$ inclusive model	$X_u l\nu$ form factors	$X_u l\nu$ branching fractions	MVA calibration
negative error	20.1%	14.5%	14.8%	4.0%
positive error	20.1%	13.4%	18.3%	4.1%
	tracking	PID correction	n$B\bar{B}$	total
negative error	0.8%	1.2%	5.0%	29.6%
positive error	0.8%	1.2%	5.0%	31.0%

Table 7.3: A summary of all the systematic uncertainties

71

8 Results and Conclusion

In this sensitivity study, an analysis framework to suppress the background in the signal region of $B^+ \to \mu^+\nu$ is established, and a fitter to extract the $B^+ \to \mu^+\nu$ signal is constructed. Pseudo data sets are generated to estimate the significance of $B^+ \to \mu^+\nu$ for different luminosity using the method described in Ch. 5. The estimated significance for different luminosity is shown in Fig. 8.1 and Table. 8.1. According to the estimation, we will need to collect data with a luminosity between 7 and 8 ab^{-1} in Belle II in order to reach a significance of 5σ and claim the discovery of $B^+ \to \mu^+\nu$, assuming that the SM prediction is correct and there is no large data fluctuation.
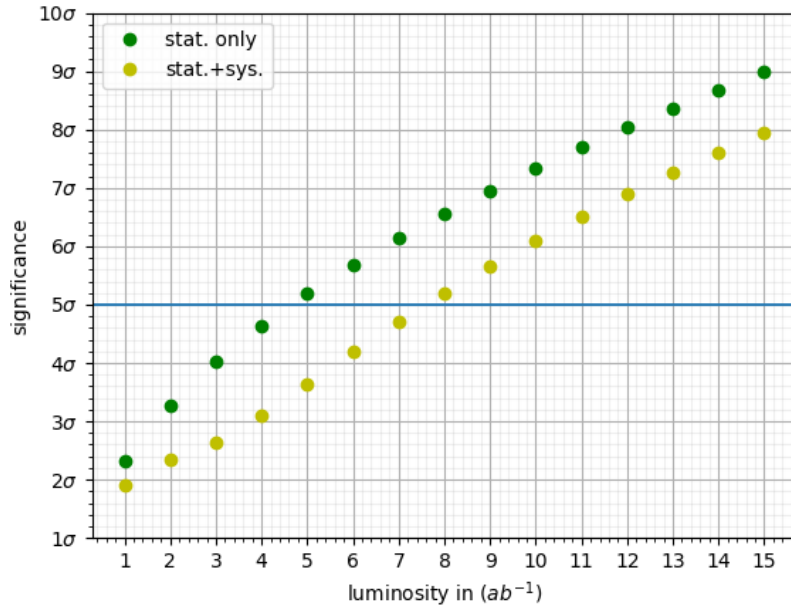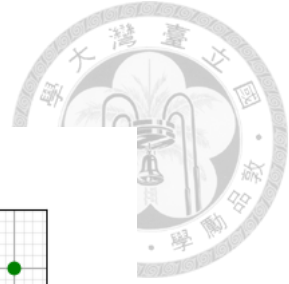
Figure 8.1: Significance estimation. The dark green points represent the significance estimation considering only statistical errors, and the light green points represent the significance estimation considering both statistical and systematic errors.

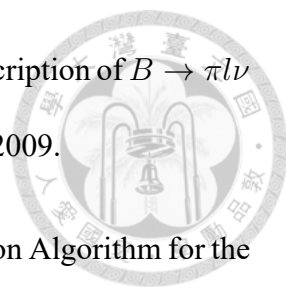| luminosity	Z score (stat. +sys.)	Z score (stat. only)
1 ab^{-1}	1.91	2.32
2 ab^{-1}	2.34	3.28
3 ab^{-1}	2.64	4.02
4 ab^{-1}	3.09	4.64
5 ab^{-1}	3.64	5.18
6 ab^{-1}	4.19	5.68
7 ab^{-1}	4.71	6.12
8 ab^{-1}	5.20	6.56
9 ab^{-1}	5.66	6.96
10 ab^{-1}	6.09	7.33
11 ab^{-1}	6.50	7.69
12 ab^{-1}	6.89	8.03
13 ab^{-1}	7.26	8.36
14 ab^{-1}	7.61	8.68
15 ab^{-1}	7.95	8.98

Table 8.1: Significance estimation

References

[1] MissMJ. Standard Model of Elementary Particles. From Wikipedia: https://en.wikipedia.org/wiki/Standard_Model.

[2] S. Stone. The discovery of the B mesons. <u>AIP Conference Proceedings</u>, 424(1):75–84, 1998.

[3] I. Heredia de la Cruz. The Belle II experiment: fundamental physics at the flavor frontier. <u>Journal of Physics: Conference Series</u>, 761:012017, 2016.

[4] B. Shwartz et al. Electromagnetic calorimeter of the Belle II detector. <u>Journal of Physics: Conference Series</u>, 928:012021, 2017.

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

[6] U. Tamponi. The TOP counter of Belle II: Status and first results. <u>Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment</u>, 952:162208, 2020.

[7] D. Weyland. Continuum Suppression with Deep Learning techniques for the Belle II Experiment. Master's thesis, KIT, Karlsruhe, ETP, 2017.

[8] K. Akai, K. Furukawa, and H. Koiso. SuperKEKB collider. <u>Nuclear Instruments and</u>

Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 907:188 – 199, 2018.
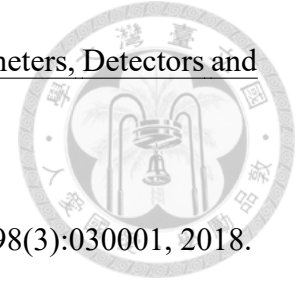
[9] M. Tanabashi et al. Review of Particle Physics. Phys. Rev. D, 98(3):030001, 2018.

[10] A. Bharucha, D. Straub, and R. Zwicky. $B \to V l^+ l^-$ in the Standard Model from light-cone sum rules. Journal of High Energy Physics, 2016:98, 2016.

[11] Jon. A. Bailey et al. $|V_{ub}|$ from $B \to \pi \ell \nu$ decays and $(2 + 1)$-flavor lattice QCD. Phys. Rev. D, 92(1):014024, 2015.

[12] T. Aushev et al. Physics at Super B Factory, 2010. arXiv:1002.5012.

[13] C. MacQueen and on behalf of the Belle II collaboration. Dark Sector Physics at Belle II. Journal of Physics: Conference Series, 1468:012047, 2020.

[14] Y. Ohnishi et al. Accelerator design at SuperKEKB. Progress of Theoretical and Experimental Physics, 2013(3):03A011, 2013.

[15] A. Abashian et al. The Belle detector. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 479(1):117–232, 2002.

[16] S. Takeda et al. Development of double-sided silicon strip detectors (DSSD) for a Compton telescope. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 579(2):859 – 865, 2007.

[17] V. Bertacchi et al. Track finding at Belle II. Computer Physics Communications, 259:107610, 2021.

[18] M. Thomson. <u>Modern Particle Physics</u>. Cambridge University Press, 2013.

[19] T. Keck et al. The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment. <u>Comput. Softw. Big Sci.</u>, 3(1):6, 2019.

[20] M. T. Prim, F. U. Bernlochner, P. Goldenzweig, et al. Search for $B^+ \to \mu^+\nu$ and $B^+ \to \mu^+ N$ with inclusive tagging. <u>Phys. Rev. D</u>, 101(3), 2020.

[21] E. Kou et al. The Belle II Physics Book. <u>Progress of Theoretical and Experimental Physics</u>, 2019(12):123C01, 2019.

[22] W.-S. Hou. Enhanced charged Higgs boson effects in $B^- \to \tau\bar{\nu}, \mu\bar{\nu}$ and $b \to \tau\bar{\nu}+X$. <u>Phys. Rev. D</u>, 48(5):2342–2344, 1993.

[23] W.-S. Hou, M. Kohda, T. Modak, and G.-G. Wong. Enhanced $B \to \mu\nu$ decay at tree level as probe of extra Yukawa couplings. <u>Physics Letters B</u>, 800:135105, 2020.

[24] W.-S. Hou, T. Modak, and G.-G. Wong. Scalar leptoquark effects on $B \to \mu\bar{\nu}$ decay. <u>Eur. Phys. J. C</u>, 79(11):964, 2019. arXiv:1909.00403.

[25] D. J. Lange. The EvtGen particle decay simulation package. <u>Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment</u>, 462(1):152 – 155, 2001.

[26] S. Jadach, B.F.L. Ward, and Z. Was. The Precision Monte Carlo event generator KK for two fermion final states in e^+e^- collisions. <u>Comput. Phys. Commun.</u>, 130:260–325, 2000.

[27] B. O. Lange, M. Neubert, and G. Paz. Theory of charmless inclusive B decays and the extraction of V_{ub}. <u>Phys. Rev. D</u>, 72:073006, 2005.

[28] C. Bourrely, L. Lellouch, and I. Caprini. Model-independent description of $B \to \pi l \nu$ decays and a determination of $|V_{ub}|$. Phys. Rev. D, 79:013008, 2009.

[29] T Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. Comput. Softw. Big Sci., 1(1):2, 2017.

[30] A. J. Bevan, B. Golob, Th. Mannel, S. Prell, B. D. Yabsley, H. Aihara, F. Anulli, N. Arnaud, T. Aushev, M. Beneke, and et al. The Physics of the B Factories. The European Physical Journal C, 74(11), 2014.

[31] D. Ferlewicz, M. Milesi, and P. Urquijo. Lepton identification and pion fake rates using $J/\psi \to \ell^+\ell^-$ and $K_S^0 \to \pi^+\pi^-$ events. 2020. https://confluence.desy.de/display/BI/Recommendations+for+LeptonID+-+ICHEP+2020.

[32] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Statist., 9(1):60–62, 1938.

[33] F. De Fazio and M. Neubert. $B \to X_u l \bar{\nu}_l$ decay distributions to order α_s. Journal of High Energy Physics, 1999(06):017–017, 1999.

[34] A. Glazov, P. Rados, A. Rostomyan, E. Paoloni, and L. Zani. Measurement of the tracking efficiency in phase 3 data using tau-pair events. 2020. https://confluence.desy.de/display/BI/Tracking+and+Vertexing+Performance.

[35] C.Cecchi, G. De Nardo, E. Manoni, and M. Merola. Study of R2 distribution and B counting in Early Phase 3 Data. 2019. https://confluence.desy.de/display/BI/Accelerator+performance+subgroup.

[36] F. U. Bernlochner. $B \to \pi \tau \bar{\nu}_\tau$ decay in the context of type II 2HDM. Phys. Rev. D, 92(11):115019, 2015.

[37] M. T. Prim, F. U. Bernlochner, and D. J. Robinson. Precision predictions for $B \rightarrow$ $\rho\tau\nu_\tau$ and $B \rightarrow \omega\tau\nu_\tau$ in the SM and beyond, 2020. arXiv:2001.06170.
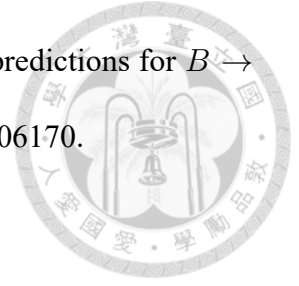
Appendix A — Special cases for maximum likelihood estimator

A.1 Simultaneous fit

A typical unbinned maximum likelihood function looks like:

$$L(X_i; \Theta_k, \theta_k) = \mathcal{P}(N | \sum_k \theta_k) \prod_{i=1}^{N} \left(\frac{\sum_k \theta_k P^k(X_i; \Theta_k)}{\sum_k \theta_k} \right) \tag{A.1}$$

, where

k labels the different physics processes,

N is the total amount of data,

X_i denotes independent observations,

θ_k is the expected number of data of the kth physics process,

$P^k(X_i; \Theta_k)$ is the PDF of the kth physics process,

Θ_k are some floated nuisance parameters of the kth PDF.

$\mathcal{P}$ is the Poisson distributions,

In a typical unbinned maximum likelihood fit, we fit to one data set by constructing and maximizing the above likelihood function to extract the parameters of interest, θ_k. On

the other hand, in a simultaneous fit, we fit to several data set at the same time and extract

some common parameters of interest for these data sets. Suppose we have two data sets,

S_1 and S_2, and we have constructed two unbinned likelihood functions for the distribution

of these data sets: $L_1(X_i; \theta_k^1, \Theta_k^1)$ and $L_2(X_i; \theta_k^2, \Theta_k^2)$, where θ_k^1 and θ_k^2 are the parameters

of interest for S_1 and S_2. In a typical maximum likelihood fit, we maximize L_1 and L_2

respectively to extract the most likely θ_k^1 and θ_k^2, denoted as $\hat{\theta}_k^1$ and $\hat{\theta}_k^2$. What if we have a

common parameter of interest, θ_k, for L_1 and L_2? In such case, we have to maximize the

multiplied likelihood:

$$L(\theta_k) = L_1(\theta_k) \times L_2(\theta_k) \tag{A.2}$$

to get the best $\hat{\theta}_k$. This is called a simultaneous fit.

In our case of the study of the control channel, we construct two likelihoods for

the samples that pass/fail the MVA selection based on the PDFs shown in Fig. 6.2.

The constructed likelihoods are denoted as $L_{\text{pass}}(\Delta E_i; N_{\text{pass}}^k, \Theta_k)$ and $L_{\text{fail}}(\Delta E_i; N_{\text{fail}}^k, \Theta_k)$.

Here $k = 1, 2$ denotes the $\overline{D}^0\pi^+$ and the $B\overline{B}$+continuum processes respectively. In this

case, the parameters of interest are the MVA selection efficiency of the yield in data,

$\epsilon_{\text{data}}^k = \frac{N_{\text{pass}}^k}{N_{\text{pass}}^k + N_{\text{fail}}^k}$, and the total yield, $N_{\text{tot}}^k = N_{\text{pass}}^k + N_{\text{fail}}^k$. We can rewrite the two likeli-

hoods in terms of these two parameters:

$$L_{\text{pass}}(\Delta E_i; N_{\text{pass}}^k, \Theta_k) = L_{\text{pass}}(\Delta E_i; N_{\text{tot}}^k \epsilon_{\text{data}}^k, \Theta_k) \tag{A.3}$$

$$L_{\text{fail}}(\Delta E_i; N_{\text{fail}}^k, \Theta_k) = L_{\text{fail}}(\Delta E_i; N_{\text{tot}}^k(1 - \epsilon_{\text{data}}^k), \Theta_k) \tag{A.4}$$

To extract $\hat{\epsilon_{\text{data}}^k}$ and $\hat{N_{\text{tot}}^k}$, we have to maximize the multiplied likelihood:

$$L(\Delta E_i; N_{\text{tot}}^k, \epsilon_{\text{data}}^k, \Theta_k) = L_{\text{pass}}(\Delta E_i; N_{\text{tot}}^k \epsilon_{\text{data}}^k, \Theta_k) \times L_{\text{fail}}(\Delta E_i; N_{\text{tot}}^k(1 - \epsilon_{\text{data}}^k), \Theta_k) \tag{A.5}$$

Then, the MVA selection efficiency of the signal yield acquired from the fit, $\hat{\epsilon}_{\text{data}}^1$, is used to derive the MVA calibration factor.

Appendix B — Differential decay rates for the resonant $B \to X_u l\nu$ modes

B.1 SM prediction of $B \to \pi l\nu$

The differential decay rate of $B \to \pi l\nu$ is given by:

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda(q^2)} \left(1 - \frac{m_l^2}{q^2}\right)^2 \left[\left(1 + \frac{m_l^2}{2q^2}\right) H_0^2(q^2) + \frac{3}{2}\frac{m_l^2}{q^2} H_t^2(q^2)\right] \quad \text{(B.1)}$$

, where

$$H_0(q^2) = \frac{\sqrt{\lambda(q^2)}}{\sqrt{q^2}} f_+(q^2), \quad \text{(B.2)}$$

$$H_t(q^2) = \frac{m_B^2 - m_\pi^2}{\sqrt{q^2}} f_0(q^2). \quad \text{(B.3)}$$

are the helicity amplitudes, $f_+(q^2)$ and $f_0(q^2)$ are the form factors, and $\lambda(q^2)$ is the Kaellen function. Assuming zero lepton mass, we can drop the H_t term. Thus, there is only one helicity amplitude H_0 and one form factor f_+ left.

B.2   SM prediction of $B \to \rho/\omega l \nu$

For $B \to \rho/\omega l \nu$, where ρ/ω are vector meson, the differential decay rate is given by:

$$
\begin{aligned}
\frac{d\Gamma}{dq^2} = |V_{ub}| \frac{G_F^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda(q^2)} \left(1 - \frac{m_l^2}{q^2}\right)^2 \\
\times \left[\left(1 + \frac{m_l^2}{2q^2}\right) \left(H_+^2(q^2) + H_-^2(q^2) + H_0^2(q^2)\right) + \frac{3}{2}\frac{m_l^2}{q^2} H_s^2(q^2)\right]
\end{aligned}
\tag{B.4}
$$

, where

$$
H_\pm(q^2) = \sqrt{\lambda(q^2)} \frac{V(q^2)}{m_B + m_V} \pm (m_B + m_V) A_l(q^2),
\tag{B.5}
$$

$$
H_0(q^2) = \frac{8 m_B m_V}{\sqrt{q^2}} A_{12}(q^2),
\tag{B.6}
$$

$$
H_s(q^2) = \frac{\lambda(q^2)}{\sqrt{q^2}} A_0(q^2)
\tag{B.7}
$$

are again the helicity amplitudes, $V(q^2)$, $A_l(q^2)$, $A_{12}(q^2)$ and $A_0(q^2)$ are the form factors, and $\lambda(q^2)$ is the Kaellen function. Assuming zero lepton mass, H_s is dropped and only 3 form factors left.

The theoretical prediction of the differential decay rates and the form factors of $B \to \pi/\rho/\omega l \nu$ are cited from Ref. [36] and Ref. [37].
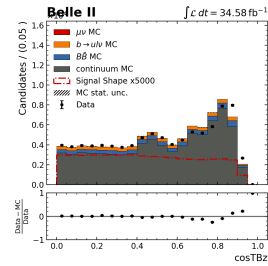
Appendix C — Plots
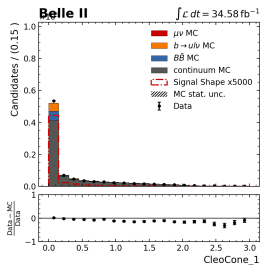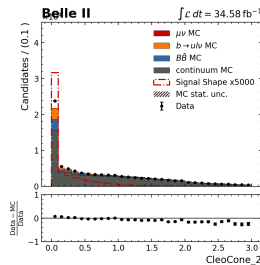
C.1 Continuum Suppression Variables
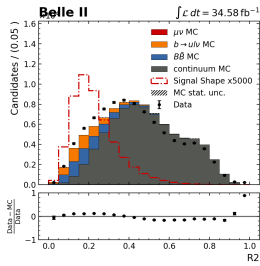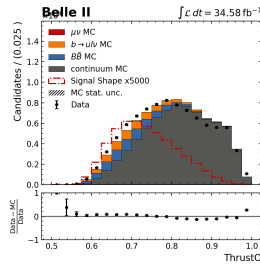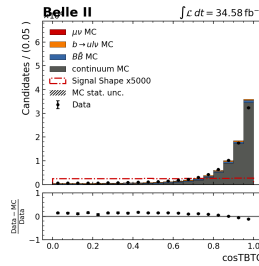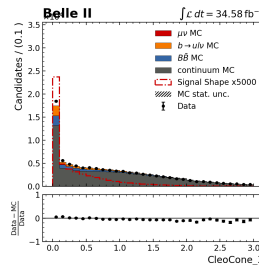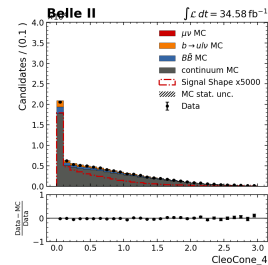

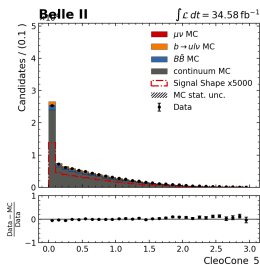(a) R2   (b) ThrustO   (c) CosTBTO   (d) cosTBz
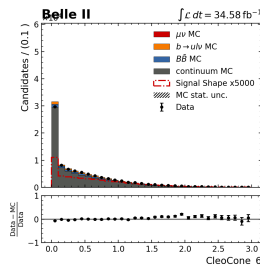
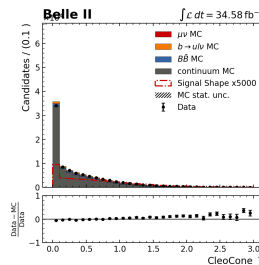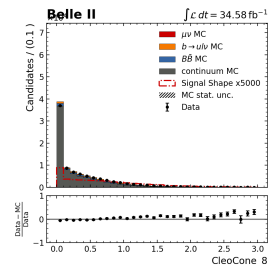(e) CC_ROE_1   (f) CC_2   (g) CC_3   (h) CC_4
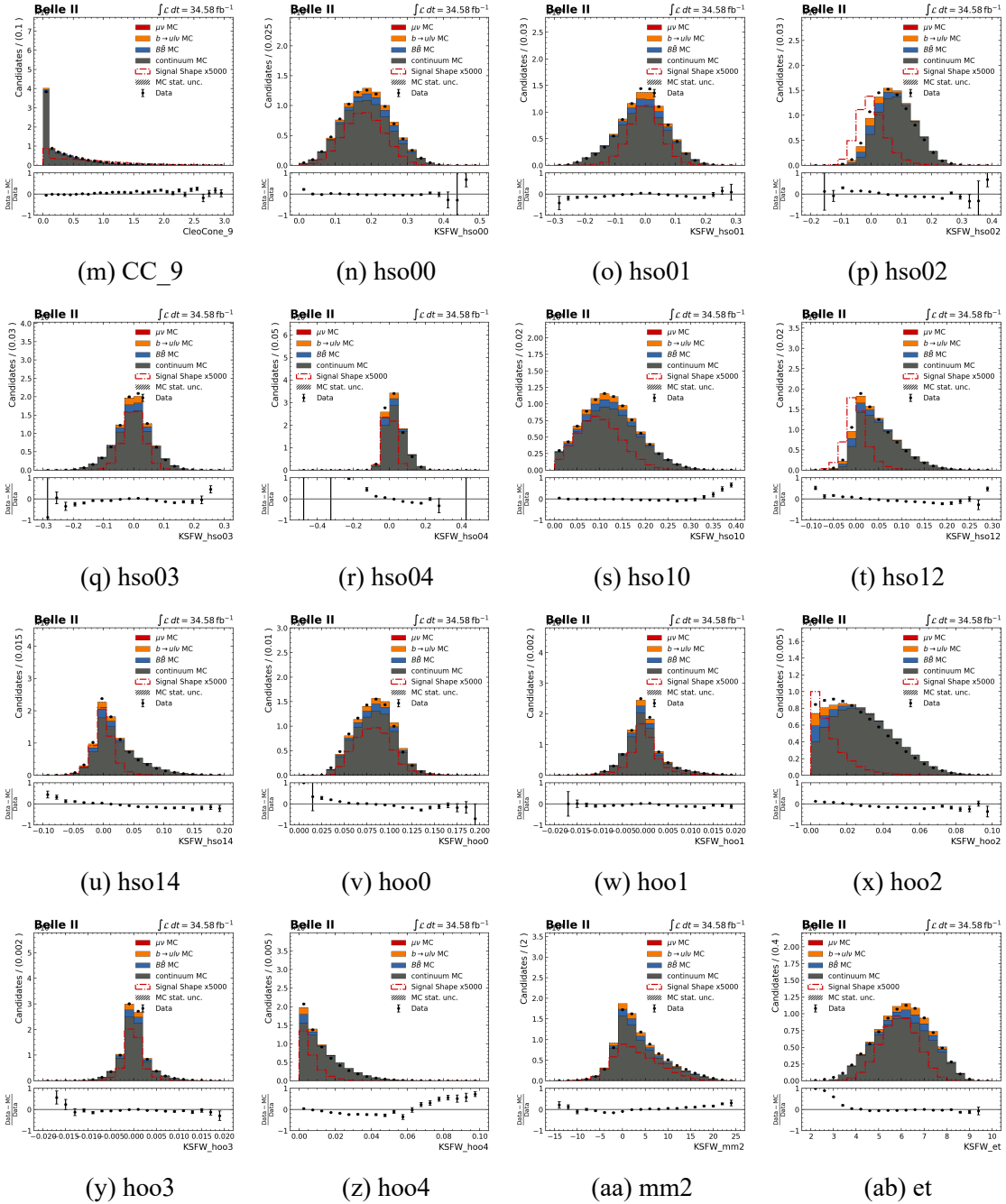
(i) CC_5   (j) CC_6   (k) CC_7   (l) CC_8

(m) CC_9   (n) hso00   (o) hso01   (p) hso02

(q) hso03   (r) hso04   (s) hso10   (t) hso12

(u) hso14   (v) hoo0   (w) hoo1   (x) hoo2

(y) hoo3   (z) hoo4   (aa) mm2   (ab) et

Figure C.1: Training variables for continuum suppression

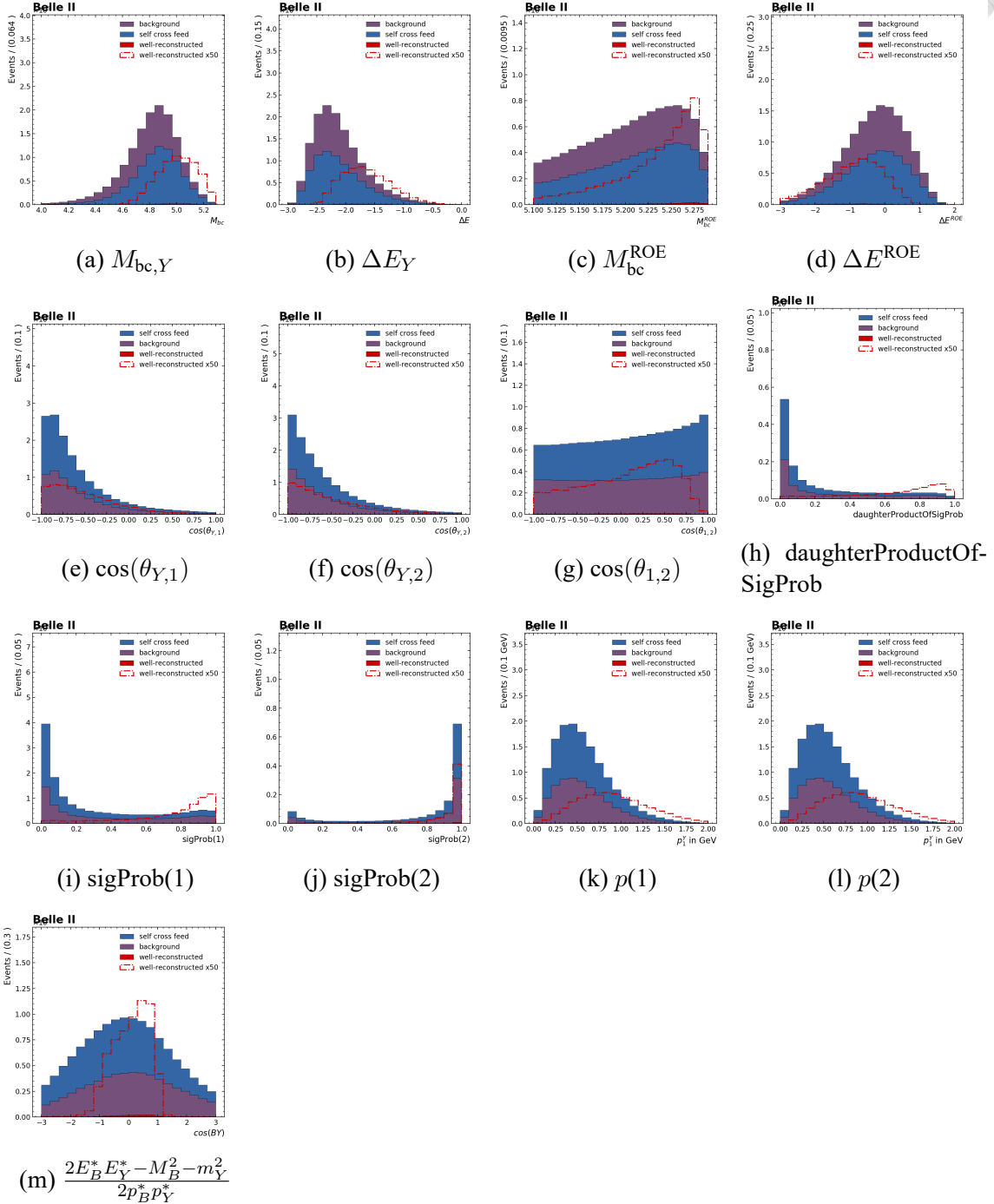## C.2 $B \to X_u l\nu$ Suppression Variables for $B^0 \to \pi^0\pi^-\mu^+\nu$

(a) $M_{\mathrm{bc},Y}$

(b) ΔE_Y

(c) $M_{\mathrm{bc}}^{\mathrm{ROE}}$

(d) ΔE^{ROE}

(e) $\cos(\theta_{Y,1})$

(f) $\cos(\theta_{Y,2})$

(g) $\cos(\theta_{1,2})$

(h) daughterProductOf-SigProb

(i) sigProb(1)

(j) sigProb(2)

(k) $p(1)$

(l) $p(2)$

(m) $\dfrac{2E_B^* E_Y^* - M_B^2 - m_Y^2}{2p_B^* p_Y^*}$

Figure C.2: $B \to X_u l\nu$ Suppression Variables for $B^0 \to \pi^0\pi^-\mu^+\nu$

C.3 Toy results and fit to Asimov's data

(a) 1 ab^{-1}

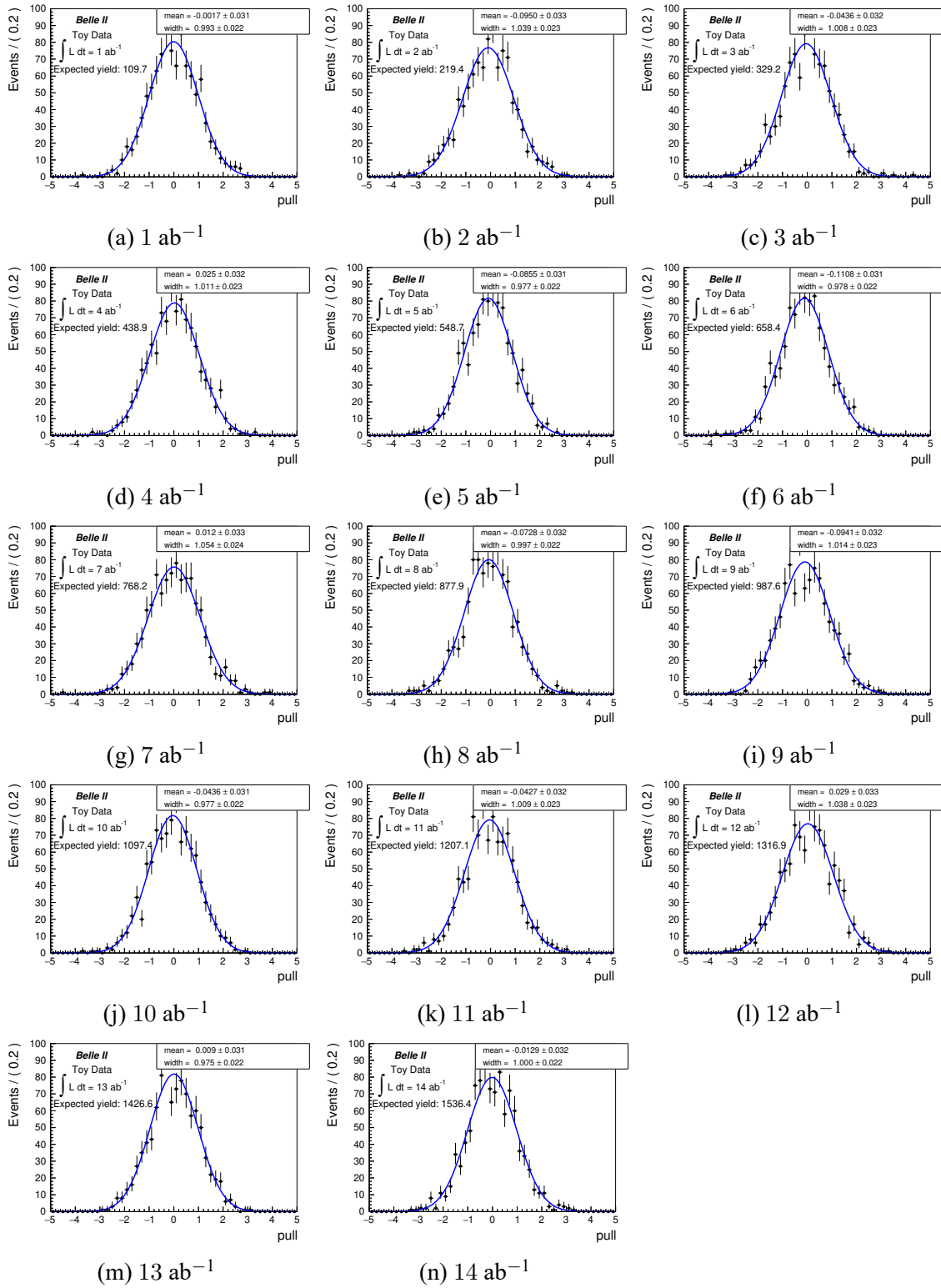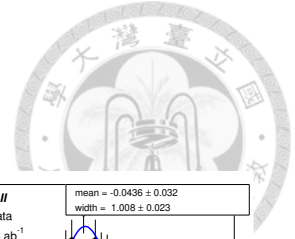(b) 2 ab^{-1}

(c) 3 ab^{-1}

(d) 4 ab^{-1}

(e) 5 ab^{-1}

(f) 6 ab^{-1}

(g) 7 ab^{-1}

(h) 8 ab^{-1}

(i) 9 ab^{-1}

(j) 10 ab^{-1}

(k) 11 ab^{-1}

(l) 12 ab^{-1}

(m) 13 ab^{-1}

(n) 14 ab^{-1}

Figure C.3: Toy results

(a) 1 ab^{-1}

(b) 2 ab^{-1}

(c) 3 ab^{-1}

(d) 4 ab^{-1}

(e) 5 ab^{-1}

(f) 6 ab^{-1}

(g) 7 ab^{-1}

(h) 8 ab^{-1}

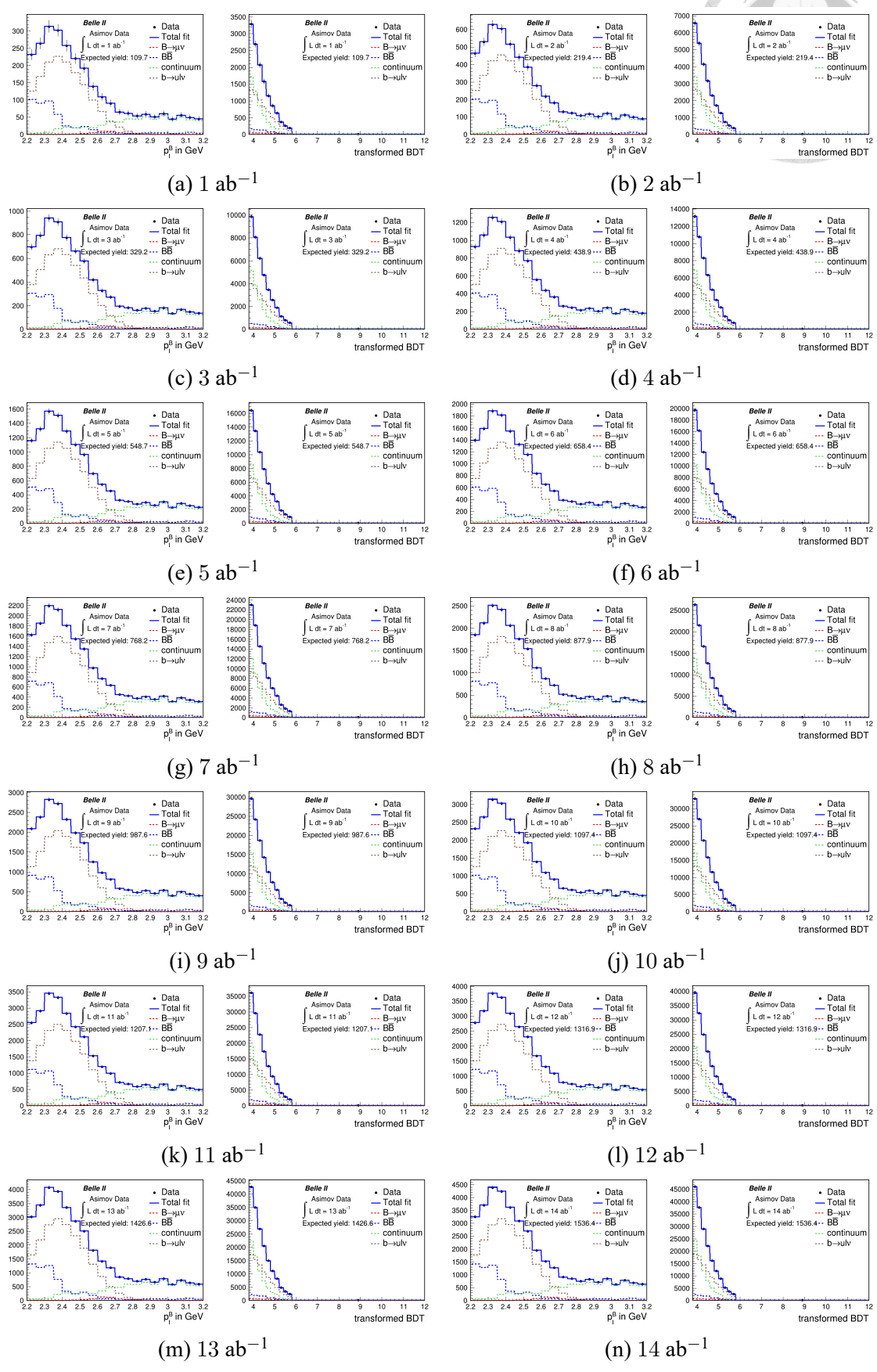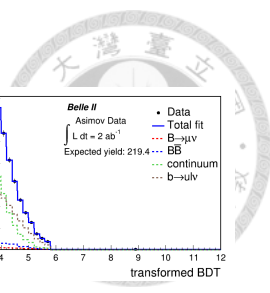(i) 9 ab^{-1}

(j) 10 ab^{-1}

(k) 11 ab^{-1}

(l) 12 ab^{-1}

(m) 13 ab^{-1}

(n) 14 ab^{-1}

Figure C.4: Fit to Asimov's data