Study of the $B^0 \to \gamma\gamma$ decay at Belle and Belle II

A thesis submitted in the fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Shubhangi Krishan Maurya

Department of Physics
Indian Institute of Technology Guwahati
Guwahati - 781039, Assam, India

April, 2024

This thesis is dedicated to
my parents for their unconditional love and motivation.

"Life is not easy for any of us. But what of that? We must have perseverance and, above all, confidence in ourselves. We must believe that we are gifted for something and that this thing, at whatever cost, must be attained."

Marie Curie

Declaration

Shubhangi Krishan Maurya
Roll No. 186121024
Department of Physics
Indian Institute of Technology Guwahati
Guwahati, India
email: shubhang@iitg.ac.in

I hereby declare that works presented in the thesis entitled **"Study of the $B^0 \to \gamma\gamma$ decay at Belle and Belle II"** have been carried out by me under the supervision of **Prof. Bipul Bhuyan** at the Department of Physics, Indian Institute of Technology Guwahati, India. The thesis has not been submitted anywhere else for any degree. Works presented in the thesis are all my own unless referenced to the contrary in the thesis.

Date: 29th April, 2024

Signature

Certificate

Prof. Bipul Bhuyan
Department of Physics
Indian Institute of Technology Guwahati
Guwahati, India
email: bhuyan@iitg.ac.in

This is certified that the work contained in the thesis entitled **"Study of the $B^0 \to \gamma\gamma$ decay at Belle and Belle II"** by Ms. Shubhangi Krishan Maurya (Roll No 186121024), a Ph.D. student in the Department of Physics, Indian Institute of Technology Guwahati, is carried out under my guidance. The results demonstrated in this thesis have not been submitted to any other Institute or University for the award of any other degree or diploma.

Bipul Bhuyan

Date: 29th April, 2024

Signature

Acknowledgements

The completion of this thesis would not have been a success without the incredible support of so many people, and I genuinely appreciate each one of them for their contributions. If I inadvertently fail to acknowledge anyone, please accept my heartfelt apologies, as every contribution has been immensely appreciated and cherished.

First of all, I express my heartfelt gratitude to Prof. Bipul Bhuyan for graciously accepting to be my thesis supervisor and guiding me into the realm of experimental high-energy physics. His invaluable advice, insightful suggestions, and continuous support have been instrumental during my Ph.D. journey. Under his mentorship, I have been granted the freedom to explore and the privilege to grow, instilling in me a profound sense of dedication to this profession.

I wish to convey my sincere gratitude to the conveners of the EWP group, Dr. Alexander Glazov, Dr. Elisa Manoni, and Dr. Akimasa Ishikawa, for their unwavering support through their valuable suggestions and comments, which have played a pivotal role in making this analysis feasible. I am truly grateful to the members of the review committee at Belle and Belle II, namely Dr. Yosuke Yusa (chairperson), Dr. Eldar Ganiev, and Prof. Karin Schönning, for their thorough review of the analysis and publication draft, and for providing invaluable insights, suggestions, and feedback. Furthermore, I express my appreciation to Prof. Jim Libby, Dr. Diego Tonelli, and Dr. Justine Serrano for their useful suggestions for the improvement of this study. I am indebted to the Belle and Belle II Collaboration for providing the platform and every resource necessary to complete this analysis.

I acknowledge all the Belle India Group members who encouraged and supported me throughout my Ph.D. journey. Special thanks to the faculty members of collaborating institutes for organizing the Belle Analysis Workshops (BAW), which helped me to grasp various technical aspects of analysis procedures, while the physics talks deepened my understanding of both experimental and theoretical aspects of particle physics.

I am thankful to my Ph.D. doctoral committee members, Dr. Debashis Borah, Dr. Meduri C Kumar, and Prof. Poulose Poulose, for giving their valuable time towards the completion of my academic responsibilities.

I am deeply grateful to my alma mater, the Indian Institute of Technology Guwahati, and to the MHRD, Government of India, for their invaluable support and cooperation, which were instrumental in helping me successfully complete my research.

It brings me immense joy to express my heartfelt thanks to my colleagues and

seniors in the EHEP group - Dr. Atanu Nath, Dr. Deepanjali Goswami, Maharnab "dada", Dibyajyoti "dada", Dr. Jyotirmoi Borah, Dr. Devender, Shivam, Shailesh, Arkodip, and Rajesh. Your friendship, support, and encouragement have truly kept me going. Whether during moments of celebration or challenges, you've always been there to listen and offer words of encouragement. Thank you for being such an essential part of my Ph.D. journey.

With immense gratitude, I thank my batchmates and cherished friends, Pushpesh, Sayan, Pushpanjali and Sudeshna, for enriching my life with memorable experiences and laughter. I will cherish the moments spent with them forever.

I am grateful for the support and care provided by my best friend, Harshita. Your understanding and empathy during my tough times, along with your positive outlook and sense of humour, were a lifeline for me. Thank you for always being there, cheering me on, and reminding me to find joy and balance in the midst of academic challenges. Your friendship means everything to me, and I am eternally thankful for having you in my life.

I am deeply grateful to my dear Mangesh for coming into my life during a critical period and making such a significant positive impact. Your support and encouragement have been invaluable and will continue to inspire me in the future.

I am deeply indebted to my parents for their unwavering love, encouragement, and sacrifices throughout my Ph.D. journey. Their endless support, belief in my abilities, and sacrifices have shaped me into who I am today. I am profoundly thankful for their guidance, patience, and unwavering faith in me, which have been my greatest source of strength and motivation. This thesis is as much theirs as it is mine, and I dedicate it to them with all my heart. Thank you, Mummy and Papa, for being my pillars of strength and for believing in me every step of the way. Big thanks to my siblings, Arushi, Vatsalya, and Vedika, for always believing in me and cheering me on. Your support means everything.

And finally, I must also thank Nature for giving us all we require to live and for teaching us vital lessons about how things move, being kind, and living in peace and balance. Its beauty reminds us to appreciate the small moments and find joy in the simple things, making life more beautiful and meaningful.

Shubhangi

List of Publications

Publication from thesis

- Search for the decay $B^0 \to \gamma\gamma$ using combined data of Belle and Belle II, S. K. Maurya, D. Kalita, B. Bhuyan *et al.* (Belle & Belle II Collaboration), Phys. Rev. D **110** L031106
- Study of the $B^0 \to \gamma\gamma$ decay at Belle and Belle II, S. K. Maurya, B. Bhuyan, Springer Proc. Physics 304, 1219-1221 (2024), Presented at XXV DAE-HEP Symposium 2022 (Conference proceedings).
- Study of the $B^0 \to \gamma\gamma$ decay at Belle and Belle II, S. K. Maurya, D. Kalita, B. Bhuyan, the Belle II internal note 2022-061.

Other publication

- Search for the decay $B_s^0 \to \pi^0\pi^0$ at Belle. Phys. Rev. D **107**, L051101, (2023).

Further publications can be accessed through the link iNSPIRE-HEP.

Abstract

We report the result of a search for the rare decay $B^0 \to \gamma\gamma$ using a combined dataset of 753×10^6 $B\bar{B}$ pairs collected by the Belle experiment and 387×10^6 $B\bar{B}$ pairs collected by the Belle II experiment at the $\Upsilon(4S)$ resonance produced in electron-positron collisions. A simultaneous fit to the Belle and Belle II data sets yields $11.0^{+6.5}_{-5.5}$ signal events. We determine the branching fraction $\mathcal{B}(B^0 \to \gamma\gamma) = (3.7^{+2.2}_{-1.8}(\text{stat}) \pm 0.5(\text{syst})) \times 10^{-8}$ with a signal significance of 2.5σ and set a 90% confidence level upper limit of $\mathcal{B}(B^0 \to \gamma\gamma) < 6.4 \times 10^{-8}$. This result improves on the previously published 90% confidence level upper limit by BABAR by a factor of about five and provides the most stringent limit to date.

Preface

The exploration of rare decays has been a captivating subject in the realm of particle physics for decades. These decays, characterized by their low occurrence rates, provide a unique window into the fundamental building blocks of our universe and offer valuable insights into the underlying theories that govern the behaviour of particles. Among these rare processes, the decay $B^0 \to \gamma\gamma$ has emerged as a particularly intriguing phenomenon. Investigating this decay is of utmost importance as it serves as a testing ground for our understanding of the Standard Model of particle physics and potentially opens the door to discovering new physics beyond its boundaries.

This thesis focuses on the search for the rare decay $B^0 \to \gamma\gamma$, conducted at both the Belle and Belle II experiments. The Belle experiment at the KEKB electron-positron collider in Japan has been at the forefront of precision measurements and groundbreaking discoveries in particle physics. Its successor, Belle II, with enhanced capabilities and higher luminosity, promises even greater sensitivity in the exploration of rare decays. In the Standard Model, the decay of $B^0 \to \gamma\gamma$ occurs through a flavor-changing neutral current (FCNC) transition involving electroweak loop diagrams. The theoretical prediction of branching fraction is $1.4^{+1.4}_{-0.8} \times 10^{-8}$. The present experimental upper limit (UL) on the branching fraction at 90% confidence limit (CL) $< 3.2 \times 10^{-7}$ set by the BABAR experiment with an integrated luminosity of 426 fb^{-1}. This analysis will be the first attempt to search for this decay using the combined dataset from the Belle and Belle II experiment, expecting to reach the SM sensitivity.

This thesis presents the results of the search for the decay $B^0 \to \gamma\gamma$ at high energy e^+e^- collisions. The search uses a combined dataset comprising an integrated luminosity of 694 fb^{-1} from the Belle and 362 fb^{-1} from the Belle II detector at $\Upsilon(4S)$ energy. No significant signal yield is observed for the decay $B^0 \to \gamma\gamma$ and in the absence of any significant signal yield, we set the 90% CL UL on its branching fraction to be $< 6.4 \times 10^{-8}$. This is an improvement by a factor of 5 compared to the previous measurement of UL by the BABAR experiment. With a dedicated run at $\Upsilon(4S)$ resonance to collect more statistics, the Belle II experiment has the potential to discover the $B^0 \to \gamma\gamma$ decay.

Chapter 1 introduces the theoretical foundations of the Standard Model, covering the dynamics of electroweak and strong interactions, along with their associated gauge transformations. We formulate the most comprehensive renormalizable Lagrangian, with a particular emphasis on its kinetic and Yukawa components. These components involve gauge interactions and the flavor-changing aspects of quarks and leptons. It then provides a brief overview of the Cabibbo-Kobayashi-Maskawa (CKM) matrix, responsible for the flavor physics in the quark sector, and its most convenient Wolfenstein parameterization. We extend our discussions on the unitarity of the CKM matrix and the angles and sides of the unitarity triangle. This chapter also highlights the motivations behind the search for $B^0 \to \gamma\gamma$ and the potential implications of its observation.

Chapter 2 offers an overview of the experimental setups and techniques employed at both the Belle and Belle II experiments. It covers the accelerator complex, the Belle and Belle II detectors, and the data acquisition and trigger system. This chapter also addresses the improvements and upgrades implemented in the transition from Belle to Belle II, enabling higher luminosity and enhanced sensitivity for rare decay searches. It includes the Belle II analysis software framework and B2BII, suitable for studying Belle II and Belle data simultaneously and the strategies employed for this analysis.

Chapter 3 focuses on the methodology used to search for the $B^0 \to \gamma\gamma$ decay at Belle and Belle II. It details the process of event generation and simulation to perform MC studies. The data samples and the various skim were used for both analyses are covered in this chapter. The chapter also discusses the selection criteria for B^0 mesons, identifying photons, and estimating and reducing background sources. This chapter also presents a detailed study of the suppression of dominant backgrounds using multivariate analysis.

Chapter 4 briefly discusses the unbinned extended maximum likelihood procedure used to extract the signal yields. We present the PDF parametrization of the signal and background for both analysis. This chapter also describes the ensemble test performed to test the stability of the fitters and their linearity among the fitting variables.

Chapter 5 provides a comprehensive study performed to validate the analysis strategy using the control mode. The decay channel, $B^0 \to K^*\gamma$ used as a control sample for this purpose. The calibration of the BDT and π^0/η veto is performed. Furthermore, a section is devoted to looking for the data-MC differences for the fitting variables of the control sample and sideband regions for the target mode, and in data off-resonance to cross-check the MC simulation of the continuum background.

Finally, **Chapter 6** concludes the thesis by summarizing the final results of both analysis utilizing a combined dataset comprising an integrated luminosity of

694 fb^{-1} from Belle and 362 fb^{-1} from Belle II. Several sources of systematic uncertainties and their estimation procedure are discussed in detail in this chapter. The estimation of 90% CL upper limit for $B^0 \to \gamma\gamma$ analysis is explained in detail in this chapter. This chapter also suggests avenues for future research and explores potential directions for further investigation.

In conclusion, this thesis aims to contribute to the ongoing search for rare decays and the exploration of physics beyond the Standard Model. By focusing on the search for the $B^0 \to \gamma\gamma$ decay at both the Belle and Belle II experiments, we strive to provide valuable insights into the behaviour of particles and potentially uncover new physics phenomena. The results presented in this thesis serve as a foundation for further investigations and pave the way for future advancements in particle physics.

Contents

List of Figures

List of Tables

Chapter 1

Theoretical Framework and the Motivation

1.0 Introduction

This chapter introduces the theoretical foundations of the Standard Model (SM), covering the dynamics of electroweak and strong interactions, along with their associated gauge transformations. We formulate the most comprehensive renormalizable Lagrangian, with a particular emphasis on its kinetic and Yukawa components. These components involve gauge interactions and the flavor-changing aspects of quarks and leptons. It then provides a brief overview of the Cabibbo-Kobayashi-Maskawa (CKM) matrix, responsible for the flavor physics in the quark sector, and its most convenient Wolfenstein parameterization. We extend our discussions on the unitarity of the CKM matrix and the angles and sides of the unitarity triangle. This chapter also highlights the motivations behind the search for $B^0 \to \gamma\gamma$ and the potential implications of its observation.

2.0 The Standard Model

The SM of particle physics stands as one of the most successful theories in modern physics, providing a comprehensive framework for understanding the fundamental building blocks of matter and the forces that govern their interactions. Developed over several decades through experimental observations and theoretical advancements, the SM encapsulates our current understanding of the fundamental particles and their interactions. The SM depicts nature as comprised of two distinct types of particles: fermions and bosons. Fermions possess half-integer spin and follow Fermi-Dirac statistics [1, 2], while bosons exhibit integer spin and follow Bose-Einstein statistics [3]. Fundamental fermions, which are spin-$\frac{1}{2}$ particles, are further categorized into quarks and leptons, each with six types, also called the *flavors*. These categories are organized into three 'generations' based on particle masses. Each generation comprises two flavors, each possessing a distinct electrical charge: up (u)

and down (d) for the first generation, charm (c) and strange (s) for the second, and top (t) and bottom (b) for the third. The electric charge of up, charm, and top quarks is $+\frac{2}{3}$, while down, strange, and bottom quarks carry an electric charge of $-\frac{1}{3}$. Similarly, lepton generations also consist of three flavors: electron (e) for the first generation, muon (μ) for the second, and tau (τ) for the third. Each of these particles corresponds to a neutrino: the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). Unlike quarks, electrons, muons, and tau leptons possess an electric charge of -1, while neutrinos are electrically neutral. Within the framework, there are four spin-1 gauge bosons and one scalar boson. These gauge bosons serve as the mediators of the fundamental forces that govern interactions among particles. Meanwhile, the only scalar boson is the Higgs boson (H), which arises as an excitation of the Higgs field. This Higgs field plays a crucial role in the SM, as it is responsible for generating mass for the fundamental particles. The three fundamental forces are each mediated by their own set of gauge bosons. The strong force is mediated by eight types of gluons (g), the electromagnetic force is mediated by the photon (γ), and the weak force is mediated by two charged W bosons ($W^\pm$) and one neutral Z boson (Z). The processes involving flavour changes are governed by the $W^\pm$ bosons. Additionally, the Z boson is a weak force carrier due to its significant mass but does not change the flavor. The fundamental particles of the SM, including the fermions (quarks and leptons) and the gauge bosons, are summarized in Figure 1.1.

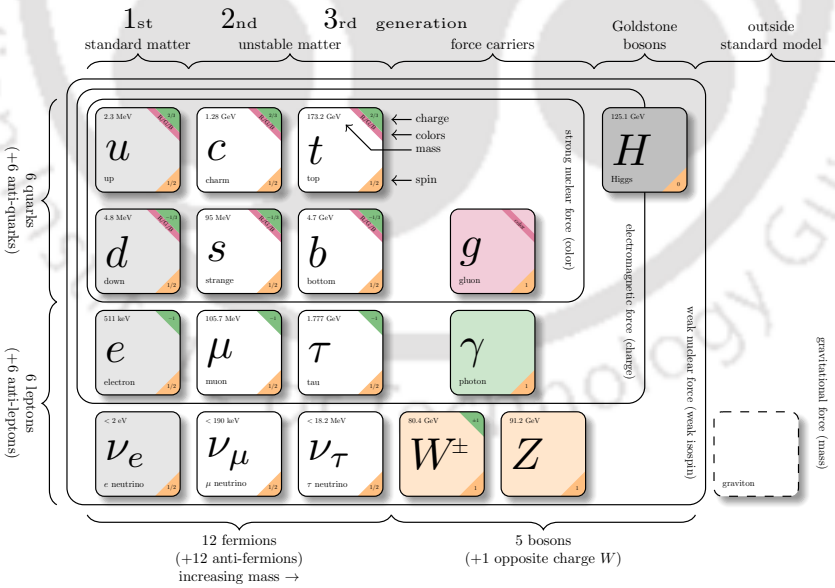


Figure 1.1: Scheme of particles and interactions in the SM. Their masses, charges, spin, weak isospin and colours are listed. Credit: Ref. [4]

The SM mathematical description is the Quantum Field Theory (QFT), composed of Quantum Mechanics, Field theory and Special relativity. It is a gauge

theory based on the group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The strong interaction is described by the symmetry group $SU(3)_C$, where SU stands for Special Unitary group and C represents color. This group defines Quantum Chromodynamics (QCD) theory, illustrating the strong interaction between quarks mediated by gluons. The group's generators are the eight gluons which mediate the strong force among the quarks. Quarks can combine to form composite hadrons such as the proton (uud), neutron (udd), or pions (e.g. $u\bar{d}$). The weak interaction is best described by the symmetry group $SU(2)_L$, where L denotes left-handedness. Together with the $U(1)_Y$ group, where Y represents hypercharge, they form the Electroweak theory with the combined symmetry group $SU(2)_L \times U(1)_Y$, which after spontaneous symmetry breaking (SSB), give rise to the mediating gauge bosons for the weak and electromagnetic interactions, ie. $W^{\pm}$, Z, and photons, respectively. The unitary group, $U(1)_Y$, represents the electromagnetic interactions upon SSB and characterizes how electrically charged particles interact by means of the exchange of photons. The weak interaction is unique in its ability to change the flavor of leptons through the exchange of W bosons [5, 6].

3.0 Theoretical Framework

The SM with its associated particles in gauge theory can be defined by the following ingredients [7]:

- The symmetry is local

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{1.1}$$

- There exist three fermion generations, each comprising five representations.

$$Q^I_{Li}(3,2)_{+\frac{1}{6}}, U^I_{Ri}(3,1)_{+\frac{2}{3}}, D^I_{Ri}(3,1)_{-\frac{1}{3}}, L^I_{Li}(1,2)_{-\frac{1}{2}}, E^I_{Ri}(1,1)_{-1}. \tag{1.2}$$

  where, the left-handed quarks, denoted by $Q^I_{Li}(3,2)_{+\frac{1}{6}}$, form triplets under $SU(3)_C$, doublets under $SU(2)_L$, and possess a hypercharge $Y = +\frac{1}{6}$. The superscript I indicates gauge interaction eigenstates, while the subscript i=1,2,3 represents the flavor (or generation) index. The notations used to represent the quark and lepton fields can be similarly extended to the other particle representations within the SM framework. For quarks, the right-handed up-type (u, c, t) and down-type (d, s, b) quark fields are denoted as U_R and D_R, respectively. These right-handed quark fields transform as triplets under the $SU(3)$ group for strong interactions and as singlets under the $SU(2)$ group.

  In contrast, left-handed lepton fields, including both charged leptons (e, μ, τ) and neutrinos $(\nu_e, \nu_\mu, \nu_\tau)$, are represented as L_L, while right-handed lepton fields, corresponding to charged leptons, are denoted as E_R. These left-handed (right-handed) lepton fields transform as singlets under the $SU(3)$ group and as doublets (singlets) under the $SU(2)$ group.

- There is a single scalar field, $\phi(1,2)_{+1}$ which acquires a vacuum expectation value, v after SSB as

$$\langle\phi\rangle = \langle 0|\,\phi\,|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}, \tag{1.3}$$

The Lagrangian of the SM can be written as the sum of three main components:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Kinetic}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}, \tag{1.4}$$

where, $\mathcal{L}_{\text{Kinetic}}$, $\mathcal{L}_{\text{Higgs}}$ and $\mathcal{L}_{\text{Yukawa}}$ denotes the kinetic, Higgs and Yukawa component of the SM Lagrangian.

The kinetic part of the Lagrangian in the equation 1.4 describes the dynamics of the gauge fields and the matter fields. It can be written as:

$$\mathcal{L}_{\text{Kinetic}} = \mathcal{L}_{\text{k}}(Q^I_{Li}) + \mathcal{L}_{\text{k}}(U^I_{Ri}) + \mathcal{L}_{\text{k}}(D^I_{Ri}) + \mathcal{L}_{\text{k}}(L^I_{Li}) + \mathcal{L}_{\text{k}}(E^I_{Ri}), \tag{1.5}$$

In this thesis, the focus is on exploring quark flavor dynamics within the SM. Therefore, we will concentrate on the quark sector and their associated Yukawa interactions, rather than the full SM Lagrangian. The expansion of the left-handed quark doublet Q^I_L is as follows

$$\mathcal{L}_k(Q^I_L) = i\bar{Q}L^I_{Li}\gamma_\mu\left(\partial^\mu + \frac{i}{2}g_s G^\mu_a\lambda_a + \frac{i}{2}gW^\mu_b\tau_b + \frac{1}{6}g'B^\mu\right)Q^I_{Li}, \tag{1.6}$$

Here, G^μ_a represents the eight gluon fields, $\frac{\lambda_a}{2}$ are the generators of $SU(3)_C$, g and g' are the coupling constants for weak and electroweak interactions respectively, W^μ_b denotes the three weak interaction bosons, $\frac{\tau_b}{2}$ represents the generators of $SU(2)_L$, and B^μ stands for the single hypercharge boson.

The second term of the equation 1.4 describes the dynamics of the Higgs field and its interactions with the gauge fields, which can be written as

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi), \tag{1.7}$$

where, $\phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}$, which is a complex scalar field and $D_\mu\phi = (\partial^\mu + \frac{i}{2}gW^\mu_b\tau_b + \frac{1}{6}g'B^\mu)$ is the covariant derivative. The covariant derivative includes the interactions of the Higgs field with the gauge bosons.

When the Higgs field acquires a non-zero vacuum expectation value v after spontaneous symmetry breaking, this Higgs potential term gives rise to the masses of the $W^\pm$ and Z bosons, as well as the masses of the fundamental fermions through the Yukawa interactions.

The third term of the equation 1.4, which describes the Yukawa interactions, is written as

$$-\mathcal{L}_{\text{Yukawa}} = Y^d_{ij}\bar{Q}^I_{Li}\phi D_{Rj} + Y^u_{ij}\bar{Q}^I_{Li}\phi' U^I_{Rj}, \tag{1.8}$$

where, Y_{ij}^d and Y_{ij}^u are the Yukawa coupling constants for up-type quarks and down-type quarks, respectively. ϕ is the Higgs field, and ϕ' is its conjugate, given by $\phi' = i\tau_2\phi^*$ in terms of the Pauli matrix τ_2. This term is the crucial part where the masses of the quarks and the phenomenon of quark flavor arise within the SM framework.

In the subsequent sections, we will delve deeper into the discussion of the CKM matrix, which serves as the origin of the quark flavor physics program within the SM.

3.1 The CKM Matrix

The CKM matrix is the matrix that connects the quark mass eigenstate basis to the weak interaction eigenstate basis. There are two important bases - the mass basis, where the quark masses are diagonal, and the interaction basis, where the $W^\pm$ interactions are diagonal. The disparity between these bases gives rise to flavor-changing interactions. Since most measurements are performed in the mass basis, the interactions are typically expressed in that basis using the CKM matrix.
Upon the SSB, the scalar field ϕ in equation 1.8 is chosen to acquire the following form,

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \tag{1.9}$$

where, h serves the purpose of the Higgs boson. We decompose the $SU(2)_L$ quark doublets into their components:

$$Q_{Li}^I = \begin{pmatrix} u_{Li}^I \\ d_{Li}^I \end{pmatrix}, \tag{1.10}$$

and then the Yukawa interactions, equation 1.8, give rise to mass terms [7]:

$$\begin{aligned} -\mathcal{L}_{\text{Yukawa}} &= \frac{v}{\sqrt{2}}[Y_{ij}^d \bar{d}_{Li}^I d_{Rj}^I + Y_{ij}^u \bar{u}_{Li}^I u_{Rj}^I] + \text{other terms} + \text{h.c.} \\ &= M_{ij}^d \bar{d}_{Li}^I d_{Rj}^I + M_{ij}^u \bar{u}_{Li}^I u_{Rj}^I + \text{other terms} + \text{h.c.} \end{aligned} \tag{1.11}$$

where considering

$$M_{ij}^d = \frac{v}{\sqrt{2}} Y_{ij}^d \tag{1.12}$$

$$M_{ij}^u = \frac{v}{\sqrt{2}} Y_{ij}^u \tag{1.13}$$

The interaction terms of the fermion fields to the Higgs field are omitted. To obtain the correct mass terms, the diagonalization of the matrices M^d and M^u is essential, a process accomplished through the principles of biunitary transformation.

$$M_{diag}^d = V_L^d M_{ij}^d V_R^{d\dagger} \tag{1.14}$$

$$M_{diag}^u = V_L^u M_{ij}^u V_R^{u\dagger} \tag{1.15}$$

where, V_L^d, V_R^d, V_L^u, and V_R^u are unitary matrices.
Using the requirement that the matrices V are unitary ($V^{d\dagger}V^d = \mathbb{1}$) the Lagrangian can now be expressed as follows:

$$\begin{aligned} -\mathcal{L}_{\text{Yukawa}} &= \bar{d}^I_{Li} V_L^{d\dagger} V_L^d M_{ij}^d V_R^{d\dagger} V_R^d d^I_{Rj} \\ &\quad + \bar{u}^I_{Li} V_L^{u\dagger} V_L^u M_{ij}^u V_R^{u\dagger} V_R^u u^I_{Rj} + \text{other terms} + \text{h.c.} \\ &= \bar{d}_{Li} (M_{ij}^d)_{\text{diag}} d_{Rj} + \bar{u}_{Li} (M_{ij}^u)_{\text{diag}} u_{Rj} + \text{other terms} + \text{h.c.} \end{aligned} \tag{1.16}$$

where the matrices V are absorbed in the quark states, resulting in the following quark mass eigenstates:

$$\begin{aligned} d_{Li} &= (V_L^d)_{ij} d^I_{Lj} & d_{Ri} &= (V_R^d)_{ij} d^I_{Rj} \\ u_{Li} &= (V_L^u)_{ij} u^I_{Lj} & u_{Ri} &= (V_R^u)_{ij} u^I_{Rj} \end{aligned} \tag{1.17}$$

Thus, the Yukawa Lagrangian, $\mathcal{L}_{\text{Yukawa}}$, can now be represented in the mass eigen basis. Using this transformation, the kinetic part of the Lagrangian from equation 1.6 becomes

$$\mathcal{L}_{\text{kinetic}} = -\frac{g}{2} \gamma_\mu W^{+\mu} \bar{u}^I_{Li} d^I_{Lj} + \text{otherterms}. \tag{1.18}$$

Now, to transform $\bar{u}^I_{Li}$ into the mass eigenbasis, we utilize equation 1.17, resulting in the transformation equation

$$\bar{u}^I_{Li} d^I_{Lj} = \bar{u}_{Li} (V_L^{u\dagger} V_L^d)_{ij} d_{Lj}, \tag{1.19}$$

where, u_L and d_L are in mass eigenbasis. Now, the equation 1.18 takes the form in mass basis:

$$\mathcal{L}_{\text{kinetic}} = -\frac{g}{2} \gamma_\mu \bar{u}_{Li} (V_L^{u\dagger} V_L^d)_{ij} d_{Lj} W^{+\mu} + \text{otherterms}. \tag{1.20}$$

where $V = V_L^{u\dagger} V_L^d$ is a 3×3 unitary matrix in the generation space called the CKM matrix. The interaction eigenstates of the down-type quarks relate to the mass eigenstates via the CKM matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.21}$$

Each element in the matrix (V_{ij}) represents the transition strength between corresponding quarks. This matrix is constructed to be unitary, denoted by $V^\dagger V = I = VV^\dagger$. A 3×3 complex matrix typically possesses 9 real and 9 imaginary parameters. However, the unitarity constraints eliminate 9 parameters, leaving three mixing angles and 6 complex phases. Of the 6 phases, 5 can be absorbed by rephasing the quark fields. As a result, the CKM matrix can be entirely described by four independent parameters: three Euler angles ($\theta_{12}, \theta_{23}, \theta_{32}$) and one complex phase (δ) [8]. Utilizing these four parameters, one can parametrize the CKM matrix, expressed as:

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.22}$$

where, $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$ with $(i, j \in [1, 3]$ and $i \neq j)$, and δ is the only phase in the CKM matrix which is responsible for the Charge-Parity (CP) violation within the SM. The CKM matrix best known from the experimental measurements is [9]

$$V_{\rm CKM} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}. \quad (1.23)$$

As shown in Equation 1.23, the experimental values of the CKM elements $|V_{ud}|$, $|V_{cs}|$, and $|V_{tb}|$ are close to 1, while the off-diagonal elements are significantly smaller, following a hierarchical pattern, for example, $|V_{ud}| >> |V_{us}| >> |V_{ub}|$. Hence, the CKM matrix is expanded around the identity matrix in terms of the relatively small parameter $\lambda = \sin\theta_c = 0.225$, where $\theta_c \equiv \theta_{12}$ represents the Cabibbo angle. The CKM matrix with unitarity condition to the order of $\mathcal{O}(\lambda^4)$ is given as

$$V_{\rm CKM} = \begin{bmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + \mathcal{O}(\lambda^4), \quad (1.24)$$

where ρ, η, and A are the real parameters. This matrix is known as the Wolfenstein parameterization of the CKM matrix [10]. The Wolfenstein parameters are defined as [9, 10, 11],

$$\lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \quad \bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}, \quad A\lambda^2 = \lambda\left|\frac{V_{cb}}{V_{us}}\right|. \quad (1.25)$$

In terms of the standard parametrization of the CKM matrix given in equation 1.22,

$$s_{12} = \lambda, \qquad s_{23} = A\lambda^2, \qquad s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta)$$
$$\rho = \frac{s_{13}e^{i\delta}}{s_{12}s_{23}}\cos\delta \qquad \eta = \frac{s_{13}e^{i\delta}}{s_{12}s_{23}}\sin\delta. \quad (1.26)$$

3.2 The Unitarity Triangle

The unitarity condition of the CKM matrix follows that

$$\sum_i V_{ij}V_{ik}^* = \delta_{jk} \quad \text{and} \quad \sum_j V_{ij}V_{kj}^* = \delta_{ik}. \quad (1.27)$$

One of the six non-diagonal equations, most relevant to the flavor physics in B-mesons, is

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0,$$

which can be written as

$$\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + 1 + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0, \tag{1.28}$$

by dividing each term by $V_{cd}V_{cb}^*$. Defining the [11]

$$\overline{\rho} + i\overline{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \quad \text{and} \quad 1 - (\overline{\rho} + i\overline{\eta}) = \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}, \tag{1.29}$$

the unitarity condition in equation 1.28 can be expressed as a triangle with vertices at $(0,0)$, $(1,0)$, and $(\overline{\rho},\overline{\eta})$ in the $\overline{\rho}-\overline{\eta}$ complex plane, known as the unitarity triangle (UT). The UT corresponding to the above equation is shown in Figure 1.2. The parameters $\overline{\rho}$ and $\overline{\eta}$ are related to the original parameters as

$$\rho + i\eta = \frac{\sqrt{1 - A^2\lambda^4}(\overline{\rho} + i\overline{\eta})}{\sqrt{1 - \lambda^2[1 - A^2\lambda^4(\overline{\rho} + i\overline{\eta})]}}. \tag{1.30}$$

The angles of the UT are

$$\begin{aligned} \beta \quad \text{or} \quad \phi_1 &= \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \\ \alpha \quad \text{or} \quad \phi_2 &= \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \\ \gamma \quad \text{or} \quad \phi_3 &= \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \end{aligned} \tag{1.31}$$

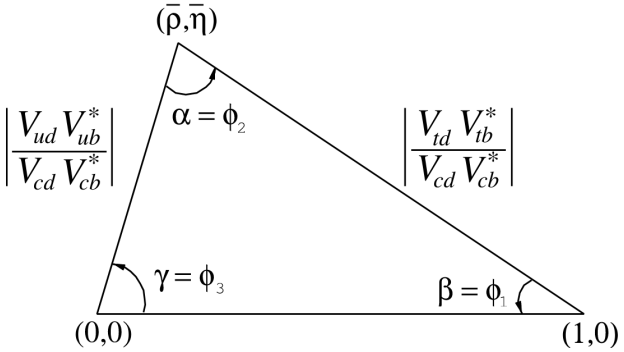


Figure 1.2: The unitarity triangle [9].

Accurate measurement of the angles and sides of the UTs through flavor-changing processes serves as a test for the unitarity of the CKM matrix. Moreover, precise measurements of UT parameters are crucial for constraining various physics models beyond the SM. The combined results of numerous independent studies in flavor physics yield the Wolfenstein parameters, as reported in [9],

$$\begin{aligned} \lambda &= 0.22650 \pm 0.00048 & A &= 0.790^{+0.017}_{-0.012} \\ \overline{\rho} &= 0.141^{+0.016}_{-0.017} & \overline{\eta} &= 0.357 \pm 0.011. \end{aligned} \tag{1.32}$$

Despite the tremendous successes of the SM in explaining a wide range of experimental data, it is regarded as an effective theory that may be part of a more comprehensive framework at higher energy scales. This is evident from its inability to address certain phenomena, such as the non-zero masses of neutrinos revealed by neutrino oscillations and the matter-antimatter asymmetry observed in the universe, which cannot be fully explained by the CP-violating phase δ alone. Additionally, the existence of dark matter and its interactions with ordinary matter remains unaccounted for within the SM.

Historically, the discovery of the c and t quarks was driven by investigations in flavor physics. Therefore, exploring new fundamental particles or interactions at high energy scales can be indirectly probed through flavor physics studies at lower energy scales, which are currently accessible in high luminosity experiments such as Belle, Belle II, LHCb, and others.

4.0 Significance of the Decay $B^0 \to \gamma\gamma$

The rare decay $B^0 \to \gamma\gamma$ provides an exceptionally clean probe of the SM and physics beyond it. In the SM, $B^0 \to \gamma\gamma$ proceeds through a flavor-changing neutral current (FCNC) transition involving electroweak loop diagrams. Such decays, forbidden at the tree level, are highly sensitive to contributions beyond the SM (see Sec 9.1 in [12]). In $B^0 \to \gamma\gamma$, there is no direct interaction between the b quark and the d quark. Instead, an effective FCNC is induced by a *one-loop* or *penguin* diagram, where a quark emits and reabsorbs a W^- gauge boson, leading to a flavor change. Possible Feynman diagrams contributing to this channel are shown in Fig 1.3. The

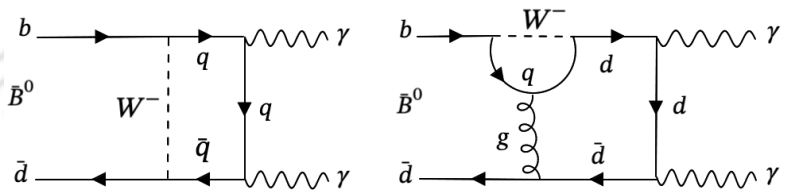


Figure 1.3: Box (left) and penguin (right) diagrams contributing to $B^0 \to \gamma\gamma$ at leading order in the SM. The symbol q represents a u, c or t quark.

decay $B^0 \to \gamma\gamma$ & $B_s \to \gamma\gamma$ are closely related, with the $b \to d\gamma\gamma$ transition being suppressed with respect to $b \to s\gamma\gamma$ by CKM factors $(|V_{td}|^2/|V_{ts}|^2) \sim 0.04$. Hadron dynamics introduces uncertainties in predicting the branching fraction for these decays and may modify the ratio away from the CKM implied value [13].

The presence of charged current mediated by W^- gauge boson can be a valuable area to search for new physics as the presence of more mediators of a different kind can eventually affect the branching fractions. A precise measurement of the Branching fraction ($\mathcal{B}$) is thus essential to validate other models associated with such a process. The charge-conjugation and parity-averaged branching fraction for

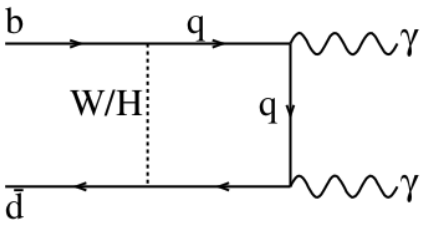


Figure 1.4: Example of possible diagram responsible for the decay $B^0 \to \gamma\gamma$. In these diagrams $q = u, c, or$ t, and H is a hypothetical charged non-standard-model Higgs boson.

the decay $B^0 \to \gamma\gamma$ is predicted in the SM to be $(1.4^{+1.4}_{-0.8}) \times 10^{-8}$ [14], including next-to-leading logarithmic and next-to-leading power (NLP) corrections. The inclusion of NLP corrections can change these leading-order branching fraction results by 15% or more. compared to just the leading power contributions. These NLP effects are systematically calculated using QCD factorization and soft-collinear effective theory (SCET). Long-distance penguin contribution is expected to be negligible. However, it can noticeably impact the CP-violating observables by approximately an amount of $\mathcal{O}(30\%)$ [15]. This mode is sensitive to new physics that could lead to an enhancement of the branching fraction due to possible contributions of non-SM heavy particles occurring in the loop of the leading-order Feynman diagrams [16, 17, 18]. This will allow the replacement of the W boson in Figure 1.4 with another charged particle, such as a charged Higgs boson. Despite their relatively small branching fractions within the SM and the challenges associated with detecting such exclusive reactions with two photons in the final states at the LHCb experiment, the radiative penguin decays $B_{d,s} \to \gamma\gamma$ have been identified as prominent channels for the Belle II program [12]. This is particularly due to the straightforward event reconstruction facilitated by the electromagnetic calorimeter.

The upper limit set by L3 collaboration on the branching fraction of this mode is 3.9×10^{-5} at 90% confidence level (CL) using 73 pb^{-1} of data [19]. The best previous experimental upper limit on the branching fraction of this mode is 3.2×10^{-7} at 90% CL set by BaBar[13] using 426 fb^{-1} of data. This decay process was also studied previously by the Belle collaboration, [20] and an upper limit was set at 6.2×10^{-7} at 90% CL using 104 fb^{-1} of data, corresponding to 111×10^6 $B\overline{B}$ pairs. The combined dataset from Belle and Belle II experiments so far provides a unique opportunity to reach the SM sensitivity.

5.0 Summary

In this chapter, we introduce SM, including electroweak and strong interactions and their gauge transformations. It emphasizes the renormalizable Lagrangian, focusing on kinetic and Yukawa components involving gauge interactions and flavor-changing properties of quarks and leptons. It highlights the CKM matrix, essential for quark

flavor physics, and its unitarity aspects. Additionally, it discusses motivations behind the search for $B^0 \to \gamma\gamma$ and its potential implications.

Chapter 2

The Belle and Belle II Experiment

This chapter gives an overview of the Belle and Belle II experiments. The study of $B^0 \to \gamma\gamma$ is carried out using the data collected by two experiments, Belle and Belle II. An overview of both the Belle and Belle II detector and their associated accelerators, KEKB and SuperKEKB, data acquisition and trigger systems are described in the following subsections. It also includes the software framework for physics analysis, physics of B^0 mesons at $\Upsilon(4S)$, and the strategy adopted for both analysis.

1.0 The Υ resonances and B meson factories

The Υ mesons are bound states of $b\bar{b}$ quarks with spin, parity and charge conjugation quantum numbers $J^{\mathcal{PC}} = 1^{--}$. After the first observation of the $\Upsilon(1S)$ in proton-nucleon collisions by the CFS (Columbia-Fermilab-Stony Brook) E288 collaboration at Fermilab [21], the existence of the Υ meson was confirmed by experiments at CESR [22] and DORIS [23, 24]. Figure 2.1 shows the total e^+e^- annihilation cross section as a function of the CM energy in the region of the Υ resonances.

The masses of the first three Upsilon resonances ($\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ are less than the sum of the mass of a B and $\bar{B}$ meson (5.279 GeV each), and therefore, they cannot decay to B mesons. Rather they can produce light hadrons (or lepton pairs) as shown in Figure 2.2.

The mass of the $\Upsilon(4S)$ resonance is ~ 20 MeV/c^2 above the mass of two B^+ or B^0 mesons; therefore, it is the lightest Υ resonance to be able to decay into a $B\bar{B}$ meson pair. In fact, it decays almost exclusively, $> 96\%$ according to [26], to B^+B^- and $B^0\bar{B}^0$ pairs. The $\Upsilon(4S)$ decay rate to B^+B^- is measured to be very consistent with the rate to $B^0\bar{B}^0$ [27]. The $\Upsilon(4S)$ mass is too small to produce excited B^* mesons or to create an extra particle. The $b\bar{b}$ production cross-section at the $\Upsilon(4S)$ resonance is about 1.1 nb and e^+e^- storage rings operating at the $\Upsilon(4S)$ resonance are called B-factories. A majority of e^+e^- collisions result in lighter $q\bar{q}$ states such as $u\bar{u}$, $d\bar{d}$, $s\bar{s}$ and $c\bar{c}$. These events are generally known as continuum events. At the $\Upsilon(4S)$ resonance, the continuum production cross-section is about 3.7 nb, more than three times larger than the $B\bar{B}$ cross-section. Therefore, although KEKB (SuperKEKB)

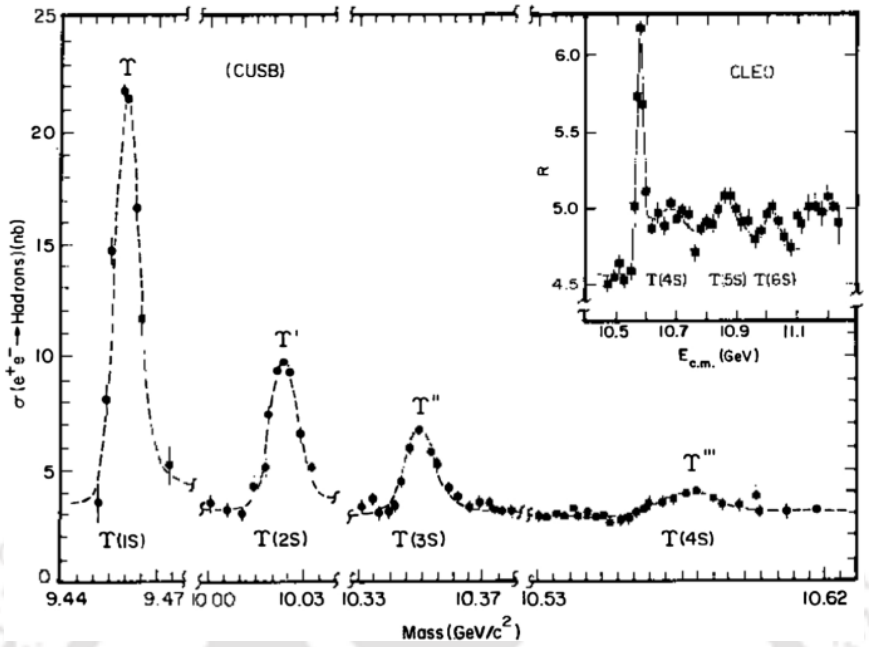


Figure 2.1: The hadronic cross-section for Υ(nS) resonances as a function of e^+e^- CM energy in nb and inset normalized to theoretical muon-pair cross-section. Results of CUSB Collaboration. [25]

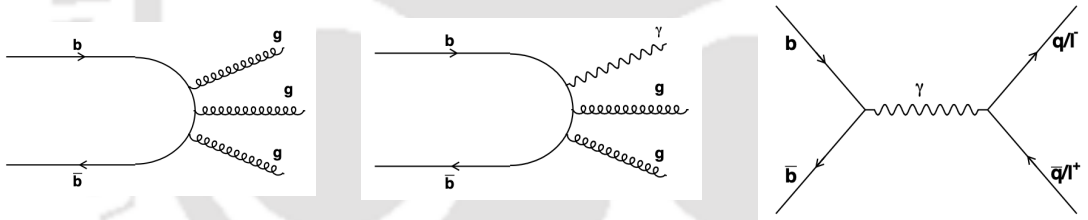


Figure 2.2: Decay processes of Υ(nS) system at first 3 resonance states, Υ(1S), Υ(2S) and Υ(3S)

.

is operating at the Υ(4S) resonance to produce $B\bar{B}$ pairs, it produces three times more $q\bar{q}$ events that constitute a major source of background to B physics. In order to study the $q\bar{q}$ contribution cleanly, KEKB (SuperKEKB) is sometimes operated at 60 MeV below the Υ(4S) resonance. The data taken at the Υ(4S) resonance and 60 MeV below are referred to as "on-resonance" and "off-resonance" respectively. The energy of the Υ(4S) resonance is sufficient for generating a pair of $b\bar{b}$ quarks while concurrently producing the lighter u or d quarks. Consequently, this process leads to the formation of pairs of $B\bar{B}$ mesons at the Υ(4S) resonance, as illustrated in Figure 2.3.

The ratio of the branching fraction of Υ(4S) to B^0B^0 and B^+B^- has also been

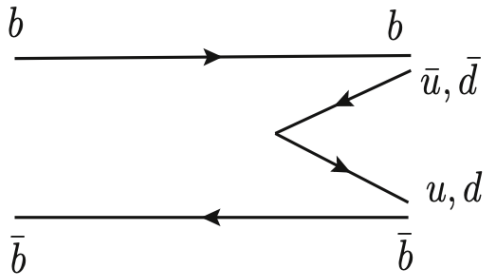


Figure 2.3: Feynman diagram for the production of $B\bar{B}$ at $\Upsilon(4S)$ resonance

measured by Belle [28] and is:

$$\frac{\mathcal{B}(\Upsilon(4S) \to B^+B^-)}{\mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0)} = 1.065 \pm 0.012 \pm 0.019 \pm 0.047$$

which is consistent with equal production rates for charged and neutral $B\bar{B}$ pairs. The discovery of $\Upsilon(5S)$ resonance (in 1985) [29] and the production of B_s^0 mesons at this resonance (in 2005) [30] was established by the CLEO collaboration. The $\Upsilon(5S)$ mass is substantially larger than the $B_s^0\bar{B}_s^0$ production threshold. Unlike the $\Upsilon(4S)$ resonance, the $\Upsilon(5S)$ resonance does not stand just above a $B\bar{B}$ threshold. B mesons are mostly produced through excited states B^* with subsequent low-energy photon de-excitation. The $b\bar{b}$ production at the $\Upsilon(5S)$ resonance is about 0.3 nb, hence more than three times smaller than at the $\Upsilon(4S)$ resonance.

The first generation of B-factories, DORIS II at DESY in Germany with the ARGUS experiment and CESR in Cornell, USA, with the CLEO detector, were symmetric colliders, where the electron and positron beams have the same energy, corresponding to one-half of the $\Upsilon(4S)$ mass. The second generation of B-factories, KEKB and SuperKEKB at KEK in Tsukuba, Japan, with the Belle and Belle II experiment and PEP-II at SLAC in Stanford, USA, with the BABAR experiment, are asymmetric colliders. The physics of B mesons is also investigated at hadron colliders, such as the Tevatron, a $p\bar{p}$ collider with $\sqrt{s} = 1.96$ TeV, located at Fermilab near Chicago (USA), with the DØ and CDF experiments. The ATLAS and CMS experiments, along with the b-devoted LHCb experiment, also investigate B mesons as the Large Hadron Collider (LHC), a proton-proton collider with $\sqrt{s} = 14$ TeV, situated at CERN near Geneva, Switzerland.

2.0 KEKB Accelerator

The KEKB accelerator is an asymmetric-energy collider of 8 GeV electrons and 3.5 GeV positrons. It comprises two storage rings: the High Energy Ring (HER) for the electrons and the Low Energy Ring (LER) for the positrons. Due to the beam asymmetry, the $\Upsilon(4S)$ particle is produced with a boost, with respect to the laboratory frame,

$$\beta\gamma = \frac{E_{HER} - E_{LER}}{E_{CM}} = 0.425 \tag{2.1}$$

along the electron beam direction (z-axis), where $E_{HER} = 8.0$ GeV, $E_{LER} = 3.5$ GeV and $E_{CM} = \sqrt{E_{HER} \times E_{LER}} = 10.58$ GeV. Since the B mesons from the $\Upsilon(4S)$ decay have very little kinetic energy in the centre of mass, they have approximately the same boost $\beta\gamma$ in the laboratory. This boost provides the opportunity to perform time-dependent analyses. B mesons generally fly 0.2 mm before decaying and the decay-length difference between the two B mesons along the z-axis (Δz) can be measured with a vertex detector. The B mesons decay-time difference (Δt) can then be estimated as

$$\Delta t \approx \frac{\Delta z}{\beta\gamma c} \tag{2.2}$$

A schematic of the KEKB accelerator complex is given in Figure 2.4. The two rings are approximately three kilometres in circumference, 11 metres below ground level, and built side-by-side in the TRISTRAN tunnel [1]. The electrons are produced from an electron gun and are accelerated through a LINAC to produce the beam. The positron beam is created with the help of a tungsten target in the path of the electron beam and is separated using a magnetic field. Then, these two beams are fed into the storage rings and injected into KEKB in the Fuji area. The LER and HER can store currents up to 2.6 A and 1.1 A, respectively; they use an RF of 508.9 MHz to accelerate the beams [31]. With the aid of dipole, quadrupole, and

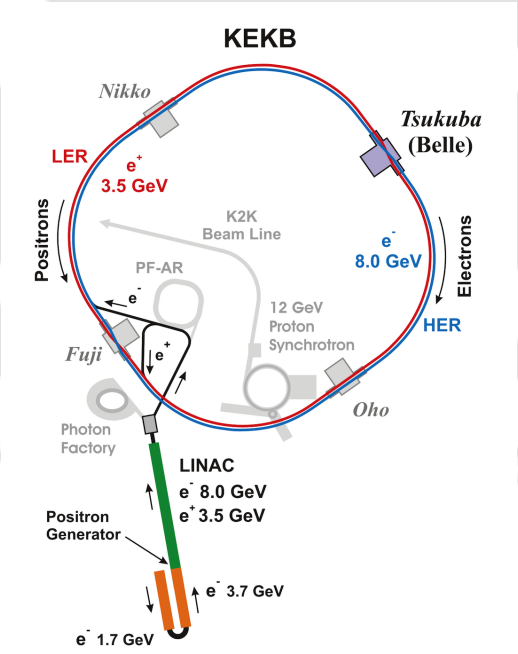


Figure 2.4: Schematic of the KEKB accelerator complex [32].

sextupole magnets, the beams are made to circulate in trajectories that maximise the instantaneous luminosity. In order to compensate for the energy loss of the beams

[1] A symmetric electron-positron particle accelerator at KEK, Japan. This was built in 1981-1986 to search for $t\bar{t}$.

due to radiation as they circulate in KEKB, two kinds of RF cavities are installed: normal conductive cavities, which are accelerator resonantly coupled with energy storage, and superconducting cavities. The two beams cross at the interaction point (IP) in the Tsukuba experimental hall at the center of the Belle detector. At the IP, the two beams collide with a finite horizontal crossing angle of ±11 mrad [33]. This angle eliminates the need for separation-bend magnets, substantially lowering the beam-related background level. Two superconducting crab cavities are employed to kick the bunches into the horizontal plane in order to achieve the maximum luminosity [34, 35]. The crab cavities have a transverse RF that allows the bunches to spin and collide with the most overlap at the IP, as shown in Figure 2.5.

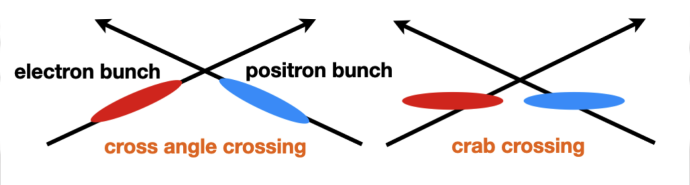


Figure 2.5: A schematic of cross-angle and crab crossings.

Data type	$\sqrt{s}$ **(GeV)**	**Luminosity**(fb^{-1})
Υ(4S) resonance	10.58	711
Υ(4S) off-resonance	10.52	89
Υ(5S) resonance	10.86	121
Υ(5S) scan	$10.58 < E_{CM} < 11.02$	7
Υ(3S) resonance & off-resonance	10.35 & 10.32	2
Υ(2S) resonance & off-resonance	10.02 & 9.99	28
Υ(1S) resonance & off-resonance	9.46 & 9.43	8
Total		966

Table 2.1: Data collected at or near Υ resonances and their corresponding luminosities.

The data samples were collected at or near centre-of-mass (c.m.) energies ($\sqrt{s}$) corresponding to different Υ resonances, as shown in Table 2.1. The data sample used in our analysis corresponds to the sample collected at the Υ(4S) resonance, which is just above the threshold for decay into $B\overline{B}$ meson pairs. The highest instantaneous luminosity achieved by KEKB was 2.11×10^{34} cm^{-2} s^{-1}, which was a world record until 2020 [36]. Figure 2.6 shows the integrated luminosity collected by the Belle and BaBar experiments over various resonances during its lifetime. After completing the physics runs in 2010, the overall integrated luminosity was around 1 ab^{-1}.

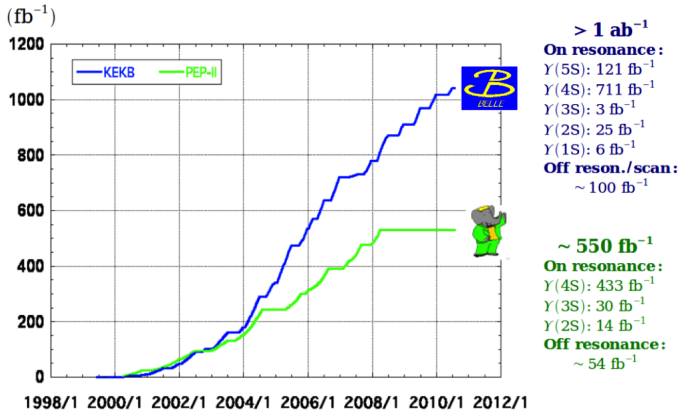


Figure 2.6: Data sets accumulated at the Belle experiment over various resonances during its lifetime. [36]

3.0 Belle Detector

The Belle detector [37, 38] is situated at the interaction region of the KEKB accelerator. As shown in Figures 2.7 and 2.8, it is a layered general-purpose detector consisting of different sub-detector systems dedicated to various aspects of event reconstruction. The innermost sub-detector system is a Silicon Vertex Detector (SVD), which precisely measures the decay vertices of B-mesons. Outside the SVD, we have a Central Drift Chamber (CDC), the main tracking system, and SVD. It also helps to distinguish the various particles using dE/dx measurements. Due to a large number of production of kaons and pions at Belle from the decay of primary particles, they behave identically at higher momentum. We have a particle identification system consisting of CDC, Aerogel Cherenkov counter (ACC) and Time of Flight (TOF) counter. It aims to separate the particle species efficiently. Outside the PID, we have an electromagnetic calorimeter (ECL) to detect electrons and photons produced in the electromagnetic interactions. The whole detector is placed in a magnetic field of 1.5 T produced by a superconducting solenoid. The outermost part is the K_L^0 and muons detection system (KLM).

Belle follows a right-handed coordinate system; the origin is the nominal IP position, the x-axis is horizontally aligned, pointing outward from the centre of the KEKB ring, the y-axis is along the vertical direction (upward) and the z-axis is aligned opposite to the positron beam direction. The polar angle θ is measured relative to the positive z-axis and the azimuthal angle ϕ with respect to the x-axis. Due to the boost of the $\Upsilon(4S)$ resonances, the components of Belle are asymmetric in z. All sub-detector systems of the Belle experiment are discussed in detail in the following subsections.

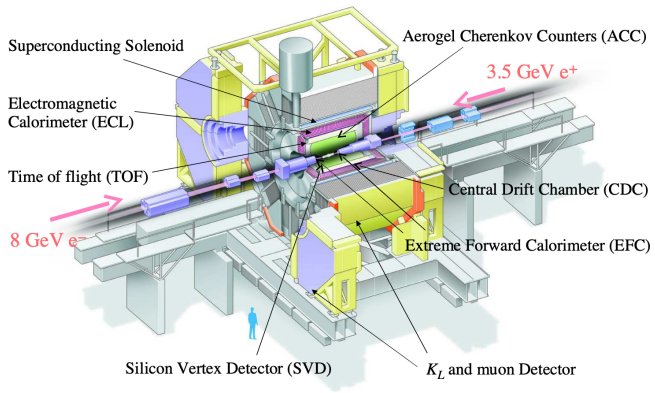


Figure 2.7: Cut away view of the Belle detector

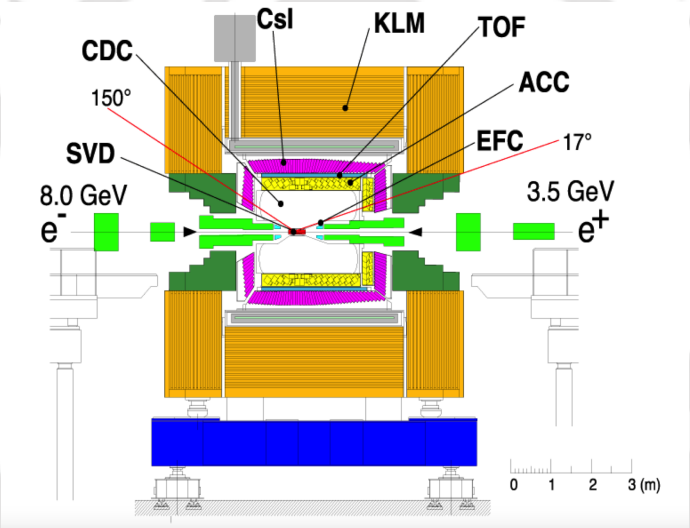


Figure 2.8: Side view of the Belle detector [37]

3.1 Beam Pipe

The beam pipe is the innermost part of the detector and closest to the IP region. The double-walled beryllium (Be) cylinder is used as the beam pipe to avoid coulomb scattering, which affects the resolution of the SVD. The beam pipe is 0.5 mm thick and the separation between both walls is 2.5 mm, through which helium gas is used as a coolant. To absorb X-rays from synchrotron radiation, the beam pipe is enveloped with a gold foil with a thickness of 20 μm. At the interaction region, the cross-section of the beam pipe is shown in Figure 2.9.

3.2 Silicon Vertex Detector (SVD)

The SVD helps in measuring the B and D meson vertices, which is essential in studying the time-dependent CP asymmetry as the averaged separation between the two B meson vertices is around $\sim$ 200 μm. Most of the particles produced in the

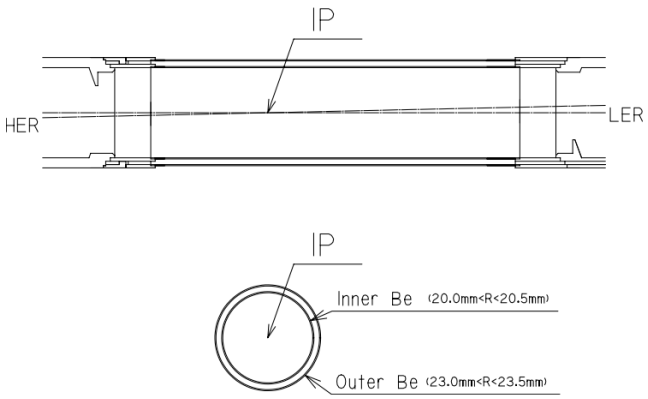


Figure 2.9: The cross-section of the beryllium beam pipe at the interaction point [37].

Belle experiment have momenta less than or equal to 1 GeV/c and multiple-coulomb scattering worsens vertex resolution. To minimize the effect of such scattering, the vertex detector should be placed close to the IP. Also, the electronics must be kept outside the tracking volume and the supporting structure should be low in mass. The SVD is also used to measure the trajectory of charged particles. In the presence of a uniform magnetic field, the shape of the particle's path is helical. Some of the charged particles are probably unable to reach the CDC, and their path becomes curved in the SVD detector. Belle has two types of SVD: a three-layer SVD1, operated till the summer of 2003, and a four-layer SVD2, operated till 2010. As displayed in Figure 2.10, SVD includes the three concentric cylindrical layers having radii of 30.0, 45.5, and 60.5 mm. These layers are made up of different ladders, each comprising a double-sided silicon-strip detector (DSSD). It is supported by boron-nitride ribs. Each DSSD has 1280 sense strips and 640 readout pads on either side. A total of 102 DSSDs are used, where each sensor has a size of 57.5×33.5 mm^2. The ϕ-strip pitch is 25 mm and the z-strip pitch is 42 mm. The readout system of SVD is based on the VA1 integrated circuit [39, 40].

SVD2 was placed in 2003 due to the damage in SVD1 by synchrotron radiation [42].

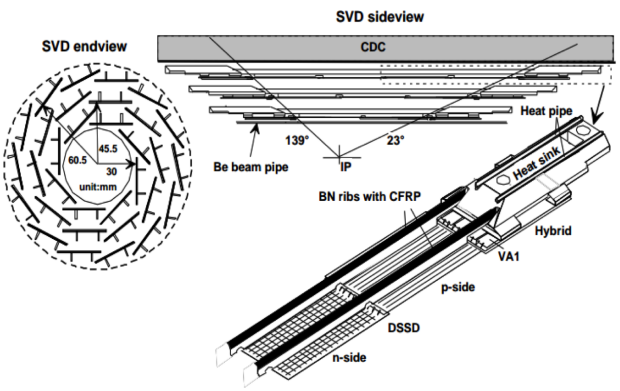


Figure 2.10: View SVD1 detector [41].

As shown in Figure 2.11, SVD2 covers 92% of the 4π solid angle and has four layers. SVD2 has an extra layer of DSSDs, which covers more area than SVD1. Therefore, SVD2 performance is obtained better than SVD1. When reconstructing K_S^0, it's typically done from the paths of its two daughter pions, which have flight lengths under 6 cm. Most K_S^0 particles decay within the SVD region into two oppositely charged pions. The hits from these pions on the SVD layers are used to reconstruct K_S^0, help in tracking low-momentum particles and reduce the multiplicity. The

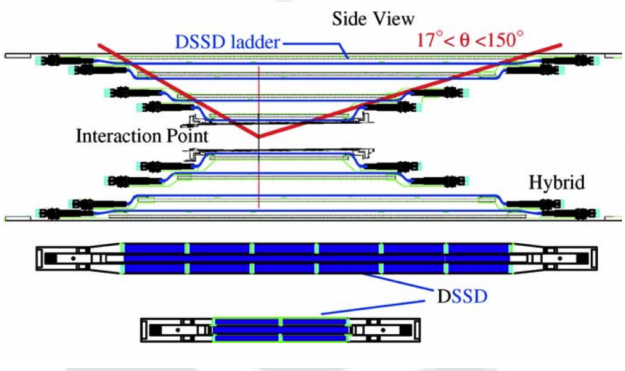


Figure 2.11: Schematic view of the SVD2 subsystem of Belle [42]

performance assessment of the SVDs was conducted using cosmic rays. The primary criterion for evaluating their performance was the resolution of the impact parameter (the closest approach distance of charged tracks to the IP) in the R-ϕ and z planes. Figure 2.12 presents the comparison of performance between SVD1 and SVD2. It was observed that SVD2 exhibited superior performance to SVD1 following the upgrades.

3.3 Central Drift Chamber (CDC)

The CDC, as a gaseous detector [44], serves as both the tracking device and SVD in the Belle experiment. It covers a wide range of polar angles, from 17° to 150°. The inner radius of the CDC is 103.5 mm, while the outer radius extends to 874.0 mm. To minimize multiple scattering and enhance momentum resolution, the CDC is filled with equal parts of helium and C_2H_6. As charged particles traverse the gaseous medium, they interact with gas molecules, causing ionization. Then, a tracking algorithm is used to reconstruct the path of charged particles from their hits in the CDC. Figure 2.13 provides a schematic representation of the CDC, comprising 50 cylindrical layers with a spatial resolution of approximately 130 μm. Within the CDC, there are 8400 drift cells, each containing a sense wire surrounded by eight parallel field wires. It is used to measure the energy loss (dE/dx) of a charged particle using the Bethe Bloch formula [45, 46]. The dE/dx measurement for identification of various charged particle types versus particle's momentum in CDC is shown in Figure 2.14. It provides a clear separation between charged particles like electrons, pions, and protons.

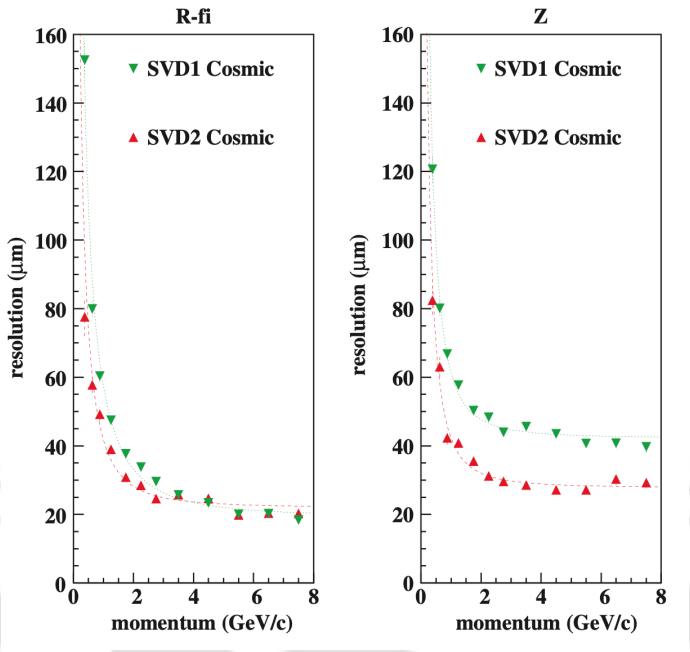


Figure 2.12: Comparision of performances between SVD1 and SVD2 in terms of impact parameter resolutions. [43]

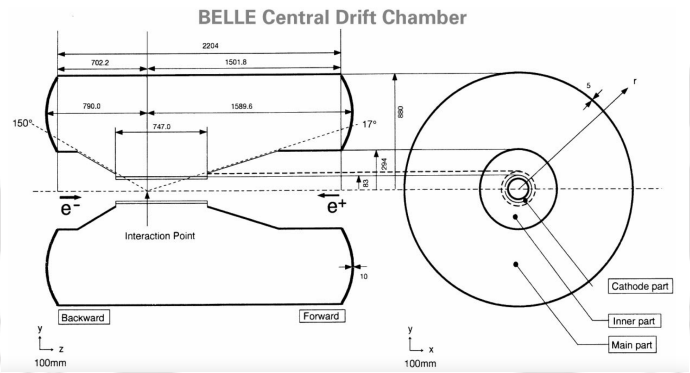


Figure 2.13: Overview of the BELLE CDC structure. The lengths in the figure are in units of mm. [44]

3.4 Aerogel Cherenkov Counter (ACC)

ACC [33] is implemented just after the CDC detector for the identification of particles such as kaons/pions within the momentum region of 1.2-3.5 GeV/c. Particle identification plays a vital role in studying CP violation in B physics. ACC working is based on the principle of Cherenkov radiation. When a particle moves in a medium greater than the velocity of light c in that medium, it emits the Cherenkov radiation. The relation between the refractive index (n) and β (velocity of the particle) is given

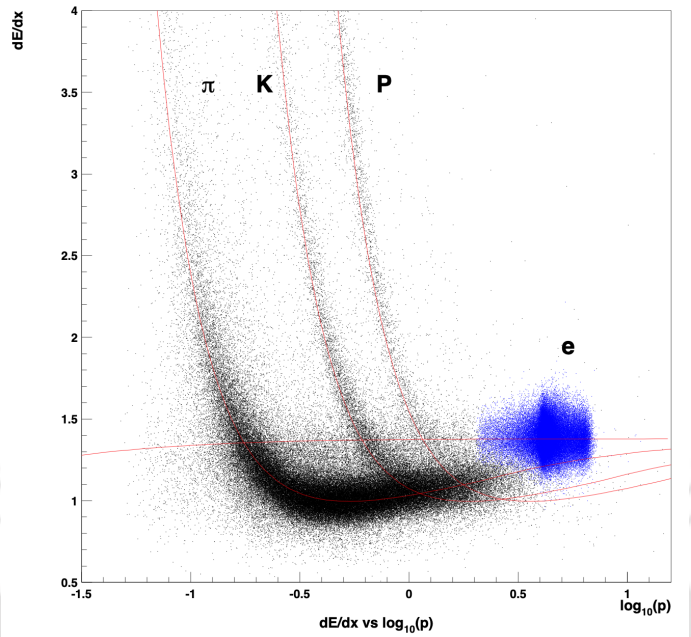


Figure 2.14: Particle identification performance for pions, kaons, protons and electrons based on the dE/dx measurement as a function of the momentum (p in GeV/c).

as:

$$n > \frac{1}{\beta} = \left(1 + \left(\frac{m}{p}\right)^2\right)^{1/2} \tag{2.3}$$

where p represents the momentum, m is the mass of the moving particle and β = v/c. We understood from the above relation that two different particles (kaon and pion) with the same momentum emit the Cherenkov light at different times. Here, n is selected so only pions emit the Cherenkov light, not kaons. Figure 2.15 illustrates the configuration of the ACC detector, consisting of 228 modules arranged in 5 concentric layers in the forward endcap part, and 960 counters segmented into 60 cells in the barrel part. The single ACC cell comprises silica aerogel tiles ($12\times12\times12$ cm^3) boxed in an aluminium box of around 0.2 mm. Photomultiplier tubes are installed within a 1.5 T magnetic field. The refractive index n of the aerogel blocks, selected between 1.01 to 1.028 for improved particle identification, is correlated with the polar angle. The ACC's angular acceptance ranges from 17° to 127° and is positioned in both the forward and barrel endcap regions.

3.5 Time of Flight counter (TOF)

The TOF [47] is the third component of the particle identification system within the Belle experiment, employing plastic scintillator counters. Its angular coverage spans from 34° to 120°, distinguishing between different particles with momenta below 1.2 GeV/c. Only particles with a minimum transverse momentum of 0.28 GeV/c

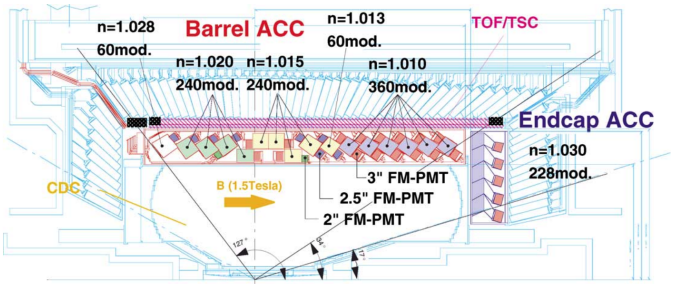


Figure 2.15: The arrangement of ACC at the central part of the Belle detector. [37]

are able to reach the TOF. Offering a rapid signal with a resolution of 100 ps, it comprises 64 trigger scintillator counters (TSC) and 128 TOF counters. A TOF module, depicted in Figure 2.16, comprises two trapezoidal-shaped TOF counters and one TSC counter, spaced 1.5 cm apart. This gap ensures protection for the TOF counters against any electrons or positrons generated in the TSC layer. The system includes 64 such modules. In the TOF, scintillator counters excite molecules, producing photons that are then detected by photomultiplier tubes. The desired time resolution in TOF is achieved with the following design:

- A fast scintillator with an attenuation length greater than 2 m is used.
- The time dispersion of scintillation photos is minimized by eliminating the light guide.
- Photo tubes have a large area of photo cathode to maximize the photon collection efficiency.

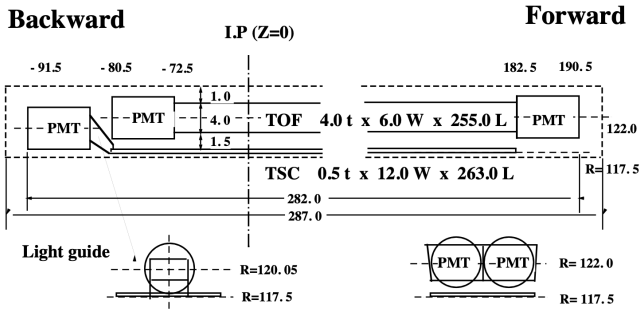


Figure 2.16: TOF detector [37]

TOF detector counts the time of the particle from IP to TOF detector, and the average time is measured around 3 ns. The measured time is expressed as:

$$t = \frac{l}{c\beta} = \frac{l}{c}\left(1 + c^2\left(\frac{m}{p}\right)^2\right)^{1/2}, \tag{2.4}$$

where t provides the information about the particle species, the mass distribution obtained for various particle types is displayed in Figure 2.17. An overall likelihood is performed using the combined measurements of the CDC, ACC, and TOF subsystem, and this information is used to identify the particle type at the reconstruction level.

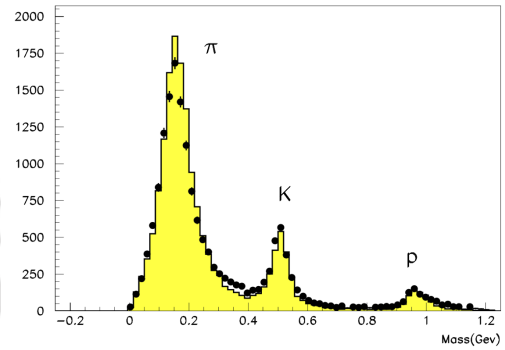


Figure 2.17: Hadron mass distributions calculated from TOF measurements for particles with momentum below 1.2 GeV/c. The black points represent data, and the shaded histogram represents the simulation sample.

3.6 Electromagnetic Calorimeter (ECL)

The Electromagnetic Calorimeter (ECL) is a critical component of Belle's particle detection system. Its primary role is to measure the energy of electromagnetic showers produced when electrons, positrons, and photons interact with the detector material. The ECL is highly regarded for its exceptional efficiency and remarkable energy resolution, making it well-equipped to handle particles across a wide range of energy levels. When an electron or photon strikes a dense absorbing material within the detector, it loses its energy through a process called bremsstrahlung (for electrons) or through the creation of e^+e^- pairs (for photons). The ECL's design is optimized to deliver top-notch performance and high resolution, enabling accurate measurements of both low and high-energy photons. Distinguishing between photons and $e^{\pm}$ involves combining the energy measurements of the showers with information from the tracking detectors. An energy cluster that is not associated with any charged particle track is likely a photon candidate, while a cluster that is linked to a track is more likely related to an electron or positron candidate. Figure 2.18 shows the overall configuration of ECL. The Electromagnetic Calorimeter (ECL) is composed of an array of scintillating crystals made of Thallium-doped Cesium Iodide (CsI(Tl)). These crystals are read out by silicon photodiodes and are installed within a superconducting solenoid magnet that generates a magnetic field of 1.5 Tesla. CsI crystals were chosen for their moderate cost, mechanical stability, and high photon yield. The ECL crystals are divided into three sections: a cylindrical barrel region with a length of 3.0 meters and an inner radius of 1.25 meters, an annular forward section located at $z = 2$ m from the interaction point (IP), and an annular backward

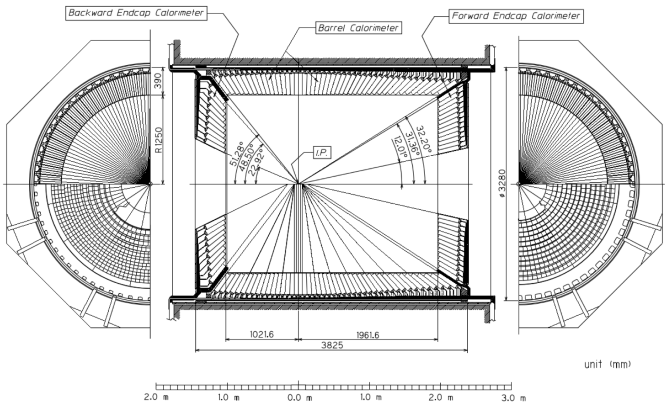


Figure 2.18: Schematic view of the ECL subsystem of Belle. [37]

section at z = -1 m from the IP. Each crystal has a tower-like shape and is arranged to point nearly towards the interaction point. In the barrel section, the crystals are tilted by a small angle of $\sim$ 1.3° in both the θ and ϕ directions to prevent photons from escaping through the gaps between crystals. The end-cap crystals in the forward and backward sections are tilted by angles of $\sim$ 1.5° and $\sim$ 4°, respectively, in the θ direction. The number of crystals and the angular coverage of each section are shown in Figure 2.18. The angular acceptance of ECL is also in the range between $17° < \theta < 150°$, corresponding to a total solid-angle coverage of 91% of 4π. The geometrical parameters of the ECL are provided in Table 2.2. In total, the ECL detector comprises 8,736 crystals and weighs 43 tons.

	θ **coverage (°)**	**No of crystals**	**position (cm)**
Forward endcap	$12.4 < \theta < 31.4$	1152	z = 196
Barrel	$32.2 < \theta < 128.7$	6624	r = 125
Backward endcap	$130.7 < \theta < 155.1$	960	z = 196

Table 2.2: The geometrical configuration of the ECL of Belle.

The ECL was designed with trigger cells to generate trigger logic decisions and capture timing information. Each trigger cell was formed by combining a group of (4×4) ECL cells or clusters. Within each trigger cell, there was a more intricate arrangement called trigger sub-cells, each composed of (4×2) ECL cells. Thus, each trigger cell contained two sub-cells. These trigger cells operated above a predefined energy threshold set by the crystals, allowing them to produce trigger logic decisions and timing details. Additionally, each trigger cell had the capability to provide three sets of time-to-digital converter (TDC) counts, corresponding to both the rising and falling edges of the signal. To distinguish between signals originating from the sub-cells within a trigger cell, the TDC count of one of the sub-cells was offset by 1046 counts, and the widths for the rising and falling edges were adjusted

differently. This deliberate alteration allowed the extraction of timing information from each of the sub-cells. The timing information was crucial in removing off-time (radiative) Bhabha and $e^+e^- \to \gamma\gamma$ events, which are background processes. A detailed description of the implementation of the timing criteria to remove these off-time QED backgrounds can be found in Ref. [48, 49].

The energy resolution of the ECL with a threshold energy of 0.5 MeV was calculated using the formula:

$$\frac{\sigma_E}{E} = \frac{0.0066}{E} \oplus \frac{1.53}{E^{1/4}} \oplus 1.18(\%), \tag{2.5}$$

where the terms in the quadratic sum are contributions from the electronic noise, fluctuations due to shower leakage from the crystals and uncertainty in crystal calibrations. A detailed description of the ECL sub-detector is described elsewhere.

3.7 Extreme Forward Calorimeter

The Extreme Forward Calorimeter (EFC) is a specialized detector component designed to detect and measure electromagnetic showers produced in the extreme forward and backward directions. The polar angle coverage of EFC, $17° < \theta < 150°$, is extended in the forward and backward regions. It covers the angular range from $6.4° < \theta < 11.5°$ in the forward region and $163.3° < \theta < 171.2°$ in the backward region. Figure 2.19 shows the schematic view of the EFC. The region around the beam pipe near the interaction point is subject to extremely high levels of radiation. As the EFC is positioned in close proximity to the interaction point, it must possess the property of radiation hardness. To meet this requirement, the EFC employs Bismuth Germanate $Bi_4Ge_3O_{12}$ (BGO) crystals, which exhibit the necessary radiation resistance. The EFC serves an additional function as a beam mask, reducing background levels for the Central Drift Chamber (CDC). To provide better position resolutions, the EFC is segmented into 5 and 32 sections in the θ and ϕ directions, respectively, for both the forward and backward end-caps. A detailed description of the EFC can be found in section 4 of Ref. [37].

3.8 K_L^0 and Muon detector (KLM)

Neutral kaons (K_L^0) and muons are particles that deposit only small amounts of energy in the subdetectors located closer to the interaction point. These particles have sufficiently long lifetimes, allowing them to decay outside the main detector volume. A specialized subdetector called the KL and Muon detector (KLM) is employed to detect and identify these long-lived particles. The KLM is a dense and massive detector composed of alternating layers of iron plates and charged particle detectors known as glass resistive-plate chambers (RPCs). When K_L^0 and muons interact with the iron plates, they create ionizing particles that can be detected by the RPCs. The KLM consists of a cylindrical barrel section and two endcap sections. The barrel section covers the polar angle range from $45° < \theta < 125°$

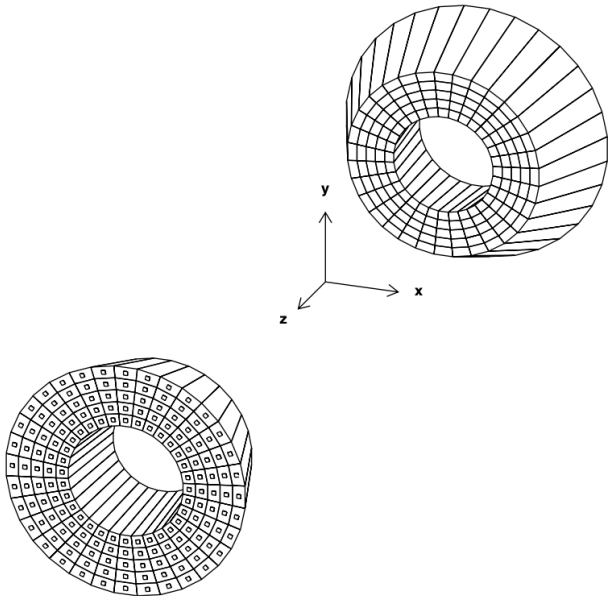


Figure 2.19: A 3D view of EFC sub-detector. [37]

and is constructed with 15 layers of RPCs interleaved with 14 cm thick iron plates. The forward and backward endcaps extend the angular coverage to the range of $20° < \theta < 155°$. Each endcap contains 14 layers of RPCs and 4.7 cm thick iron plates. The KLM layers are organized into "super-layers," as illustrated in Figure 2.20. A super-layer consists of θ and ϕ cathode strips surrounding two RPCs. RPCs are gaseous detectors with two parallel plate electrodes separated by a gas-filled gap. When an ionizing particle passes through the gap, it initiates a gas streamer, resulting in a local discharge. This discharge creates a signal on the external cathode strips, allowing the location and time of the ionization to be recorded.

The iron plates in the KLM provide approximately 3.9 interaction lengths of material. When a (K_L^0) interacts with the iron, it induces a shower of particles. By assuming that the K_L^0 originated from the interaction point, the direction of the K_L^0 can be determined from the observed shower, although no energy information can be extracted. Muons are identified as clusters in the KLM that can be associated with a charged particle track. The muon identification relies on the length and dispersion of the shower induced by the muon. Muons typically travel much farther and experience smaller deflections within the KLM than other charged particles, such as pions and kaons. For muons with momenta above 1.5 GeV/c, the identification efficiency is typically around $\sim$ 90% with a fake rate of $\sim$ 3%. A more detailed description of the KLM detector can be found in Ref. [37].

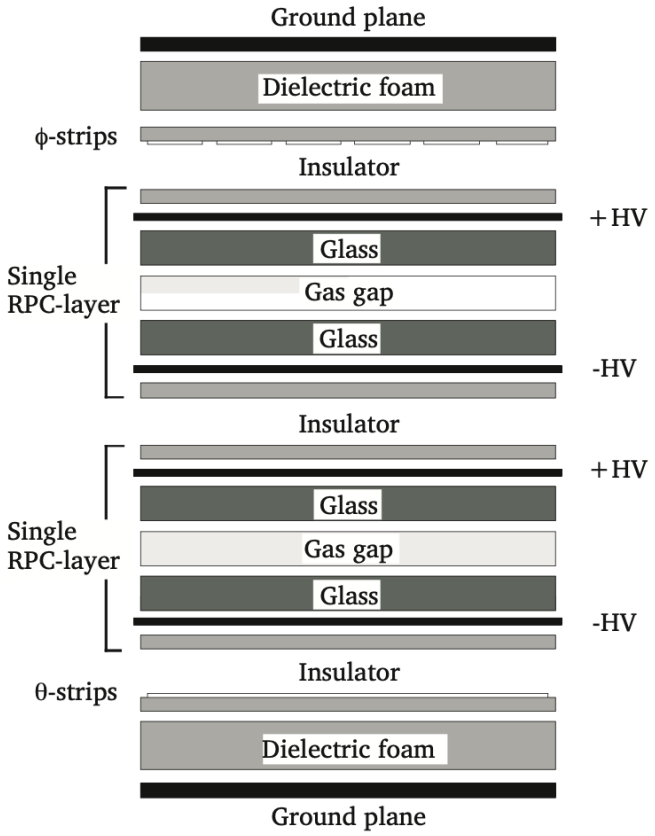


Figure 2.20: Cross-section of a super-layer of the KLM detector. [37]

3.9 Trigger system

The purpose of the trigger system in the Belle detector is to determine when the various subsystems should record and store the data from a particle collision event. When a specific particle interaction meets the predetermined trigger conditions, data from all subsystems is retrieved and saved for later analysis. The term "event" is used interchangeably to refer to either the physical particle collision itself or the recorded data representing that interaction. The decision to read out and store an event is based on carefully designed criteria aimed at eliminating background events while efficiently preserving events of interest. The major sources of background events include undesirable interactions between the electron and positron beams, collisions of a beam particle with residual gas molecules or the beam pipe, and synchrotron radiation from the beams. Events of interest primarily include hadronic events containing B mesons, photon-induced events like Bhabha scattering ($e^+e^- \to e^+e^-$) and μ pairs events ($e^+e^- \to \mu^+\mu^-$), which are used for luminosity monitoring and detector calibration. At the Belle experiment, various physics processes were produced. Table 2.3 provides an example of such processes at the $\Upsilon(4S)$ resonance, including their cross-sections and the expected rates at which the trigger system was designed to operate at an instantaneous luminosity of $\mathcal{L}_{inst} = 10^{34}$ cm^{-2}s^{-1}.

Table 2.3: Total cross-section and trigger rates with $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} from various physics processes at $\Upsilon(4S)$.

Physics Process	**Cross-section (nb)**	**Rate (Hz)**
$\Upsilon(4S) \to B\overline{B}$	1.2	12
Hadron Production from continuum	2.8	28
$\mu^+\mu^- + \tau^+\tau^-$	1.6	16
Bhabha ($\theta_{lab} \geq 17°$)	44	4.4*
$\gamma\gamma$ ($\theta_{lab} \geq 17°$)	2.4	0.24*
2γ process ($\theta_{lab} \geq 17°$, $p_t \geq$0.1 GeV/c)	$\sim$ 15	$\sim$ 17†
Total	$\sim$ 67	$\sim$ 96

*Indicates the values pre-scaled by a factor 1/100
†Indicates the restricted condition of $p_t \geq 0.3$ GeV/c

At the design luminosity of 10^{34} cm^{-2}s^{-1}, both physics events of interest and background events occur at a rate of roughly 100 Hz each. However, to accommodate the possibility of higher luminosities and increased beam backgrounds, the trigger system is designed to operate at a rate of 500 Hz.
The Belle trigger data flow is depicted in Figure 2.21. It consists of three main levels: a Level-1 (L1) hardware trigger, a Level-3 (L3) software trigger implemented by an online computer farm, and a Level-4 (L4) trigger that operates on the Belle computing system offline and performs more complex background reduction based on full event reconstruction.

The L1 trigger comprises sub-trigger systems and a central trigger system called the global decision logic (GDL). All sub-trigger signals are designed to arrive at the GDL within 1.85 μs of the event occurring. The final L1 trigger signal is issued 2.2 μs after the event crossing. The L1 trigger receives track information from the CDC, ACC, and TOF subsystems. The energy trigger is obtained from the ECL based on the energy deposition in the CsI(Tl) crystals. These triggers are evaluated in parallel before being sent to the GDL. The timing of the trigger signals must be accurate since the trigger signal determines the readout timing. The timing of the final trigger is determined by the TOF trigger or, in its absence, by the ECL trigger. For hadronic events, the L1 trigger logic has a 98% efficiency; a schematic of the L1 trigger is shown in Figure 2.22.

The L3 trigger aims to minimize the number of events to be stored. It first verifies the L1 trigger information but passes certain categories of events, such as Bhabha events and random trigger events. If an event does not belong to these categories, the L3 trigger performs a quick reconstruction and rejects the event if it has no track with $|z| < 5$ cm from the IP. This procedure eliminates a substantial fraction of the beam background events. The L4 trigger filters events just before the full event reconstruction takes place. The basic strategy of this trigger is to reject tracks that originate away from the IP. Events rejected by the L4 trigger remain in the raw

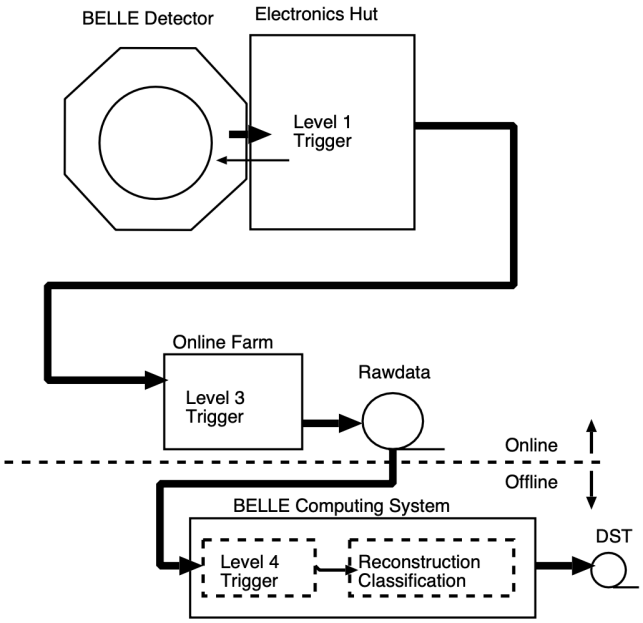


Figure 2.21: An overview of the Belle trigger system. [37]

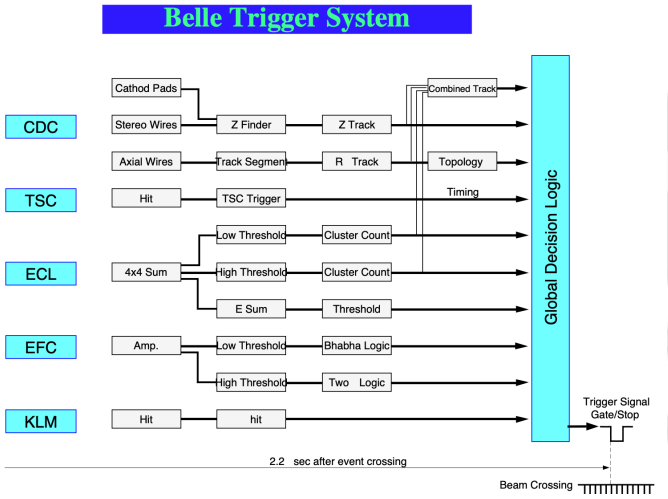


Figure 2.22: Belle level-1 (L1) trigger schematic. [50]

data. Thus, the L3 trigger (together with the L1 trigger) reduces the data size to be recorded, while the L4 trigger reduces only the CPU time for data summary tape (DST) production. The L4 trigger rejects about 78% of triggered events while keeping nearly 100% of B meson events.

3.10 Data Acquisition system (DAQ)

The DAQ system is configured to handle data at a rate of up to 500 Hz. The typical data size for a hadronic event is around 30 kB, and the maximum data transfer rate is 15 MB/s. Figure 2.23 provides an overview of the Data Acquisition system. It reads data from all subdetectors.

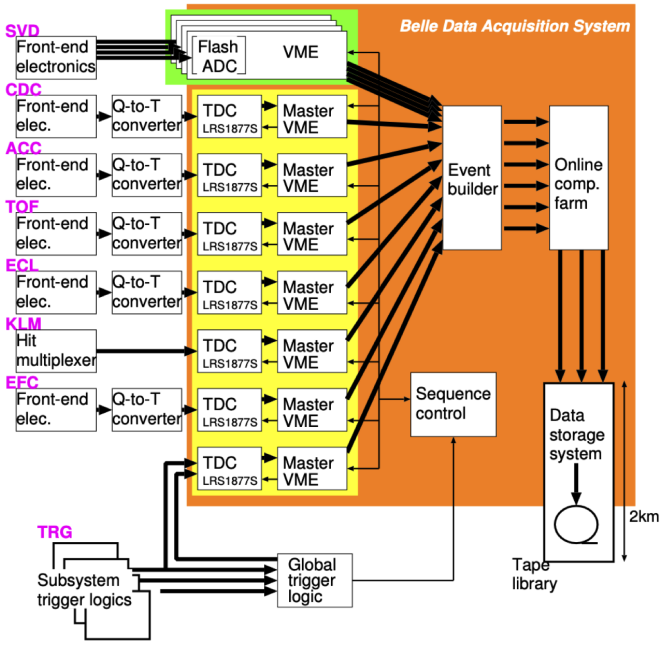


Figure 2.23: A schematic of the Belle DAQ system. [37]

In most subdetectors, the signals are directly related to the energy deposited in the individual detector components. These analogue signals are converted into digital form using time-to-digital converters (TDCs). The SVD is an exception, as it uses flash analog-to-digital converters (FADCs). When the GDL receives a trigger signal, the event builder combines the data from each subdetector into a single event, converting the "detector-by-detector" parallel data streams into "event-by-event" data. The output data from the event builder is then passed to the online computer farm, which comes from the L3 trigger system. Finally, the information is transferred to the mass storage system at the KEK computing facility, where it is saved on tape for further analysis.

4.0 SuperKEKB accelerator

The new generation of B-Factory is represented by the SuperKEKB accelerator, which is an upgrade of the KEKB. The improvement program began in 2010 with the major objective of increasing the machine's instantaneous luminosity from 2.1×10^{34} $\text{cm}^{-2}\text{s}^{-1}$ (achieved by KEKB) to 8×10^{35} $\text{cm}^{-2}\text{s}^{-1}$. This luminosity enhancement aims to meet the physics goals of the Belle II experiment. The luminosity increase

is achieved by employing a higher beam current and a smaller beam dimension at the IP, while keeping the beam emittance as low as possible, using the nano-beam scheme [51]. SuperKEKB is an e^+e^- asymmetric collider, with an energy of 7 GeV (4 GeV) for the electron (positron) beam. The resulting centre-of-mass energy in the relativistic limit is $\sqrt{s} = \sqrt{4Ee^+Ee^-} = 10.58$ GeV. The beam asymmetry generates a Lorentz boost of $\beta\gamma = 0.28$ for the B mesons in the laboratory frame, resulting in their average flight distance of 130 μm. The electron and positron beams are produced in the same manner as in KEKB.
To achieve the factor of 40 increase in luminosity, the nano-beam scheme, further elaborated in [[52], sec. 2], is employed. In this scheme, the colliding beams are focused by additional final focusing quadrupoles near the IP, obtaining a beam size of $\sigma_y^* = 0.08$ μm. Compared to KEKB, this reduces the vertical beta function at the IP $\beta_{y\pm}^*$ for both beams by a factor of almost 20. Table 2.4 provides an overview of the important machine parameters of SuperKEKB, with a comparison to the parameters of KEKB. Figure 2.24 is a representation of the beam collisions, where d represents the size of the region of overlap. Reducing the value of d, which represents the size of the overlap region between the colliding electron and positron beams, allows for the squeezing of the vertical beta functions (β_y^*) of each beam at the interaction point. This squeezing of the vertical beta functions has a significant impact on increasing the resultant instantaneous luminosity $\mathcal{L}$. The effect of this reduction in the vertical beta functions can be estimated using equation 6.1, where the assumption of flat beams and equal horizontal and vertical beam sizes for both the e^+ and the e^- beam at the IP is made [[52], sec. 2].

$$\mathcal{L} = \frac{\gamma_\pm}{2er_e}\left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)\frac{I_\pm \xi_{y\pm}}{\beta_{y\pm}^*}.\frac{R_L}{R_{\xi_y}}, \tag{2.6}$$

Table 2.4: Comparison of achieved KEKB machine parameters and design values of SuperKEKB [[52], sec. 2].

Machine Parameter	**KEKB**		**SuperKEKB**	
	HER	LER	HER	LER
Beam current (A)	1.64	1.19	3.60	2.61
Energy (GeV)	8.0	3.5	7.0	4.0
$\beta_y^*(mm)$	5.9	5.9	0.27	0.47
Crossing angle (mrad)	22		83	
Beam emittance (μm)	300	1200	20	10
Beam lifetime (min)	200	150	10	10
Luminosity (10^{34} cm^{-2}s^{-1})	2.11		80	

The luminosity is indirectly proportional to the vertical beta functions, so reducing these functions leads to an enhancement in luminosity. Additionally, the luminosity depends on several factors, including the relativistic Lorentz factor $\gamma_\pm$ of

the beam, the absolute value of the electron charge (e), the classical radius of the electron (r), the widths of the beam bunch at the IP in the transverse plane σ_x^* and σ_y^*, the beam current (I), the vertical betatron function at the IP β_y^*, the vertical beam-beam parameter ξ_y, and reduction factors due to the non-vanishing crossing angle R_L and R_{ξ_y}. Moreover, an increase in the beam currents by a factor of about two results in an overall 40 times higher instantaneous luminosity than KEKB.

To accomplish these ambitious design goals, extensive hardware modifications were

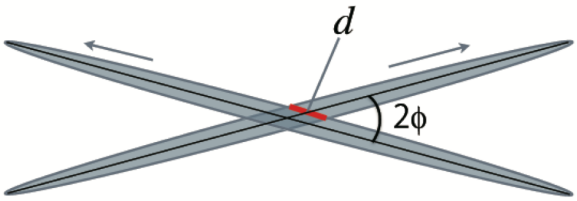


Figure 2.24: A depiction of the interaction between an electron and positron beam at SuperKEKB. The grey areas represent the bunches of colliding particles, with the size of the beam overlap d and the crossing angle ϕ minimised with the Nano-Beam design [52].

required, including a redesign of the IP region. One crucial change was increasing the crossing angle to 83 milliradians to accommodate the final focusing quadrupoles, which are vital in achieving the desired beam characteristics. To generate the low-emittance electrons and positrons necessary for the nano-beam scheme, the injector complex underwent significant upgrades. A new radio frequency gun was installed for the production of low-emittance electrons, and a new tungsten positron target was implemented. However, the positrons produced from e^- collisions with this target initially had high emittance. To dramatically reduce their emittance compared to the previous KEKB accelerator (see Table 2.4), the addition of a positron damping ring, shown in Figure 2.25, was essential.

One side effect of the small vertical beta function at the IP and the higher beam current is an increased rate of intra-bunch scattering, known as the Touschek effect [54] This effect leads to a lower beam lifetime. Furthermore, the amount of beam-induced background significantly increases due to the Touschek effect and other sources, such as beam-gas scattering and synchrotron radiation. To mitigate the impact of the Touschek effect, its proportionality to the inverse cube of the beam energy E_{Beam}^{-3} was exploited, and the energies of the LER and HER were adjusted to 4.0 and 7.0 GeV, respectively. The Belle II detector also undergoes significant upgrades to manage the anticipated rise in beam background and to accommodate technological advancements. However, certain components of the original detector remain largely unchanged. For instance, the superconducting solenoid and its iron return yoke are retained. Only minor enhancements are made to the KLM instrumentation on the yoke, such as replacing resistive plate counters with plastic scintillator strips in the endcaps and the inner two layers of the barrel, to counteract increased background rates in these areas [[52], sec. 10][55]. Similarly, most of

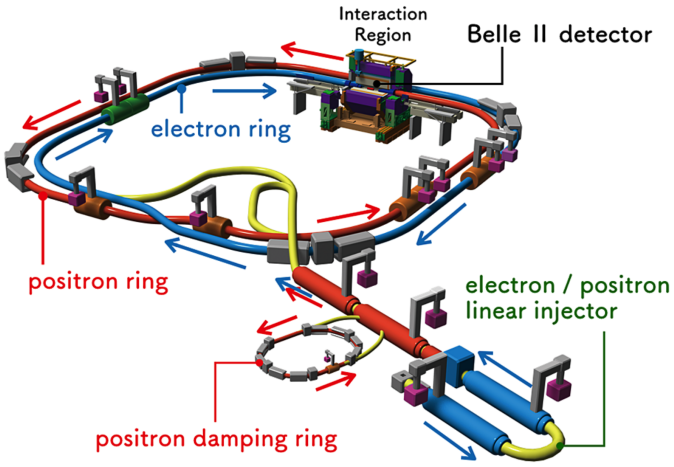


Figure 2.25: Sketch of the SuperKEKB collider with its high energy electron ring (HER, blue) and low energy positron ring (LER, red), as well as the added positron damping ring and the Belle II detector [53].

the ECL crystals are reused, with the possibility of replacing crystals in the endcap regions with pure CsI crystals offering a shorter scintillation time [56]. This change is particularly important for regions experiencing higher beam background. Additionally, the ECL's electronics will be upgraded to a faster readout with waveform sampling, further enhancing the performance of the entire ECL system under the new conditions.
Figure 2.26 presents a comparison between the Belle and Belle II detectors, highlighting significant innovations in colour. Alongside the new beryllium beam pipe with a 12 mm radius, the quadrupoles, and the potential ECL upgrade mentioned earlier, the tracking detectors and PIDs see notable changes due to the upgrade.
Figure 2.27 shows an overview of the Belle II detector, along with its coordinate system. The z-axis roughly aligns with the electron beam's direction, the y-axis points upward, and the x-axis extends radially from the centre of the storage ring. At the IP, where particle collisions occur, is the origin. The xy-plane, known as the transverse plane, plays a crucial role in particle detection. Spherical coordinates measure the azimuthal angle ϕ from the x-axis and the polar angle θ from the z-axis. Like its predecessor Belle, Belle II maintains symmetry around the z-axis and features an asymmetric acceptance range, covering polar angles from 17° to 150° to accommodate for the particle boost. The sub-detectors are discussed slightly more thoroughly in the following subsections.

4.1 Vertex Detector (VXD)

The Vertex Detectors (VXD), comprising the two-layered PiXel Detector (PXD) and the four-layered Silicon Vertex Detector (SVD), play a crucial role in measuring vertices and tracking charged particles with low momenta. They replace the SVD2 and are positioned closest to the IP – the innermost PXD sensor layer, with a ra-

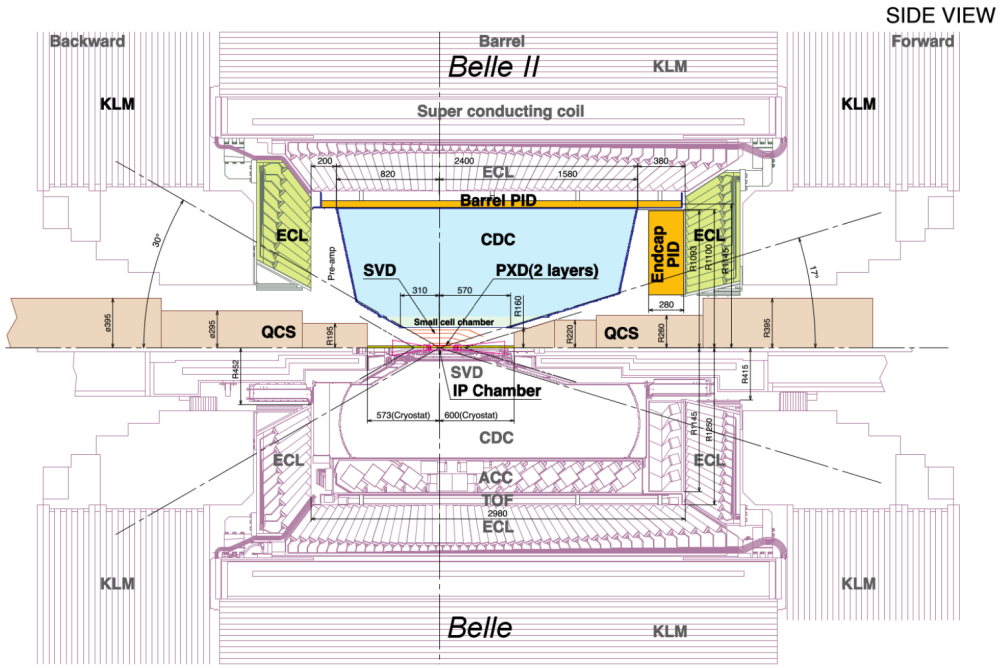


Figure 2.26: Comparison of Belle (lower half) and Belle II (upper half); the most significant changes are marked in colour [52].

dius of 14 mm, sits right on the beam line. With the second layer having a radius of 22 mm, the PXD boasts over eight million depleted field effect transistor pixels (DEPFET), ensuring a low occupancy of the detector despite the higher background rate. Furthermore, the DEPFET technology enables a thin design of the sensor elements, reducing multiple scattering in these detector regions. Figure 2.28 illustrates the structure of the Belle II VXD.

The PiXel Detector (PXD) is accompanied by the Silicon Vertex Detector (SVD), which surrounds it with radii ranging from 38 to 140 mm. To ensure coverage within the same acceptance range of 17° to 150° in θ as the rest of the PXD, the three outermost layers of the SVD are tilted towards the beamline in the forward region, with an angle of approximately 15 degrees [[52], p. 142]. This tilt reduces the amount of material needed for the double-sided-readout silicon strip detector elements. The readout strips on both sides of the elements are arranged nearly perpendicular to each other, providing complete positional information about the hits made by charged tracks. These elements also have a faster readout time than the PXD, making them less susceptible to beam background. The SVD serves as a link between the PXD and CDC, helping extrapolate reconstructed tracks to their respective vertex. Moreover, it provides only the tracking information for charged particles that don't reach the CDC due to low momentum and assists the CDC in other scenarios.

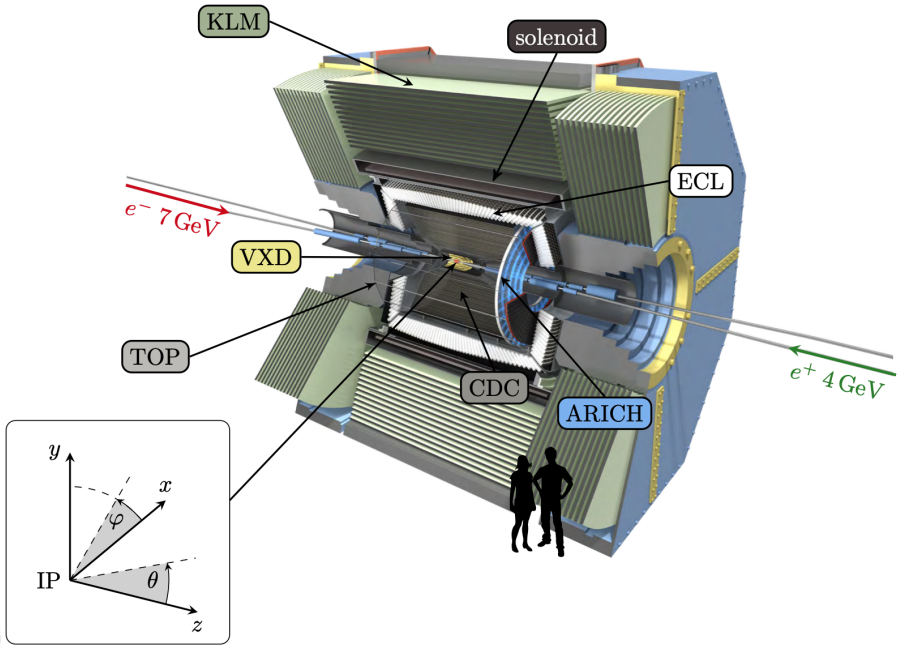


Figure 2.27: Overview of the Belle II detector with the vertex detector (VXD), the central drift chamber (CDC), the time of propagation counter (TOP), the Aerogel RICH detector (ARICH), the electromagnetic calorimeter (ECL), the solenoid, the K_L^0 and μ detector (KLM) and a schematic of the Belle II coordinate system. The origin corresponds to the interaction point (IP) [52].

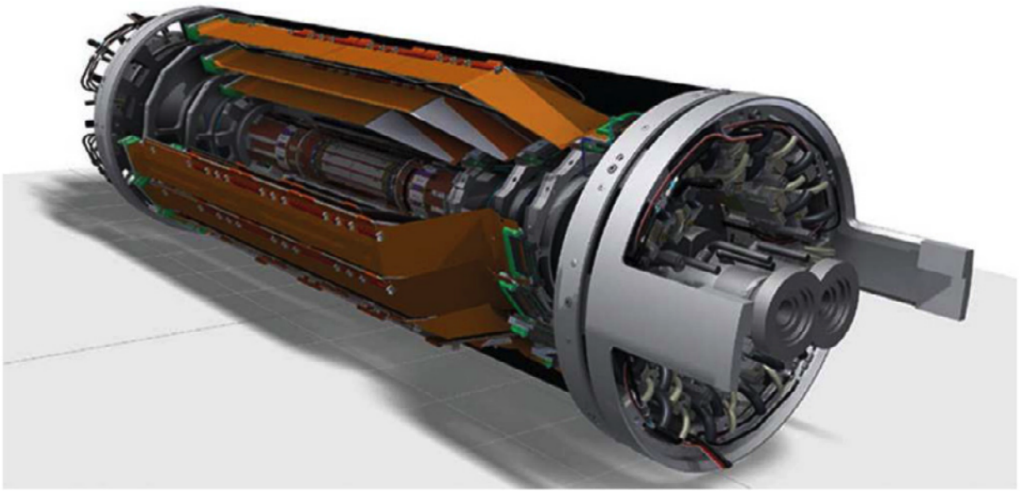


Figure 2.28: Belle II Vertex Detector: the beam pipe, PXD, SVD, and the shield material are assembled in a single structure [52].

4.2 Central Drift Chamber (CDC)

The CDC serves as the main tracking tool in the Belle II detector. The CDC reconstructs the paths of charged particles to determine their momentums precisely,

and it relies on the energy they lose within their volume to identify particles in the low-momentum range. It also helps identify particles with low momentum by measuring how much energy they lose as they pass through the CDC. Additionally, it gives us reliable signals to trigger when charged particles are detected. The CDC is built on the same design as its predecessor, using the same materials and gas combination. The most significant changes concern:

- Upgrading the readout electronics to handle increased trigger rates;
- Adjusting the innermost cylinder radii to accommodate the larger SVD and modifying the outermost radius for the compact barrel detector required for PID. This results in radii of 160 mm and 1130 mm, respectively, with corresponding adjustments to the wiring arrangement. The polar angular coverage spans from 17° to 150°;

The CDC consists of 56 layers grouped into nine super-layers. The cells within these layers are square-shaped, varying in length from 10 mm for the innermost super-layer to 18.2 mm for the outermost one. Inside, there are a total of 14,336 tungsten sensing wires, each with a diameter of 30 μm, immersed in a gas mixture of 50% helium and 50% ethane. To create the necessary electric field gradient, aluminium wires with a diameter of 126 μm are used. The CDC achieves a spatial resolution of around 100 μm, with a relative accuracy of approximately 12% on dE/dx measurements for particles approaching a 90° angle.

4.3 Particle Identification system

Belle II features two particle identification systems: the TOP counter in the barrel region and the ARICH counter in the endcap. While both subsystems rely on the Cherenkov effect to identify charged particles passing through the spectrometer, their operational principles differ. The Ring-Imaging Cherenkov detectors determine particle velocities by measuring the angle θ_C of photons created when charged particles travel through a radiator material at high speed. This velocity (β) is then calculated using the equation $\cos\theta_C = 1/n\beta$, where n is the refractive index of the material. To find out the mass of the particle, data from the PID detector is combined with the momentum measured by the tracking devices and the energy loss data measured by the CDC.

The TOP counter estimates how long it takes for Cherenkov photons to travel through a quartz bar and be reflected internally, as illustrated in Figure 2.29. By combining the arrival time of these photons with the position they arrive at, along with the x coordinate of the quartz bar (as shown in Figure 2.29), the TOP counter can figure out the Cherenkov angle θ_C. This angle helps determine the speed of the particle and allows for the calculation of probabilities for different mass possibilities. While the main goal of the TOP detector is to improve the separation between K and π particles, it also provides useful information about other charged particles. The TOP counter accepts angles within the range of $31° < \theta < 128°$.

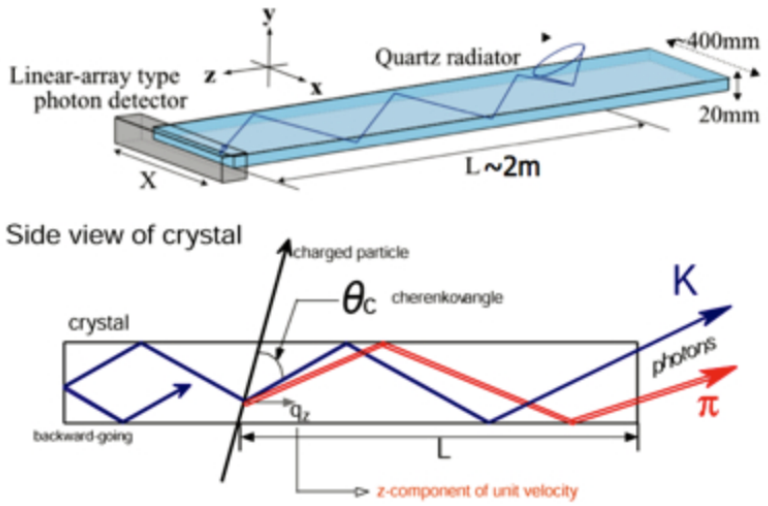


Figure 2.29: Conceptual representation of the time-of-propagation counter, which measures both the arriving time and position of the internally reflected Cherenkov photons [57].

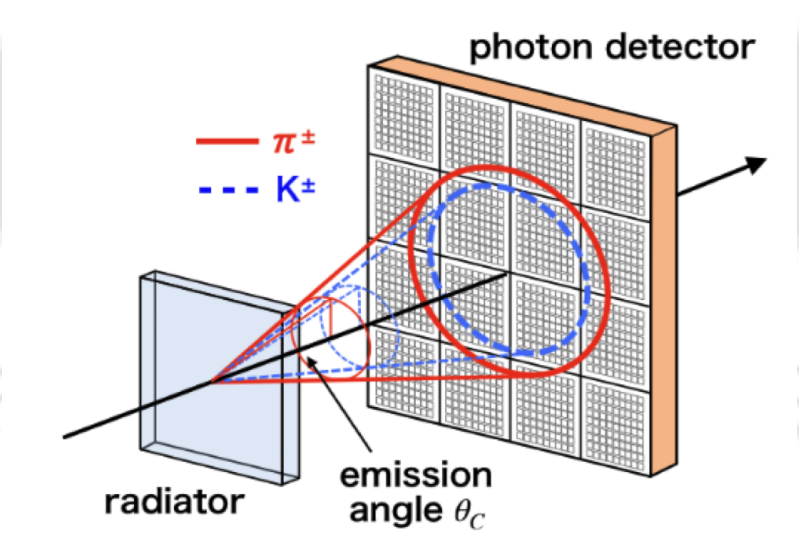


Figure 2.30: The principle of K/π identification for the ARICH counter. The solid line and dotted line cones illustrate the emitted Cherenkov light for a pion and a kaon, respectively.

In the forward endcap of the detector, the ARICH plays a crucial role in providing information for PID. Figure 2.30 illustrates how ARICH operates. Unlike the TOP detector, which utilizes a quartz radiator, ARICH employs an aerogel radiator. ARICH functions as a more traditional Cherenkov detector, utilizing ring-imaging of the Cherenkov cone to identify particles. In contrast, the TOP detector operates based on time-of-propagation (although this time-of-propagation is related to the

Cherenkov angle). The reason for the different PID systems in different regions of the detector is primarily due to geometric limitations. Furthermore, there isn't a dedicated PID system in the backward endcap. This absence is mainly because decays in Belle II are forward-boosted due to the asymmetric beam energies. Consequently, the lack of a dedicated PID system in the backward endcap is expected to minimise the detector's capabilities. Overall, only about 10% of the CDC angular coverage falls outside the coverage provided by TOP and ARICH. The ARICH has an angular acceptance range of $14° < \theta < 30°$.

4.4 Electromagnetic Calorimeter (ECL)

In Belle II, the ECL plays several crucial roles. Firstly, it efficiently detects photons and provides information about their energy and angle across a wide range, from 20 MeV to 4 GeV. Secondly, it identifies electrons and aids in the detection of K_L^0 particles. Thirdly, it generates a photon trigger both in hardware and software. Additionally, the ECL monitors the luminosity of SuperKEKB, both in real-time and offline. While the Belle II ECL utilizes the same crystals and configuration as Belle's calorimeter, the readout electronics had to be entirely upgraded to handle the increased luminosity of SuperKEKB. The geometrical setup remains unchanged from Belle's configuration, as detailed in Table 2.2. The energy resolution achieved with this calorimeter at Belle is discussed in Section 3.6.

A similar performance is also expected at Belle II in the absence of background. However, the increased background levels are expected to result in overlapping pulses from neighbouring background events due to the prolonged decay time of CsI(Tl) scintillations. To address this issue of pile-up noise, photodetectors are fitted with waveform-sampling readout electronics. These electronics store ADC samples in Field Programmable Gate Array (FPGA) internal buffers and facilitate waveform fitting to differentiate signals from hits that occur off-timing.

For electron identification, the ECL was utilized to create a likelihood ratio. t helped align the position of the track with the energy cluster, calculated the E/p ratio[2] and the information on the transverse shower shapes. The variable used for electron identification is defined as

$$\mathrm{eID} = \frac{\Pi_i \mathcal{L}_i^{\mathrm{e}}}{\Pi_i \mathcal{L}_i^{\mathrm{e}} + \Pi_i \mathcal{L}_i^{\pi}}$$

where all the discriminant variables mentioned earlier contribute to the products. The most recent performance measurements, with an eID greater than 0.9 selection in Belle II data, are depicted in Figure 2.31. These measurements are computed using many different channels, as listed in the legend. The efficiency and fake rate closely resemble those observed in Belle.

[2]An electron is expected to release all its energy in the calorimeter, yielding $\mathrm{E/p} \sim 1$.

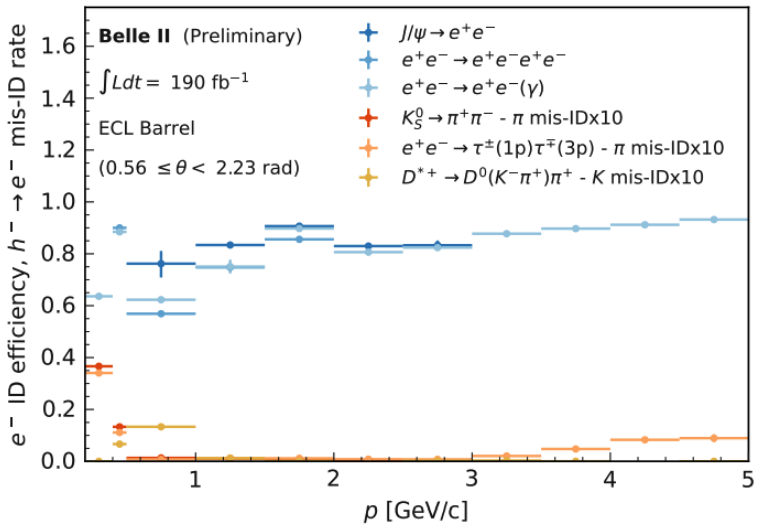


Figure 2.31: Electron identification and hadron-electron misidentification rates as a function of momentum in the ECL barrel region, with eID > 0.9 selection. The misidentification rate is multiplied by a factor of 10 for illustration purposes. Taken from Ref. [58].

4.5 K_L^0-muon detector (KLM)

Electrons, unlike muons, release all their energy in the calorimeter, while muons tend to travel farther without undergoing bremsstrahlung radiation. To capture both muons and K_L mesons, Belle II experiments incorporate a specialized detector in the outermost layer beyond the solenoid. Muons must possess a minimum momentum to reach this detector, with identification efficiency notably high above 600 MeV/c. The detector consists of two primary regions: a barrel-shaped section covering an angular range from 45° to 125°, and end-caps extending from 20° to 155°. Charged particles pass through the KLM in a straight trajectory due to the lack of a magnetic field, either escaping (with p $>$ 1.5 GeV/c) or losing energy. Comprising 4.7 cm thick iron plates and active detectors, it resides outside the superconducting solenoid and aims to determine the momentum of long-lived particles. The iron serves as the magnetic flux return for the solenoid and allows K_L^0 particles to shower within it, owing to a material budget of 3.9 interaction lengths. In regions with anticipated higher neutron backgrounds, such as the end-caps and innermost layers of the barrel region, RPCs have been replaced by layers of scintillator strips with wavelength-shifting fibres. This upgrade is tolerant to higher rates and addresses the prolonged dead time of the RPCs due to electric field recovery after discharge, which impacts detection efficiency at high rates. For muon identification, CDC tracks are extrapolated to the KLM and compared against reconstructed hits using ΔR (difference between measured and expected track range) and χ_r^2 statistics (constructed from transverse deviations of associated hits normalized by the number of hits). Likelihoods for muon, pion, and kaon hypotheses are established based on probability density functions (PDFs)

in ΔR and χ^2_r. The normalized ratio:

$$\mu\text{ID} = \frac{\mathcal{L}_\mu}{\mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_\text{K}}$$

was then used as a discriminating variable. To identify muons with momentum below 600 MeV/c, which do not reach the KLM, machine learning methods are trained using data from the inner detectors. This means that muon identification efficiency remains nonzero even at low momenta but exhibits significantly lower purity, with a pion fake rate of approximately 7%, as depicted in Figure 2.32.

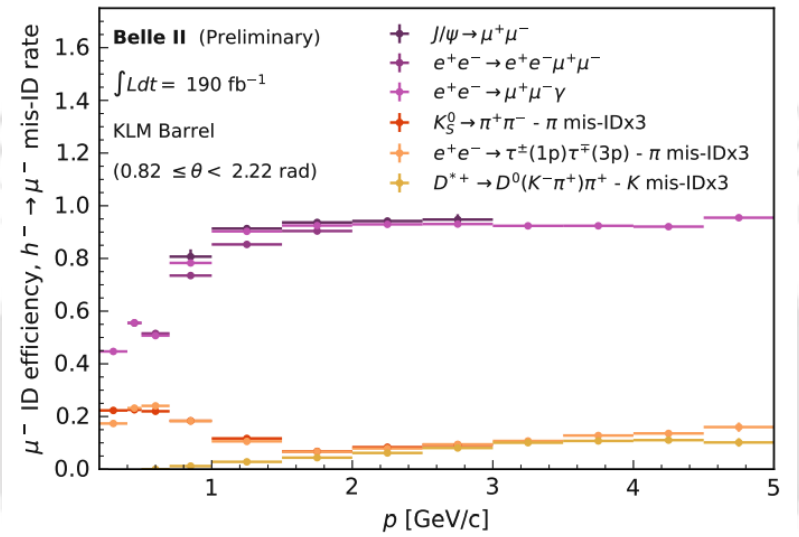


Figure 2.32: Muon identification and hadron-muon misidentification rates as a function of momentum, with $\mu\text{ID} > 0.9$ selection. The misidentification rate is multiplied by a factor of 3 for illustration purposes. Taken from Ref. [58].

4.6 Trigger System

The Belle II trigger system consists of a hardware-based Level 1 trigger (L1) [59] and a software-based High-Level Trigger (HLT) [60]. Its main goal is to ensure efficient capture of hadronic events originating from $\Upsilon(4S) \to B\bar{B}$ decays and continuum processes. With the higher instantaneous luminosity in Belle II compared to Belle, we expect more backgrounds. Thus, an effective and advanced trigger system is crucial to mitigate the high L1 trigger rate. The Belle II trigger system follows the Belle triggering scheme [37] but incorporates new technologies. Outdated components have been replaced with newer ones, additional trigger lines have been introduced, and each component now features a FPGA, allowing for configurable trigger logic rather than being hard-wired. Sub-triggers from various sub-detectors provide input to the global decision logic, which makes the final determination. The HLT trigger primarily filters out physically uninteresting decays online to primarily

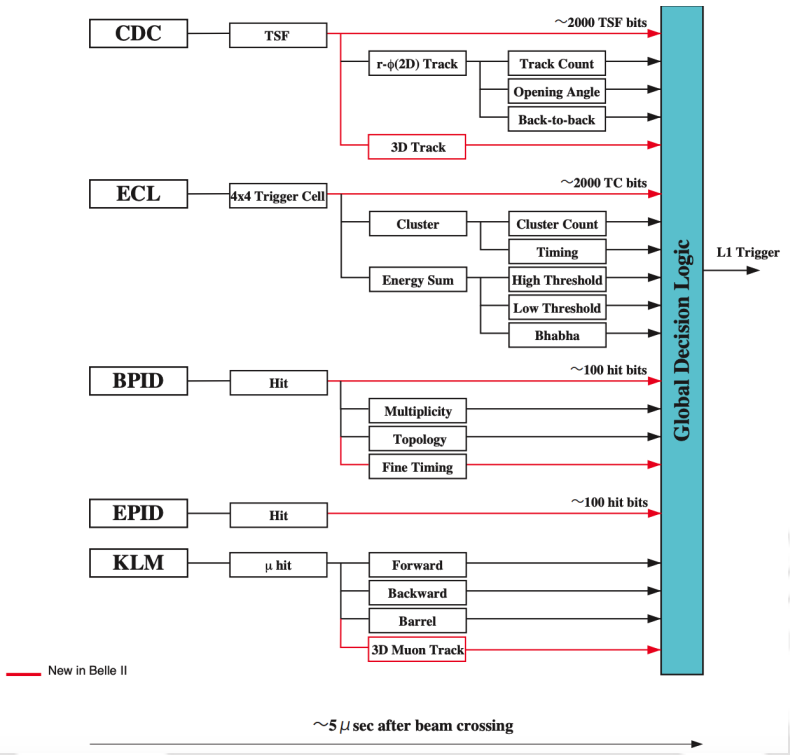


Figure 2.33: Schematic overview of the L1 trigger system. The output from the sub-trigger systems goes to the Global Decision Logic (GDL). The final trigger decision is made in the GDL. The red lines are newly added information paths relative to Belle. Taken from [52].

address storage concerns. Figure 2.33 illustrates the schematic overview of the Belle II trigger system. The CDC sub-trigger utilizes the Track Segment Finding method to provide detailed information about 2D and 3D charged tracks. ECL-based triggers are designed for neutral and charged track-oriented physics events, selecting events based on energy-deposited information from total or isolated ECL clusters. The Barrel PID (BPID) and Endcap PID (EPID) sub-triggers provide precise timing and hit topology information. The KLM sub-trigger supplies data concerning muon tracks. The Global Decision Logic (GDL) gather information from all sub-triggers and makes the final decision, ensuring the correct trigger signal is sent.

5.0 Belle II Program

The commissioning and operation of Belle II and SuperKEKB were divided into three phases.

- In 2016, Phase 1 was undertaken to commission SuperKEKB. This involved conducting tests on the accelerator ring, performing vacuum scrubbing, and conducting hardware checks. However, no collisions were generated during

this phase. Instead of installing the Belle II detector, a system of detectors to measure the beam background was positioned in the interaction region [61].

- In 2018, Phase 2 commenced, during which the Belle II detector, equipped with only an octant of the vertex detector, was relocated to its operational position. The first collision events were captured on April 26, 2018. Throughout this phase, comprehensive background studies were conducted, and a data sample corresponding to an integrated luminosity of 500 pb^{-1} was accumulated for early physics analysis.

- Phase 3 commenced in 2019 and is ongoing until the end of summer 2022. Throughout this period, the PXD and SVD were installed and the full data-taking operations of Belle II started. As of June 2022, the total recorded integrated luminosity stands at 428 fb^{-1}, with a record peak luminosity of 3.81×10^{34} cm^{-2} s^{-1}. Figure 2.34 provides a summary of the data collected during Phase 3.

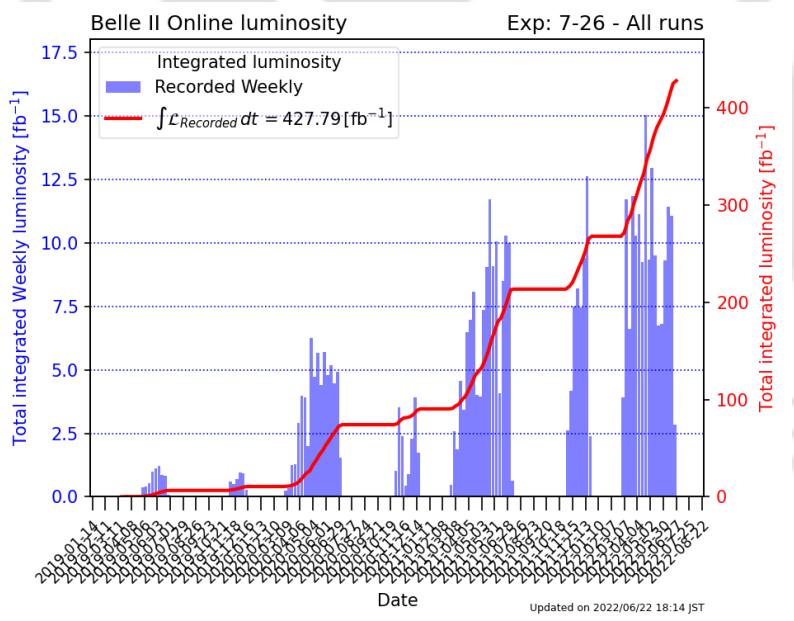


Figure 2.34: Integrated luminosity per week in Phase 3 until June 2022. The current total integrated luminosity corresponds to 428 fb^{-1}. Credit: Belle II Collaboration.

In the near future, the aim is to boost the instantaneous luminosity to 2×10^{35} cm^{-2} s^{-1} using the existing accelerator setup while enhancing the Belle II detector. Luminosity projections until 2026 are shown in Figure 2.35. Additionally, an upgrade program is in progress to achieve the target peak luminosity of 6.5×10^{35} cm^{-2} s^{-1}. This schedule spans the next decade and is outlined in Ref. [62]. It is organised in three major time scales:

- 2022: Long Shutdown 1, commencing in July 2022, will extend for roughly 15 months. During this period, the full installation of PXD, along with new TOP PMTs, will be completed.
- 2026-27: During Long Shutdown 2, the detector will undergo an update to improve its interaction region and various subdetectors. This is aimed at making the detector more robust against higher backgrounds.
- $>$ 2032: Upgrades are being evaluated to enable beam polarization and achieve ultra-high peak luminosity exceeding 1×10^{36} cm^{-2}s^{-1}.

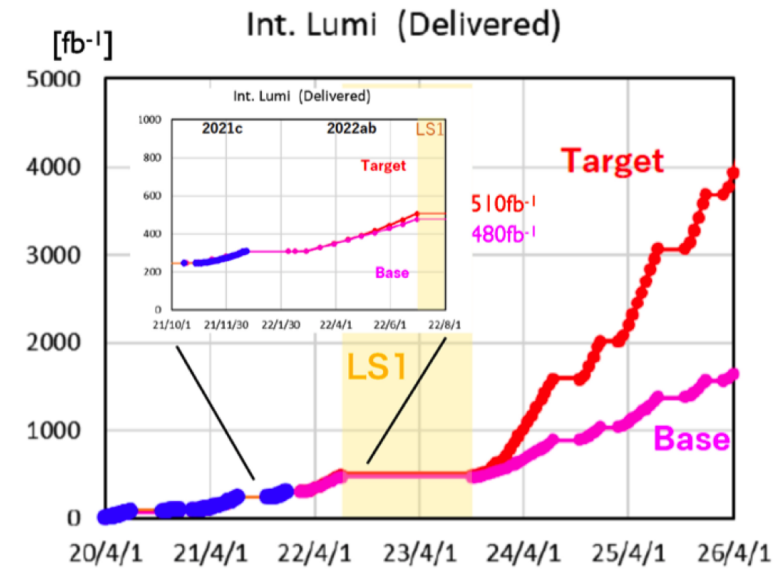


Figure 2.35: Projections of integrated luminosity until 2026, when Long Shutdown 2 is scheduled, are shown. The target scenario (red) is based on extrapolations, including expected Belle II and SuperKEKB improvements, while the base scenario (magenta) is a more conservative extrapolation. (Credit: Belle II Collaboration)

5.1 The Belle II analysis software framework

The Belle II Analysis Software Framework (`BASF2`) [63, 64], developed by the Belle collaboration, serves as the software for data-taking and offline analyses. It is designed to fulfil various tasks within the experiment, including data-taking, offline reprocessing, reconstruction, and data analysis. `BASF2` relies on independent processing blocks known as `modules`, which can be written in either `C++` or `Python`. These modules are organized within a `Python steering` script and executed linearly along a defined path. They interact with a common object store called `DataStore`, which manages relationships between objects in each event, such as particle objects with tracks or vertex information. The data are stored in a reduced format based on `ROOT TTree` [65], containing a collection of `BASF2` objects intended for analysis

and their fundamental relationships. Figure 2.36 provides a schematic illustration of the typical data processing workflow in BASF2.

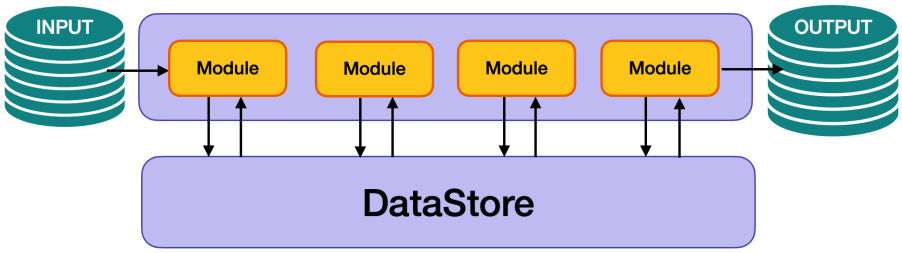


Figure 2.36: Schematic illustration of the data processing flow in BASF2.

5.2 The Belle-to-Belle II Conversion (b2bii)

Enabling the utilization of Belle data within the Belle II software facilitates conducting Belle analysis using new tools and validating the framework. Conversion of Belle's data stored in PANTHER table format [66] into the ROOT format accessible by BASF2 benefits users unfamiliar with the Belle software. This conversion enables Belle II users unfamiliar with the BASF (Belle AnalysiS Framework) software to conduct physics analyses using data collected with the Belle detector. Moreover, it ensures the preservation of Belle data in a format independent of the Belle software, which may become outdated and incompatible with modern computer systems over time. The B2BII conversion and analysis workflow are depicted in Figure 2.37. B2BII conversion contains three BASF2 modules which are as follows:

- B2BIIMdstInput module opens the PANTHER-based Belle mDST(mini-data summary tapes) files and reads the data event-by-event into the main memory.
- B2BIIFixMdst module applies various calibration factors to the PANTHER tables, including those related to beam energy, track momenta and error matrices, energy deposition in the ECL, and particle identification information from the CDC and TOF. Additionally, it implements a general filter called HadronB(J) [67, 68], which includes cuts imposed on recorded Belle data to discard unwanted QED background and contaminations from processes such as $\tau^+\tau^-$, $\gamma\gamma$, and beam gas. This filter must also be applied to MC data to ensure conditions similar to those of recorded data.
- B2BIIConvertMdst module converts the information stored in the Belle PANTHER tables and writes it to the Belle II DataStore. It also extracts beam energy and IP profile information, storing them in the Belle II BeamParameters object and condition database.

The BeamParameters, including the run-dependent beam energies of the high- and low-energy ring, the resulting centre-of-mass energy, and the beam angle, are obtained from a Belle database server or retrieved directly from the Belle mDST file in

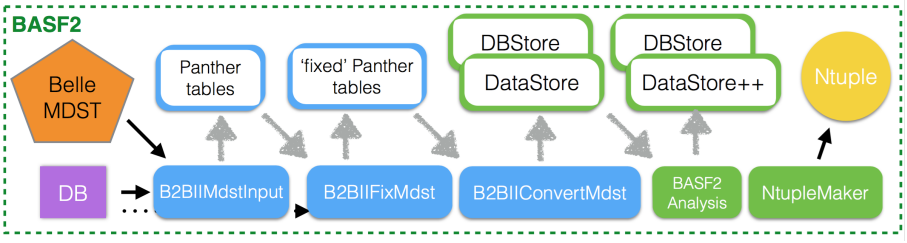


Figure 2.37: Workflow of the conversion process of Belle to Belle II `mDST` files using the `BASF2` modules provided by the `b2bii` package.

the case of privately produced MC data. Further technical information about the `B2BII` conversion process can be found in Ref. [69].

6.0 Analysis Strategy

We are using two different frameworks, `BASF2` and `B2BII`, which are suitable for studying Belle II and Belle data simultaneously. In this way, we can compute the efficiency from both Belle and Belle II, allowing a unified analysis of the Belle and Belle II data sets. The steps that we follow in these analysis can be summarized as follows:

- We perform a blind analysis by optimizing the signal selection criteria and parameterizing the signal and background shapes using Monte-Carlo (MC) samples of signal and background processes.
- Generate signal MC samples to optimize the signal selection and to study the signal shape. Study the dedicated background MC samples for parameterizing the background.
- Optimize the selection criteria to discard the random combination of the background.
- We employ multivariate analysis to reduce continuum backgrounds ($e^+e^- \to q\bar{q}, q = u, d, s, c$) by leveraging the differences in event topology between these backgrounds and the signal.
- Perform an unbinned extended multi-dimensional maximum likelihood fit to extract the signal and background yields.
- After confirming our analysis strategy in the target sample ($B^0 \to \gamma\gamma$), we perform a GSIM (Generation and simulation) test. GSIM involves the random extraction of signal events following their generation and simulation using MC data. This test is conducted to validate the fitting procedure and assess any potential bias in the signal extraction process.

- Use $B^0 \to K^{0*}\gamma$ as a control sample to calibrate the MC for the analysis and study associated systematic uncertainties. Further, Data-MC comparison studies are performed to account for discrepancies among fitting variables.
- Estimate the statistical and systematic uncertainties.
- We extract the signal yield from the data and estimate the $\mathcal{B}$ and Upper Limit (UL) based on the significance of the observed signals.
- Combine the results obtained from the two studies.

A detailed description of the steps are discussed in the subsequent chapters.

Summary

This chapter discussed the physics of the Υ resonances produced at the various CM energies, particularly looking at the $\Upsilon(4S)$ resonance. We briefly discussed the Belle and Belle II experiments, the KEKB and SuperKEKB collider, and the Belle and Belle II detector. We explored the trigger, the data acquisition system and analysis software framework in both experiments. In the end, we concluded the chapter by outlining the strategy adopted for this analysis. The next chapter introduces event generation and simulation studies to perform the analysis.

Chapter 3

Event Generation and Monte Carlo studies

In this chapter, we will outline the procedures for event generation and simulation, essential for conducting Monte Carlo studies. The preliminary event selection criteria will be explained, outlining the process of choosing the appropriate event type for analysis. Additionally, we will delve into the reconstruction of our B^0 signal candidates from neutral photons, implementing optimized selection criteria to distinguish them from the background. Further, we will perform the background analysis using Monte Carlo samples and formulate strategies for their suppression.

1.0 Blind Analysis

A blind analysis is employed in our analysis to prevent bias in the analysis process. This approach is particularly crucial when researchers are searching for rare signals within a dataset.

The key idea is to conceal certain aspects of the data or analysis details from the researchers until after the analysis strategy has been defined, the criteria for signal extraction have been established, and potential sources of bias have been considered. Researchers might "blind" the analysis by withholding certain information, such as specific data ranges until the analysis plan is finalized.

2.0 Monte Carlo Event Generation

Before the analysis can be performed using the real data, the reconstruction, the other analysis tools, and techniques must be tested on the simulated MC data. MC simulated data is necessary to ensure proper reconstruction of the kinematical variables, such as the mass of the final state particles, beam-constrained mass ($M_{\rm bc}$), and the energy difference between the center-of-mass (CM) energy and the beam energy, etc. MC production process involves two steps: event generation and modeling of detector response. Events are generated according to various physics processes

using a decay table, which specifies the decay models, modes, branching fractions, etc., for all possible particles involved in the decay chain. The event generation is done using the `EvtGen` [70] event generator package, designed for B-meson decays based on particle properties such as mass, width, lifetime, charge, etc. The generated particles are then passed through `GEANT4` [71] for Belle II and `GEANT3` [72] for Belle to simulate the detector response after adding the background. MC data files generated through these processes match the real data formats. A schematic representation of the steps involved in the event generation and detector simulation at Belle (Belle II) is shown in Figure 3.1.

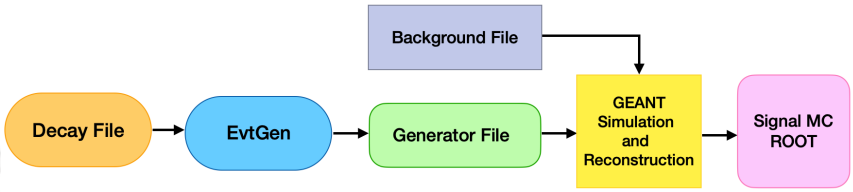


Figure 3.1: Schematic diagram of steps followed in event generation and simulation

The initial stage of event generation involves generating a decay file that outlines the signal decay chain, along with generic decay chains (other than the signal) based on the physics model. These decay models are fed to the `EvtGen` to generate the signal events for $B^0 \to \gamma\gamma$ which can be summarized as below:

- Vector to Scalar and Scalar using $B^0\bar{B}^0$-like coherent mixing (`VSS_BMIX dm`) to generate $\Upsilon(4S) \to B^0\bar{B}^0$. The two possible daughter particles must be charge conjugates with the same lifetime. Their mass difference is supplied as an argument to the model, in units of $\hbar/s$.
- N-Body Phase Space (`PHSP`) to generate $B^0 \to \gamma\gamma$
- N-Body Phase Space (`PHSP`) to generate $\overline{B^0} \to \gamma\gamma$

Table 3.1 provides a summary of the decay models utilized to generate the decay $B^0 \to \gamma\gamma$.

Decay Table	
Decay Mode	Decay Model
$\Upsilon(4S) \to B^0\overline{B^0}$	`VSS_BMIX dm`
$B^0 \to \gamma\gamma$	`PHSP`
$\overline{B^0} \to \gamma\gamma$	`PHSP`

Table 3.1: Decay Table for the $B^0 \to \gamma\gamma$ mode

The decay models used to generate the signal MC events for the control sample ($B^0 \to K^{0*}\gamma$) are summarised in Table 3.2 as follows:

Decay Table	
Decay Mode	Decay Model
$\Upsilon(4S) \to B^0\overline{B^0}$	`VSS_BMIX dm`
$B^0 \to K^{0*}\gamma$	`SVP_HELAMP 1.0 0.0 1.0 0.0`
$\overline{B^0} \to K^{0*}\gamma$	`SVP_HELAMP 1.0 0.0 1.0 0.0`

Table 3.2: Decay Table for the control sample ($B^0 \to K^{0*}\gamma$)

`SVP_HELAMP` model is used to describe $b \to s\gamma$ transitions. It describes the decay of a scalar to a vector and a photon, which is parameterized by the helicity amplitudes. The parameters used in the model '`1.0 0.0 1.0 0.0`' are the magnitudes and phases of the complex amplitudes of the particles, which are specified in pairs. The amplitudes are ordered, starting with the highest allowed helicity of the first particle.

For Belle, the background events for $e^+e^- \to q\bar{q}$ are referred as continuum background, where $q = u, d, s, c$, are generated by `PYTHIA` [73]. For Belle II, the $e^+e^- \to q\bar{q}$ sample is generated using the `KKMC` [74] generator interfaced with `PYTHIA`. The `EvtGen` package also simulates the decay of short-lived particles. A sample of $e^+e^- \to \tau^+\tau^-$ events is generated with `TAUOLA` [75]. Background samples corresponding to 1 ab^{-1} or more, are used. Belle simulation includes the effect of beam background by overlaying data taken that is unrelated to e^+e^- collisions (random triggers). QED background processes, like $e^+e^- \to e^+e^-(\gamma)$ and $e^+e^- \to \gamma\gamma(\gamma)$ are simulated by `BABAYAGA.NLO` [76, 77]. Belle II simulation samples include the effect of simulated beam-induced background caused by the Touschek effect (scattering and loss of beam particles) and by beam-gas scattering, as well as luminosity-dependent backgrounds caused by Bhabha scattering and two-photon quantum electrodynamic processes [61]. Cross-sections of the main physics processes related to the analysis used for the MC generation are given in Table 3.3.

Physics Process	**Cross-section (nb)**
$\Upsilon(4S)$	1.11
$u\bar{u}$	1.61
$d\bar{d}$	0.40
$s\bar{s}$	0.38
$c\bar{c}$	1.30
$\tau\bar{\tau}$	0.92

Table 3.3: Total production cross-section of various physics processes from collisions at $\sqrt{s} = 10.58$ GeV.

3.0 Data Samples

- Data: Target data is the same as LS1; Proc13 + Buckets 26 to 36, which corresponds to about 362 fb^{-1}. Here, the Proc13 is the processing number used for the official production of the data. The Bucket label is used for "prompt" productions, and these are a group of runs to be processed together. We also study off-resonance data from experiments 8 and 12 processed under the `Proc13` (9.5 fb^{-1}) processing campaign and `Bucket 20, 24 and 34` (32.8 fb^{-1}) from experiment 18 and 25 to check the agreement between data and background MC. We use good runs only(as defined by the data production group).

  * For the Belle study, the target sample consists of the Υ(4S) CaseB dataset collected by the Belle detector during its more than 10 years of running. It consists of 694 fb^{-1} of integrated luminosity which corresponds to $(753 \pm 10) \times 10^6$ $B\bar{B}$ pairs. The number of $B\bar{B}$ events is smaller than the one corresponding to the entire Belle dataset $(772 \pm 11) \times 10^6$, as we only use the data containing timing information. We have studied the off-resonance data from experiments 7-65, consisting of 89.50 fb^{-1}.
  * For the control sample study, we have processed full data sample (`362 fb`$^{-1}$) of on-resonance data for Belle II. For Belle, we have used the data from experiments 7-65 with the integrated luminosity of `694 fb`$^{-1}$. The permission is given to use the full dataset for DATA-MC comparison and the calibration of BDT and π^0/η selection to assign systematic uncertainty.

In this analysis, we have used the MC samples from the 15th official Belle II MC campaign (MC15ri_b) located on the Belle II grid computing system. The two types of MC samples are used in our analysis - signal, and background.

- Signal MC: We have processed officially generated 100,000 signal events using the decay file illustrated in Table 3.1. For the Belle study, we have generated 100,000 signal MC events using the same decay file illustrated in Table 3.1.

- Background MC: The dominant background in this analysis arises from light quark continuum events ($u\overline{u}$, $d\overline{d}$, $s\overline{s}$ and $c\overline{c}$). There are also contributions from generic (`mixed and charged`), QED events, $qq\gamma$ events, τ pairs and "two-photon" processes resulting from the π^0 and η decays. MC samples for these processes are taken from MC15ri_b campaign generated by the MC production team with BtoXgamma skim. We have processed 400 fb^{-1} equivalent of $e^+e^- \to \gamma\gamma$ samples, 90 fb^{-1} of $e^+e^- \to e^+e^-$ and 1 ab^{-1} of continuum, generic and τ pairs samples for the background study.
  For the Belle, we have used the background MC datasets which are centrally generated and stored in KEKCC (Computing Centre at KEK, Japan). We have used the six streams of the generic MC simulation sample, having luminosity in 711 fb^{-1} of each stream. Samples called "generic MC" consist of

continuum and $B\overline{B}$ events. Multiple streams of these samples are available, which is very useful to make more in-depth studies of the background nature but also to have independent samples of the same size to validate Machine Learning techniques.
We have also investigated the $B\bar{B}$ rare MC, comprising of both charged and mixed rare backgrounds, employing a dataset that is 50 times larger than real data. $B\bar{B}$ rare decay means one B meson decays via some rare processes (electroweak and radiative penguins, charmless).

- Control sample MC: We have also generated 100,000 signal MC events for $B^0 \to K^{0*}\gamma$ for the control sample study. The decay models used to generate the signal MC events for $B^0 \to K^{0*}\gamma$ are summarised in Table 3.2 in Sec. 2.0. We have processed 1 ab^{-1} of generic MC for the background study.
For the Belle study, we used the 1 stream of background MC samples for experiments 7-65 corresponding to 711 fb^{-1}.

3.1 Software Version

The MC samples are produced with `release-06-00-08` in the fifteen official MC campaign (MC15ri_b). The reconstruction of B^0 (discussed in Sec. 6.0) and the background study (discussed in Sec. 7.0) uses `light-2303-iriomote` release.
`Proc13` uses `release-06-00-12` with the global tag `data_reprocessing_proc13` for the reconstruction of particles and the calibration. Bucket 26 to 36 uses `release -06-00-05` and later release with the global tag data_reprocessing_prompt.
For Belle analysis, we are using the recommended CaseB data which is reprocessed with the latest Belle library `b200901247_0910`.

4.0 Skim Selection

Skims refers to a set of loose selection criteria applied to the raw data or Monte Carlo (MC) samples to retain only the events of interest for a specific analysis. These criteria are designed to enhance the signal-to-background ratio, reduce the volume of data to be processed, and focus on events relevant to the particular physics process under investigation. Skims are a crucial step in data preprocessing before detailed analysis and interpretation. The BtoXgamma skimmed samples are used for this analysis. The following selection criteria are used:

- Energy of γ in the CM frame $1.4 < E_\gamma(\mathrm{CM}) < 3.4$ GeV;
- Ratio of the energy deposited in the array of the central 3×3 calorimeter cells to that of 5×5 cells crystals around the central crystal without corners. $(E_{9/21}) > 0.9$.
- Reject the events with less than 3 good tracks. $\mathtt{n_{tracks}} >= 3$; and

- The ratio of the 2nd to 0th order Fox Wolfram moment `R2` < 0.7.

These selection criteria are summarised in Table 3.4.

Cuts	Variables	Selection Criteria
Photon(γ) cuts	$\texttt{E}_{\gamma}$(CM)	$1.4 < E_{\gamma}(\text{CM}) < 3.4$ GeV
	$\texttt{E}_{9/21}$	> 0.9
Events cuts	$\texttt{n}_{\texttt{tracks}}$	$>= 3$
	`foxwolframR2`	< 0.7

Table 3.4: BtoXgamma Skim Selection Criteria

For Belle analysis, we select only those events which pass the `HadronB` skim[1] and rad_b selection criteria shown in Table 3.5. In fact, at the same time of the `b2bii` conversion, the `HadronB` skim is applied which is prior to the full event reconstruction algorithm to make the necessary corrections.

Variables	**Selection Criteria**
E_{γ}(CM)	$1.4 < E_{\gamma}(\text{CM}) < 3.4$ GeV
$E_{9/25}$	> 0.9

Table 3.5: rad_b selection criteria.

The `B2BIIFixMdst` module (discussed in Sec. 5.2) in `b2bii` corrects the difference between the MC and Data and rejects events which do not pass the L4 trigger (discussed in Sec. 3.9) and `HadronB` selection criteria. The signal efficiency due to the `HadronB` skim is 70%. The drop in the signal efficiency is due to the `B2BIIFixMdst` module which includes the `HadronB` requirement. This skim selection is used to select the $B\overline{B}$ events; however, since the data processing was already completed, the loss could not be recovered. Events are discarded based on track multiplicity and visible energy: the event must have at least three charged tracks with $p_T > 0.1$ GeV/c and the visible energy (sum of the energy of charged tracks and reconstructed photons) must be greater than 20% of $\sqrt{s}$. Further selection criteria remove the majority of the beam-gas background and two-photon events; the first being reduced by requiring that the primary vertex position of the event be close to the IP. Background events from QED and tau-pairs are suppressed by cutting on the total energy measured in the ECL and on the HJM variable[2]. Such combined cuts do not remove light quark pair production events ($e^+e^- \to q\overline{q}$ with $q = u, d, s, c$) but are not very efficient for inclusive ψ events. Therefore, the events with J/ψ and ψ(2S) candidates

[1]The detailed selection criteria are explicitly listed in Appendix (A).

[2]The Heavy Jet Mass is the invariant mass of particles found in hemispheres perpendicular to the event thrust axis.

are explicitly added to `HadronB(J)`. In summary, by using the pre-selection, or skim, described above we expect our MC data samples to be almost free from low-multiplicity processes (QED and $\tau^+\tau^-$). Even though the exclusive reconstruction of hadronic B-decays would naturally discard those events, the skim prevents us from running the reconstruction on them which would have increased the processing time. A detailed description of skims is available at Ref. [67]

5.0 Event Selection

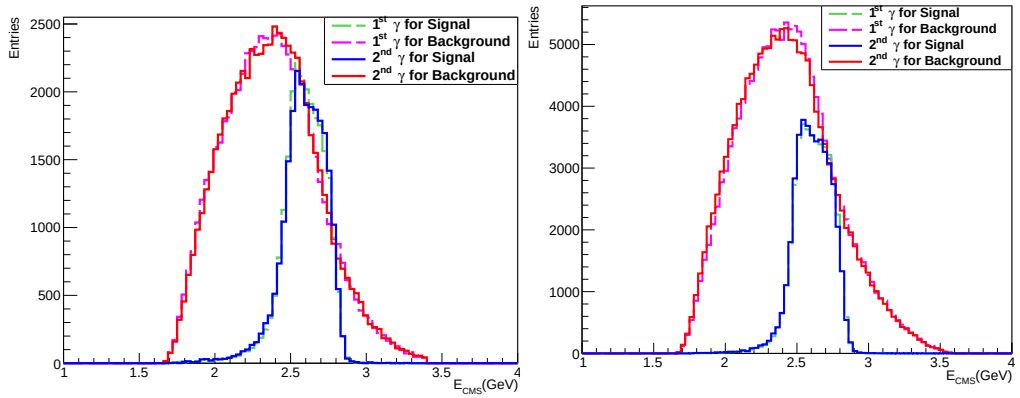
5.1 Photon Selection

The $B^0 \to \gamma\gamma$ event gives two back-to-back highly energetic photons in the CM frame. These events are reconstructed using four-momentum conservation by identifying the γ candidates which are detected in the ECL of the detector. The selection variables used in the Belle Study are the same except those specific to Belle II. The requirements for the selection cuts are different because both detectors have different performances with varying efficiency. The first step of this analysis is to select true candidate photons.

The final signal states only consist of photons, so it is essential to accept as many true photons while rejecting false photons as possible. Two highly energetic photons are selected by using a selection on their energy at the center of the mass frame. We use a selection cut, $E_\gamma(\mathrm{CM}) > 1.4$ GeV, to select the photons. This cut is quite effective in removing all the low-energy photons, making the B meson reconstruction easier without worrying about the ordering of the photon candidates. The energy distribution of the two photons for signal and background MC in the CM frame is shown in Figure (3.2a) for Belle and (3.2b) for Belle II. The two-photon candidates are mostly back-to-back kinematically, it isn't very likely to have one gamma in the barrel region and another in the one end cap. Since the end-cap region has a large amount of gamma from the beam background, we require "`clusterReg`" to be 2 to exclude gamma candidates in end-cap. Here, `clusterReg` describes the ECL region of a cluster: 1 for forward, 2 for barrel, and 3 for backward.

The ClusterReg distributions of two candidate photons for signal and background MC are shown in Figure 3.3 for Belle and Figure 3.4 for Belle II.

Photons are identified as energy depositions in the ECL that are not matched to a charged track. To discriminate electromagnetic showers and hadron showers in ECL, we have "$\mathtt{E}_{9/25}$" variable in the b2bii package, which is the ratio of the energy deposited in 3×3 CsI crystals and in 5×5 crystals. Photons are detected in the ECL sub-detector by the amount of energy deposited in ECL cells. To separate these showers from hadronic showers produced by a hadron, the ratio ($\mathtt{E}_{9/25}$) should be near about 1 since electromagnetic showers are narrower than hadronic ones. We use a selection of $\mathtt{E}_{9/25} \geq 0.95$. Figure 3.5 shows the distribution of the ratio ($\mathtt{E}_{9/25}$) for γ candidates for the signal and background MC.



(a) (b)

Figure 3.2: (a) Energy of two γ_s at the CM frame for Belle and (b) for Belle II

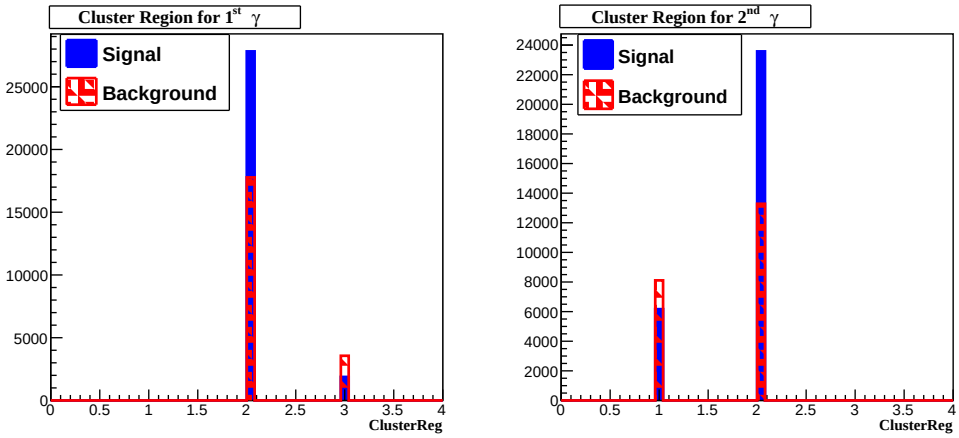


Figure 3.3: `ClusterReg` distribution of two γ candidates for signal and background MC for Belle.

For Belle II, we have three cluster variables: "$\texttt{E}_{9/21}$" which is the ratio of the energy deposited in 3×3 CsI crystals and in 5×5 crystals around the central crystal without corners, `clusterZernikeMVA` [78] which is the MVA output trained on Zernike moments [79] to separate photon and K_L^0 showers, and the number of crystals in the cluster (`clusterNHits`).

To separate electromagnetic showers from hadronic showers produced by a hadron, we use a selection of $\texttt{E}_{9/21} \geq 0.95$. Figure 3.6 shows the distribution of the ratio ($\texttt{E}_{9/21}$) should be near about 1 since electromagnetic showers are narrower than hadronic ones. To select gammas from the radiative B decay, we require `clusterNHits` ≥ 15. Figure 3.7 and 3.8 show the distribution of `clusterNHits` and `clusterZernikeMVA` for γ candidates from both signal and background MC. The value of `clusterZernikeMVA` is close to 1 for showers coming from a photon and 0 for a K_L^0. We apply a selec-

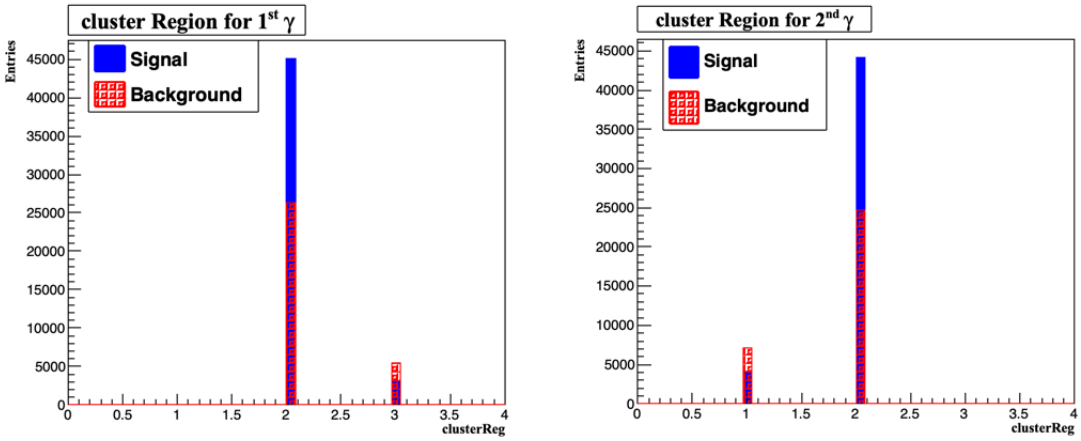


Figure 3.4: `ClusterReg` distribution of two candidate photons(γ_1 and γ_2) for signal and background for Belle II.

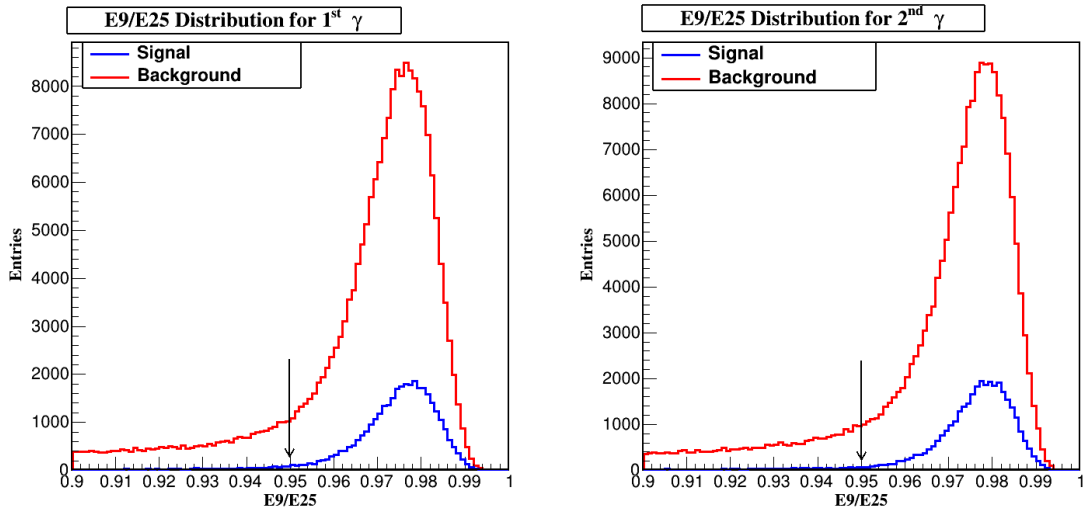


Figure 3.5: $E_{9/25}$ ratio of γ candidate for signal and background MC in Belle study

tion `clusterZernikeMVA` > 0.75 to suppress contamination from K_L^0 clusters. The normalized distribution of `clusterZernikeMVA` coming from the photon and K_L^0 clusters is shown in Figure 3.9. We optimize the selection cuts by using the punzi figure of merit [80], which is defined as, FOM $= \frac{e_{\text{sig}}}{(1.5+\sqrt{\text{B}})}$, where e_{sig} is the efficiency of the signal selection and B are the expected background events. To find the best value of the selection cut, we have used the selection variable one at a time, fixed each to an optimal value and repeated for another selection variable. The vertical arrow lines show the optimized cut values in Figures 3.5 and 3.6. The FOM plots for selection variables can be found in `Appendix (B)`.

The Timing information of ECL clusters is also used, which is `abs(clusterTiming)`

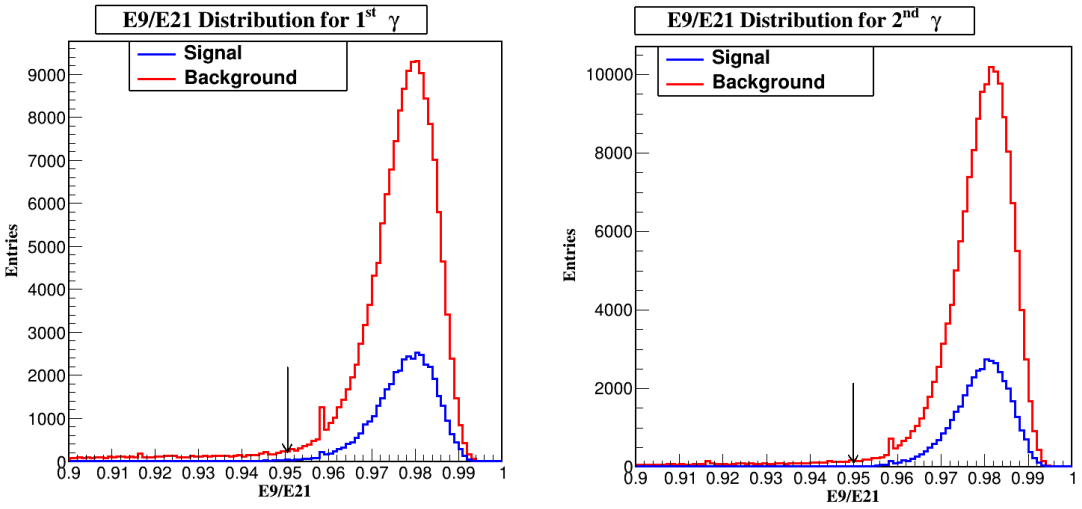


Figure 3.6: $\mathtt{E_{9/21}}$ ratio of γ candidate for signal and background MC in Belle II study.

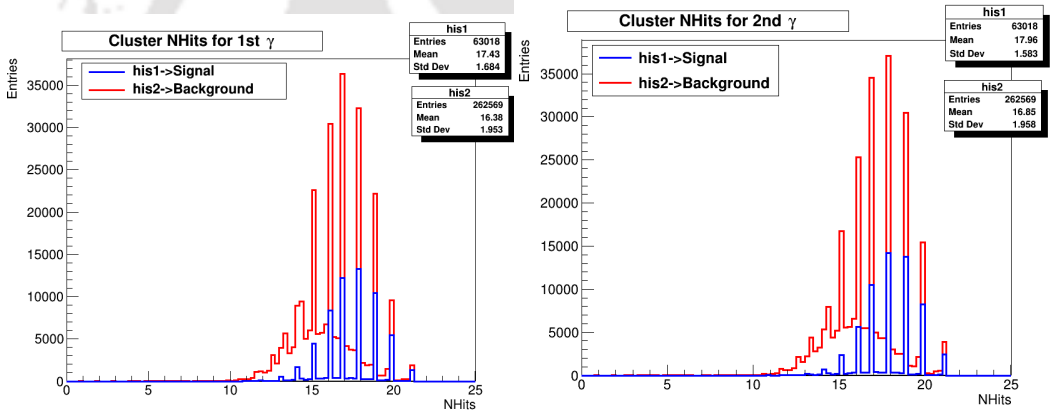


Figure 3.7: `ClusterNHits` of γ candidate for signal and background MC in Belle II study.

`< 200 ns` and `abs(clusterTiming/clusterErrorTiming) < 2.0`. Here, `clusterTiming` is defined as the difference between photon time and event time, and `clusterErrorTiming` is the error on `clusterTiming` variable. For photons coming from an e^+e^- collision event, the value of `clusterTiming` variable is consistent with zero. These selections are optimized to reject out-of-time beam backgrounds while maintaining good DATA/MC agreement.

To suppress gammas coming from π^0 and η decays, we use the "π^0/η veto" tool [81]. Rejection of π^0 and η meson is of primary importance in a search for a pure radiative rare decay of the B^0 meson. Photons coming from asymmetric decays of π^0 or η meson to di-gamma final states constitute a significant background for the analysis. A photon emitted along the flight direction of $\pi^0(\eta)$ can be mis-reconstructed as a signal side hard photon candidate. The signal side hard photon is paired with all the photon candidates in the event to create a $\pi^0(\eta)$ meson. The tool then assigns a weight (pi0_prob or eta_prob) on the signal side photon candidate. For photons

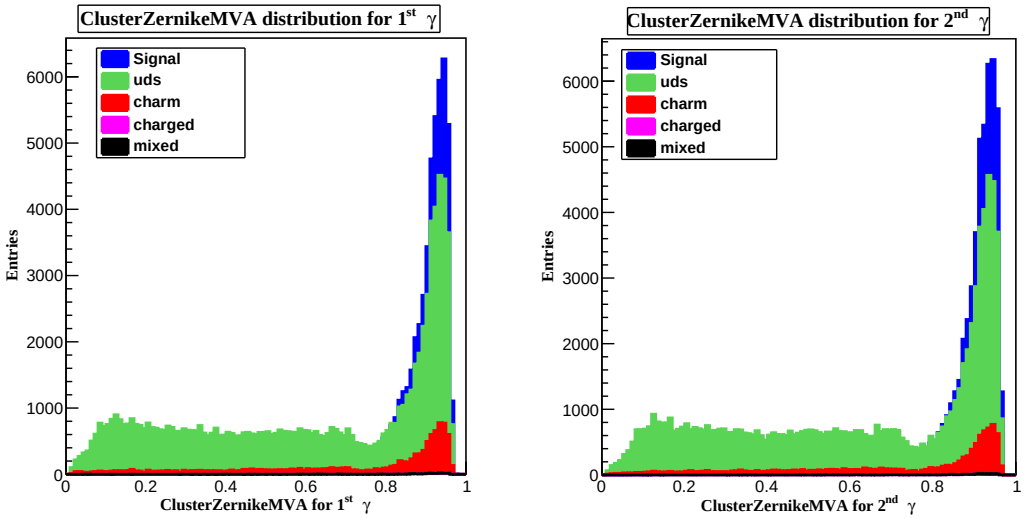


Figure 3.8: `ClusterZernikeMVA` distribution of γ candidate for signal and background MC

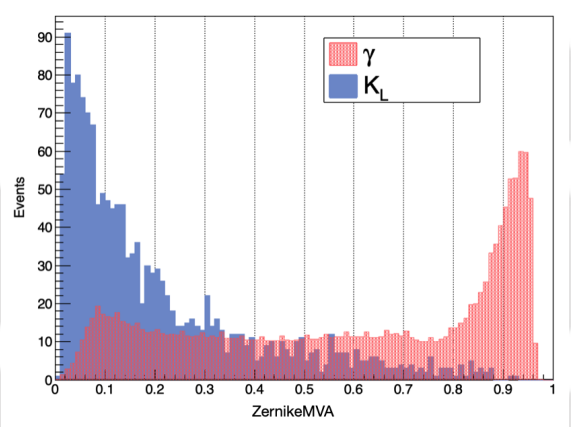


Figure 3.9: The normalised plot for `ClusterZernikeMVA` distribution obtained for K_L^0 and photons

coming from the decay of $\pi^0(\eta)$ the weight is close to 1, whereas, for other scenarios, it's close to 0. We require π^0 probability and η probability to be less than 0.71 and 0.62, respectively, as shown in Figure 3.10 and 3.11. In order to find the optimal cuts, firstly, we applied $q\overline{q}$ suppression (see Sec. 7.4 for detail), $C_{\text{BDT}} > 0.45$, to reduce the obvious $q\overline{q}$ backgrounds. Then, the cut values of the π^0/η veto are optimized by Punzi FOM. Various selection cuts for candidate photons are listed in Table 3.6 for Belle II analysis.

For the Belle data set, another FastBDT is similarly trained using Belle MC samples. Note that some variables used in [81] are not included in Belle due to the different format of the dataset. The photons coming from π^0/η decays are rejected

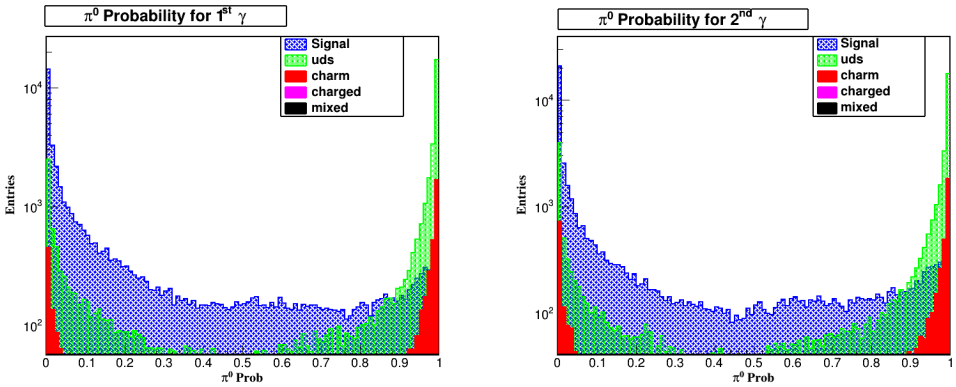


Figure 3.10: π^0 probability for signal and background MC in Belle II study.

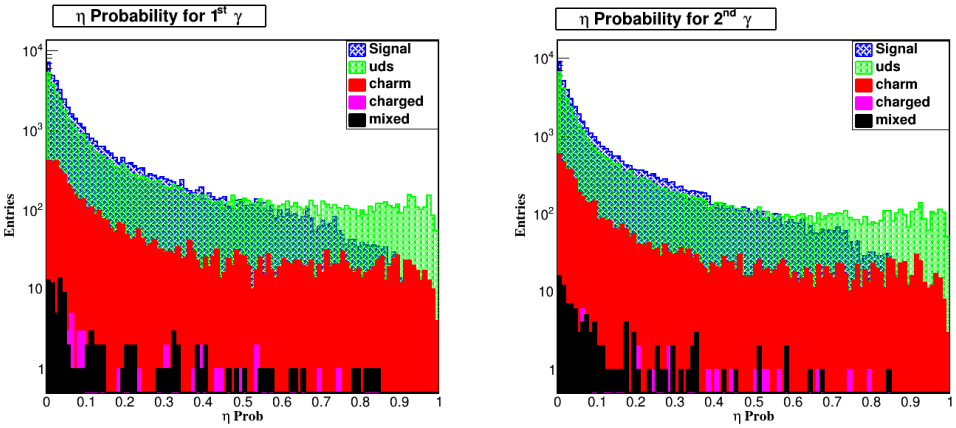


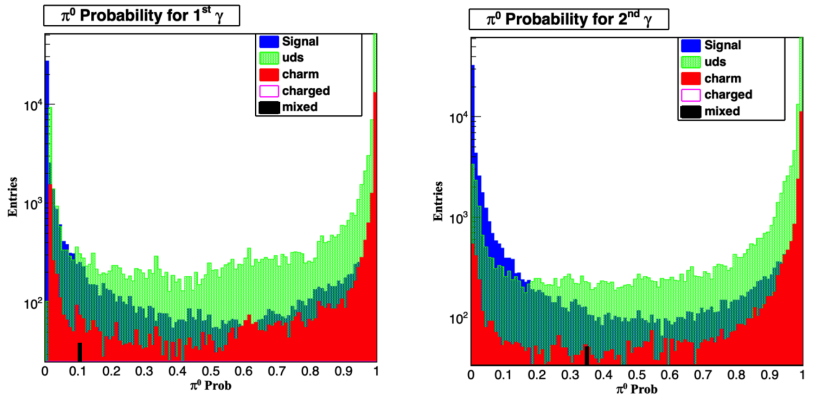
Figure 3.11: η probability for signal and background MC in Belle II study.

Particle	Variables	Selection Criteria
Photon(γ)	E_γ(CM)	$1.4 < E_\gamma(\text{CM}) < 3.4$ GeV
	$E_{9/21}$	> 0.95
	clusterTiming	`abs(clusterTiming)` < 200 ns
		`abs(clusterTiming/clusterErrorTiming)` < 2.0
	$\theta(\gamma)$	`clusterReg` $== 2$
	ClusterNHits	≥ 15
	ClusterZernikeMVA	> 0.75
	π^0 Prob	< 0.71
	η Prob	< 0.62

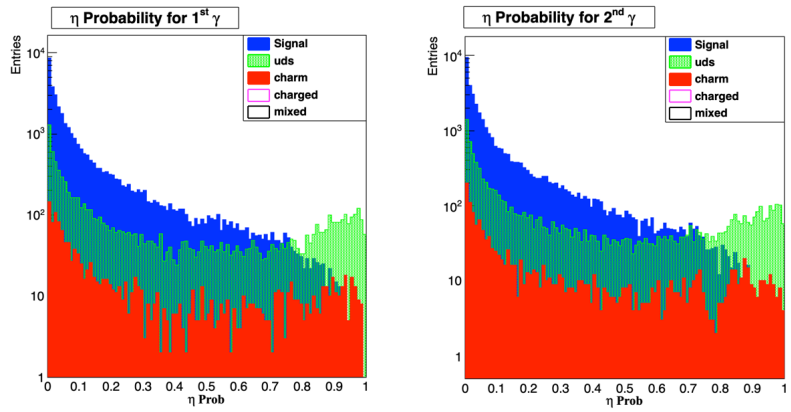
Table 3.6: Selection Cuts for candidate photon for Belle II

based on likelihood information $P_{\pi^0}(\gamma)$ and $P_\eta(\gamma)$ obtained by using the energy and polar angle of photons, di-photon invariant mass, $E_{9/25}$, clusterNHits and cosine of helicity angle of π^0/η. $P_{\pi^0}(\gamma)$ and $P_\eta(\gamma)$ denotes the maximum probability that a candidate photon can be combined with another photon in the event to form a π^0 and η particle, respectively. We require π^0 and η probability to be less than 0.50 and 0.65, respectively as shown in Figure 3.12. Similar to Belle II, we applied $q\bar{q}$

suppression (see Sec. 7.3 in detail), $C_{\rm BDT} > 0.55$, to reduce the $q\bar{q}$ background, and then the cut values of π^0/η veto are optimized. Various selection cuts for candidate photons are listed in Table 3.7 for the Belle analysis.



(a) π^0 Probability



(b) η Probability

Figure 3.12: (a) π^0 and (b) η Probability Distribution of γ candidate for signal and background MC in Belle study.

Particle	Variables	Selection Criteria
Photon(γ)	E_γ(CM)	$1.4 < E_\gamma(\mathrm{CM}) < 3.4$ GeV
	$E_{9/25}$	> 0.95
	$\theta(\gamma)$	$clusterReg == 2$
	$P_{\pi^0}\gamma$	< 0.50
	$P_\eta\gamma$	< 0.65

Table 3.7: Selection Cuts for candidate photon for Belle analysis.

6.0 Event Reconstruction

Reconstruction of events is carried out through variables like beam-constrained mass $M_{\rm bc}$, and energy difference (ΔE), which are essentially based on the energy conservation principle. B^0 meson candidates are formed by combining two photons candidates selected based on the beam-constrained mass $M_{\rm bc}$ and energy difference ΔE, defined as:

$$M_{\rm bc} = \sqrt{(E^{\rm CM}_{\rm beam})^2 - (p^{\rm CM}_{B^0})^2}$$

$$\Delta E = E^{\rm CM}_{B^0} - E^{\rm CM}_{\rm beam}$$

where, $E^{\rm CM}_{\rm beam}$ is the beam energy, E_{B^0} and p_{B^0} are the energy and momentum of the B^0 candidate, all evaluated in the e^+e^- CM frame. B candidates are produced almost at rest in the CM frame, carrying half the total energy. For correctly reconstructed B candidates ΔE should be close to zero. This variable is especially helpful in discriminating backgrounds arising due to particle misidentification since it is sensitive to the respective mass hypothesis employed while reconstructing daughters of B candidates. Second, is the beam-energy-constrained mass, a fairly independent kinematic variable related to the vector sum of momenta of B daughters. Here the B-meson energy is replaced with beam energy as the latter has a better resolution.

For correctly reconstructed B candidates, the $M_{\rm bc}$ distribution peaks at the nominal B mass. For $B^0 \to \gamma\gamma$ events, the $M_{\rm bc}$ distribution will peak at the B-meson mass, 5.279 GeV/c^2. Figures 3.13 show the $M_{\rm bc}$ and ΔE distributions after applying the selection criteria to the signal MC. The selection criteria are summarized in Table 3.8 for $B^0 \to \gamma\gamma$ mode. The ΔE distribution is asymmetric, and the peak near zero with the tail to the negative ΔE side due to photon energy loss outside the active volume of ECL is shown by Figure 3.13 for signal MC. We also checked the number of B^0 candidates per event, which is given by multiplicity as shown in Figure 3.14. Only one candidate B^0 per event exists.

Variables	Selection Criteria
$M_{\rm bc}$	$5.24 < M_{\rm bc} < 5.29$ GeV/c^2
ΔE	-0.6$< \Delta E <$ 0.2 GeV

Table 3.8: Table for Selection Criteria

For Belle II analysis, we used the same selection criteria summarized in Table 3.8 for the $B^0 \to \gamma\gamma$ mode. Figure 3.15 shows the $M_{\rm bc}$ and ΔE distributions after applying the selection criteria to the signal MC. The multiplicity of B^0 per event is also found to be 1.

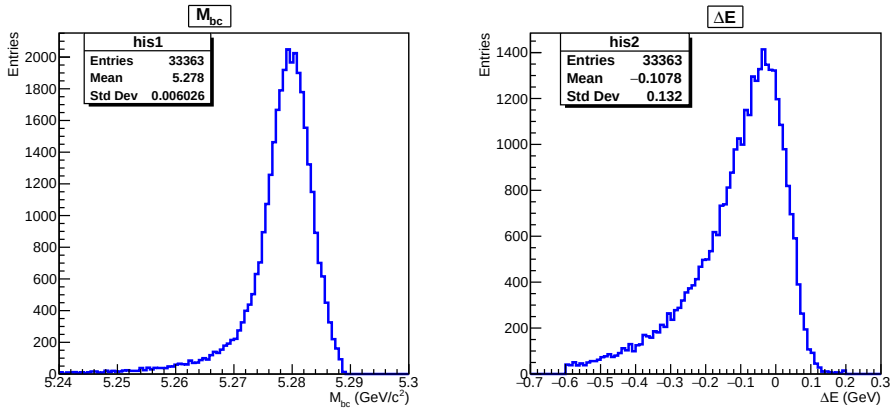


Figure 3.13: M_{bc} and ΔE distribution of reconstructed B^0 for Belle.

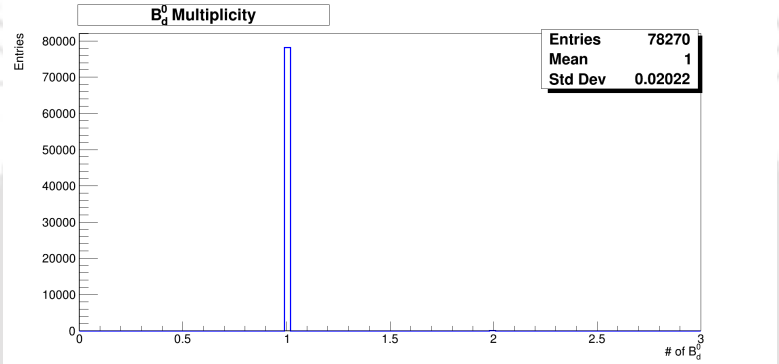


Figure 3.14: Multiplicity of B_d per event for Belle.

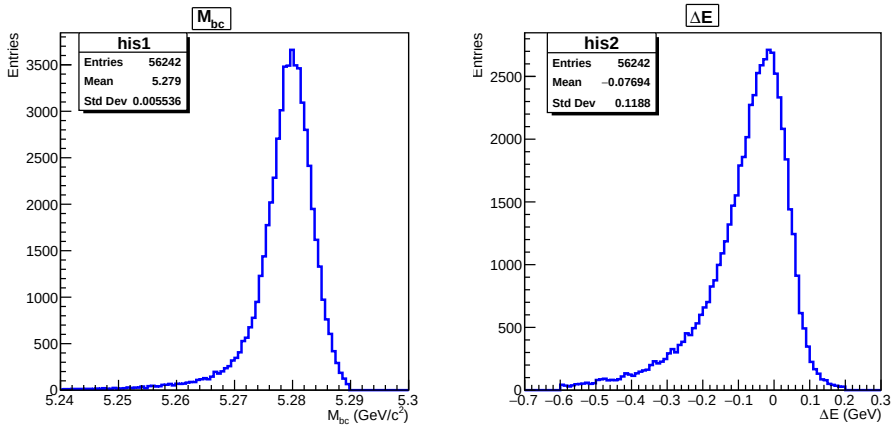


Figure 3.15: M_{bc} and ΔE distribution of reconstructed B^0 for Belle II.

6.1 $K^{\pm}$ and $\pi^{\pm}$ selection for Contol Sample ($B^0 \to K^{0*}\gamma$)

Charged tracks with impact parameters $|dr| < 0.5$ cm and $|dz| < 3$ cm is considered to be a kaon candidate if it satisfies:

$$\mathcal{L}(K/\pi) \equiv \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi} > 0.6$$

Rest of the tracks are considered to be coming from pions. Here, $\mathcal{L}_K$ and $\mathcal{L}_\pi$ are the likelihood of the track to be coming from a kaon and pion, respectively.

6.2 K^{0*} reconstruction

The K^{0*} candidate is reconstructed by combining a kaon with a pion. We retain K^* candidates inside the invariant-mass window $M_{K^*} \in (0.817, 0.968)$ GeV/c^2, which corresponds to about $\pm 2\sigma$ around the nominal mass of K^* meson. To remove the multiple K^{0*} candidates, K/π vertex fit is performed, and the events with the least χ^2 value are chosen as the best K^{0*} candidate.

6.3 $B^0 \to K^{0*}\gamma$ Reconstruction

K^{0*} candidates are combined with the photons candidates to form the B^0 candidates by applying the selection mentioned in Table 3.8 for the $B^0 \to K^{0*}\gamma$. After applying all the selection criteria, we are often left with more than one reconstructed B candidate per event. The candidate multiplicity is 1.08 after all selection cuts. Out of the multiple reconstructed B candidates, we retain the candidate having the highest value of FastBDT output.

7.0 Background Study

7.1 Backgrounds in the Belle analysis

The backgrounds for the analysis can be classified into the following two categories:

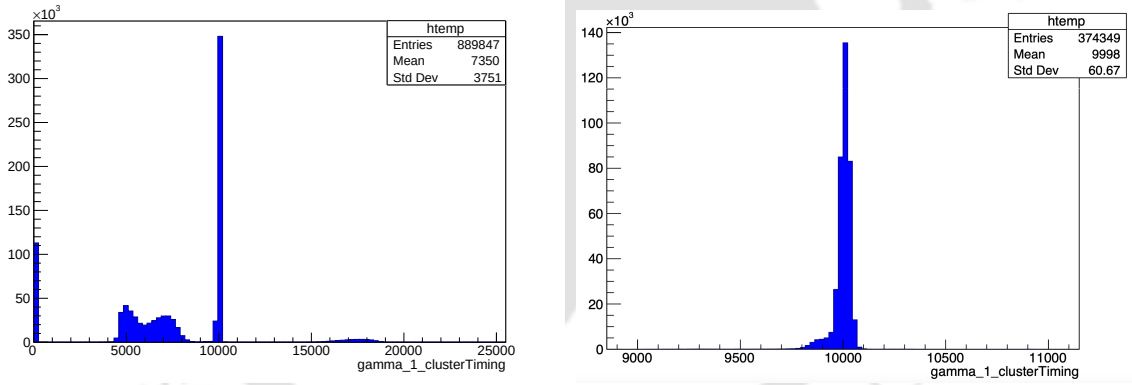
- **Continuum background**: The dominant source of background is $e^+e^- \to q\bar{q}$ continuum events with $q \in (u, d, s, c)$. These quarks have a mass significantly lower than B mesons, hence in the CM frame, they are produced with high momenta. The daughters from such $q\bar{q}$ events exhibit a jetlike event shape.

- **Generic $B\bar{B}$ background**: The decay of $B\bar{B}$ events, where the daughters of decaying B candidates pass our allowed kinematic constraints, can fake signal events. These fake signal events can mimic our signal and hinder the calculation of signal yield via the fit model. These are called the peaking backgrounds or the higher resonance backgrounds.

The dominant background in this analysis arises from light quark continuum events ($e^+e^- \to u\bar{u}, d\bar{d}$, $s\bar{s}$ and $c\bar{c}$). Continuum and generic MC samples are processed, and candidate events are selected using the selection criteria discussed in Tables 3.7 and 3.8 earlier.

Due to the long decay time of the CsI crystals, QED processes like Bhabha scattering or $e^+e^- \to \gamma\gamma$ events can leave two perfectly back-to-back clusters in the ECL (due to pile-up) and can very well mimic the $B^0 \to \gamma\gamma$ events. These off-time

events act as the beam background in MC. It gives an unusual peaking component in the signal region of $M_{\rm bc}$ distribution. This off-time QED background observed for the first time in $B^0 \to \gamma\gamma$ analysis [49], are removed with 100% efficiency by using the timing information of the Tigger Cells associated to the ECL cluster. And, as timing information is not simulated, a selection based on MC truth information is used to remove such background from the simulated samples with equal efficiency. In the B2BII framework, we have used photonID cut `mcPDG!=nan, genMotherID and genMotherPDG!=0` to reject all the peaking structures from the MC samples and corresponding to that we have used the `clusterTiming` in the real data. The peaking effect of the off-time QED background is completely removed after applying the cuts requiring that the photons hit the ECL cluster within 9 to 11 microseconds [82] of the collision, and the remaining distributions are smooth, as one would expect from generic continuum background. In Fig 3.16(a), the peak at 0 should correspond to the beam background, and the events around 10,000 ns should correspond to clusters produced at the interaction point in time with the event, and the rest are off-time. Fig 3.16(b) shows the distribution after applying the selection criteria having only in-time events.

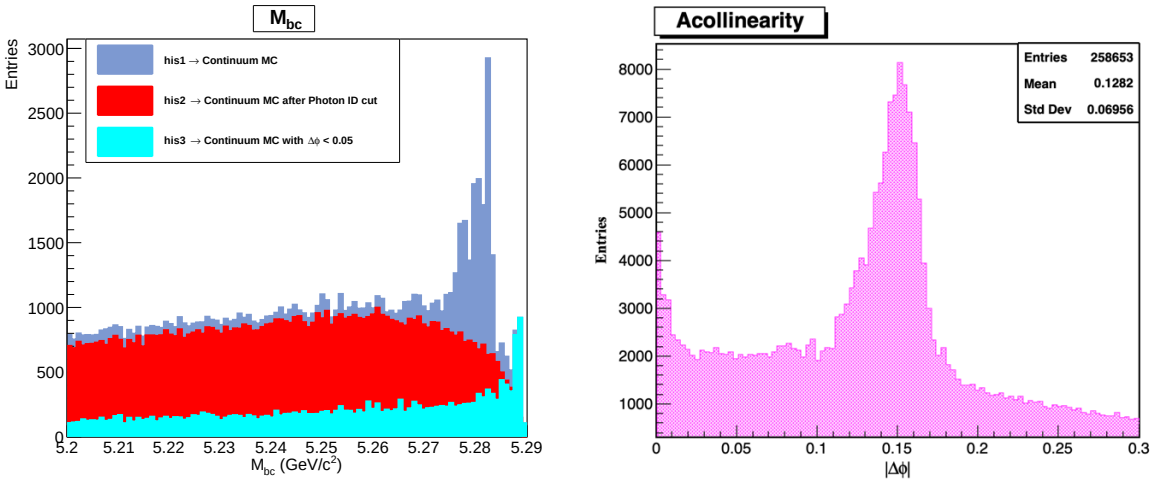


(a) (b)

Figure 3.16: (a) ECL timing of all the events in real data. (b) ECL timing after the application of selection cuts

Figure 3.17a shows the $M_{\rm bc}$ distributions before and after applying the MC truth information to the background MC. In Figure 3.17a, one can clearly distinguish two peaks. The first one, centered roughly at 5.28 GeV/c^2, is due to Bhabha events, while the second one, centered at the beam energy ($\approx$ 5.29 GeV/c^2), is due to $e^+e^- \to \gamma\gamma$ events. The two types of events are separated by looking at their ϕ acollinearity, defined as $\Delta\phi \equiv |\phi_1 - \phi_2| - \pi$, where ϕ_1 and ϕ_2 are the two photons (or misidentified electrons) azimuthal angles. In Figure 3.17b, $|\Delta\phi|$ is plotted for off-time events; a cut at $|\Delta\phi| < 0.05$ selects $e^+e^- \to \gamma\gamma$ events, as shown by the (blue) solid histogram.

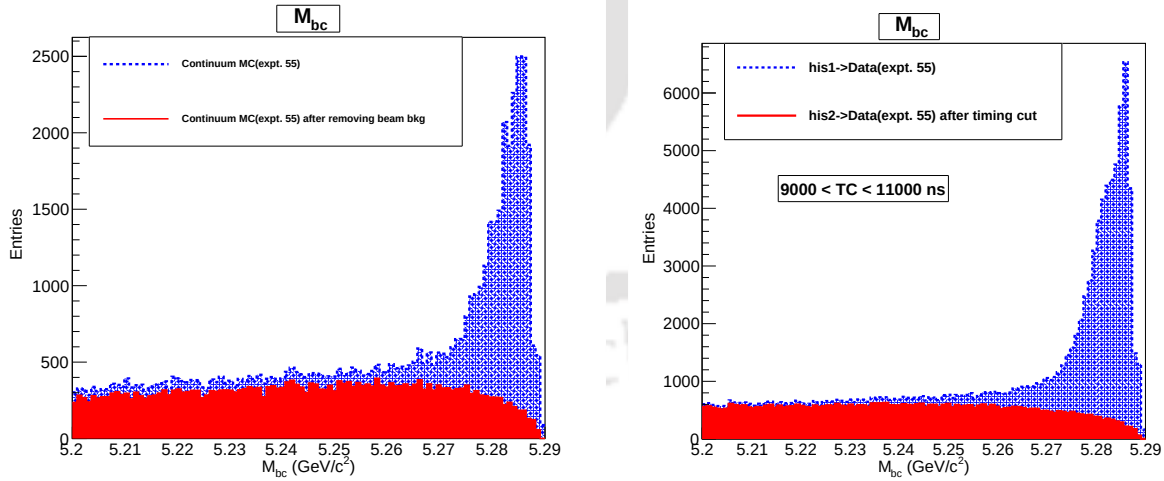
Figures 3.18 show the $M_{\rm bc}$ distributions before and after applying the MC truth



(a) (b)

Figure 3.17: (a) $M_{\rm bc}$ distribution of reconstructed B^0 before and after photon ID cut. The shaded histograms (blue) show the off-time events with $\Delta\phi < 0.05$, which correspond to $e^+e^- \to \gamma\gamma$ events.
(b) Distribution of acollinearity $|\Delta\phi|$ between the two photons for the off-time events. The peak at $|\Delta\phi| = 0$ is due to $e^+e^- \to \gamma\gamma$ events, while the one at $|\Delta\phi| \simeq 0.13$ is formed by mis-reconstructed Bhabha events.

information and timing criteria to background MC and real data.



(a) (b)

Figure 3.18: $M_{\rm bc}$ distributions before and after applying the (a) MC truth information to background MC and (b) timing criteria to real data.

Figures 3.19 shows the $M_{\rm bc}$ and ΔE distributions for the background MC.

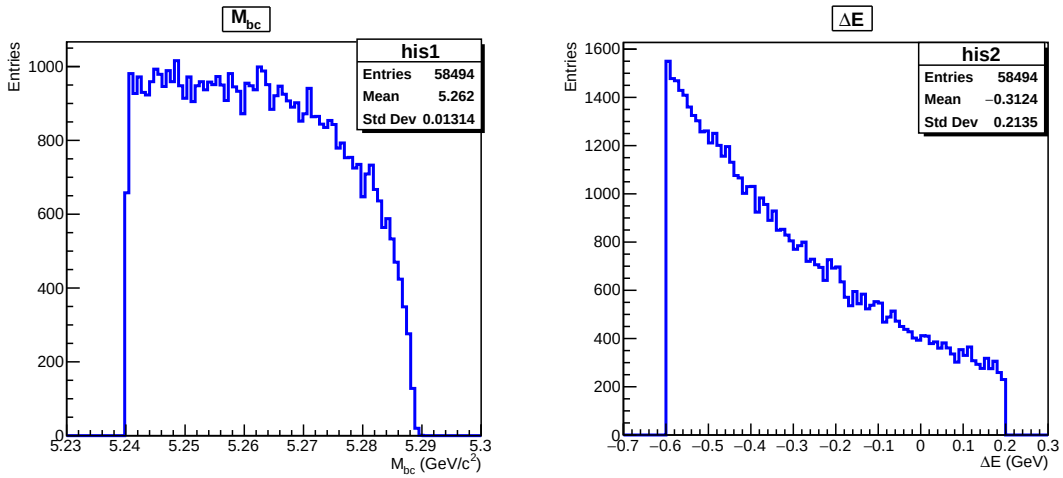


Figure 3.19: $M_{\rm bc}$ and ΔE distribution for udsc background for Belle study

The background contribution from charged and mixed processes is significantly less. Almost all the events were rejected after applying the candidate region cut. Now, after applying all the selection criteria, the normalized plots of ΔE and $M_{\rm bc}$ distribution for signal and background MC are shown in Figure 3.20.

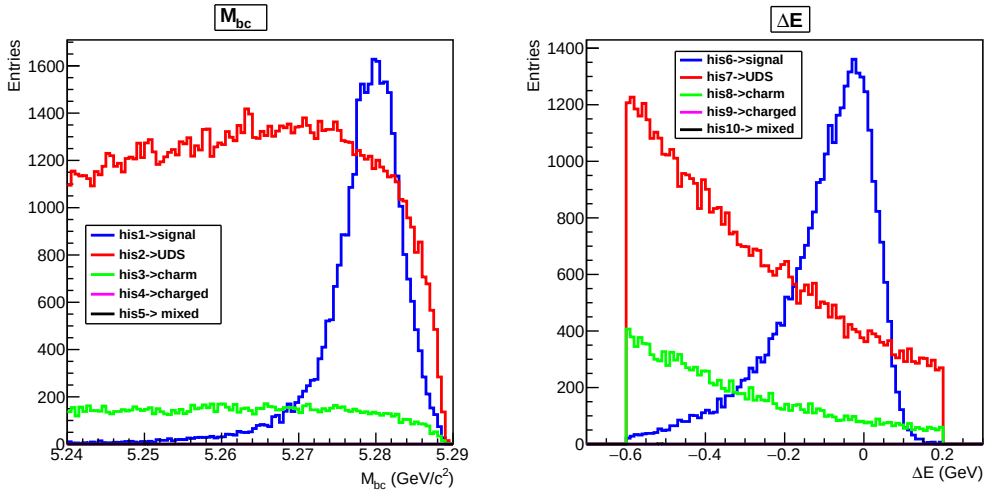


Figure 3.20: Normalized ΔE and $M_{\rm bc}$ distribution for signal and background MC for Belle study.

7.2 Backgrounds in Belle II analysis

Continuum and generic MC samples are processed, and candidate events are selected using the same selection criteria as discussed in Table 3.6 and 3.8 earlier. Figures

3.21 show the $M_{\rm bc}$ and ΔE distributions for the background MC. For the background events, we expect $M_{\rm bc}$ to have a continuous distribution of randomly combined final state particles and ΔE to peak towards the negative values.

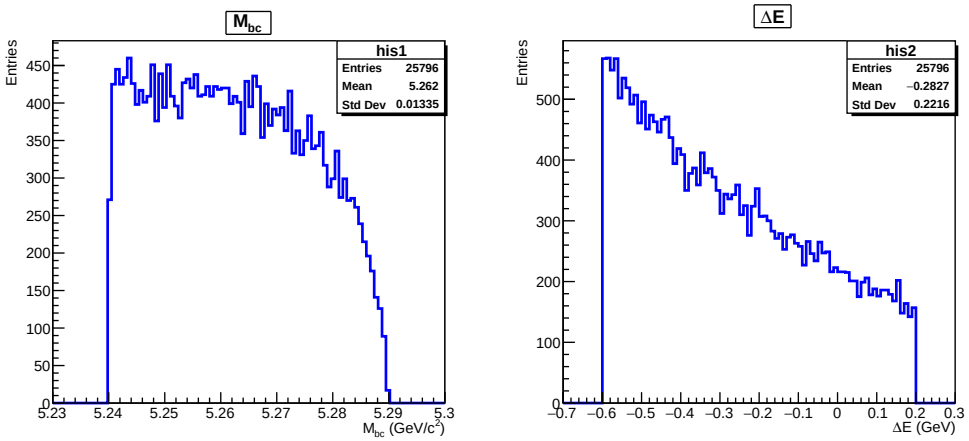


Figure 3.21: $M_{\rm bc}$ and ΔE distribution for udsc background MC for Belle II study.

The background contribution from charged and mixed processes is much less. Figures 3.22 and 3.23 show the $M_{\rm bc}$ and ΔE distributions for the charged and mixed background MC, respectively.

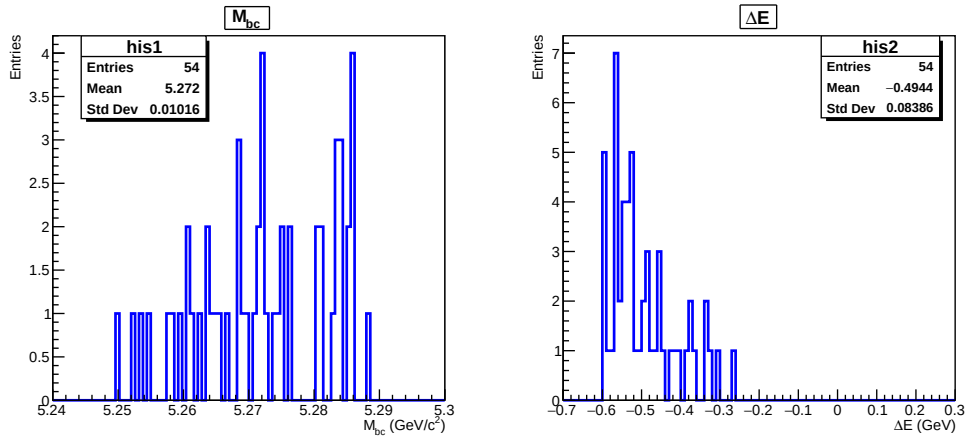


Figure 3.22: $M_{\rm bc}$ and ΔE distribution for Charged Background MC for Belle II study.

Now, after applying all the selection criteria, the normalized plots of ΔE and $M_{\rm bc}$ distribution for signal and background MC are shown in Figure 3.24.

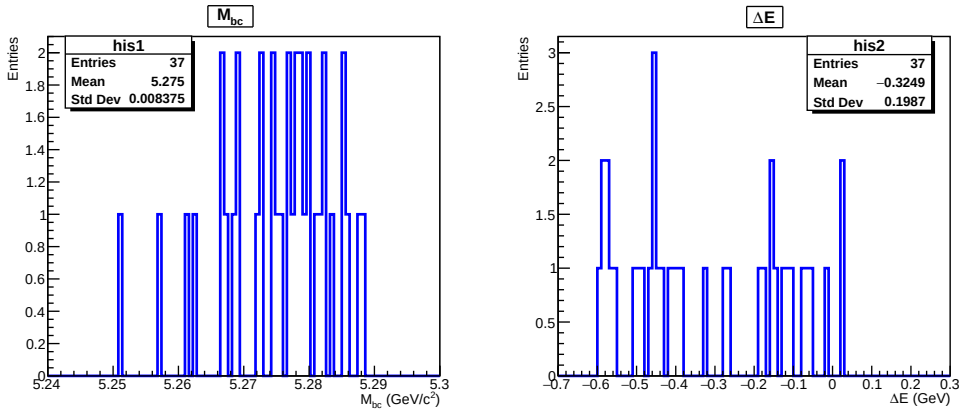


Figure 3.23: $M_{\rm bc}$ and ΔE distribution for Mixed Background MC for Belle II study.

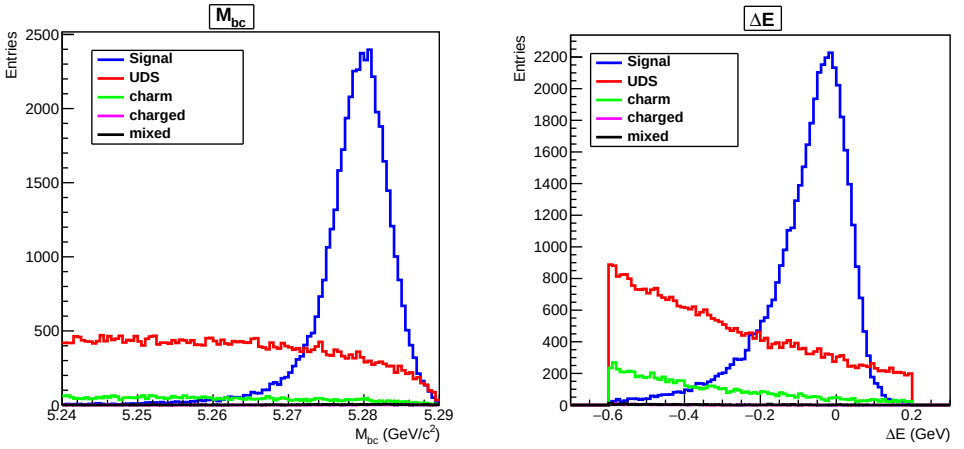


Figure 3.24: Normalized $M_{\rm bc}$ and ΔE distribution for signal and background MC for Belle II study.

Also, we have thoroughly examined the low multiplicity gamma-gamma and τ pairs contributions and evaluated the impact of selection cuts in rejecting these events. We observed minimal contribution from these processes once all the selection cuts are applied. We processed 400 fb^{-1} equivalent of $e^+e^- \to \gamma\gamma$ samples, 90 fb^{-1} of $e^+e^- \to e^+e^-$ and 1 ab^{-1} of $e^+e^- \to \tau\tau$ samples, which resulted in the rejection of 99%, 100% and 100% of these processes, respectively, confirming the effectiveness of the selection cuts. In the upcoming sections, we shall describe the strategies deployed to suppress the aforementioned backgrounds.

7.3 Continuum Suppression Using Multivariate Analysis

Continuum background events can be suppressed over signal B events using their production mechanism at $\Upsilon(4S)$ resonance. In a $B\bar{B}$ event, both B mesons are produced almost at rest in $\Upsilon(4S)$ frame, as $\Upsilon(4S)$ mass is barely above the $B\overline{B}$ production threshold. As a result, the B decay products are distributed isotropically in the $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ rest frame. Since spin-0 $B\bar{B}$ comes from spin-1 $\Upsilon(4S)$, the angular distribution is $\sin^2\theta_B$, where θ_B is the angle between B meson and beam axis. On the other hand, for continuum events, the light quarks are produced with large initial momentum, resulting in a back-to-back fragmentation into two jets of light hadrons. Spin-$\frac{1}{2}$ fermions resembling jet-like $q\bar{q}$ pairs are produced from e^+e^- collisions, with their distribution following $1+\cos^2\theta_B$, as illustrated in Figure 3.25.

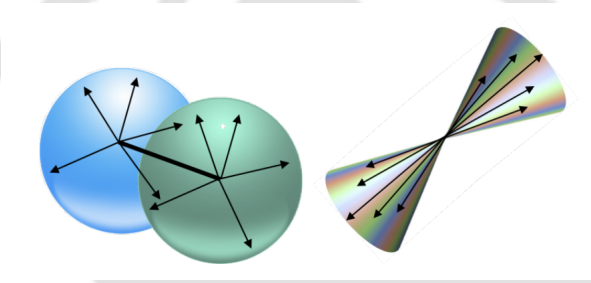


Figure 3.25: Event shapes for $B\overline{B}$(Left) and $q\overline{q}$(Right) events

Continuum events can thus be distinguished from B decays by their event topology. We train a single multivariate analyzer [83] to suppress background coming from continuum and $B\bar{B}$ events. We use FastBDT [84], the default method from the MVA package of the BASF2 framework, as the MVA method. This Multivariate Analysis can be much more powerful than applying the individual cut on variable separately and can be a much more efficient way to keep signal events while reducing continuum backgrounds. In general, the task of a multivariate method is to map an n-dimensional input variable space into a single scalar that contains all the information, including correlations between the input variables. The single variable can be used separately as a discriminant instead of all input variables. To use a multivariate classifier, it has to be trained first. For such training, one needs a large set of data for which truth is known. Using a dataset of signal and background events where the truth for every entry is known, the parameters of the classifier are determined in such a way that the classifier output is as close to the known truth as possible. The performance of a classifier is characterized by its efficiency E and purity P, defined as:

$$\mathrm{E} = \frac{N_{\mathrm{selected}}}{N_{\mathrm{S}}}$$

$$\mathrm{P} = \frac{N_{\mathrm{selected}}}{N_{\mathrm{total}}}$$

where N_{selected} is the number of selected signal events, N_{S} is the total number of signal events, and N_{total} is the total number of selected signal and background events.

Any classifier should ideally have high efficiency as well as high purity. To suppress continuum background, we add the following event-shape variables to the Multivariate Analysis(MVA). Those variables are loosely or uncorrelated with $M_{\rm bc}$ and ΔE and kinematic-fit quality.

- **Kakuno-Super-Fox-Wolfram(KSFW) moments**: We use modified Fox-Wolfram [85] (KSFW) moments H_{xl}^{so} and H_{l}^{oo}. These moments are designed to capture the principal geometric features of the spatial distributions of the final state particles. The particles are divided into two categories: reconstructed B candidate daughters denoted as 's', and those from the rest-of-event (ROE) denoted as 'o'. H_{xl}^{so} is further decomposed into three categories: charged ($x = c$), neutral ($x = n$), and missing ($x = m$), where the missing momentum is treated as a particle. The KSFW moments are then defined using the relation:

$$H_{xl}^{so} = \sum_{i} \sum_{j,x} |\overline{P_{jx}}| P_l(\cos\theta_{i,jx})$$

where i runs over the B daughters; j, x runs over the ROE in the category x; $|\overline{P_{jx}}|$ is the three-momentum of particle jx; and $P_l(\cos\theta_{i,jx})$ is the $l^{\rm th}$ order Legendre polynomial of the cosine of the angle between particles i and jx.

$$H_{l}^{oo} = \sum_{j} \sum_{k} |\overline{P_j}||\overline{P_k}| P_l(\cos\theta_{j,k}) \ (l = even)$$

$$H_{l}^{oo} = \sum_{j} \sum_{k} Q_j Q_k |\overline{P_j}||\overline{P_k}| P_l(\cos\theta_{j,k}) \ (l = odd)$$

Here j, k run over the ROE and Q_j, Q_k are charges of $j^{\rm th}$ and $k^{\rm th}$ particles, respectively. Thus, we have a total of 16 modified moments. In addition to these 16 moments, we have,

$\sum_{n}^{N_{trk}} p_T$ *or* et : *the sum of transverse momentum of all visible particles*

and '*the apparent mass that is missing from the event*' defined as:

$$mm^2 = (E_{\Upsilon(4S)} - \sum_{n}^{N_{trk}} E_n)^2 - \sum_{n}^{N_{trk}} |p_n|^2,$$

where $E_{\Upsilon(4S)}$ is the energy of $\Upsilon(4S)$ corresponding to CM energy and E_n and p_n are the energy and momentum of n, respectively. The final state particles that 'n' iterates over are generally the $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$, $e^{\pm}$, $\mu^{\pm}$ and γ. We use in total 18 KSFW moments as follows: `hso00, hso02, hso01, hso03, hso04, hso10, hso12, hso14, hso20, hso22, hso24, hoo0, hoo1, hoo2, hoo3, hoo4, et and mm`^{2}.

- `cosTBTO`: Cosine of the angle between the thrust axis of the signal B candidate and the thrust axis of the rest of the event. The thrust axis is defined as the direction which maximizes the sum of the longitudinal momenta of the particles in the decay. In $B\bar{B}$ events, the particles are almost at rest, so the thrust vectors are uniformly distributed. Therefore, `cosTBTO` will also be uniformly distributed between 0 and 1. In $q\bar{q}$ events, the particles are collimated, and the thrust axes point back-to-back, leading to a peak at high values of `cosTBTO`.
- $\cos\theta_B$: cosine of the angle θ_B between the reconstructed momentum of the B candidate in the $\Upsilon(4S)$ reference frame and the beam axis.
- `cosTBz`: cosine of the angle between the thrust axis of B-meson and z-axis.
- `qrOutput(FBDT)`: The flavour tagger output, which determines the flavour of the tag side B meson. Here `q` = +1(-1) is defined as b-flavor of tag side B^0 ($\bar{B}^0$). The dilution factor for flavor determination, denoted as r, ranges from 0 to 1, with r = 0 indicating no information available to determine the tag side b-flavor, and r = 1 indicating a perfectly tagged b-flavor. In events involving $B\bar{B}$, a pronounced peak at 1 is observed when tag side B information is available, whereas events involving $q\bar{q}$ tend to peak at 0.
- nTracks: The total number of tracks in the event.

The performance of classifier was first checked with 23 input variables(18 KSFW moments, `cosTBTO`, $\cos\theta_B$, `cosTBz`, `qrOutput(FBDT)` and nTracks). We removed the variables with zero weight in training, hence the least contribution towards the separation of background and signal. KSFW(`hso01`) and KSFW(`hso03`) are identically zero for our case. A total of 21(16 KSFW + 3 cosine angle + `qrOutput(FBDT)` + nTracks) variables are then fed into the MVA for discrimination. Figures 3.26 show three of them, and as expected, they have discriminating power to separate the signal and background-like events. Refer to `Appendix` (C) for the rest of the plots of event shape variables. The importance of MVA input variables is tabulated in `Appendix` (D).

7.3.1 Validation and Overtraining Check

To create the training and testing samples, we take signal and continuum background MC samples after applying all the selection criteria listed in Table 3.7 and 3.8 discussed in Sec. 5.1 and 6.0, respectively. The FBDT is trained to classify signal from the background using 50% of sample. The performance is tested on the remaining 50% of the sample. At the end of the training, we can print out analysis information such as linear correlation matrix of the input variables, evaluation plots, purity vs. efficiency plots, ROC plots, and overtraining plot for the training sample. The correlations between input variables in the signal and background are shown in Figure 3.27. We check the correlation between the input training variables in a sample set and weed out the variables having a high degree of correlation amongst themselves.

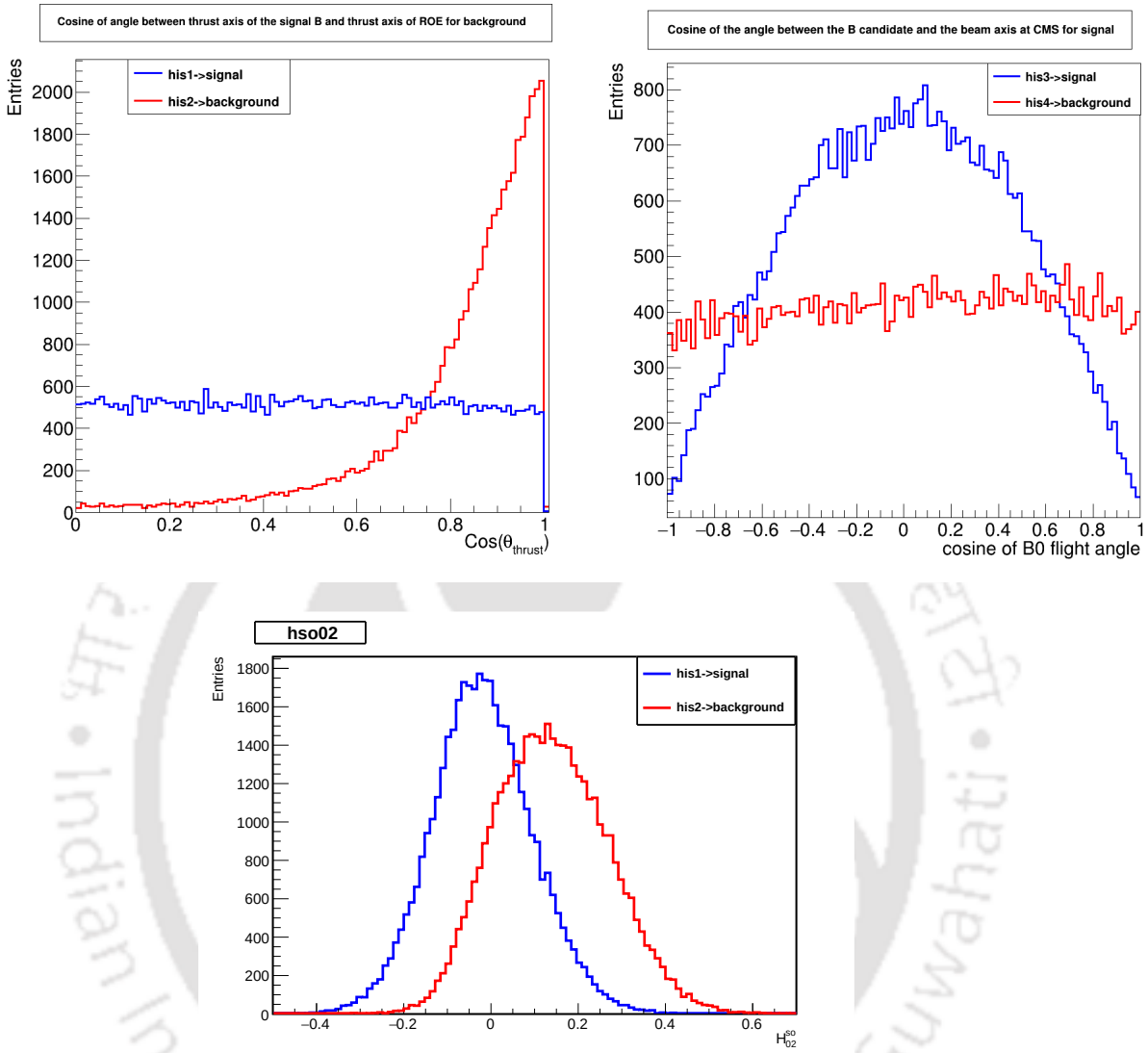
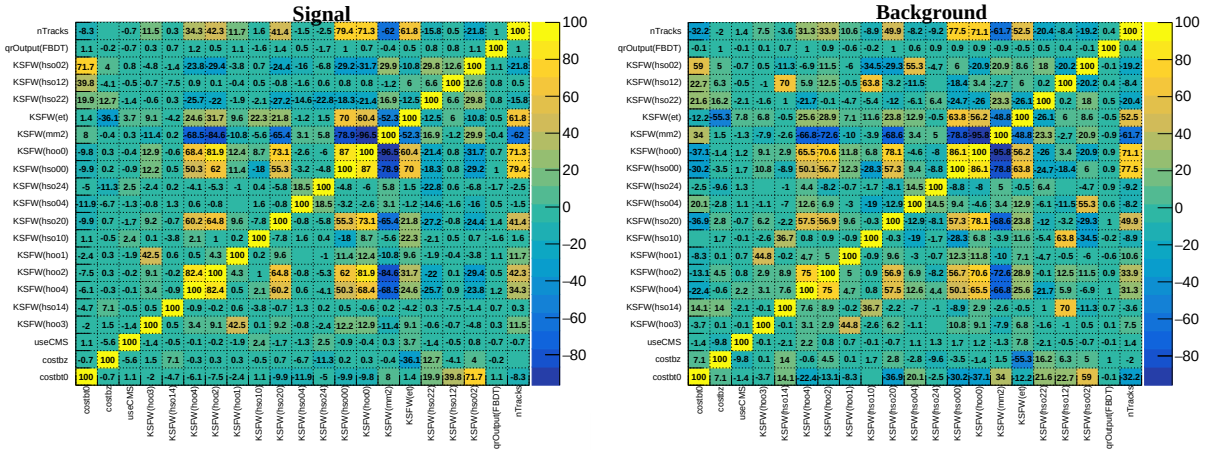


Figure 3.26: Continuum suppression variables for signal and background

As a check for possible overfitting, Figure 3.28 shows the Receiver-Over-Operator (ROC) curve of the classifiers' outputs for the training samples. Since the curves for the training and validation data looks very similar and have less gap, it can be concluded that the BDTs generalize well to unobserved data. Also, for a well-trained network, the purity as a function of Classifier output should be on the diagonal shown in Figure 3.29.

To ensure that the MVA output distributions of the training and test samples exhibit similar shapes, we conducted a Kolmogorov-Smirnov test [86] between the distributions. In Figure 3.30, we present a comparison of the MVA output distributions between the training and test samples. Additionally, the figures display the corresponding Kolmogorov-Smirnov probabilities. If the distributions follow different shapes, these probabilities would be less than 0.05. However, in our case, all



(a) (b)

Figure 3.27: (a) Correlation plot for the (a) signal and (b) background for Belle.

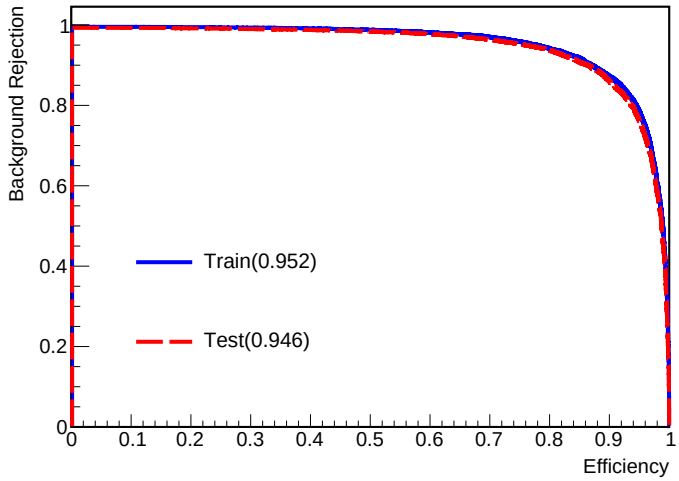


Figure 3.28: ROC curves of the continuum suppression classifier for the datasets of decay mode. The number in parentheses corresponds to the area under the curve (AUC) of the curve. The closer it is to one, the better the performance of the classifier.

probabilities (both for signal and background distributions) are greater than 0.05, indicating no bias between the training and test samples. We find acceptable consistency between the training and validation samples, which signifies no significant overtraining.

7.4 Optimizing the BDT Output selection

The trained classifier, as discussed in the previous section, is applied to the remaining dataset statistically independent from training. The value of the classifier output (BDT output) is recorded for each event. Rather than using a figure of merit

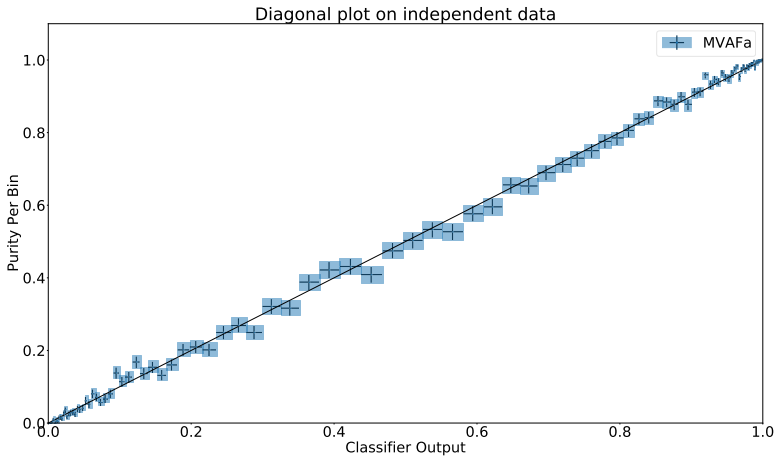


Figure 3.29: Purity vs. Classifier output for Belle.

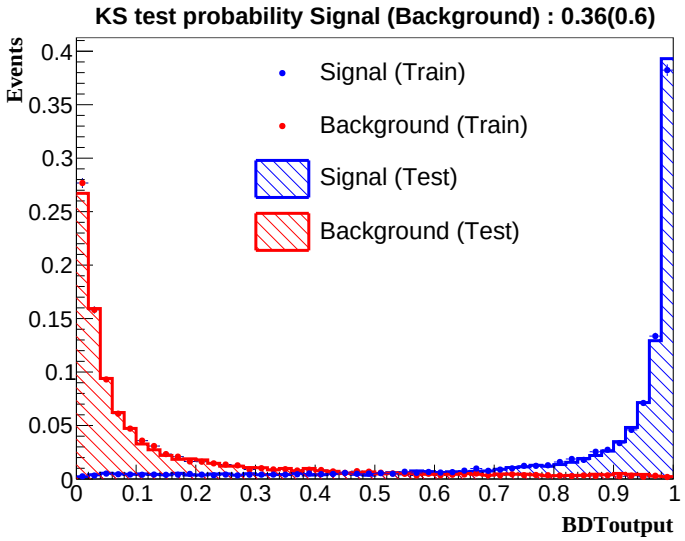


Figure 3.30: Superimposed plot of FastBDT output for training and validation samples

to determine the optimal BDT Output selection, we perform ToyMC studies for different selections varying from 0.40 to 1.00 and select the cut that minimises the signal yield error. The ToyMC contains the signal, continuum and $B\bar{B}$ components with expected yields in 694 fb^{-1} depending on BDT output selection cut as shown in Figure 3.31. We find that a selection of 0.55 with a lower signal yield error, rejects 93% of the continuum and retains 86% of the signal events. We obtain a signal MC purity of approximately 93% with this choice of BDT selection criterion. We applied a selection cut on the BDT output. Then we used the transformed BDT Output ($C'_{\rm BDT}$) as a third variable to perform the unbinned extended maximum likelihood fit to extract the signal yield for $B^0 \to \gamma\gamma$ channel.

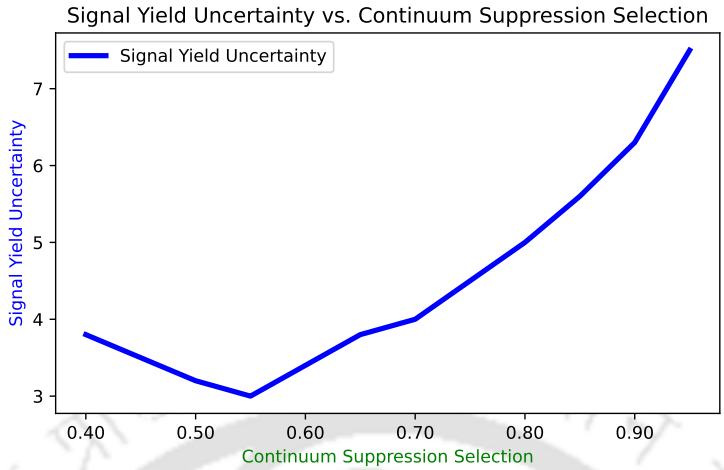


Figure 3.31: The mean signal yield uncertainty (blue) from ToyMCs as a function of continuum suppression output for Belle study.

7.4.1 Transformation of BDT Output

The output obtained from the FastBDT classifier (C_{BDT}) tends to peak at 1 for signal-like events and 0 for background-like events, as shown in Figure 3.30 in subsection 7.3.1. It is difficult to model the FastBDT output distribution with a simple analytic function. Thus to improve the modeling, we apply the cut on the BDT output(C_{BDT})$>$ 0.55 and then transform it using the $\mu - transformation$ (probability integral transformation) [87], where the signal shape is flat and the background shape is exponential, as shown in Figure 3.32.

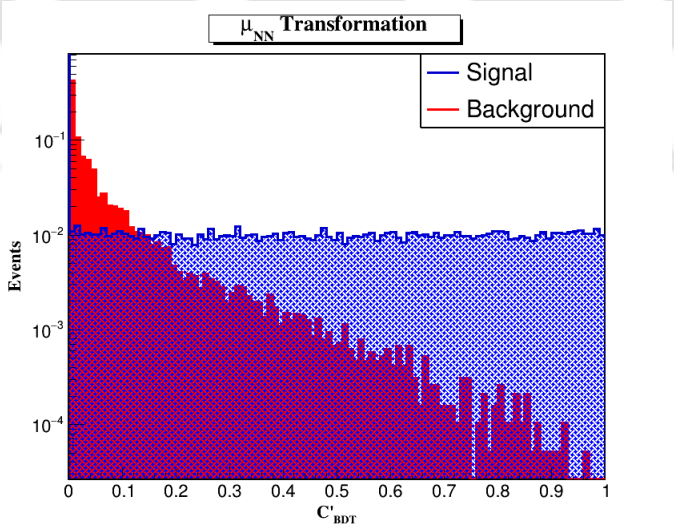


Figure 3.32: Transformed FBDT output for signal and background in log scale for Belle study.

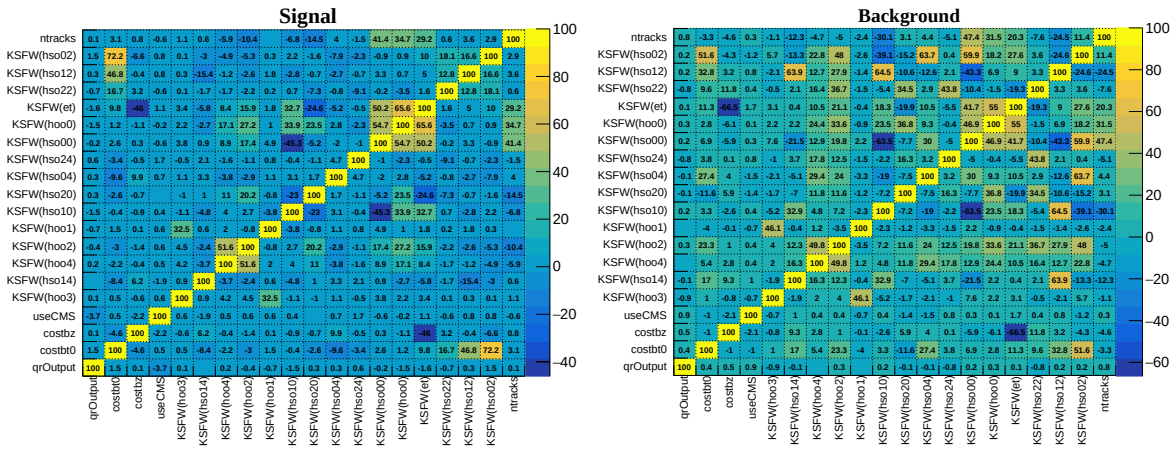
The $\mu - transformation$ $\mathcal{R}(y)$ of a classifier y is given by the integral [88]

$$\mathcal{R}(y) = \int_{-\infty}^{y} \hat{y_B}(y')\, dy',$$

where $\hat{y_B}$ is the original signal distribution normalized to unity and it is defined such that $\mathcal{R}_{(y_B)}$ for signal events is uniformly distributed between 0 and 1, while background events cluster towards 0. There are several advantages of using this transformation compared with the usual Gaussian transformation; the signal and background shapes are easily distinguishable and can be described by the analytic PDFs with less parameters in the fit. In comparison, the Gaussian transformation can be difficult to describe and depends on the maximum value of the output.

7.5 Continnum Suppression for Belle II analysis

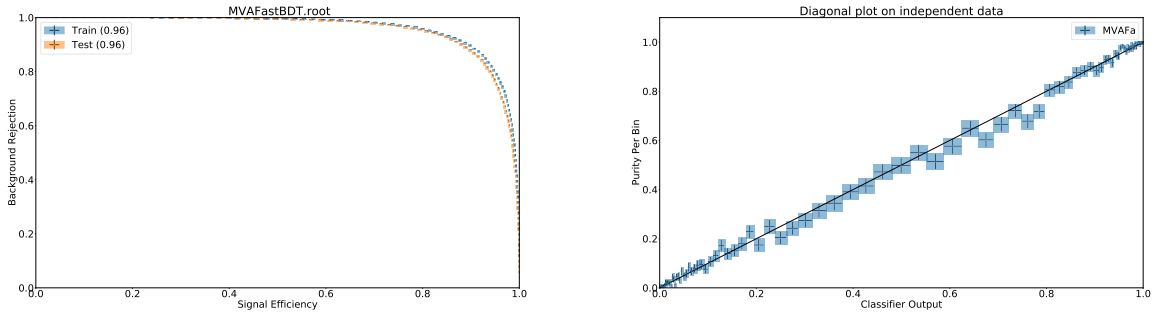
Continuum suppression was performed using the same variables in Sec 7.3. For training and testing the MVA, we used two independent datasets with an equal number of correctly reconstructed signals and misrecontructed continuum background events. The performance of the classifier was first checked with 23 input variables (18 KSFW moments, $\cos_{\theta_B}$, `cosTBTO`, `cosTBz`, nTracks and `qrOutput(FBDT)`). The correlations between input training variables in signal and background are shown in Figure 3.33. Refer to `Appendix` (C) for the plots of event shape variables. The importance of MVA input variables is tabulated in `Appendix` (D).



(a) (b)

Figure 3.33: The correlation matrix of input variables for Belle II.

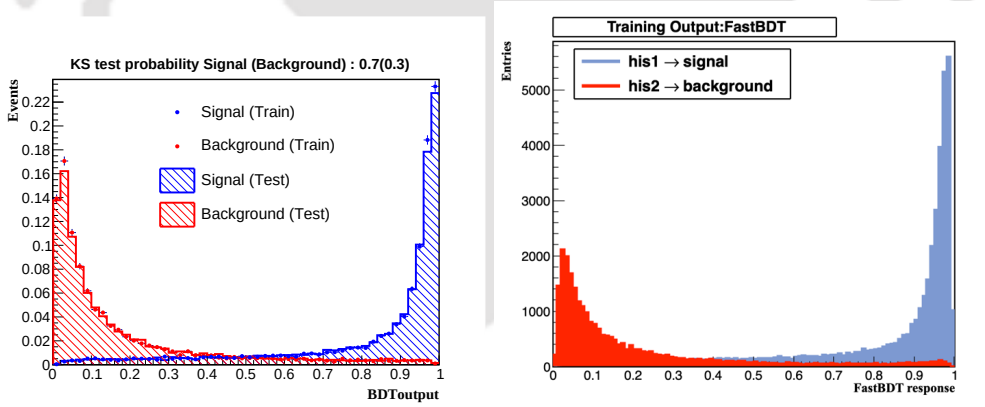
As a check for possible overfitting, Figure 3.34a shows the Receiver-Over-Operator (ROC) curve of the classifiers' outputs for the training samples. Also, for a well-trained network, the purity as a function of classifier output should be on the diagonal, as shown in Figure 3.34b.



(a) (b)

Figure 3.34: (a) ROC curves of the continuum suppression classifier. (b) Purity vs. Classifier output

To ensure that the MVA output distributions of the training and test samples exhibit similar shapes, we conducted a Kolmogorov-Smirnov test between the distributions. We find acceptable consistency between the training and validation samples, as shown in Figures 3.35, which signifies no significant overtraining.



(a) (b)

Figure 3.35: (a) Superimposed plot of FastBDT output for training and validation samples and (b) Distribution of FastBDT output on simulated validation sample.

A large continuum background is rejected by applying a cut on the FastBDT output. The selection could be chosen from the ToyMCs studies defined in Sec 7.4. The ToyMC contains the signal, continuum and $B\bar{B}$ components with expected yields in 362 fb^{-1} depending on the BDT Output selection as shown in Figure 3.36. The different selections are varied from 0.40 to 0.87, and we select the cut that minimises the signal yield error. We found the cut at 0.45 with lower signal yield error rejects 87% of the background and retains 89% of signal events. We obtain a signal MC purity of approximately 92% with this choice of BDT selection criterion. We would take Transformed Network Output(C'_{BDT}) as a third variable

to perform an unbinned extended maximum likelihood fit to extract the signal yield for $B^0 \to \gamma\gamma$ channel.

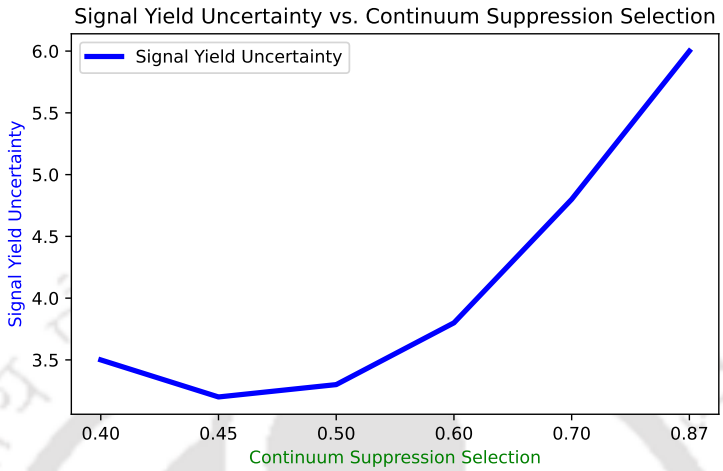


Figure 3.36: The mean signal yield uncertainty (blue) from ToyMCs as a function of continuum suppression output

Figure 3.37 shows the C'_{BDT} distributions for signal and background MC for Belle II.

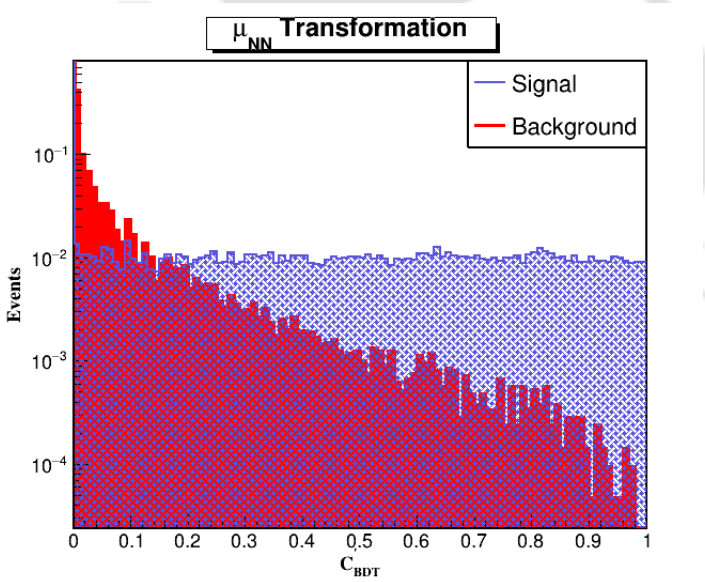


Figure 3.37: Transformed FBDT output for signal and background in log scale.

7.6 Peaking Background from Belle Study

Decay modes similar to $B^0 \to \gamma\gamma$ may arise from rare B (i.e., mixed and charged) decay backgrounds and can get misidentified as a signal candidate. This background

peaks at similar values of $M_{\rm bc}$ but has ΔE shifted to the negative values due to energy losses from the lost particle. We investigate rare backgrounds employing a dataset that is 50 times larger than the real data. To know the sources of peaking backgrounds, we have used `TopoAna tool` [89] to print all the decay trees and some specific information in pdf/tex/root file. `TopoAna` is a generic tool for the topology analysis of the inclusive MC samples. We can quickly get the decay trees or tables of all the MC sample events. Applying the identical selection criteria discussed earlier, we find that a few events still pass these criteria, as outlined in Table 3.9. Figures 3.38 and 3.39 show the $M_{\rm bc}$ and ΔE distribution of charged and mixed rare backgrounds after applying all the selection cuts mentioned in Table 3.9.

Variable	Requirement	charged Rare MC	mixed Rare MC
Precut + HardronB cut + Photon ID	$1.4 < E_\gamma(\mathrm{CM}) < 3.4$ GeV	12587	10909
$E_{9/21}$	> 0.95	10929	9422
$M_{\rm bc}$	$5.24 < M_{\rm bc} < 5.29$ GeV/c^2	8192	7383
ΔE	$-0.6 < \Delta E < 0.2$ GeV	975	1364
$\theta(\gamma)$ Barrel region	$clusterReg == 2$	786	1093
FastBDT output	> 0.55	331	520
$P_{\pi^0\gamma}$	< 0.50	45	53
$P_{\eta\gamma}$	< 0.65	21	30

Table 3.9: Table for cut-flow for charged and mixed Rare MC samples. The table is not scaled to integrated luminosity.

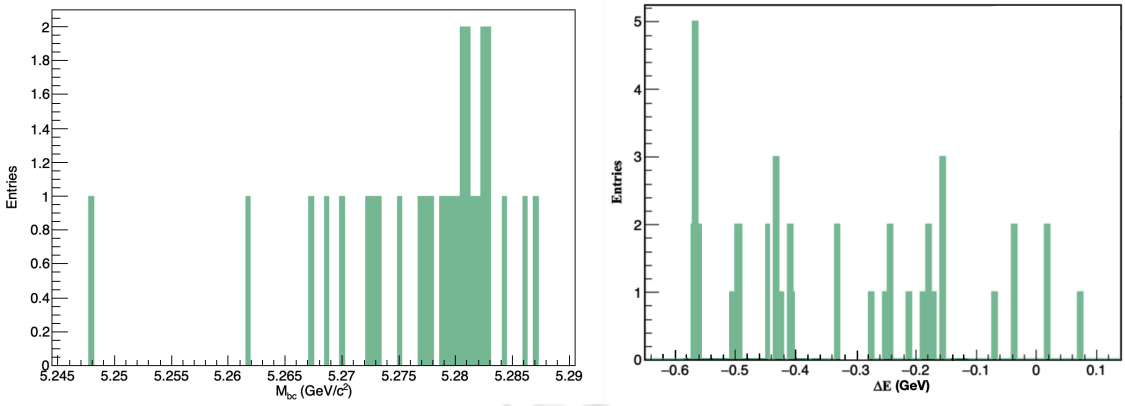


Figure 3.38: $M_{\rm bc}$ (left) and ΔE (right) distributions for mixed rare background.

The selection criteria outlined in Table 3.9, along with the signal region cuts, are applied to rare MC files to investigate potential exclusive backgrounds. We found only three channels give a contribution after the final selection, $B^0 \to \pi^0\pi^0$, $B^0 \to \eta\eta$ and $B^0 \to \eta\pi^0$ as shown in Figure 3.40. For these three channels, large (1M) samples are explicitly generated and simulated to get a more precise estimate of their effect. Assuming the measured $B^0 \to \pi^0\pi^0$ branching fraction [90] and the existing limit on the $B^0 \to \eta\pi^0$ [91] and $B^0 \to \eta\eta$ [92] branching fraction, an estimate of the expected level of these backgrounds can be obtained. Based on it, in

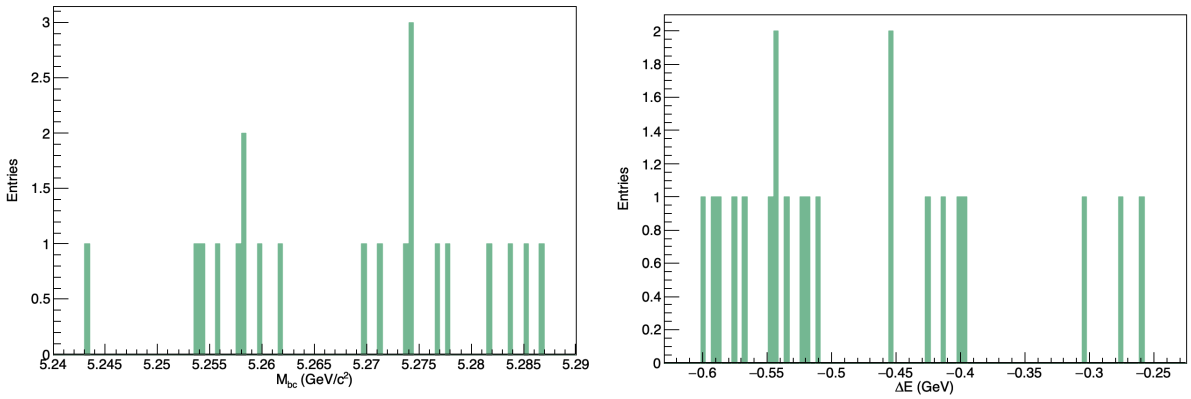


Figure 3.39: $M_{\rm bc}$ (left) and ΔE (right) distributions for charged rare background.

the signal region 0.049 events for $B^0 \to \pi^0\pi^0$, 0.007 events for $B^0 \to \eta\eta$ and 0.007 events for $B^0 \to \eta\pi^0$ are expected. The expected number of events at 694 fb^{-1} are obtained using the following formula,

$$N_{exp} = \mathcal{B} \times N_{B\bar{B}} \times 2 \times f^{00} \times \mathcal{E}$$

where $\mathcal{B}$ is the branching fraction, the total number of $B\bar{B}$, $N_{B\bar{B}} = 753 \times 10^6$, $f^{00} = (48.4 \pm 0.6)\%$ indicates the ratio of $\Upsilon(4S)$ decaying into $B^0\bar{B}^0$ and $\mathcal{E}$ is the efficiency of events which have survived the selection criteria. The results are summarised in Table 3.10.

$B\overline{B}$ **decay**	**Branching Ratio**	**Efficiency**	**Expected events (694 fb^{-1})**
$B^0 \to \pi^0\pi^0$	$1.83 \pm 0.21(\text{stat.}) \pm 0.13(\text{syst.})] \times 10^{-6}$ [90]	$(0.004 \pm 0.003)\%$	0.049
$B^0 \to \eta\pi^0$	$4.1 \pm 1.7 \times 10^{-7}$ [91]	$(0.003 \pm 0.003)\%$	0.007
$B^0 \to \eta\eta$	$< 1.0 \times 10^{-6}$ at 90% CL[92]	$(0.001 \pm 0.002)\%$	0.007

Table 3.10: Estimation of number of events in 694 fb^{-1} for $B^0 \to \pi^0\pi^0$, $B^0 \to \eta\pi^0$ and $B^0 \to \eta\eta$ from efficiency determined from MC.

As the expected number of $B^0 \to \pi^0\pi^0$, $B^0 \to \eta\pi^0$ and $B^0 \to \eta\eta$ background events is 0.049, 0.007 and 0.007, which is approximately 50 times, 342 times and 342 times less than the expected number of signal (2.4) by SM in Belle, we do not expect these backgrounds processes to contribute significantly.

7.7 Peaking Background from Belle II Study

We investigate the rare backgrounds, comprising both charged and mixed MC backgrounds. After applying all the selection cuts, we have only two events from the mixed MC samples and zero events from the charged MC samples in the Belle II study. The two B decay modes were identified as potential background sources. These peaking backgrounds are coming mainly from decay of $B^0 \to \pi^0\pi^0$ and $B^0 \to \eta\pi^0$. We have also checked the $B^0 \to \eta\eta$ decay mode as it could also be

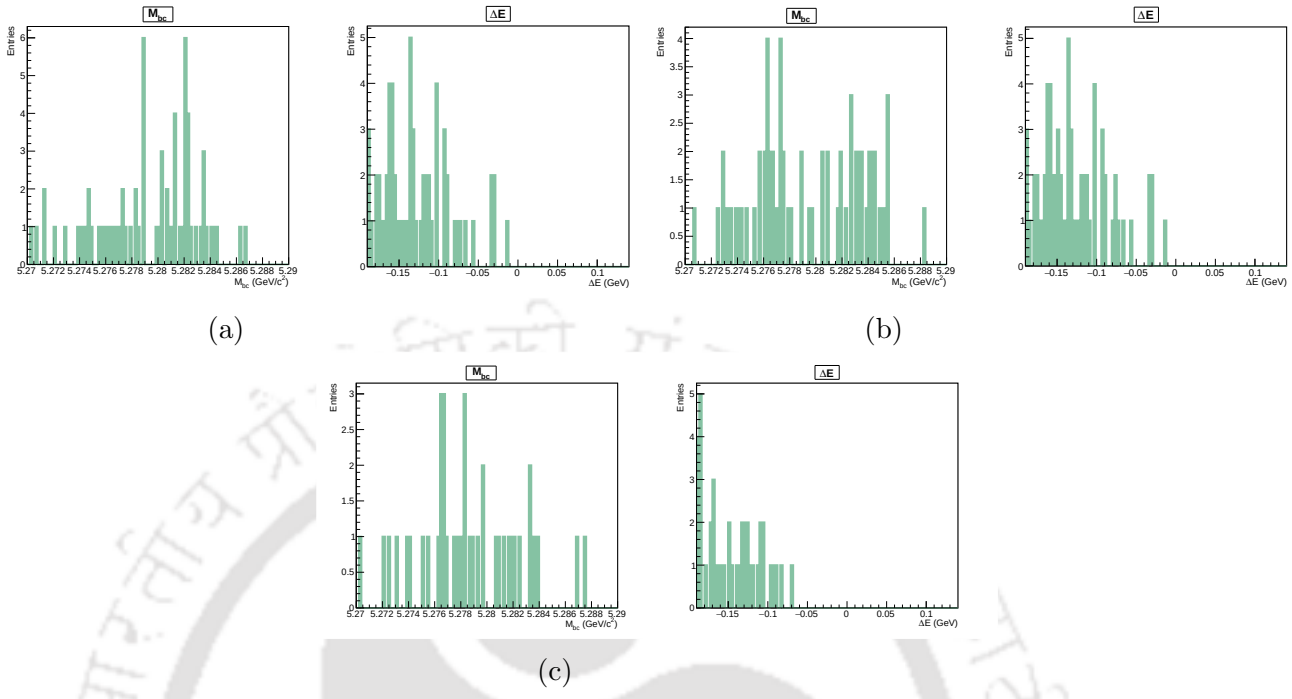


Figure 3.40: $M_{\rm bc}$ and ΔE distribution for (a) $B^0 \to \pi^0\eta$ channel (b) $B^0 \to \pi^0\pi^0$ and (c) $B^0 \to \eta\eta$ channel for Belle analysis

a source of peaking background. We perform a study on generic mixed samples to check these peaking contributions and generate 1M MC events of these peaking backgrounds. The optimized event selection is applied to each of these samples is shown in Figure 3.41, and the estimated number of background events expected in on-resonance data is determined from the latest branching fractions. After scaling the yields of these modes to the luminosity of the on-resonance data, it is estimated that they contribute a 0.021 for $B^0 \to \pi^0\pi^0$, 0.003 events for $B^0 \to \eta\eta$ and 0.004 events for $B^0 \to \eta\pi^0$ in the signal region. The results are summarised in Table 3.11.

$B\overline{B}$ **decay**	**Branching Ratio**	**Efficiency**	**Expected events (362 fb^{-1})**
$B^0 \to \pi^0\pi^0$	$1.83 \pm 0.21(\text{stat.}) \pm 0.13(\text{syst.})] \times 10^{-6}$ [90]	$(0.003 \pm 0.004)\%$	0.021
$B^0 \to \eta\pi^0$	$4.1 \pm 1.7 \times 10^{-7}$ [91]	$(0.002 \pm 0.003)\%$	0.003
$B^0 \to \eta\eta$	$< 1.0 \times 10^{-6}$ [92]	$(0.001 \pm 0.002)\%$	0.004

Table 3.11: Estimation of number of events in 362 fb^{-1} for $B^0 \to \pi^0\pi^0$, $B^0 \to \eta\pi^0$ and $B^0 \to \eta\eta$ from efficiency determined from MC.

As the expected number of $B^0 \to \pi^0\pi^0$, $B^0 \to \eta\eta$ and $B^0 \to \eta\pi^0$ background events is 0.021, 0.004 and 0.003, which is approximately 80 times, 425 times and 533 times less than the expected number of signal (1.6) by the SM, we do not expect these background processes to contribute significantly.

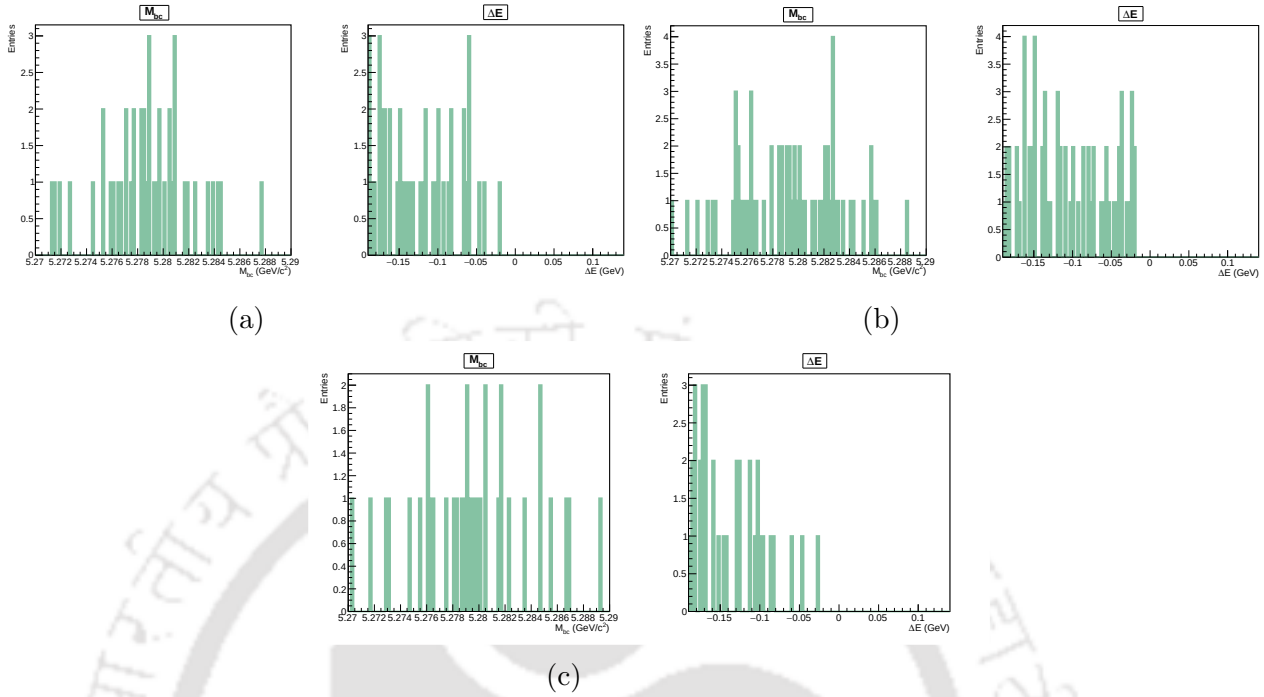


Figure 3.41: $M_{\rm bc}$ and ΔE distribution for (a) $B^0 \to \pi^0\eta$ channel (b) $B^0 \to \pi^0\pi^0$ and (c) $B^0 \to \eta\eta$ channel for Belle analysis for Belle II

8.0 Selection Summary

8.1 Selection Summary for Belle Study

The cut flow table for the signal and backgrounds at each selection step are shown in Table 3.12 for the Belle study.

Variable	Requirement	Signal Events	uds	charm	charged	mixed
Precut + HardronB cut + Photon ID	$1.4 < E_\gamma(\mathrm{CM}) < 3.4$ GeV	47912	1423000	196067	305	303
$E_{9/25}$	> 0.95	43443(91%)	1146513(25%)	154766(25%)	244(20%)	249(18%)
$M_{\rm bc}$	$5.24 < M_{\rm bc} < 5.29$ GeV/c^2	42929(99%)	640027(45%)	80511(48%)	206(16%)	210(16%)
ΔE	$-0.6 < \Delta E < 0.2$ GeV	42198(98%)	220817(66%)	24391(70%)	0(0%)	2(99%)
$\theta(\gamma)$ Barrel region	$clusterReg == 2$	33494(79%)	149012(33%)	17397(29%)	0	1(50%)
FastBDT output	> 0.55	28937(86%)	12901(93%)	1860(90%)	0	1(0%)
$P_{\pi^0\gamma}$	< 0.50	24599(85%)	2957(78%)	530(73%)	0	0(100%)
$P_{\eta\gamma}$	< 0.65	23308(95%)	2812(5%)	507(5%)	0	0

Table 3.12: The figures in the parenthesis indicates the survival(blue) and loss(red) percentage in signal and background MC samples, respectively. The table is not scaled to 694 fb^{-1}

Assuming $\mathcal{B}(B^0 \to \gamma\gamma) = 1.4 \times 10^{-8}$ and overall signal efficiency 23.3%, we can expect about 2.4 signal and 553 total background events in 694 fb^{-1} of data in the candidate region.

8.2 Selection Summary for Belle II Study

The cut flow table for the signal and backgrounds at each selection step are shown in Table 3.13 for the Belle II study.

Variable	Requirement	Signal Events	uds	charm	charged	mixed
Precut+skim	$1.4 < E_\gamma(\mathrm{CM}) < 3.4$ GeV	72016	1173443	128608	1597	1667
|clustertiming| / |clusterTiming/clusterErrorTiming|	< 200 ns < 2	68677(95.3%)	1094631(7%)	119969(7%)	1543(3%)	1589(5%)
$E_{9/21}$	> 0.95	67111(97.7%)	1018805(8%)	110648(8%)	1565(5%)	1589(5%)
$\theta(\gamma)$ Barrel region	$clusterReg == 2$	56824(84.6%)	778932(23%)	85739(23%)	1245(15%)	1307(14%)
M_{bc}	$5.24 < M_{\mathrm{bc}} < 5.29$ GeV/c^2	56392(99.2%)	434045(44.7%)	45665(47%)	957(23%)	1012(23%)
ΔE	$-0.6 < \Delta E < 0.2$ GeV	55651(98.6%)	166266(62%)	14624(68%)	91(90%)	151(85%)
clusterNHits	clusterNHits>= 15	47135(84.6%)	83666(50%)	9060(38%)	73(20%)	128(15%)
ClusterZernikeMVA	ClusterZernikeMVA> 0.75	42273(89.7%)	22103(74%)	3397(63%)	53(27%)	86(33%)
FastBDT Output	> 0.45	37395(89%)	2810(87%)	556(84%)	46(14%)	68(21%)
$P_{\pi^0\gamma}$	< 0.71	32117(81.6%)	714(75%)	229(59%)	12(74%)	23(66%)
$P_{\eta\gamma}$	< 0.62	30829(96%)	567(20%)	176(27%)	11(9%)	22(5%)

Table 3.13: The figures in the parenthesis indicate the survival(blue) and loss(red) percentage in signal and background MC samples, respectively. The table is not scaled to 362 fb^{-1}

Assuming $\mathcal{B}(B^0 \to \gamma\gamma) = 1.4 \times 10^{-8}$ and overall signal reconstruction efficiency 30.8%, we can expects about 1.6 signal and 286 total background events in 362 fb^{-1} of data in the candidate region.

8.3 Comparison between Belle and Belle II performances

In this section, we have compared the performances between Belle and Belle II in terms of efficiency. We have processed 100,000 signal MC for both Belle and Belle II. The comparison of the performances of the selection criteria for Belle and Belle II is shown in Table 3.14. At the precut level (skim), there is a difference in the efficiency of Belle and Belle II, which is 47.9% and 72%, respectively. The lower efficiency of Belle is due to the HadronB skim cuts applied in the B2BIIfixMdst and B2BIIConvertMdst modules, which perform correction to Belle mdst data and Belle mdst to Belle II mdst conversion. There is not much difference in the performance of cuts in Belle and Belle II, as shown in the parenthesis, except for the clusterReg cut, which retains 79% (84.6%) of signal events in Belle (Belle II).
We have also compared the signal efficiency with the previous Belle results, which are shown in the table 3.15. The efficiency improvement in our Belle analysis is twice as compared to the previous Belle analysis. This is mostly due to the usage of advanced data analysis techniques such as the FastBDT MVA method, likelihood-based analysis and a new set of optimized selection criteria which improve the overall reconstruction efficiency.

Cuts	Belle II	Cuts	Belle
skim	72016	skim	47912
Timing cuts	68677(95.3%)	Photon ID	47910(99%)
$E_{9/21} > 0.95$	67111(97.7%)	$E_{9/25} > 0.95$	43443(91%)
clusterReg==2	56824(84.6%)	clusterReg==2	34379(79%)
$5.24 < M_{\rm bc} < 5.29$ GeV/c^2	56392(99.2%)	$5.24 < M_{\rm bc} < 5.29$ GeV/c^2	34011(99%)
$-0.6 < \Delta E < 0.2$ GeV	55651(98.6%)	$-0.6 < \Delta E < 0.2$ GeV	33494(98%)
clusterNHits>=15	47135(84.6%)	-	-
clusterZernikeMVA>0.75	42273(89.7%)	-	-
FastBDT output>0.45	37117(89%)	FastBDT output>0.55	28937(86%)
Pi0_Prob<0.71	32117(81.6%)	Pi0_Prob<0.50	24599(85%)
Eta_Prob<0.62	30829(96%)	Eta_Prob<0.65	23308(95%)

Table 3.14: The figure in the parenthesis indicates the survival percentage in the signal MC sample.

	Belle	Belle II	Previous Belle
Luminosity	694 fb^{-1}	362 fb^{-1}	104 fb^{-1}
Signal Efficiency	23.3%	30.8%	11.7% [20]

Table 3.15: Performance comparison with previous Belle result.

Summary

In this chapter, we explored the steps involved in event generation and simulation. We examined the initial event selection requirements utilized to identify the specific event type for our analysis. Furthermore, we delved into the selection criteria employed to distinguish signals from background events. Through background Monte Carlo studies, we observed the prevalence of the continuum background over other types. To mitigate the impact of the continuum background, we utilized multivariate analysis, specifically employing the FastBDT method. Additionally, we conducted studies on peaking and rare backgrounds, determining that their contributions were not significantly influential. Finally, we presented our expectations regarding the number of signals and background events from the complete dataset, showcasing a performance comparison between Belle and Belle II.

Chapter 4

PDF Parametrization and Fit Validation

1.0 Introduction

This chapter details the procedure of parametrizing both the signal and background MC using the variables $M_{\rm bc}$, ΔE, and $C'_{\rm BDT}$. We aim to create a model that can effectively extract signal yields from real data. We also conduct ensemble studies to understand any potential biases associated with the fit model.

2.0 Maximum Likelihood Method

The Maximum Likelihood (ML) fit is a technique for estimating the values of the unknown parameters of a probability distribution function (PDF) that best describes the observed data. Consider an experiment where the outcome is a continuous random variable x, whose possible values form the sample space Ω. The probability that the measurement x falls in the range $[x, x + dx]$ is given by $f(x; \alpha)dx$, where $f(x; \alpha)$ is the PDF and α are the parameters characterizing the PDF.

If the experiment is repeated N times, measuring a vector of n continuous variables $x = (x_1, ..., x_n)$ each time, the total sample can be represented as $(x_1^1, ..., x_n^1)$, ..., $(x_1^N, ..., x_n^N)$. Assuming the measurements are independent and distributed according to the PDF $f(x1, ..., x_n; \alpha)$, the probability of observing this entire sample is the product of the individual probabilities:

$$L(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N; \boldsymbol{\alpha}) = \prod_{i=1}^{N} f(x_1^i, \ldots, x_n^i; \boldsymbol{\alpha}), \tag{4.1}$$

The function L is called the likelihood function. The ML estimators of the unknown parameters α are the values that maximize this likelihood function, or equivalently, the logarithm of the likelihood function:

$$\ln L(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N; \boldsymbol{\alpha}) = \sum_{i=1}^{N} \ln f(x_1^i, \ldots, x_n^i; \boldsymbol{\alpha}). \tag{4.2}$$

The solutions to the Equation 4.3 give the maximum likelihood (ML) estimators of unknown parameters,

$$\frac{\partial L(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N; \boldsymbol{\alpha})}{\partial \alpha_i} = 0, \tag{4.3}$$

where the estimators are represented with hats $\hat{\boldsymbol{\alpha}} = \{\hat{\alpha}_1, \ldots, \hat{\alpha}_m\}$ to distinguish them from the actual values $(\alpha_1, \ldots, \alpha_m)$. The error $(\sigma(\hat{\alpha}))$ on the estimator distribution is given by:

$$\sigma[\hat{\alpha}_j] = \int \cdots \int (\alpha_j - \hat{\alpha}_j)^2 \prod_{i=1}^{N} f(x_1^i, \ldots, x_n^i; \boldsymbol{\alpha}) dx_1^i \ldots dx_n^i, \tag{4.4}$$

where α_j is the true value of the j^{th} parameter. The covariance of the different parameter estimators is defined as,

$$\sigma[\hat{\alpha}_{jk}] = \int \cdots \int (\alpha_j - \hat{\alpha}_j)(\alpha_k - \hat{\alpha}_k) \prod_{i=1}^{N} f(x_1^i, \ldots, x_n^i; \boldsymbol{\alpha}) dx_1^i \ldots dx_n^i. \tag{4.5}$$

In general, ML estimators exhibit are consistent, unbiased and efficient for large samples. An estimator is one where the deviation between the estimated value and the true value vanishes as the sample size increases, i.e.,

$$\lim_{N \to \infty} \hat{\alpha}_j = \alpha_j \tag{4.6}$$

An estimator is said to be unbiased if its expectation value is equal to the true value i.e.,

$$\langle \hat{\alpha}_j \rangle = \alpha_i \tag{4.7}$$

An estimator is efficient if its variance is equal to or less than the Cramér-Rao bound [93, 94] on the variance. Further detail on the ML method can be found in the references [95, 96].

2.1 Extended Maximum Likelihood

The standard maximum likelihood method can be extended to incorporate the total number of observed events N as an additional parameter to be estimated.

In an experiment repeated N times, the number of observed events N itself can be considered a random variable, typically following a Poisson distribution with mean μ, representing the expected total number of events.

The extended likelihood function then becomes:

$$L(\mathbf{x}_1, \ldots, \mathbf{x}_N, N; \boldsymbol{\alpha}, \mu) = \frac{\mu^N}{N!} e^{-\mu} \prod_{i=1}^{N} f(\mathbf{x}_i; \boldsymbol{\alpha}) \tag{4.8}$$

Here, the first term accounts for the Poisson distribution of the total number of events N, while the second term represents the likelihood of observing the measurement values $\mathbf{x}_1, \ldots, \mathbf{x}_N$ given the PDF parameters $\boldsymbol{\alpha}$.

In general, the total mean μ can be expressed as the sum of signal and background event yields, $\mu = \mu_s + \mu_b$, where μ_s and μ_b are the unknown parameters to be estimated along with the PDF parameters $\boldsymbol{\alpha}$.

The log-likelihood function, after dropping the term for N (since it is independent of the PDF parameters), becomes:

$$\ln L(\boldsymbol{\alpha}, \mu_s, \mu_b) = -\mu_s - \mu_b + \sum_{i=1}^{N} \ln(\mu_s f_s(\mathbf{x}_i; \boldsymbol{\alpha}) + \mu_b f_b(\mathbf{x}_i; \boldsymbol{\alpha})) \tag{4.9}$$

This extended maximum likelihood formulation allows for the simultaneous estimation of the signal and background event yields (μ_s and μ_b) along with the parameters $\boldsymbol{\alpha}$ that characterize the signal and background PDFs (f_s and f_b).

The analysis in this thesis uses this unbinned extended maximum likelihood method to estimate the unknown parameters, taking advantage of the RooFit package [97] within the ROOT framework [98]. The values of parameters, along with their statistical uncertainties, are determined using the MINUIT package [99, 100]. This package utilizes the MIGRAD algorithm to identify parameter values that minimize the negative log-likelihood function. Additionally, it employs the MINOS algorithm to calculate the uncertainties associated with these parameters. This unbinned extended maximum likelihood approach is used to estimate the unknown parameters of the signal and background PDFs, as the limited statistics of the dataset make the use of binned data less desirable.

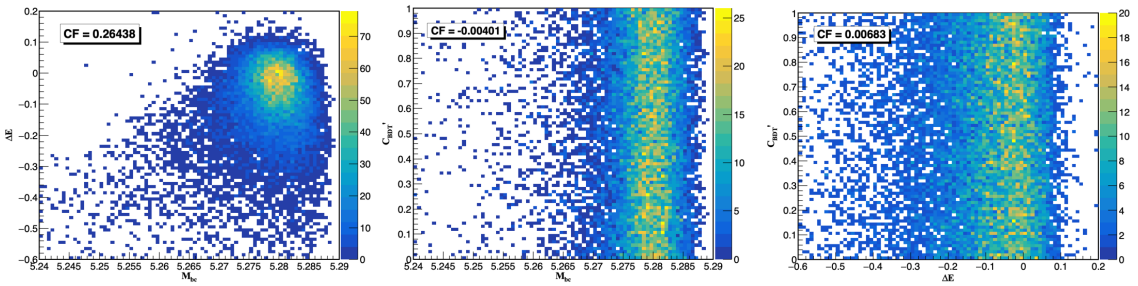
In the following sections, we discuss the PDF parameterization of the signal and background distributions of $M_{\rm bc}$ ΔE, and $C'_{\rm BDT}$.

We will perform a 3D unbinned extended ML fit with $M_{\rm bc}$, ΔE, and $C'_{\rm BDT}$ to extract the signal yield. The variables used in the fit should not have a high correlation among themselves. Before performing the fit, we need to ensure that the fit variables $M_{\rm bc}$, ΔE, and $C'_{\rm BDT}$ are not correlated significantly. Otherwise, the correlation will reduce the signal identification capability of the fit variables. Table 4.1 summarizes the correlation factors between different variables, and the correlation plots are shown as 4.1 and 4.2 for the Belle.

Fit Variables	$M_{\rm bc}$ **vs** ΔE	$M_{\rm bc}$ **vs** $C'_{\rm BDT}$	ΔE **and** $C'_{\rm BDT}$
Signal MC	0.26438	-0.00401	0.00683
Continuum	0.01449	0.00317	0.03473

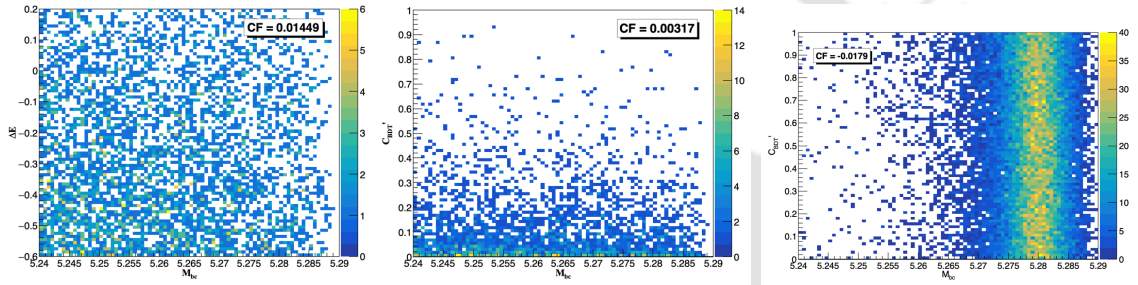
Table 4.1: Linear correlation factors between the fit variables in $B^0 \to \gamma\gamma$ mode for Belle study.

For Belle II, the correlation plots between $M_{\rm bc}$, ΔE and $C'_{\rm BDT}$ are shown in Figure 4.3 and 4.4. Table 4.2 summarizes the correlation factors between different variables.



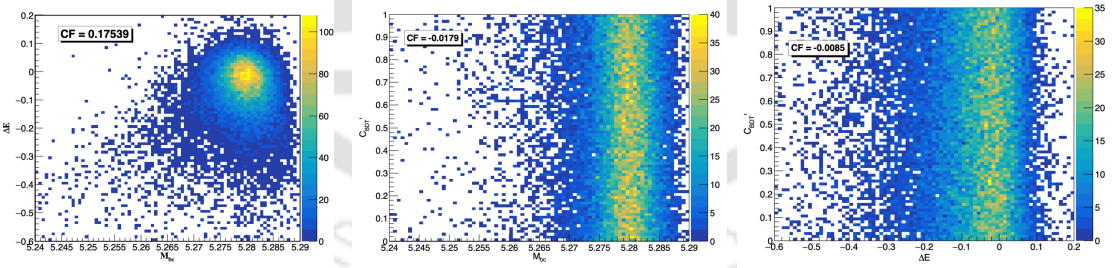
(a) (b) (c)

Figure 4.1: Correlation between (a) $M_{\rm bc}$ and ΔE (b) $M_{\rm bc}$ and $C'_{\rm BDT}$ (c) ΔE and $C'_{\rm BDT}$ for the signal $B^0 \to \gamma\gamma$ for Belle.



(a) (b) (c)

Figure 4.2: Correlation between (a) $M_{\rm bc}$ and ΔE (b) $M_{\rm bc}$ and $C'_{\rm BDT}$ (b) ΔE and $C'_{\rm BDT}$ for the background $B^0 \to \gamma\gamma$ for Belle.



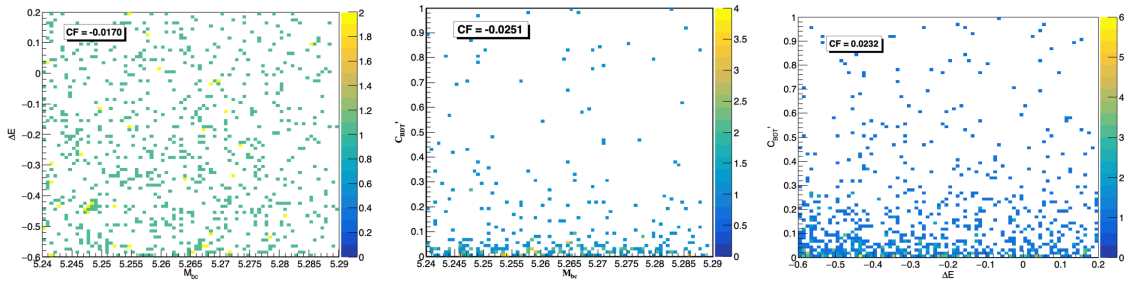
(a) (b) (c)

Figure 4.3: Correlation between (a) $M_{\rm bc}$ and ΔE (b) $M_{\rm bc}$ and $C'_{\rm BDT}$ (b) ΔE and $C'_{\rm BDT}$ for the signal $B^0 \to \gamma\gamma$ for Belle II.

2.2 PDF Modelling for Belle Study

Initially, the 3-dimensional PDFs are modeled as the product of 1-dimensional PDFs:

$$\mathcal{P}(M_{\rm bc}, \Delta E, C'_{\rm BDT}) = \mathcal{P}(M_{\rm bc}) \times \mathcal{P}(\Delta E) \times \mathcal{P}(C'_{\rm BDT})$$



(a) (b) (c)

Figure 4.4: Correlation between (a) $M_{\rm bc}$ and ΔE (b) $M_{\rm bc}$ and $C'_{\rm BDT}$ (b) ΔE and $C'_{\rm BDT}$ for the background $B^0 \to \gamma\gamma$ for Belle II.

Fit Variables	$M_{\rm bc}$ **vs** ΔE	$M_{\rm bc}$ **vs** $C'_{\rm BDT}$	ΔE **and** $C'_{\rm BDT}$
Signal MC	0.1723	-0.0179	-0.0085
Continuum	-0.0170	-0.0251	0.0232

Table 4.2: Linear correlation factors between the fit variables in $B^0 \to \gamma\gamma$ mode for Belle II.

However, significant correlations were found between M_{bc} and ΔE in the signal component, with 26% and 17% correlations for Belle and Belle II, respectively. This invalidates the factorized form for the signal component. Hence, the 3-dimensional PDF for the signal component was adjusted using a 2-dimensional joint PDF for the correlated variables and a separate 1-dimensional PDF for the remaining variable:

$$\mathcal{P}(M_{\rm bc}, \Delta E, C'_{\rm BDT}) = \mathcal{P}(M_{\rm bc}, \Delta E) \times \mathcal{P}(C'_{\rm BDT})$$

This modified approach accommodates the observed correlation between M_{bc} and ΔE while maintaining the independence assumption for C'_{BDT}. For the signal, the correlation between $M_{\rm bc}$ and ΔE is taken into account with a 2D Kernel Density Estimation (KDE) PDF [101].

The PDFs used to parameterize the signal are tabulated in Table 4.3 for Belle and Belle II.

Fit Variables	**Signal**	**Parameter**
$M_{\rm bc}$	2D KDE	-
ΔE	2D KDE	-
$C'_{\rm BDT}$	$0^{\rm th}$ Polynomial	-

Table 4.3: PDFs for signal MC for Belle and Belle II

Continuum Background PDF

The 1D PDFs used to fit the background ($q\bar{q}$) MC are summarized in Table 4.4, for Belle. These function is briefly described in `Appendix` (G).

- $M_{\rm bc}$ is fitted with an Argus function [102] with an endpoint fixed to 5.289 GeV/c^2.
- ΔE is fitted with 1$^{\rm st}$ order Chebychev polynomial.
- $C'_{\rm BDT}$ is fitted with 2 exponential function.

Fit Variables	**Background**	**Parameter**
$M_{\rm bc}$ (GeV/c^2)	Argus	Curvature(argpar) Endpoint fixed
ΔE (GeV)	1$^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)
$C'_{\rm BDT}$	2 exponential	lamda(λ_1) lamda2(λ_2)

Table 4.4: PDFs for continuum background for Belle. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

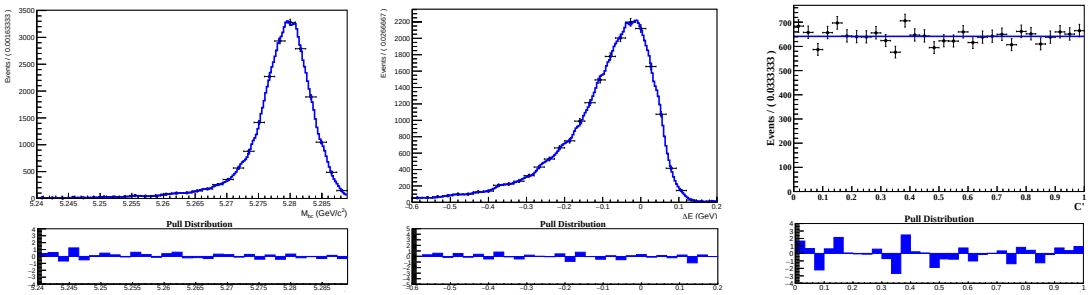
2.3 Fixed and Floated parameters for 3D Fit model

To extract the signal and background yields from the three-dimensional fit model, we fix the $C'_{\rm BDT}$ background exponential parameters (λ_1 and λ_2) and the endpoint of the ARGUS function and float all the other background parameters in the final fit model. The signal PDFs have no parameters. In addition, we float the signal and the background yields. Table 4.5 summarizes the signal and background PDF parameters used in the 3D Fit model.

Variable		**Function**	**Parameter**	**MC Value**	**Comment**
$M_{\rm bc}$ (GeV/c^2)	Signal	2D KDE	-	-	No parameter
	Background	Argus	Endpoint	5.289	Fixed
			Curvature(argpar)	-33.06 ± 3.5	Floated
ΔE (GeV)	Signal	2D KDE	-	-	No parameter
	Background	1$^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)	-0.6373 ± 0.025	Floated
$C'_{\rm BDT}$	Signal	0$^{\rm th}$ order Polynomial	-	-	No parameter
	Background	Exponential	λ_1	-5.434	Fixed
			λ_2	-55.80	Fixed
			frac	0.827 ± 0.019	Floated

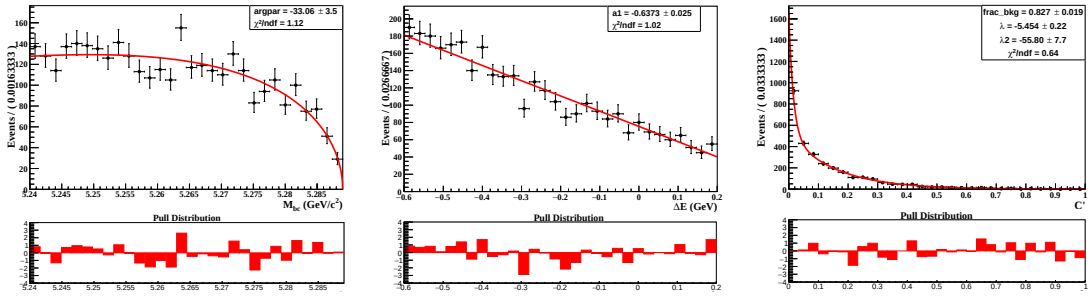
Table 4.5: PDF used in the 3D Fit of Belle $B^0 \to \gamma\gamma$ sample and parameter information. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

Fitted signal and background MC distributions are shown in Figure 4.5 and 4.6, respectively, for Belle.



(a) (b) (c)

Figure 4.5: Signal 1D PDFs parametrizations for Belle (a) $M_{\rm bc}$ (b) ΔE (c) $C'_{\rm BDT}$



(a) (b) (c)

Figure 4.6: Background 1D PDFs parametrizations for Belle (a) $M_{\rm bc}$ (b) ΔE (c) and $C'_{\rm BDT}$

2.4 PDF Modelling for Belle II Study

The PDFs used to parameterize the signal are tabulated in Table 4.3, which is the same as Belle.

Continuum Background PDF

The 1D PDFs used to fit the background ($q\bar{q}$) MC are given in Table 4.6.

- $M_{\rm bc}$ is fitted with an Argus function with an endpoint fixed to 5.29 GeV/c^2.
- ΔE is fitted with 1$^{\rm st}$ order Chebychev polynomial.
- $C'_{\rm BDT}$ is fitted with two exponential function.

2.5 Fixed and Floated parameters for 3D Fit model

To extract the signal and background yields from the three-dimensional fit model, we fix the $C'_{\rm BDT}$ background exponential parameters (λ_1 and λ_2) and the endpoint

Fit Variables	Background	Parameter
$M_{\rm bc}$ (GeV/c^2)	Argus	Curvature(argpar) Endpoint fixed
ΔE (GeV)	1$^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)
$C'_{\rm BDT}$	2 exponential	lamda(λ_1) lamda2(λ_2) fraction(frac_bkg)

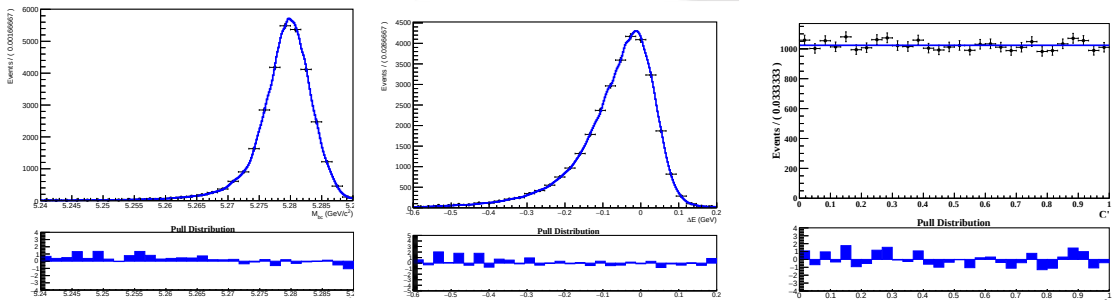
Table 4.6: PDFs for continuum background for Belle II. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

of the ARGUS function and float all the other background parameters in the final fit model. The signal PDFs have no parameters. Table 4.7 summarizes the signal and background PDF parameters used in the 3D Fit model.

Variable		Function	Parameter	MC Value	Comment
$M_{\rm bc}$ (GeV/c^2)	Signal	2D KDE	-	-	No parameter
	Background	Argus	Endpoint	5.29	Fixed
			Curvature(argpar)	-34.56 ± 7.1	Floated
ΔE (GeV)	Signal	2D KDE	-	-	No parameter
	Background	1$^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)	-0.3338 ± 0.058	Floated
$C'_{\rm BDT}$	Signal	0$^{\rm th}$ order Polynomial	-	-	No parameter
	Background	Exponential	λ_1	-4.434	Fixed
			λ_2	-22.21	Fixed
			frac	0.585 ± 0.097	Floated

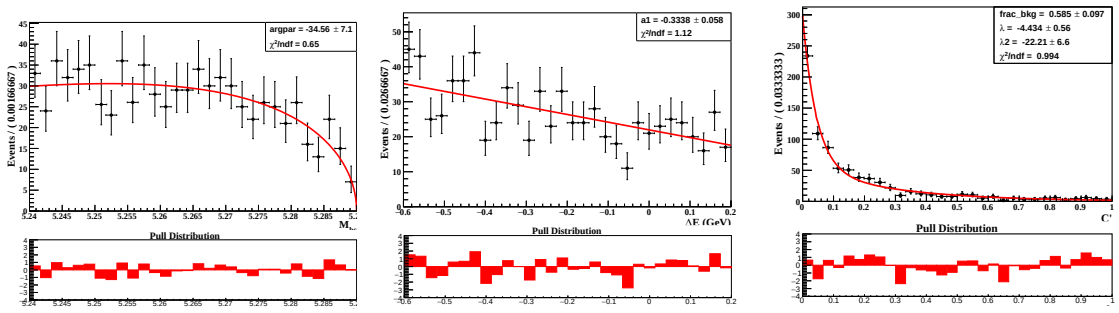
Table 4.7: PDF used in the 3D Fit of Belle II $B^0 \to \gamma\gamma$ sample and parameter information. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

Fitted signal and background MC distributions are shown in Figure 4.7 and 4.8, respectively, for Belle II.



(a) (b) (c)

Figure 4.7: Signal 1D PDFs parametrization for Belle II (a) $M_{\rm bc}$ (b) ΔE (c) $C'_{\rm BDT}$.



(a) (b) (c)

Figure 4.8: Background 1D PDFs parametrizations for Belle II (a) $M_{\rm bc}$ (b) ΔE (c) and $C'_{\rm BDT}$.

3.0 Fit Validation

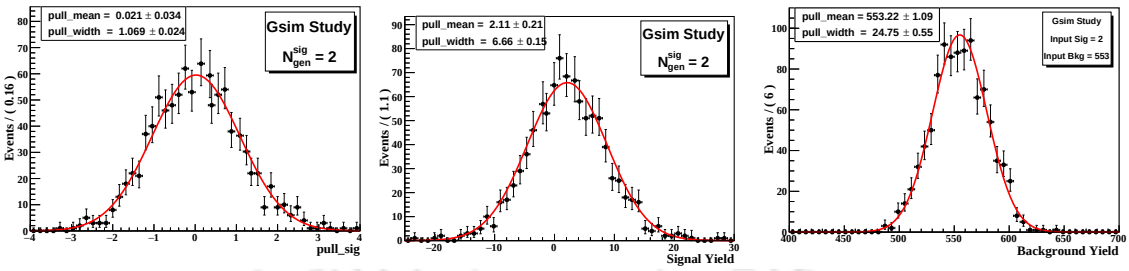
3.1 Gsim Study for Belle

We perform the GSim (Generation and simulation) study to validate the fit procedure and test the bias in signal yield extraction. It is the process of random extraction of signal events after its generation and simulation. In order to get a realistic estimation of these biases, each pseudo-experiment must be statistically equivalent to the real data sample. For this, a set of 1000 samples is generated for the different pseudo-experiments having $N^{\rm sig}_{\rm gen}$ = (0, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20) events for Belle and Belle II. We generate and simulate the signal MC using EvtGen and pick the signal events randomly using Poisson's distribution. In contrast, the background events are extracted randomly from the 3D PDFs corresponding to the $M_{\rm bc}$, ΔE and $C^{'}_{\rm BDT}$ by incorporating Poissonian fluctuations. For different values of signal yield, we then run our fit function over these pseudo experiments to extract pulls and yields. Pull in the signal yield is calculated as follows:

$$\text{Signal Pull} = \frac{N^{\rm sig}_{\rm yield} - N^{\rm sig}_{\rm gen}}{N^{\rm sig}_{\rm error}}$$

where $N^{\rm sig}_{\rm yield}$ refers to the signal yield from the fit function, $N^{\rm sig}_{\rm gen}$ is the input signal to the fit function and $N^{\rm sig}_{\rm error}$ refers to the uncertainty of the signal yields. If the parameter is unbiased, the corresponding pull distribution is a Gaussian of zero mean and unity width. Otherwise, the pull distribution gives the bias in units of the statistical error. For the yield distribution, we expect the yield from the fit to match with the input signal events within MC statistical uncertainty.
For the GSIM study in both analysis, we fix the $C'_{\rm BDT}$ background exponential parameters (λ_1 and λ_2) and the endpoint of the ARGUS function and float all the other background parameters in the final fit model. We expect to observe at most 2 signal events in real data for our Belle MC study. Figure 4.9 shows the pull, signal

yield and background yield from the fit model for $N^{\rm sig}_{\rm gen} = 2$, where $N^{\rm sig}_{\rm gen}$ is the input number of signal events given to the fit.



(a) (b) (c)

Figure 4.9: (a) Signal Pull, (b) Signal yield and (c) background yield distributions for $N^{\rm sig}_{\rm gen}$=2.

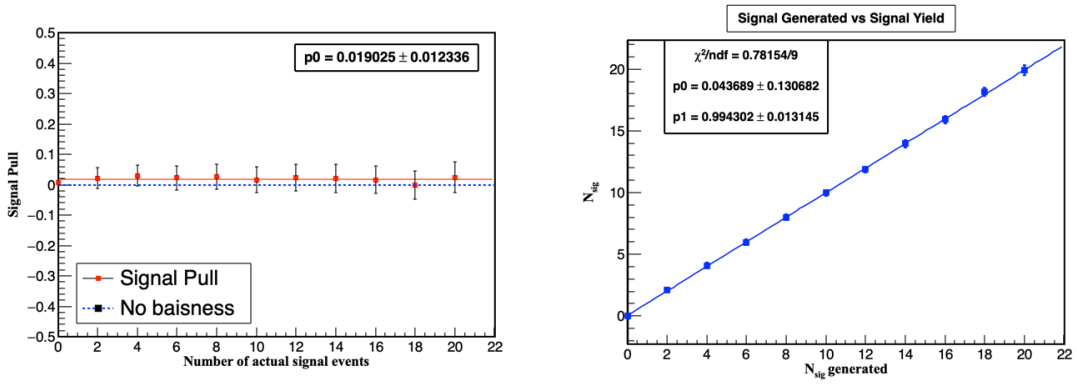
The mean of the pull distribution, shown in Figure 4.9 shows that the fitter is stable. Here mean of the pull distribution agrees with a standard normal distribution, whereas the output yield of the signal and background agrees with the input yield. The pull and signal yield distributions for $N^{\rm sig}_{\rm gen} = 0$, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 are shown in `Appendix` (H).

Figures 4.10a show the average pull distributions for $N^{\rm sig}_{\rm gen} = 0$, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20. We plan to assign the deviation of the average pull distribution from the $0^{\rm th}$ order polynomial as a source of systematic uncertainty on the fit model. We obtain a fit bias of 1.90% (0.05 event) from the fitted signal pull, which will be included as an additive systematic uncertainty.

Figure 4.10b shows the linearity of the fit model. The fitted signal yield demonstrates a linear behaviour with a slope of one and an intercept of 0, as expected. Based on this test, the fit bias is expected to be not more than 0.13 event, which will be assigned as a source of additive systematic uncertainty.
We have assigned the combined systematic of +0.14 event from the fit bias and linearity test.

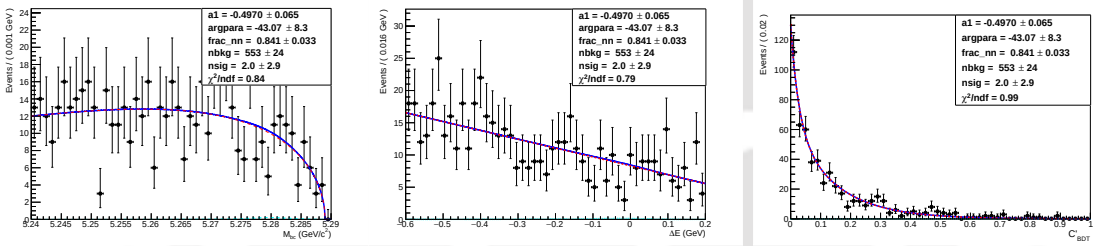
3.1.1 Quality of fit in simulated data for Belle

To check the performance of fit in the realistic background conditions, we also use the simulated data. We have performed the 3D fit using the simulated dataset for a size expected in the real data set and a larger sample than the expected one in data. The background sample is from $q\overline{q}$ events, which is the dominating background. We found the quality of fits is good in both configurations. The fit plots shown Figures 4.11 and 4.12 are for similar and larger datasets, respectively.



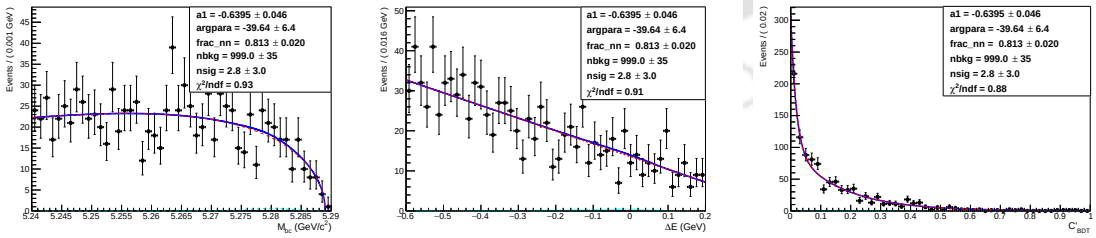
(a) (b)

Figure 4.10: (a) Average pull distribution from MC ensemble study: input signal yield versus pull mean fitted with a straight line (red solid line). The (blue) dashed line shows the behaviour of pull means in the ideal case. (b) Results of the linearity test for signal extraction: input versus output signal yield fitted with first-order polynomial (Blue solid line)



(a) (b) (c)

Figure 4.11: (a) $M_{\rm bc}$, (b) ΔE and (c) $C'_{\rm BDT}$ distributions for $B^0 \to \gamma\gamma$ mode in MC where the signal and background data is similar to the expected ones in the real data. frac_nn is the fraction of two exponential functions.

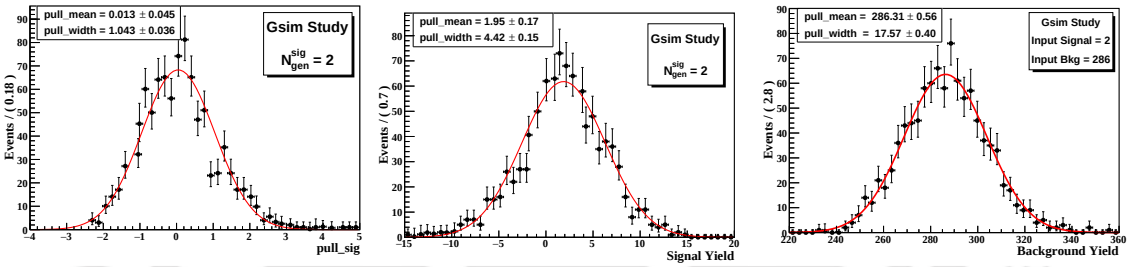


(a) (b) (c)

Figure 4.12: (a) $M_{\rm bc}$, (b) ΔE and (c) $C'_{\rm BDT}$ distributions for $B^0 \to \gamma\gamma$ mode in MC where the signal size is as per expectation in the real data but the background is a factor of 2 higher than the expectation in the real data. frac_nn is the fraction of two exponential functions.

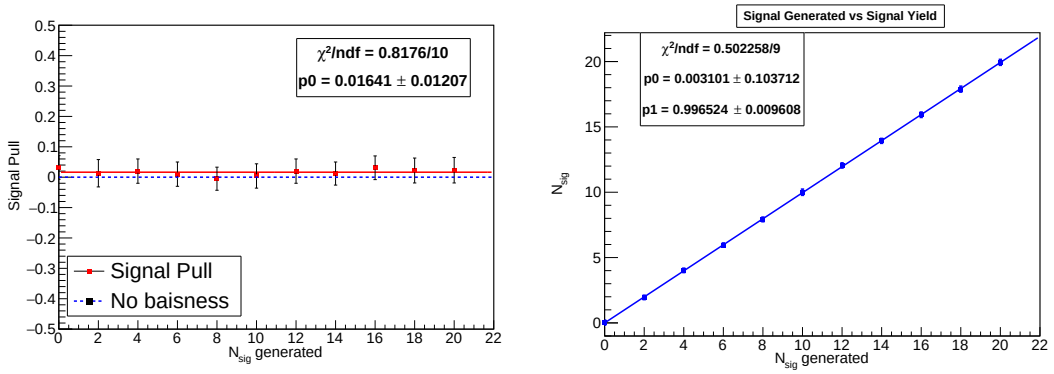
3.2 Gsim Study for Belle II analysis

The pull and signal yield distributions for $N_{\text{gen}}^{\text{sig}}$ = 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20 are shown in `Appendix` (H).
For our MC Study, we do not expect to observe more than 2 signal events in the real data. Figure 4.13 shows the pull, signal MC yields and background yields from the fit model for $N_{\text{gen}}^{\text{sig}}$ = 2. Figure 4.14a shows the average pull distribution for $N_{\text{gen}}^{\text{sig}}$ = (0, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20). We obtain a fit bias of 1.64% (0.03 event) from the average pull distribution in Figure 4.14a, which will be included as an additive systematic uncertainty. To check for linearity, which is shown in Figure 4.14b. We find the fit model to be linear. The fitted signal yield demonstrates a linear behaviour with a slope of one and an intercept of 0, as expected. Based on this test, the fit bias is expected to be not more than 0.10 event, which will assigned as a source of additive systematic uncertainty. We have assigned the quadrature sum of uncertainty from the fit bias and linearity test, which is a +0.10 event.



(a) (b) (c)

Figure 4.13: (a) Signal Pull, (b) Signal yield and (c) background yield distributions for $N_{\text{gen}}^{\text{sig}}$=2 for Belle II.

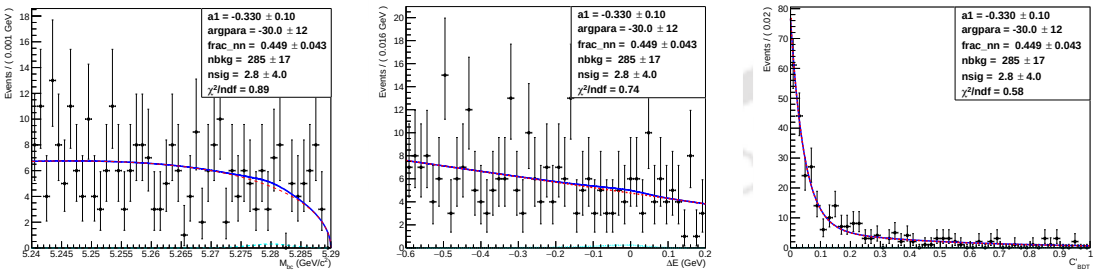


(a) (b)

Figure 4.14: (a) Average pull distribution from Belle II MC ensemble study: input signal yield versus pull mean fitted with a straight line (red solid line). The data points are the Signal Pull mean for the input signal ($N^{\rm sig}_{\rm gen} = 0, 2, 4....20$), and the error bars are the uncertainty on the Pull mean. The (blue) dashed line shows the behaviour of pull means in the ideal case. (b) Results of the linearity test for signal extraction: input versus output signal yield fitted with first-order polynomial (Blue solid line)

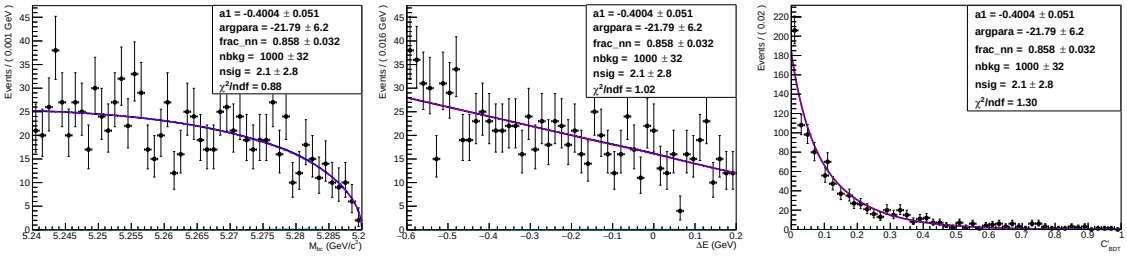
3.2.1 Quality of fit in simulated data for Belle II

We have performed the 3D fit in a similar way as in the case of Belle using the simulated dataset for a size expected in the real data set and a larger sample than the expected one in data. We found the quality of fits is good in both configurations. The fit plots shown in Figures 4.15 and 4.16 are for similar and larger datasets, respectively.



(a) (b) (c)

Figure 4.15: (a) $M_{\rm bc}$ (b) ΔE and (c) $C'_{\rm BDT}$ distributions for $B^0 \to \gamma\gamma$ mode in Belle II MC where the signal and background data is similar to the expected ones in the real data. frac_nn is the fraction of two exponential functions.



(a) (b) (c)

Figure 4.16: (a) $M_{\rm bc}$ (b) ΔE and (c) $C'_{\rm BDT}$ distributions for $B^0 \to \gamma\gamma$ mode in Belle II MC where the signal size is as per expectation in real data but the background is a factor of 3 higher than the expectation in the real data. frac_nn is the fraction of two exponential functions.

4.0 Summary

In this chapter, we have discussed the maximum likelihood fit method to estimate the PDF parameters and the PDF modeling of signal and background distributions. Further, we have investigated the large MC toy ensembles to estimate the bias from the fitted PDF model. We also checked the performance of the fit using the simulated dataset for the size expected in the real data and a larger sample than the expected one in the data. We found the quality of fit is good in both configurations.

Chapter 5

Analysis Validation

1.0 Introduction

This chapter provides a comprehensive study performed to validate the analysis strategy using the control mode. The decay channel, $B^0 \to K^*\gamma$ is used as a control sample for this purpose. Using the control sample, the calibration of the BDT and π^0/η veto is performed. Furthermore, a section is devoted to looking for the data-MC differences for the fitting variables of the control sample and side-band regions for the target mode. Moreover, off-resonance data is used to cross-check the MC simulation of the continuum background. We refer $B^0 \to \gamma\gamma$ throughout this chapter as a "target" mode to differentiate from the control channel.

2.0 Control Sample Study

We perform a control sample study using a decay channel with the closest decay topology and final state particles to calibrate the BDT and the π^0/η selection criteria. For this purpose, we use $B^0 \to K^*\gamma$. We apply photon selection criteria similar to the target analysis ($B^0 \to \gamma\gamma$) mentioned in the section 5.1. The selection criteria for the $B^0 \to K^*\gamma$ mode are summarized in Table 5.1 for Belle and Table 5.2 for Belle II, respectively. For Belle, we use the dataset corresponding to 694 fb^{-1}, while for Belle II, we use 362 fb^{-1} of data for the control sample study.

3.0 Fitting

3.1 Belle data sample

The PDFs used to model $M_{\rm bc}$, ΔE, and $C'_{\rm BDT}$ for the control sample are similar to the ones used in the target mode. The control sample had a 36% correlation between $M_{\rm bc}$ and ΔE, which is taken care of with a 2D KDE, while the $C'_{\rm BDT}$ PDF is modelled with the $0^{\rm th}$ order polynomial function. The continuum component in $M_{\rm bc}$ distribution is modelled by ARGUS function with endpoint fixed at 5.289 GeV/c^2,

Particle	Selection Criteria
Photons	$9\ \mu\text{sec} < (\Delta T)_{ECL\ hit-collision} < 11\ \mu\text{sec}$ $1.4 < E_\gamma(\text{CM}) < 3.4$ GeV $E_{9/21} > 0.95$ $clusterReg(\theta(\gamma)) == 2$ $P_{\pi^0}(\gamma) < 0.50$ $P_\eta(\gamma) < 0.65$
	$dr < 0.5$ cm $\lvert dz\rvert < 2$ cm
K/π	$\mathcal{L}(K/\pi) > 0.6$ for selecting kaons, rest considered pions $817 < M_{K^{0*}} < 967$ MeV/c^2 Best K^{0*} chosen based on minimum χ^2 of $K\pi$ vertex fit
B^0	$5.24 < M_{\text{bc}} < 5.29$ GeV/c^2 $-0.6 < \Delta E < 0.2$ GeV

Table 5.1: Table for Selection Criteria for $B^0 \to K^*\gamma$ Belle analysis.

Particle	Selection Criteria
Photons	$1.4 < E_\gamma(\text{CM}) < 3.4$ GeV $E_{9/21} > 0.95$ $clusterReg(\theta(\gamma)) == 2$ $clusterNHits \geq 15$ $\lvert clusterTiming\rvert < 200$ ns $\lvert clusterTiming/clusterErrorTiming\rvert < 2.0$ ClusterZernikeMVA> 0.75 $P_{\pi^0}\gamma < 0.71$ $P_\eta\gamma < 0.62$
	$dr < 0.5$ cm $\lvert dz\rvert < 2$ cm
K/π	$\mathcal{L}(K/\pi) > 0.6$ for selecting kaons, rest considered pions $817 < M_{K^{0*}} < 967$ MeV/c^2 Best K^{0*} chosen based on minimum χ^2 of $K\pi$ vertex fit
B^0	$5.24 < M_{\text{bc}} < 5.29$ GeV/c^2 $-0.6 < \Delta E < 0.2$ GeV

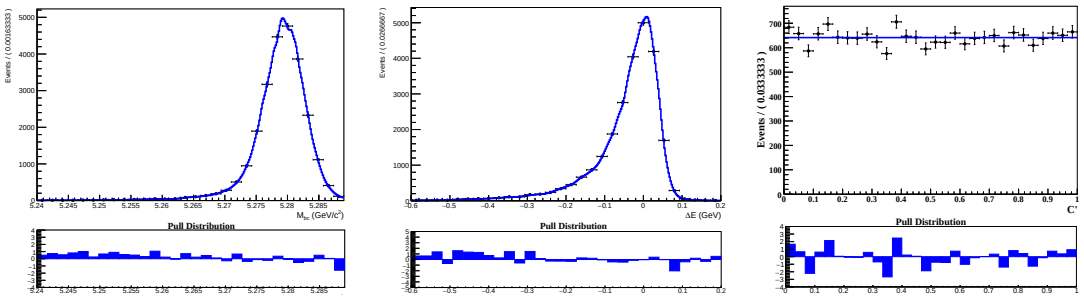
Table 5.2: Table for Selection Criteria for $B^0 \to K^*\gamma$ Belle II analysis.

ΔE uses a 1$^{\text{st}}$ order Chebychev function and C'_{BDT} with two exponential functions. Table 5.3 shows the PDF models used to parametrize signal and background MC.

Fit Variables	Signal	Background
$M_{\rm bc}$	2D KDE	Argus
ΔE	2D KDE	$1^{\rm st}$ Order Chebyshev polynomial
$C'_{\rm BDT}$	$0^{\rm th}$ Polynomial	2 exponential

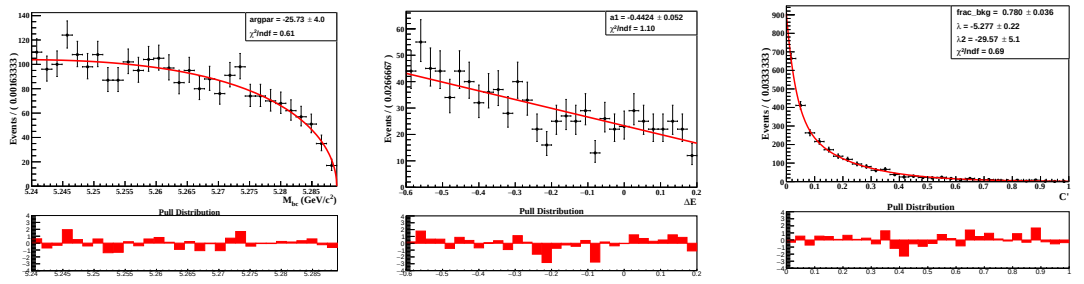
Table 5.3: List of PDFs used for parametrizing the signal and background for both Belle and Belle II control samples.

Figure 5.1 and 5.2 shows 1D PDF parametrizations of the fit variables, $M_{\rm bc}$, ΔE, $C'_{\rm BDT}$ for signal and background MC respectively.



(a) (b) (c)

Figure 5.1: Signal PDFs for (a) $M_{\rm bc}$ (b) ΔE (c) $C'_{\rm BDT}$ for Belle control mode.



(a) (b) (c)

Figure 5.2: Background PDFs for (a) $M_{\rm bc}$ (b) ΔE (c) and $C'_{\rm BDT}$ for Belle control mode.

3.1.1 Real data results

To extract the signal and background yields from the three-dimensional fit model, we fix the $C'_{\rm BDT}$ background exponential parameters (λ_1 and λ_2) and the endpoint of the ARGUS function and float all the other background parameters in the final fit model. The signal PDFs have no parameters. In addition, we float the signal

and the background yields. Table 5.4 summarizes the signal and background PDF parameters used in the 3D fit model.

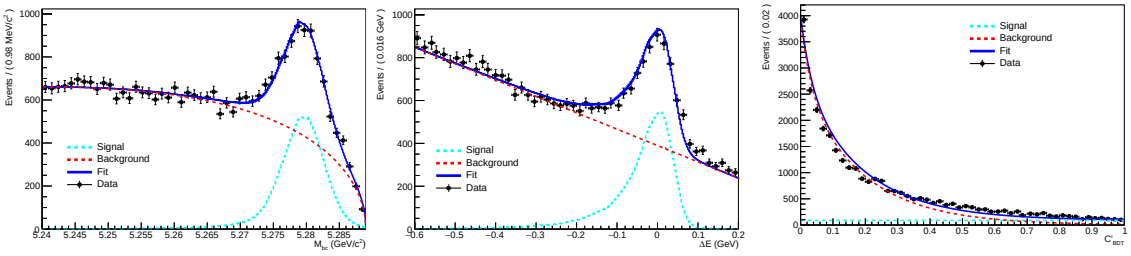
Variable		Function	Parameter	MC Value	Comment
$M_{\rm bc}$ (GeV/c^2)	Signal	2D KDE	-	-	No parameter
	Background	Argus	Endpoint	5.289	Fixed
			Curvature(argpar)	-25.73 ± 4.0	Floated
ΔE (GeV)	Signal	2D KDE	-	-	No parameter
	Background	1$^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)	-0.4424 ± 0.052	Floated
$C'_{\rm BDT}$	Signal	0$^{\rm th}$ order Polynomial	-	-	No parameter
	Background	Exponential	λ_1	-5.277	Fixed
			λ_2	-29.57	Fixed
			frac	0.780 ± 0.036	Floated

Table 5.4: PDF used in the 3D Fit of Belle $B^0 \to K^*\gamma$ sample and parameter information. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

The fit results shown in figure 5.3 give a signal yield of 4600 ± 75 from the Belle real data. We calculate the branching fraction using this signal yield as,

$$\mathcal{B}(B^0 \to K^*\gamma) = \frac{N_{\rm sig}}{\mathcal{B}(K^* \to K^+\pi^-) \times N_{B\bar{B}} \times \epsilon_{\rm rec} \times 2 \times f^{00}}$$

where, $N_{\rm sig} = 4600 \pm 75$ is the signal yield from the fit, $N_{B\overline{B}} = (753 \pm 10) \times 10^6$ is the number of $B\bar{B}$ pairs at $\Upsilon(4S)$ resonance, $f^{00} = (48.4 \pm 1.2)\%$ the ratio of $\Upsilon(4S)$ going to $B^0\bar{B}^0$, $\mathcal{B}(K^* \to K^+\pi^-)$ is the branching fraction $K^* \to K^+\pi^-$ and $\epsilon_{\rm rec} = 16.6\%$ is the signal reconstruction efficiency. The branching fraction for the decay $B^0 \to K^*\gamma$ is found to be $(3.81 \pm 0.09) \times 10^{-5}$. This result is within the 1.62σ range of the previous result by Belle ($(3.96 \pm 0.07 \pm 1.4) \times 10^{-5}$) [103] and it is 1.44σ deviated from the world average value, which is $(4.18 \pm 0.25) \times 10^{-5}$.



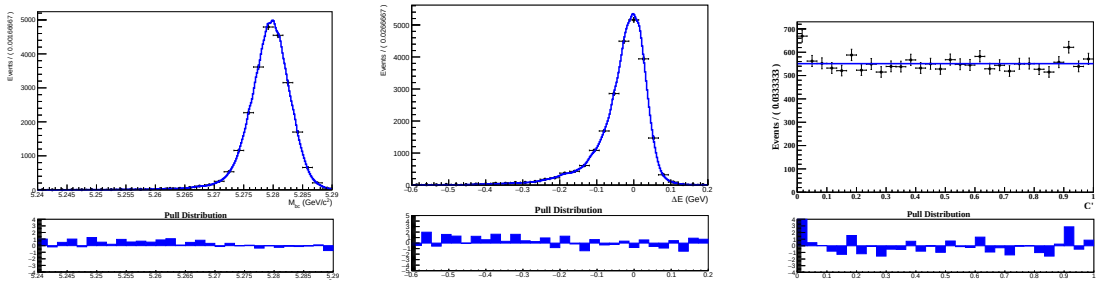
(a) $M_{\rm bc}$ (b) ΔE (c) $C'_{\rm BDT}$

Figure 5.3: Fit results of Belle real data for $B^0 \to K^*\gamma$ control mode.

3.2 Belle II data sample

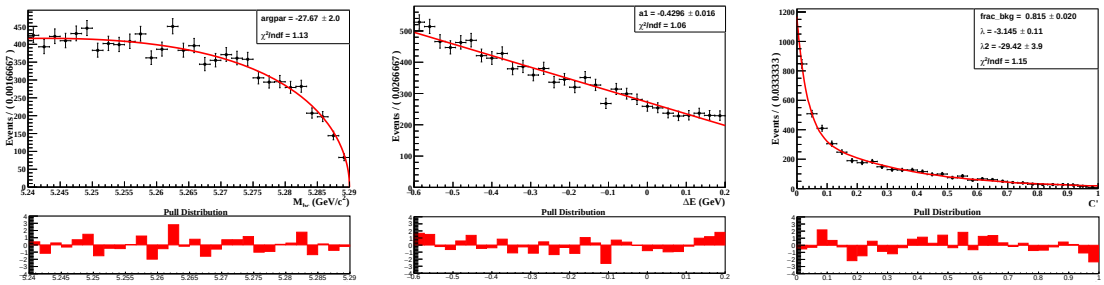
The various PDF models used for parametrizing the signal and background MC are summarized in Table 5.3, consistent with Belle. The control sample had a 44%

correlation between $M_{\rm bc}$ and ΔE, which is taken care of with a 2D Kernel Density Estimation. Figure 5.4 shows 1D PDF parametrizations of the fit variables, $M_{\rm bc}$, ΔE and $C'_{\rm BDT}$ for the signal MC. Figure 5.5 shows 1D PDF parametrizations of the fit variables, $M_{\rm bc}$, ΔE and $C'_{\rm BDT}$ for the background MC.



(a) (b) (c)

Figure 5.4: Signal 1D PDFs for Belle II (a) $M_{\rm bc}$ (b) ΔE (c) $C'_{\rm BDT}$ for $B^0 \to K^*\gamma$ control mode.



(a) (b) (c)

Figure 5.5: Background 1D PDFs for Belle II (a) $M_{\rm bc}$ (b) ΔE (c) and $C'_{\rm BDT}$ for $B^0 \to K^*\gamma$ control mode.

3.2.1 Real data results

To extract the signal and background yields from the three-dimensional fit model, we fix the $C'_{\rm BDT}$ background exponential parameters (λ_1 and λ_2) and the endpoint of the ARGUS function and float all the other background parameters in the final fit model. The signal PDFs have no parameters. In addition, we float the signal and the background yields. Table 5.5 summarizes the signal and background PDF parameters used in the 3D fit model.

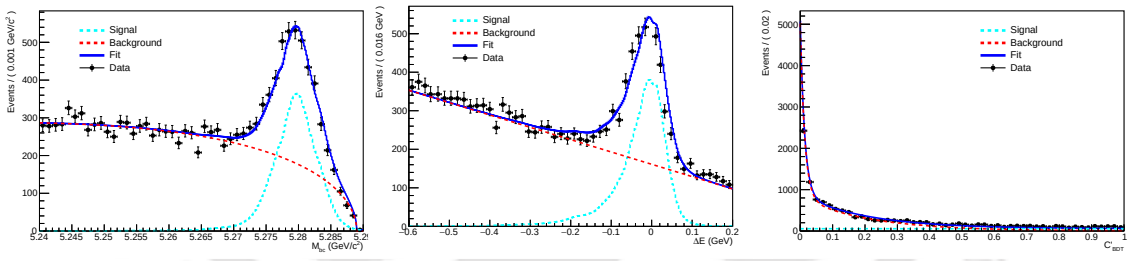
The fit results shown in figure 5.6 give a signal yield of 2790 ± 59 from the real Belle II data. We calculate the branching fraction using this signal yield as,

$$\mathcal{B}(B^0 \to K^*\gamma) = \frac{N_{\rm sig}}{\mathcal{B}(K^* \to K^+\pi^-) \times N_{B\bar{B}} \times \epsilon_{\rm rec} \times 2 \times f^{00}}$$

Variable		Function	Parameter	MC Value	Comment
$M_{\rm bc}$ (GeV/c^2)	Signal	2D KDE	-	-	No parameter
	Background	Argus	Endpoint	5.29	Fixed
			Curvature(argpar)	-27.67 ± 2.0	Floated
ΔE (GeV)	Signal	2D KDE	-	-	No parameter
	Background	$1^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)	-0.4296 ± 0.016	Floated
$C'_{\rm BDT}$	Signal	$0^{\rm th}$ order Polynomial	-	-	No parameter
	Background	Exponential	λ_1	-3.145	Fixed
			λ_2	-29.42	Fixed
			frac	0.815 ± 0.020	Floated

Table 5.5: PDF used in the 3D Fit of Belle II $B^0 \to K^*\gamma$ sample and parameter information. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

where, $N_{\rm sig} = 2790 \pm 59$ is the signal yield from the fit, $N_{B\bar{B}} = (387 \pm 6) \times 10^6$ is the number of $B\bar{B}$ pairs at $\Upsilon(4S)$ resonance, $f^{00} = (48.4 \pm 1.2)\%$ the ratio of $\Upsilon(4S)$ going to $B^0\bar{B}^0$ [104], $\mathcal{B}(K^* \to K^+\pi^-)$ is the branching fraction of $K^* \to K^+\pi^-$ and $\epsilon_{\rm rec} = 18.4\%$ is the signal reconstruction efficiency. The branching fraction for the decay $B^0 \to K^*\gamma$ is found to be $(4.05 \pm 0.09) \times 10^{-5}$. This result is within the 0.79σ range of the previous result by Belle ($(3.96 \pm 0.07 \pm 1.4) \times 10^{-5}$) [103] and it is within 0.49σ of the world average value, which is $(4.18 \pm 0.25) \times 10^{-5}$.



(a) $M_{\rm bc}$ (b) ΔE (c) $C'_{\rm BDT}$

Figure 5.6: Fit results of Belle II real data for $B^0 \to K^*\gamma$ control mode.

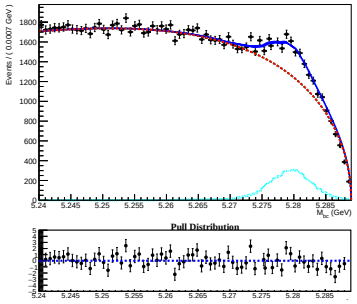
3.3 BDT Calibration

The MVA outputs are calculated using the same training weight files obtained in the $B^0 \to \gamma\gamma$ analysis. Then, 2D $M_{\rm bc}$ and ΔE distributions are fitted above and below the same cut as in $B^0 \to \gamma\gamma$ analysis to obtain efficiency. Here, the number of events selected by the cut is $N_{\rm BDT>}$ and rejected by the cut is $N_{\rm BDT<}$, where BDT is the MVA output for the control mode. The efficiency of the BDT cut can be calculated as $\epsilon = \frac{N_{\rm BDT>}}{N_{\rm BDT<} + N_{\rm BDT>}}$. Finally, the efficiency ratio between data and MC is used to estimate the calibration factor, which is defined as $R_{\rm Data/MC} = \epsilon_{\rm Data}/\epsilon_{\rm MC}$, where $\epsilon_{\rm Data}$ and $\epsilon_{\rm MC}$ are the efficiency of the cut for Data and MC, respectively.

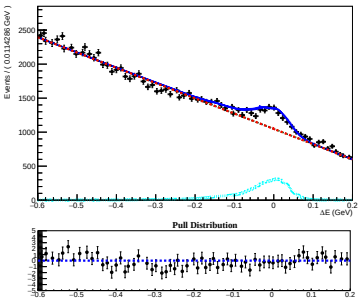
The selection criteria are summarized in Table 5.1 and 5.2 for $B^0 \to K^*\gamma$ control sample Belle and Belle II, respectively. $M_{\rm bc}$ and ΔE distributions above and below

the BDT cuts are fitted by the 2D KDE, where background distributions are fitted by the Argus and 1st Order Chebyshev Polynomial functions.

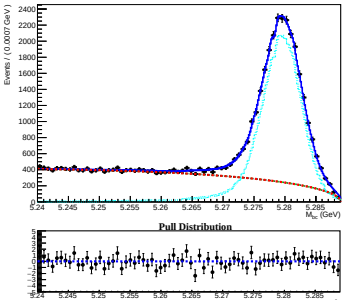
Figures 5.7 and 5.8 show the $M_{\rm bc}$ and ΔE fitting results where MC is not scaled to DATA luminosity for Belle and Belle II, respectively.



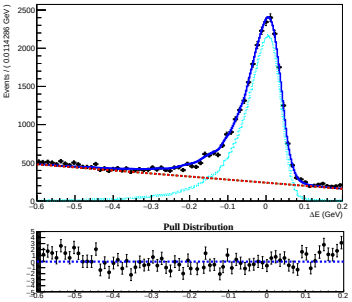
(a) $M_{\rm bc}$ Distribution for MC events rejected by BDT cut



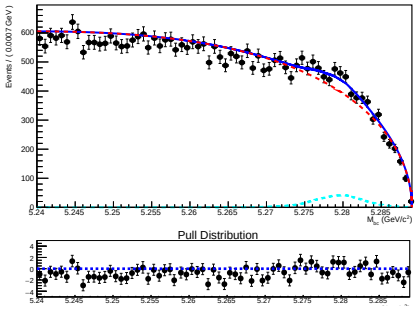
(b) ΔE Distribution for MC events rejected by BDT cut



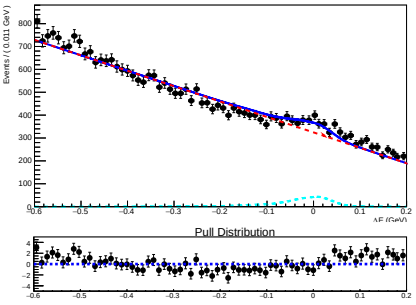
(c) $M_{\rm bc}$ Distribution for MC events selected by BDT cut



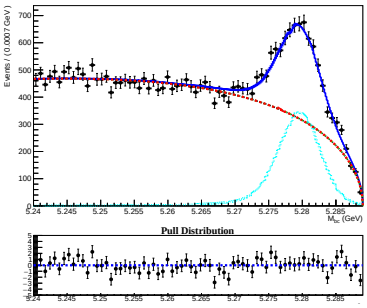
(d) ΔE Distribution for MC events selected by BDT cut



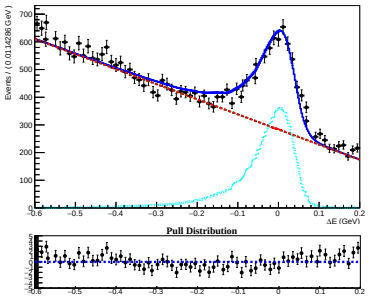
(e) $M_{\rm bc}$ Distribution for Data events rejected by BDT cut



(f) ΔE Distribution for Data events rejected by BDT cut



(g) $M_{\rm bc}$ Distribution for Data events selected by BDT cut



(h) ΔE Distribution for Data events selected by BDT cut

Figure 5.7: 2D $M_{\rm bc}$ and ΔE fits for Belle $B^0 \to K^{0*}\gamma$ control sample remove and selected by BDT cut. [a], [b], [c], and [d] are MC, while [e], [f], [g] and [h] are Data samples. Here MC events are not scaled to Data luminosity.

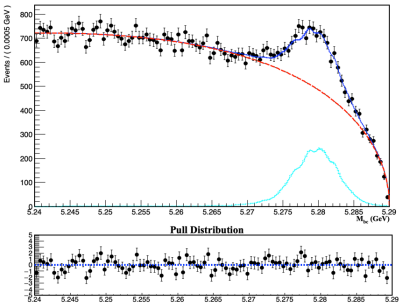
Tables 5.6 and 5.7 summarize the efficiencies and DATA/MC ratio of cut efficiency for Belle and Belle II, respectively. The calibration will be applied to correct the BDT signal efficiency in the analysis, and the uncertainty of the calibration will be taken as systematics.

$N_{\rm BDT>}$ (MC)	$N_{\rm BDT<}$ (MC)	$N_{\rm BDT>}$ (Data)	$N_{\rm BDT<}$ (Data)	$\epsilon_{\rm DATA}$	$\epsilon_{\rm MC}$	$R_{\rm DATA/MC}$
19007±140	2870±57	4093±67	498±25	0.892±0.004	0.869±0.002	1.026±0.004

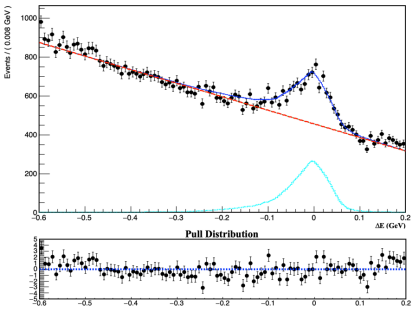
Table 5.6: Efficiency ratio of the control sample between DATA and MC for Belle.

$N_{\rm BDT>}$ (MC)	$N_{\rm BDT<}$ (MC)	$N_{\rm BDT>}$ (Data)	$N_{\rm BDT<}$ (Data)	$\epsilon_{\rm DATA}$	$\epsilon_{\rm MC}$	$R_{\rm DATA/MC}$
21086±150	4650±72	2250±48	479±22	0.824±0.007	0.819±0.003	1.006±0.009

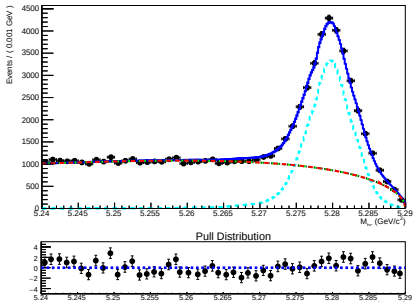
Table 5.7: Efficiency ratio of the control sample between DATA and MC for Belle II.



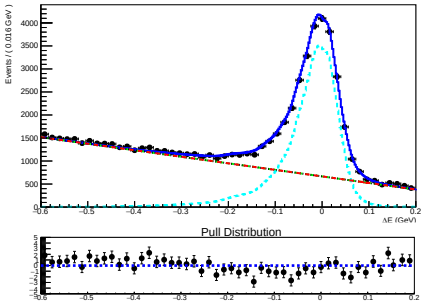
(a) $M_{\rm bc}$ Distribution for MC events rejected by BDT cut



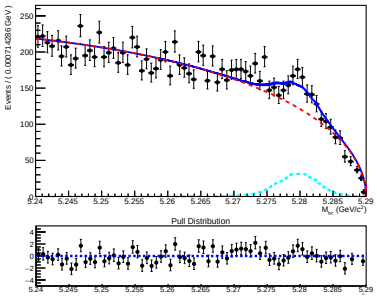
(b) ΔE Distribution for MC events rejected by BDT cut



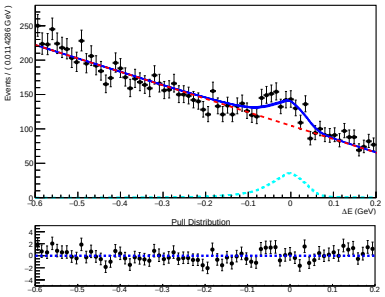
(c) $M_{\rm bc}$ Distribution for MC events selected by BDT cut



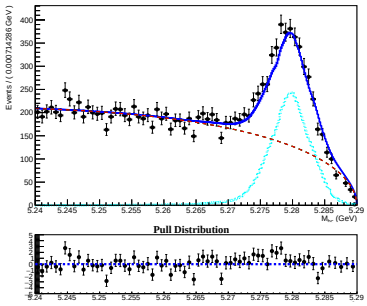
(d) ΔE Distribution for MC events selected by BDT cut



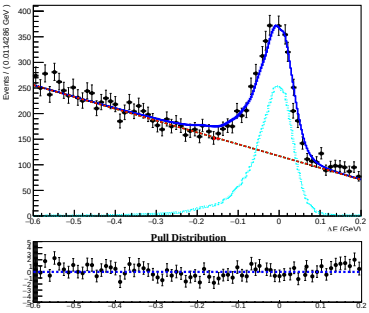
(e) $M_{\rm bc}$ Distribution for Data events rejected by BDT cut



(f) ΔE Distribution for Data events rejected by BDT cut



(g) $M_{\rm bc}$ Distribution for Data events selected by BDT cut



(h) $M_{\rm bc}$ Distribution for Data events selected by BDT cut

Figure 5.8: 2D $M_{\rm bc}$ and ΔE fits for Belle II $B^0 \to K^{0*}\gamma$ control sample remove and selected by BDT cut. [a], [b], [c], and [d] are MC, while [e], [f], [g], and [h] are Data samples. Here MC events are not scaled to Data luminosity.

3.4 π^0/η veto Calibration

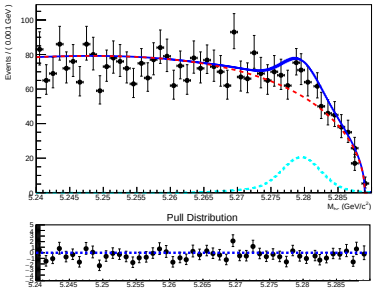
π^0/η veto calibration is performed using $B^0 \to K^*\gamma$ decay mode where $M_{\rm bc}$ and ΔE distributions before and after the π^0/η veto cuts are fitted by the 2D KDE, where background distributions are fitted by the Argus and 1$^{\rm st}$ Order Chebyshev Polynomial functions, respectively. We calculate the relative selection efficiency and the efficiency disparity between MC and data to estimate the uncertainty. The uncertainties of the calibration factors are taken into account as systematics for the Belle and Belle II study. Here, the number of events selected by the cut is $N_{\pi^0/\eta<}$ and rejected by the cut is $N_{\pi^0/\eta>}$. Then, the efficiency of the π^0/η cut can be calculated as $\epsilon = \frac{N_{\pi^0/\eta<}}{N_{\pi^0/\eta<}+N_{\pi^0/\eta>}}$. Figures 5.9 and 5.10 show the $M_{\rm bc}$ and ΔE fitting results for Belle and Belle II, respectively. The calibration factors of the π^0/η veto are summarized in Table 5.8 and 5.9 for Belle and Belle II, respectively.

$N_{\pi^0/\eta<}$ (MC)	$N_{\pi^0/\eta>}$ (MC)	$N_{\pi^0/\eta<}$ (Data)	$N_{\pi^0/\eta>}$ (Data)	$\epsilon_{\rm DATA}$	$\epsilon_{\rm MC}$	$R_{\rm DATA/MC}$
16690±135	295±22	2512±53	40±9	0.984±0.003	0.983±0.001	1.001±0.003

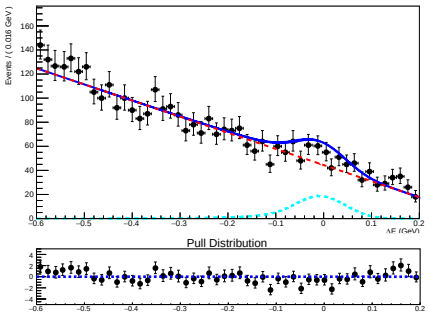
Table 5.8: Efficiency ratio of the control sample between DATA and MC for Belle.

$N_{\pi^0/\eta<}$ (MC)	$N_{\pi^0/\eta>}$ (MC)	$N_{\pi^0/\eta<}$ (Data)	$N_{\pi^0/\eta>}$ (Data)	$\epsilon_{\rm DATA}$	$\epsilon_{\rm MC}$	$R_{\rm DATA/MC}$
18406±140	275±21	2150±50	35±9	0.984±0.004	0.985±0.001	1.002±0.004

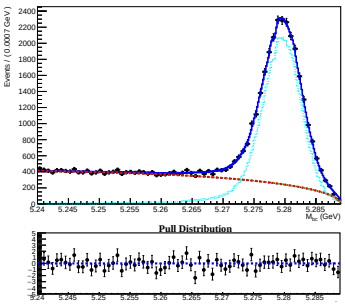
Table 5.9: Efficiency ratio of the control sample between DATA and MC for Belle II.



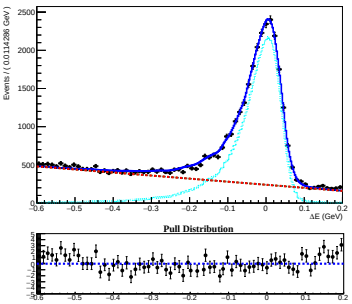
(a) $M_{\rm bc}$ Distribution for MC events rejected by π^0/η cut



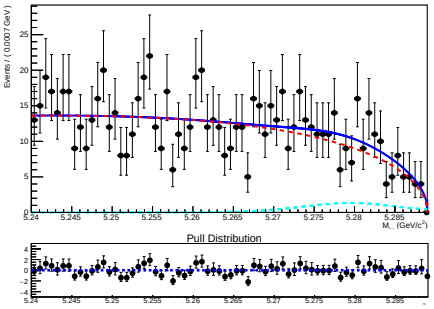
(b) ΔE Distribution for MC events rejected by π^0/η cut



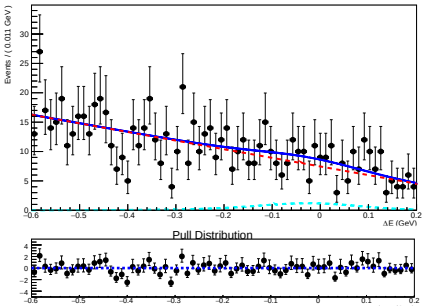
(c) $M_{\rm bc}$ Distribution for MC events selected by π^0/η cut



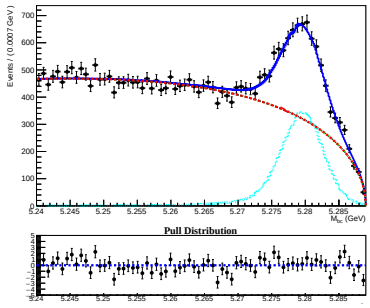
(d) ΔE Distribution for MC events selected by π^0/η cut



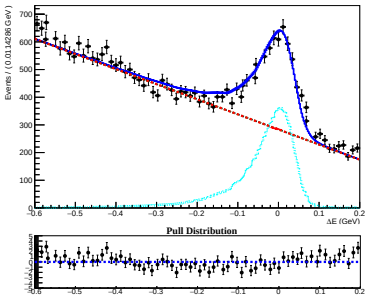
(e) $M_{\rm bc}$ Distribution for Data events rejected by π^0/η cut



(f) ΔE Distribution for Data events rejected by π^0/η cut

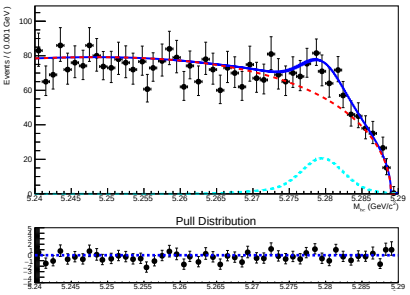


(g) $M_{\rm bc}$ Distribution for Data events selected by π^0/η cut

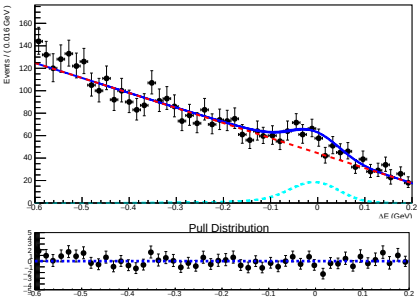


(h) ΔE Distribution for Data events selected by π^0/η cut

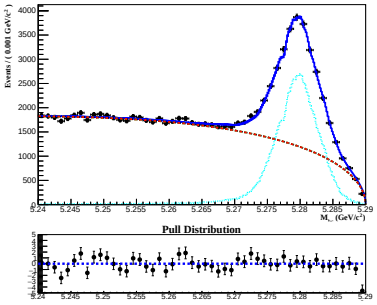
Figure 5.9: 2D $M_{\rm bc}$ and ΔE fits for Belle $B^0 \to K^{0*}\gamma$ control sample remove and selected by π^0/η veto cut. [a], [b], [c], and [d] are MC, while [e], [f], [g], and [h] are Data samples. Here MC events are not scaled to Data luminosity.



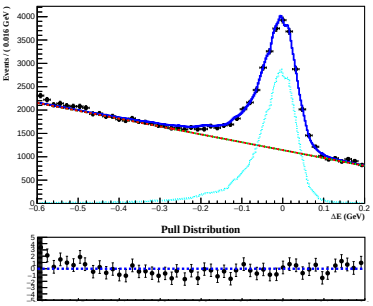
(a) $M_{\rm bc}$ Distribution for MC events rejected by π^0/η cut



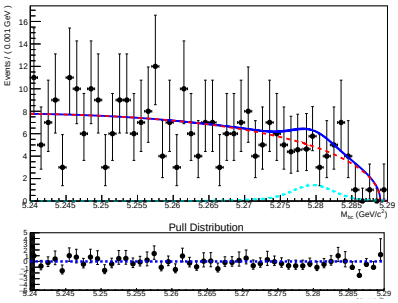
(b) ΔE Distribution for MC events rejected by π^0/η cut



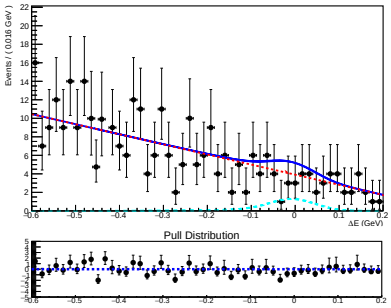
(c) $M_{\rm bc}$ Distribution for MC events selected by π^0/η cut



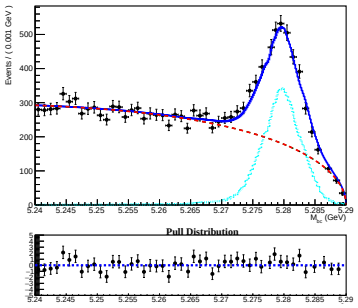
(d) ΔE Distribution for MC events selected by π^0/η cut



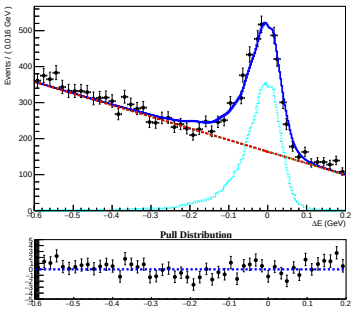
(e) $M_{\rm bc}$ Distribution for Data events rejected by π^0/η cut



(f) ΔE Distribution for Data events rejected by π^0/η cut



(g) $M_{\rm bc}$ Distribution for Data events selected by π^0/η cut



(h) ΔE Distribution for Data events selected by π^0/η cut

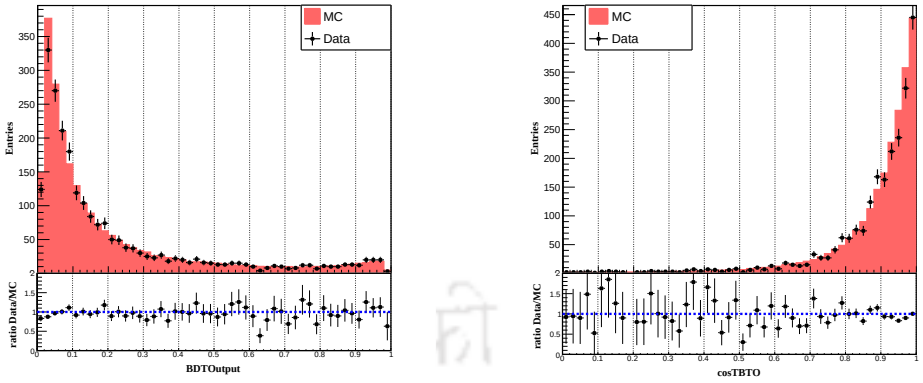
Figure 5.10: 2D $M_{\rm bc}$ and ΔE fits for Belle II $B^0 \to K^{0*}\gamma$ control sample remove and selected π^0/η veto cut. [a], [b], [c], and [d] are MC, while [e], [f], [g], and [h] are Data samples. Here MC events are not scaled to Data luminosity.

3.5 DATA-MC Agreement

3.5.1 DATA-MC Agreement with Off-resonance sample

This subsection describes off-resonance data studies using the entire dataset with an integrated luminosity of 42.3 fb^{-1}. Data recorded below the $\Upsilon(4S)$ energy is often referred to as off-resonance data. It consists of light quark events($u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$). Apart from that, QED events, $qq\gamma$ events, and two-photon processes are also present. The off-resonance data can be used to check the modeling of the input variables of the BDT for the continuum background suppression, as well as photon selection variables. The photon energy bias correction has been applied for the comparison. We normalize the MC to the number of events in the data. Figure 5.11 shows the DATA-MC comparison plots for the BDT output and `cosTBTO`. The remaining DATA-MC comparison plots can be found in `Appendix` (F).

For Belle, we compared the data samples for experiments 7-65, consisting of 89.50 fb^{-1} integrated luminosity and continuum background sample corresponding



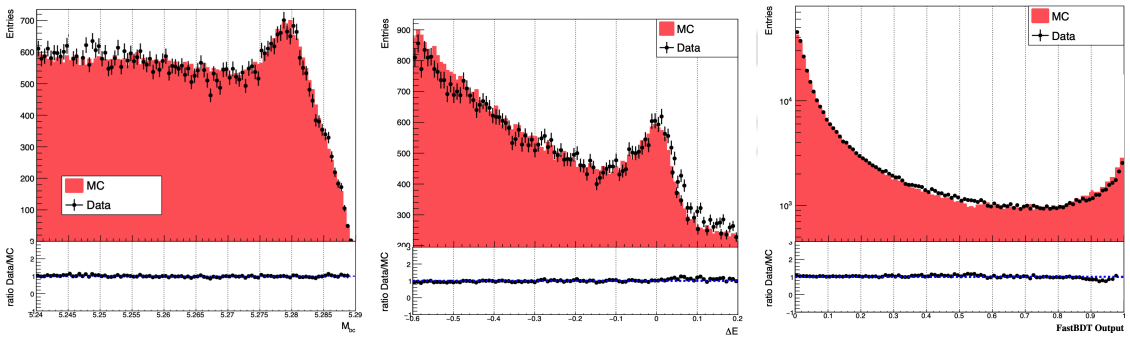
(a) BDT response (b) `cosTBTO`

Figure 5.11: DATA-MC Comparison for Belle II BDT response, and Event shape variable.

to one stream of MC sample. We observe a good data-MC agreement for all the distributions. The plots can be found in `Appendix` (E).

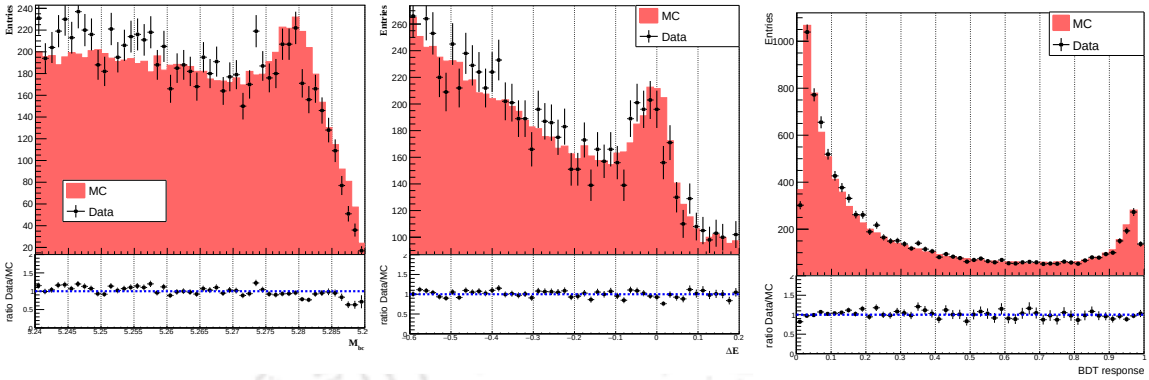
3.5.2 DATA-MC Agreement for Control Sample

A Data-MC agreement check was performed for $M_{\rm bc}, \Delta E$ and BDT output distributions using the control sample. We have applied the MVA training of the target mode to the control mode and then compared the BDT output distributions of the control mode. The data-MC comparison plots for $M_{\rm bc}, \Delta E$, and BDT output distributions are shown in Figure 5.12 and 5.13 for Belle and Belle II, respectively. After examining the shapes of the distributions, we will apply a systematic uncertainty to account for their discrepancy, which is described in sec 2.2.3.



(a) (b) (c)

Figure 5.12: DATA-MC comparison for (a) $M_{\rm bc}$ (b) ΔE (c) and $C_{\rm BDT}$ for Belle $B^0 \to K^*\gamma$ control mode.



(a) (b) (c)

Figure 5.13: DATA-MC comparison for (a) $M_{\rm bc}$ (b) ΔE (c) and $C_{\rm BDT}$ for Belle II $B^0 \to K^*\gamma$ control mode.

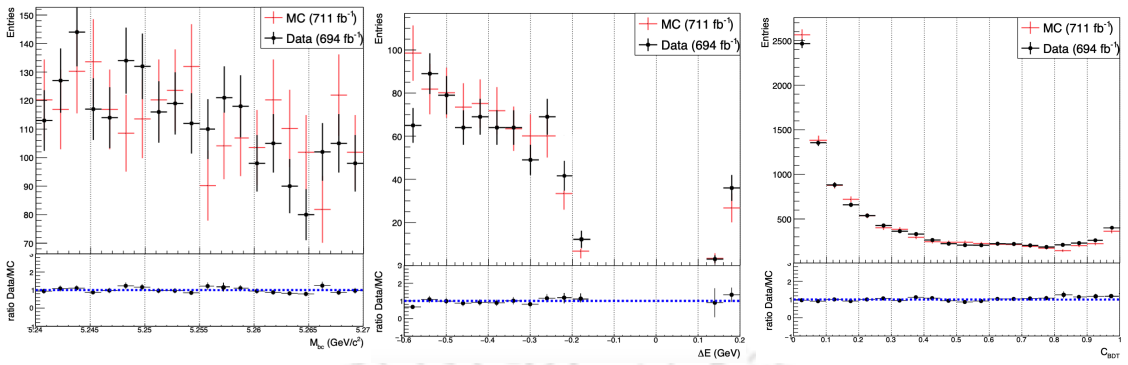
4.0 Data-MC difference in the side-bands for $B^0 \to \gamma\gamma$ in data

We have also looked at the full data sample in the sideband region for further data-MC checks. Table 5.10 defines the data sideband region, determined using $M_{\rm bc}$ and ΔE.

	Sideband region	
Fit Variable	Belle	Belle II
$M_{\rm bc}$ (GeV/c^2)	$5.24 < M_{\rm bc} < 5.27$	$5.24 < M_{\rm bc} < 5.27$
ΔE (GeV)	$-0.19 > \Delta E \| \| \Delta E > 0.14$	$-0.19 > \Delta E \| \| \Delta E > 0.15$

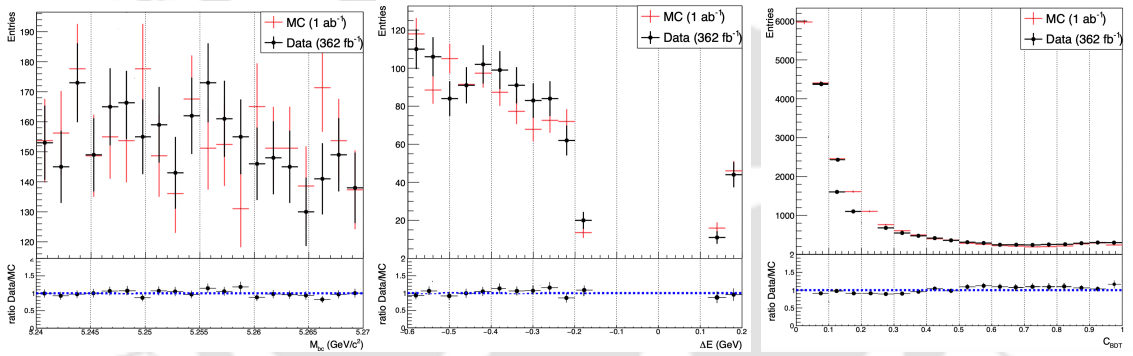
Table 5.10: Data-sideband for Belle and Belle II.

Figures 5.14a, 5.14b and 5.14c refer to the data-MC comparison for $M_{\rm bc}$, ΔE and $C_{\rm BDT}$ for the Belle full dataset, respectively. Figures 5.15a, 5.15b and 5.15c refer to the data-MC comparison for $M_{\rm bc}$, ΔE and $C_{\rm BDT}$ for the Belle II full dataset, respectively. From the data-MC comparison distributions, we do not observe any serious differences, though the results are limited by the statistics.



(a) $M_{\rm bc}$ (b) ΔE (c) $C_{\rm BDT}$

Figure 5.14: Data-MC difference plots for $M_{\rm bc}$, ΔE and $C_{\rm BDT}$ variable when Belle MC is considered for $\mathcal{L} = 711$ fb^{-1} and Data for $\mathcal{L} = 694$ fb^{-1}. The red and black (filled with circular) points with error bars represent MC and Data, respectively. The MC histogram is scaled down to the number of events in the data histogram.



(a) $M_{\rm bc}$ (b) ΔE (c) $C_{\rm BDT}$

Figure 5.15: Data-MC difference plots for $M_{\rm bc}$, ΔE and $C_{\rm BDT}$ variable when Belle II MC is considered for $\mathcal{L} = 1$ ab^{-1} and Data for $\mathcal{L} = 362$ fb^{-1}. The red and black (filled with circular) points with error bars represent MC and Data, respectively. The MC histogram is scaled down to the number of events in the data histogram.

5.0 Summary

In this chapter, we have discussed the validation methods adopted to crosscheck the analysis procedure and PDF parametrization. In the validation method, the control sample study is performed using $B^0 \to K^*\gamma$ to calibrate the BDT and π^0/η veto. The measured branching fraction for $B^0 \to K^*\gamma$ was found to be within 1.62σ and 1.44σ of the previous Belle result and PDG value, respectively. Furthermore, the data-MC comparison was performed for the fitting variable of the control sample, off-resonance data sample and side-band regions for the target mode and good data-MC agreement were observed.

Chapter 6

Result and Conclusion

1.0 Introduction

This chapter presents the final results for the decay $B^0 \to \gamma\gamma$ using the collision data at $\Upsilon(4S)$ resonance from the Belle and Belle II experiments. It provides a brief introduction to the systematic uncertainties and their impact on the results of the analysis. The estimation of signal significance, branching fraction calculation and the estimation of Bayesian upper limit (UL) at 90% confidence level (CL) for $B^0 \to \gamma\gamma$ analysis are explained in detail. Finally, we conclude with a brief discussion of the possible implications of the results of this analysis and the possibility of observing this decay with a larger dataset at Belle II.

2.0 Sources of the Systematic Uncertainty

Systematic errors consistently affect all measurements in the same way, leading to correlated results across repeated trials. Consequently, these errors cannot be corrected by any number of measurements. Such systematic uncertainties in experiments are due to imperfections associated with instrument calibration, incomplete knowledge of experimental configurations, acceptance and sensitivity of detector components, as well as uncertainties in the theoretical and experimental parameters utilized in the experiments, etc.

The systematic uncertainties can be classified into two types: additive and multiplicative. The multiplicative uncertainties do not affect the signal yield and signal significance of the decay channel but affect its $\mathcal{B}$. These uncertainties arise due to the uncertainty in signal reconstruction efficiency, the number of B mesons and the fraction $f_{B^0\bar{B}^0}$. The additive uncertainties reduce the signal significance of the observed peak and change the $\mathcal{B}$ of the decay. The additive systematics arise due to the fit procedure, which includes the uncertainty due to PDF parametrization and fit bias, respectively. All the sources of uncertainty are added in the quadrature to calculate the final uncertainty.
All the possible sources of the systematic uncertainties are discussed below:

2.1 Uncertainty on the Signal Reconstruction Efficiency

The systematic uncertainty attributed to signal reconstruction efficiency (ϵ_{rec}) is considered multiplicative. It is determined by summing the systematic uncertainties associated with photon reconstruction, π^0/η veto, BDT cut, number of $B\bar{B}$ pairs, timing cut and MC statistics in quadrature. Detailed descriptions of these uncertainties are provided in the subsequent subsections.

2.1.1 Photon Reconstruction Efficiency

Data-MC correction of photon efficiency for Belle II is studied in [105]. The data-MC selection efficiency is calculated using dedicated weight tables. The data-MC correction factor for the 1st and 2nd γ comes out to be $\epsilon_{\mathrm{DATA/MC}(\gamma_1)} = 1.004 \pm 0.013$ and $\epsilon_{\mathrm{DATA/MC}(\gamma_2)} = 1.005 \pm 0.014$, respectively. This uncertainty is correlated for the two γs. A total systematic uncertainty of 2.7% is assigned.

For Belle, the uncertainty on the photon reconstruction efficiency is computed to be 2.0% per photon by comparing the detection efficiency of the real data and MC samples of radiative Bhabha events [106]. This uncertainty is correlated for the two γs a total systematic uncertainty of 4.0% is assigned.

2.1.2 Uncertainty due to π^0/η Veto

Efficiency due to the π^0/η veto is calibrated by $B^0 \to K^*\gamma$ control samples. We calculate the relative selection efficiency and the efficiency disparity between MC and data. The uncertainties of the calibration factors are taken into account as systematics for the Belle and Belle II study. Table 5.8 and 5.9 summarize the efficiencies and DATA/MC ratio for Belle and Belle II, respectively. We assign a systematic uncertainty of 0.4% and 0.6% for two photons for Belle and Belle II, respectively.

2.1.3 Uncertainty due to FastBDT cut

The systematic uncertainty due to MVA criteria are studied using the control sample $B^0 \to K^*\gamma$. We compare the selection efficiency of MVA for the control sample using data and MC to assign a systematic. The selection efficiency is calculated using the number of control sample events in data/MC before and after applying the FastBDT cut. The error on the data-MC ratio ($\epsilon_{\mathrm{Data/MC}}$) for selection rejection efficiency is considered a systematic uncertainty for the MVA selection. Table 5.6 and 5.7 summarize the efficiencies and DATA/MC ratio of cut efficiency for Belle and Belle II, respectively. We assign systematic uncertainty of 0.4% and 0.9% for Belle and Belle II, respectively.

2.1.4 Uncertainty due to MC Statistics

The uncertainty on ϵ is calculated from the signal MC sample using the following formula, which obeys a binomial distribution,

$$\Delta\epsilon = \sqrt{\frac{\epsilon \times (1-\epsilon)}{\mathrm{N}}}$$

where N is the total number of signal MC events at the generator level. In our case for Belle II, $\epsilon = 30.8\%$ and N= 100,000. The signal efficiency is computed as $\epsilon = (30.8 \pm 0.1)\%$. Thus the uncertainty due to MC statistics ($\frac{\Delta\epsilon}{\epsilon}$) is computed to be 0.3%.
Similarly, for Belle, $\epsilon = 23.3\%$ and N= 100,000. The signal efficiency is computed as $\epsilon = (23.3 \pm 0.1)\%$. Thus the uncertainty due to MC statistics is computed to be 0.4%.

2.1.5 Uncertainty on the Number of $B\bar{B}$ pairs

The number of $B\bar{B}$ pairs corresponding to the LS1 dataset as provided by the performance group is $N_{B\bar{B}} = (387 \pm 6) \times 10^6 \pm 1.45\%$. The number of neutral B meson pairs is define as $N_{B^0\bar{B}^0} = f^{00} \times N_{B\bar{B}}$, where f^{00} is the fraction of $b\bar{b}$ events that hadronize to $B^0\bar{B}^0$. f^{00} have a relative uncertainty of around 2.5% [104]. We have assigned the combined systematic uncertainty of 2.9% for Belle II.

For Belle, the number of $B\bar{B}$ was measured to be $(753 \pm 10) \times 10^6$ $B\bar{B}$ pairs in real data (exp.7-65), and the systematic error is 1.3% [107]. f^{00} have a relative uncertainty of around 2.5%. The two errors are combined in quadrature and computed to be 2.8%.

2.1.6 Timing Cut for Belle analysis

We calculate the efficiency of data when selection criteria on timing cut are applied using the control sample $B^0 \to K^*\gamma$ in Belle analysis. 2D M_{bc} and ΔE distribution is fitted with and without the timing cuts to obtain the signal yields. The efficiency of Data and MC is defined as $\epsilon_{\mathrm{DATA}} = \frac{N^{\mathrm{yield}}_{\mathrm{DATA}}(\text{With timing cut})}{N^{\mathrm{yield}}_{\mathrm{DATA}}(\text{without timing cut})}$ and $\epsilon_{\mathrm{MC}} = \frac{N^{\mathrm{yield}}_{\mathrm{MC}}(\text{With PhotonID cut})}{N^{\mathrm{yield}}_{\mathrm{MC}}(\text{without PhotonID cut})}$, respectively.

DATA/MC	Selection	Yield	ϵ
DATA	Without Timing cut	4325 ± 68	0.9669 ± 0.0201
	With Timing cut	4182 ± 66	
MC	Without ID cut	28700 ± 150	0.9998 ± 0.0001
	With ID cut	28695 ± 165	

Table 6.1: Table to estimate systematic uncertainty due to Timing cut in Belle analysis.

Table 6.1 shows the efficiency value obtained using this procedure. There is a disparity between the efficiency of a timing cut observed in real data, which is 97% and the efficiency reported by a MC simulation, which is nearly 100%. We use a more conservative estimate for the cut's efficiency based on the real data rather than an artificial 100% efficiency for the MC simulation. The efficiency and uncertainty for two photons is 0.935 ± 0.026. We will use the central value to calibrate the signal efficiency from the data and will assign 2.8% systematic uncertainty for the timing cut.

2.2 Uncertainty due to fit Procedure

Uncertainty due to the fit procedure affects the signal yield. It includes the uncertainty due to PDF parametrization and fit bias and is estimated using the quadratic sum of these uncertainties. This uncertainty is additive in nature.

2.2.1 Uncertainty due to PDF Parametrization

We assign systematic uncertainty to the signal yield by varying all the fixed parameters by ± 1 standard deviation about the fixed values obtained from the MC study. The difference in the new and nominal yield is considered as a systematic uncertainty. We have used an analytical PDF shape Crystal Ball function for $M_{\rm bc}$ and ΔE to obtain systematics due to the signal component of fit, which uses Kernel Density PDF. For signal $C'_{\rm BDT}$ component is changed from a $0^{\rm th}$ order polynomial to $1^{\rm st}$ order polynomial. Table 6.2 and 6.3 list the signal yields by varying the fixed MC values by $\pm\sigma$ for Belle and Belle II, respectively. We add the yield deviations in quadrature to evaluate the total uncertainty associated with PDF parameterization of $^{+0.56}_{-0.48}$ and $^{+0.30}_{-0.32}$ events for Belle and Belle II, respectively.

Variable		Function	Parameter	MC Value	Yield($+\sigma$)	Yield($-\sigma$)	ΔYield
$M_{\rm bc}$	Signal	CB	μ	5.279679 ± 0.000038	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
			σ	0.003439 ± 0.000033	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
			α	1.201 ± 0.036	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
			η	3.18 ± 0.21	9.3 ± 5.0	9.5 ± 5.1	$^{+0.10}_{-0.10}$
	Background	Argus	Endpoint	5.289355 ± 0.000045	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
ΔE	Signal	CB	μ	-0.02376 ± 0.0013	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
			σ	0.05621 ± 0.00086	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
			α	0.4254 ± 0.0093	9.3 ± 5.0	9.5 ± 5.1	$^{+0.10}_{-0.10}$
			η	54 ± 34	9.5 ± 5.1	9.3 ± 5.0	$^{+0.10}_{-0.10}$
$C'_{\rm BDT}$	Signal	$1^{\rm st}$ order Polynomial	a0	2.0 ± 2.5	9.6 ± 5.2	9.2 ± 4.9	$^{+0.20}_{-0.20}$
	Background	Exponential	λ_1	-5.454 ± 0.22	9.9 ± 5.2	9.1 ± 4.9	$^{+0.50}_{-0.40}$
			λ_2	-55.80 ± 7.7	9.4 ± 5.0	9.4 ± 5.0	$^{+0.00}_{-0.00}$
Total systematic due to the fixed PDF parameters (events)							$^{+0.56}_{-0.48}$

Table 6.2: List of the parameters fixed to MC values to extract the signals from the real data for Belle.

Variable		Function	Parameter	MC Value	Yield(+σ)	Yield(-σ)	ΔYield
$M_{\rm bc}$	Signal	CB	μ	5.279848 ± 0.000028	2.0 ± 3.4	2.0 ± 3.4	$^{+0.00}_{-0.00}$
			σ	0.003382 ± 0.000024	2.0 ± 3.4	2.0 ± 3.4	$^{+0.00}_{-0.00}$
			α	1.283 ± 0.031	1.9 ± 3.4	2.0 ± 3.4	$^{+0.10}_{-0.00}$
			η	3.80 ± 0.23	1.9 ± 3.4	2.0 ± 3.4	$^{+0.10}_{-0.00}$
	Background	Argus	Endpoint	5.28987 ± 0.00029	2.0 ± 3.4	2.0 ± 3.4	$^{+0.00}_{-0.00}$
ΔE	Signal	CB	μ	-0.021776 ± 0.00089	2.0 ± 3.4	2.0 ± 3.4	$^{+0.00}_{-0.00}$
			σ	0.05658 ± 0.00059	2.0 ± 3.4	2.0 ± 3.4	$^{+0.00}_{-0.00}$
			α	0.579 ± 0.016	1.9 ± 3.4	2.0 ± 3.4	$^{+0.10}_{-0.00}$
			η	33 ± 13	2.1 ± 3.4	2.0 ± 3.4	$^{+0.10}_{-0.00}$
$C'_{\rm BDT}$	Signal	$1^{\rm st}$ order Polynomial	a0	1.439 ± 6.7	2.0 ± 3.4	1.9 ± 3.4	$^{+0.00}_{-0.10}$
	Background	Exponential	λ_1	-4.434 ± 0.55	1.8 ± 3.3	2.3 ± 3.4	$^{+0.20}_{-0.30}$
			λ_2	-22.21 ± 6.6	2.0 ± 3.4	2.0 ± 3.4	$^{+0.00}_{-0.00}$
Total systematic due to the fixed PDF parameters (events)							$^{+0.28}_{-0.32}$

Table 6.3: List of the parameters fixed to MC values to extract the signals from the real data for Belle II.

2.2.2 Fit Bias

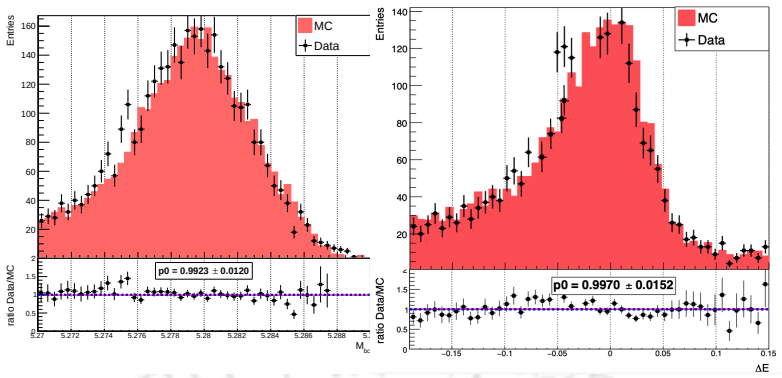
For Belle, as mentioned in section 3.1, we have obtained an average fit bias of 0.10 events, which is assigned as the systematic uncertainty due to fit bias. From the linearity test, we have added a conservative systematic uncertainty of 0.13 event. We have assigned the combined systematic of +0.16 event.

Similarly, as mentioned in section 3.2, we have obtained an average fit bias of 0.03 events, which is assigned as the systematic uncertainty due to fit bias. From the linearity test, we have added a conservative systematic uncertainty of 0.1 event. We have assigned the combined systematic of +0.10 event from the fit bias and linearity test.

2.2.3 Uncertainty due to Shape Modeling

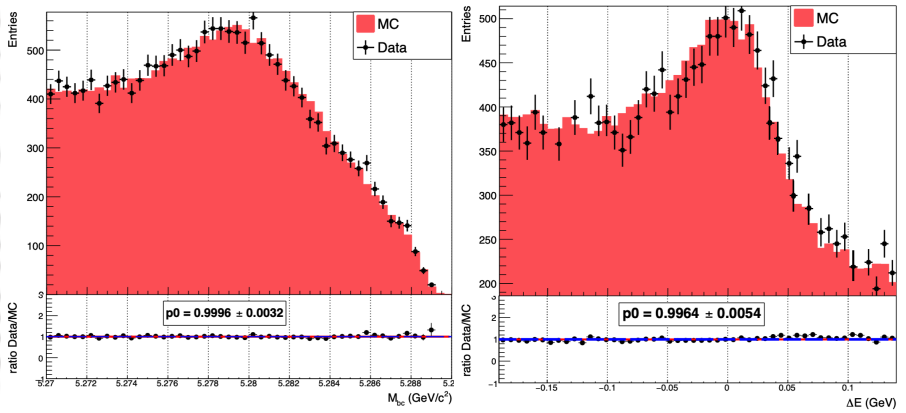
The systematic due to shape modeling between Data and MC is studied using a control sample $B^0 \to K^*\gamma$. in section 3.5.2. The Figures 6.1 show the $M_{\rm bc}$ and ΔE distribution after applying all selections, including the BDT cut for $M_{\rm bc} > 5.27$ GeV/c^2 and $-0.19 < \Delta E < 0.15$ GeV region. The deviation from unity in the data/MC ratio with uncertainty will be considered as a source of systematic uncertainty, which is 1.2% for $M_{\rm bc}$ and 1.5% for ΔE. The quadrature sum of systematic uncertainty is determined to be 2% (0.04 event) and will be incorporated as a systematic uncertainty in the analysis.

For Belle, The Figures 6.2 show the $M_{\rm bc}$ and ΔE distribution after applying all selections, including the BDT cut. The deviation from unity in the data/MC ratio with uncertainty will be considered as a source of systematic uncertainty, which is 0.32% for $M_{\rm bc}$ and 0.65% for ΔE. The quadrature sum of systematic uncertainty is determined to be 0.72% (0.06) and will be incorporated as a systematic uncertainty in the analysis.



(a) (b)

Figure 6.1: DATA-MC comparison for (a) $M_{\rm bc}$ (b) ΔE for $B^0 \to K^*\gamma$ control mode for Belle II.



(a) (b)

Figure 6.2: DATA-MC comparison for (a) $M_{\rm bc}$ (b) ΔE for $B^0 \to K^*\gamma$ control mode for Belle.

Table 6.4 and 6.5 summarizes all the multiplicative and additive systematic uncertainties associated with Belle analysis.

Systematic uncertainties	**Fractional uncertainties(%)**
Photon Efficiency	4.0
π^0/η veto	0.4
FastBDT cut	0.4
Reconstruction Efficiency	0.4
Number of $B\overline{B}$ pairs	2.8
Timing Cut	2.8
Sum in quadrature	5.7

Table 6.4: Summary of multiplicative systematic uncertainties for Belle

Systematic uncertainties	uncertainties(events)
Pdf parameterization	$^{+0.56}_{-0.48}$
Fit Bias	$+0.14$
Shape Modeling	$+0.06$
Sum in quadrature	$^{+0.58}_{-0.48}$

Table 6.5: Summary of additive systematic uncertainties for Belle

Table 6.6 and 6.7 summarize all the multiplicative and additive systematic uncertainties associated with Belle II analysis.

Systematic uncertainties	Fractional uncertainties(%)
Photon Efficiency	2.7
π^0/η veto	0.6
FastBDT cut	0.9
Reconstruction Efficiency	0.5
Number of $B\overline{B}$ pairs	2.9
Sum in quadrature	4.1

Table 6.6: Summary of multiplicative systematic uncertainties for Belle II

Systematic uncertainties	uncertainties(events)
Pdf parameterization	$^{+0.28}_{-0.32}$
Fit Bias	$+0.10$
Shape discrepancy	$+0.04$
Sum in quadrature	$^{+0.30}_{-0.32}$

Table 6.7: Summary of additive systematic uncertainties for Belle II

3.0 Data results for $B^0 \to \gamma\gamma$

In this section, we report the results obtained from the real data for Belle and Belle II experiments at $\Upsilon(4S)$, corresponding to integrated luminosities of 694 fb^{-1} and 362 fb^{-1}, respectively.

To combine the results from the Belle and Belle II datasets, we perform a simultaneous unbinned maximum likelihood fit to the $M_{\rm bc}$, ΔE, and $C'_{\rm BDT}$ distributions. To extract the signal and background yields from the simultaneous fit model, we fix the $C'_{\rm BDT}$ background exponential parameters (λ_1 and λ_2) and the endpoint of the ARGUS function and float all the other background parameters in the final fit model. The signal PDFs have no parameters. In addition, we float the signal and background yields. The PDFs and parameter information used in the simultaneous fit are given in Table 6.8. The fixed parameters obtained from the MC are mentioned in section 2.3 and 2.5 for Belle and Belle II, respectively. We have parameterized

the Belle and Belle II yields as:

$$N_{\rm b1} = 2 \times \mathcal{B} \times N_{B\bar{B}_{\rm b1}} \times \epsilon_{\rm b1} \times f^{00}$$

and

$$N_{\rm b2} = 2 \times \mathcal{B} \times N_{B\bar{B}_{\rm b2}} \times \epsilon_{\rm b2} \times f^{00}$$

where $f^{00} = (48.4 \pm 1.2)\%$ and $\mathcal{B}$ parameters are shared between Belle and Belle II datasets while performing the fit. $N_{B\bar{B}_{\rm b1}}$ and $N_{B\bar{B}_{\rm b2}}$ are the number of of $B\bar{B}$ pairs at the Υ(4S) resonance and $\epsilon_{\rm b1}$ and $\epsilon_{\rm b2}$ are efficiencies of Belle and Belle II, respectively.

Variable		**Function**	**Parameter**	**Comment**
$M_{\rm bc}$	Signal	2D KDE	-	No parameter
	Background	Argus	Endpoint	Fixed
			Curvature(argpar)	Floated
ΔE	Signal	2D KDE	-	No parameter
	Background	1$^{\rm st}$ Order Chebyshev polynomial	Coefficient(a1)	Floated
$C'_{\rm BDT}$	Signal	0$^{\rm th}$ order Polynomial	-	No parameter
	Background	Exponential	λ_1	Fixed
			λ_2	Fixed
			frac	Floated

Table 6.8: PDF used in the Fit model of $B^0 \to \gamma\gamma$ sample and parameter information. Parameter λ_1 and λ_2 represents the slopes of the exponential PDF.

The fit projections obtained from a simultaneous fit in the signal regions are shown in Figure 6.3 for the fit variables, $M_{\rm bc}$, ΔE and $C'_{\rm BDT}$. We determine the total signal (background) yield to be $11^{+6.5}_{-5.5}$ (931±31) events, where the uncertainties are statistical only. From the fits to the two independent datasets, we obtain $9.1^{+5.6}_{-4.4}$ ($1.9^{+4.2}_{-2.8}$) signal events and 615 ± 25 (317 ± 18) background events for Belle (Belle II), respectively. The data-MC comparison of yields is shown in Table 6.9.

	Belle		Belle II	
Type	MC	Data	MC	Data
Signal	$2.4^{+4.1}_{-2.3}$	$9.1^{+5.6}_{-4.4}$	$1.7^{+3.9}_{-2.2}$	$1.9^{+4.2}_{-2.8}$
Background	553±24	615±25	286±17	317±18

Table 6.9: Comparing the signal and background yield from the MC and Data.

3.1 Branching Fraction and significance level

We calculate the branching fraction using the formula,

$$\mathcal{B}(B^0 \to \gamma\gamma) = \frac{N_{\rm sig}^{\rm fit}}{N_{B\bar{B}} \times 2 \times f^{00} \times \epsilon_{\rm rec}}, \tag{6.1}$$

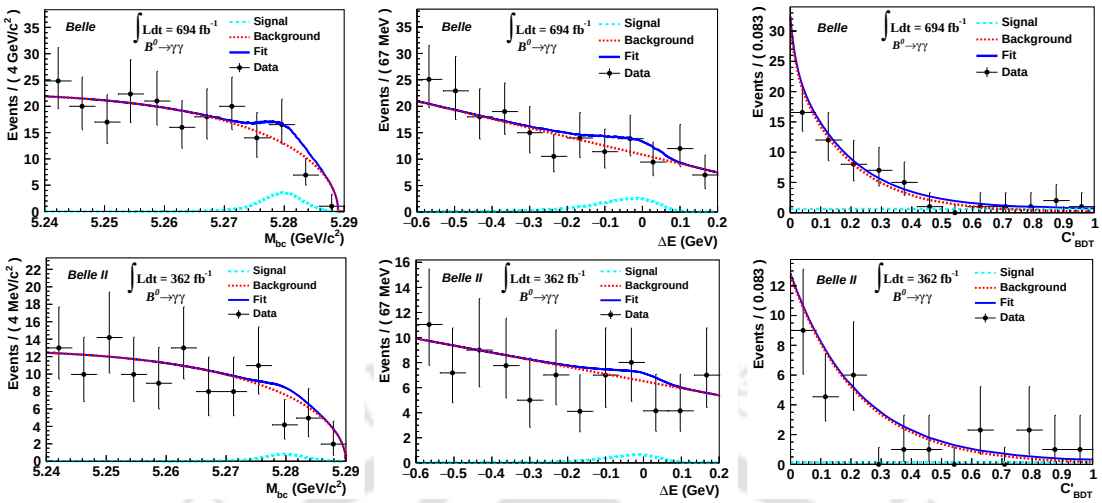


Figure 6.3: Signal enhanced projections of $M_{\rm bc}$ (left), ΔE (middle), and $C'_{\rm BDT}$ (right) for the $B^0 \to \gamma\gamma$ analysis using combined Belle (top) and Belle II (bottom) dataset. Each plot is generated by applying the signal region selection criteria on the variables other than the plotted variable. The signal regions for the two variables are as follows, 5.27 GeV/c^2 $< M_{\rm bc} <$ 5.29 GeV/c^2 and -0.19 GeV $< \Delta E <$ 0.14 GeV for Belle and 5.27 GeV/c^2 $< M_{\rm bc} <$ 5.29 GeV/c^2 and -0.19 GeV $< \Delta E <$ 0.15 GeV for Belle II. The cyan (dashed), red (dashed), and blue (solid) color distributions represent the signal, continuum background, and total fit function, respectively. Points with error bars represent data.

where $N_{\rm sig}^{\rm fit}$ represents the signal yield obtained from the fit, $N_{B\bar{B}} = (753 \pm 10) \times 10^6$ ($(387 \pm 6) \times 10^6$) denotes the number of $B\bar{B}$ pairs at the $\Upsilon(4S)$ resonance for Belle (Belle II), $f^{00} = (48.4 \pm 1.2)\%$ indicates the ratio of $\Upsilon(4S)$ decaying into $B^0\bar{B}^0$, and $\epsilon_{\rm rec} = 23.3\%$ (30.8%) represents the signal reconstruction efficiency for Belle (Belle II).

The statistical uncertainty associated with the $\mathcal{B}$ is calculated using the formula,

$$\Delta\mathcal{B} = \frac{\Delta N_{\rm sig}^{\rm fit}}{N_{\rm sig}^{\rm fit}} \times \mathcal{B}, \tag{6.2}$$

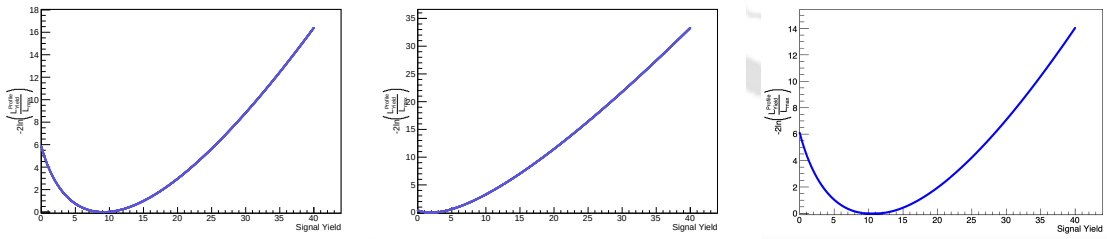
where, $\Delta N_{\rm sig}^{\rm fit} = {}^{+5.6}_{-4.4}$ ($^{+4.2}_{-2.8}$) and $N_{\rm sig}^{\rm fit} = 9.1$ (1.9), is the uncertainty associated with the signal yield obtained from the fit for Belle (Belle II), respectively. From equation 6.2, the statistical uncertainty on the branching fraction is found to be $^{+3.3}_{-2.6}\times10^{-8}$ ($^{+3.7}_{-2.4}\times10^{-8}$) for Belle (Belle II). The systematic uncertainty is the quadratic sum of multiplicative and additive systematics reported in Table 6.4 (6.6) and 6.5 (6.7) for Belle (Belle II), respectively. This corresponds to a systematic uncertainty of 0.5×10^{-8} (0.3×10^{-8}) on the branching fraction for Belle (Belle II).

We have measured the BF for Belle and Belle II to be $(5.4^{+3.3}_{-2.6} \pm 0.5) \times 10^{-8}$ and $(1.7^{+3.7}_{-2.4} \pm 0.3) \times 10^{-8}$, respectively. The combined study yields a measured BF of $(3.7^{+2.2}_{-1.8} \pm 0.5) \times 10^{-8}$. The first uncertainty is statistical while the second is system-

atic.

The signal significance is calculated as $Z = \sqrt{-2\ln\lambda}$, where $\lambda = \frac{\mathcal{L}_0}{\mathcal{L}_{\max}}$; $\mathcal{L}_0$ is the plain likelihood corresponding to background only hypothesis and $\mathcal{L}_{\max}$ is the maximum value of plain likelihood value of nominal fit. The observed significance using the combined dataset is 2.5σ, which is less than 3σ. Therefore, we calculate the UL considering the Bayesian approach.

Figure 6.4 shows the profile likelihood as a function of signal yield for Belle, Belle II, and combined datasets, respectively.



(a) (b) (c)

Figure 6.4: Profile Likelihood as a function of signal yield (a) Belle, (b) Belle II, and (c) Combined datasets.

3.2 Upper Limit on the branching fraction

The process of obtaining the UL for Belle, Belle II, and the combined datasets is enumerated below:

1. Construct the profile likelihood, $(-2\frac{\mathcal{L}_{\rm i}}{\mathcal{L}_{\max}})$ where $\mathcal{L}_{\rm i}$ is the plain likelihood with a hypothesis of i signal yield and $\mathcal{L}_{\max}$ is is the maximum value of plain likelihood value of nominal fit.

2. The profile likelihood ratio, $\lambda(x)$ is convoluted with a Gaussian function of zero mean and width equal to the total systematic uncertainties which can be expressed as follows:

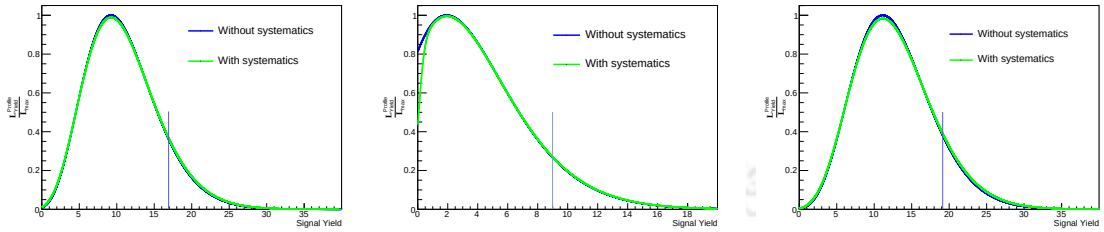
$$f(x) = \int \lambda(u)\frac{1}{\sqrt{2\pi}w}e^{-\frac{(x-u)^2}{2w^2}}\,du$$

   Here, w = represents the width of the Gaussian function, which incorporates the total systematic uncertainty for Belle, Belle II, and the combined dataset.

3. Integrate the convoluted plain likelihood up to 90% of its area to estimate the upper limit(x_{upper}) at 90% CL:

$$\int_0^{x_{upper}} g(x)\,dx = 0.9$$

Figure 6.5 shows the comparison between unconvoluted and convoluted plain likelihood ratios as a function of signal yield obtained from the real data for (a) Belle, (b) Belle II, and (c) Combined datasets.



(a) (b) (c)

Figure 6.5: Comparison between unconvoluted and convoluted plain likelihood ratios as a function of signal yield obtained from the real data for (a) Belle, (b) Belle II, and (c) Combined datasets. The violet line represents the 90% CL.

To estimate the UL on $\mathcal{B}$ the following formula is used:

$$\mathcal{B}(B^0 \to \gamma\gamma) = \frac{N_{\rm sig}^{\rm fit}(@90\%\ {\rm CL})}{N_{B\bar{B}} \times 2 \times f^{00} \times \epsilon_{\rm rec}} \tag{6.3}$$

where, $N_{\rm sig}^{\rm fit}$ is the signal yield at 90% CL = 19.2 (with systematics), $N_{B\bar{B}} = (1.1 \pm 12) \times 10^9$ is the number of $B\bar{B}$ pairs at $\Upsilon(4S)$ resonance for the combined datasets, $f^{00} = (48.4 \pm 1.2)\%$ the ratio of $\Upsilon(4S)$ going to $B^0\bar{B}^0$ and $\epsilon_{\rm rec} = 28\%$ is the signal reconstruction efficiency.

We obtain a UL on the BF of less than 6.3×10^{-8} at 90% CL (without systematics) and less than 6.4×10^{-8} at 90% CL (with systematics). This is an improvement by a factor of five from the previous upper limit set by Babar ($< 3.2 \times 10^{-7}$) [13].

The measured BF and the resulting upper limits on $\mathcal{B}(B^0 \to \gamma\gamma)$ at a 90% CL, with the systematic uncertainties, are summarized in Table 6.10.

	$\mathcal{B}(B^0 \to \gamma\gamma)$	$\mathcal{B}(B^0 \to \gamma\gamma)$ (at 90% CL)
Belle	$(5.4^{+3.3}_{-2.6} \pm 0.5) \times 10^{-8}$	$< 9.9 \times 10^{-8}$
Belle II	$(1.7^{+3.7}_{-2.4} \pm 0.3) \times 10^{-8}$	$< 7.4 \times 10^{-8}$
Combined	$(3.7^{+2.2}_{-1.8} \pm 0.5) \times 10^{-8}$	$< 6.4 \times 10^{-8}$

Table 6.10: Summary of results on $\mathcal{B}(B^0 \to \gamma\gamma)$ measurement and UL at 90% CL.

3.3 Summary and Conclusion

In summary, we present the results for the first Belle and Belle II measurement using the combined dataset of 753×10^6 $B\bar{B}$ pairs collected by the Belle experiment and 387×10^6 $B\bar{B}$ pairs collected by the Belle II experiment. The combined study yields a measured BF of

$$\mathcal{B}(B^0 \to \gamma\gamma) = (3.7^{+2.2}_{-1.8} \pm 0.5) \times 10^{-8}$$

where the first uncertainty is statistical and the second is systematic. Since the significance is 2.5σ, we set 90% C.L. upper limit of $\mathcal{B}(B^0 \to \gamma\gamma) < 6.4 \times 10^{-8}$ using the Bayesian approach. This is the most stringent UL estimated for this decay representing an improvement by a factor of five compared to the previous limit (3.2×10^{-7}) set by the BaBar experiment [13]. The measured branching fraction is in good agreement within 0.9σ of theoretical expectations [14]. The observation of this decay process is imminent at the Belle II experiment with larger datasets, which started the Run2 operation in February 2024 after the long shutdown (LS1).

3.4 Outlook

During the LS1, Belle II received detector upgrades that improved photon energy resolution and vertex resolution for better background suppression and tracking capabilities. We can also enhance our analysis techniques by developing advanced machine-learning algorithms for signal/background separation and refining calibration and alignment procedures. The current measurement benefits from more data and improved detector understanding, resulting in greater precision. We plan to collect about 4 ab^{-1} in Run2 [108]. With a signal significance of 2.5σ, we expect to observe the decay with an integrated luminosity of 6 ab^{-1} and hope to find evidence of the decay in the Run2 sample. After the observation, the direct CP violation can be measured using the flavor tagger. With the entire Belle II dataset of 50 ab^{-1} it should be possible to measure $A_{CP}(B^0 \to \gamma\gamma)$ (direct CP asymmetries) with the precision of about 10%.

Appendix A

HadronB skimming

Collision of e^+e^- at the interaction point (IP) leads to the production of many physics processes such as hadronic events, continuum events, Bhabha events, two-photon processes, τ pairs, and beam backgrounds. HadronB skims $B\overline{B}$ pairs at about 99% and continuum events with high efficiencies from the two-photons, τ-pairs, and radiative and non-radiative Bhabha events. The following variables are constructed to skim the events:
Good tracks: An event must have $nTrk \geq 3$ good charged tracks. Charged tracks with the closest approach near the IP; $|dr| < 2$ cm, $|dz| < 4$ cm, and transverse momentum $P_t > 100$ MeV/c.

Good cluster: Clusters produced due to Bhabha events have shallow angles. To remove these events, two-photon events, and beam-gas events, we require the ECL to be greater than 1 within the fiducial volume $17° < \theta < 150°$.

Momentum balance (P_z): Sum of the z-components of the track and cluster momentum. We require $\sum |p_z| < 0.5\sqrt{s}$.

Visible energy (E_{vis}): The visible energy (E_{vis}) is the sum of the 'good track momenta' and 'good photon' energies in an event, where a 'good photon' refers to energy depositions in the ECL that do not match with charged tracks in the CDC. $E_{vis} \geq 0.2\sqrt{s}$ condition is used to reject the two-photon events.

Energy sum (E_{sum}): The calorimeter energy sum (E_{sum}), i.e., the sum of energies of 'good clusters' in the ECL is required to satisfy $0.18 < E_{sum}\sqrt{s} < 0.8$. 'Good clusters' have energy deposit > 100 MeV in the ECL. This requirement suppresses the Bhabha, two-photon, beam-gas, and τ events. These selection criteria are highly efficient ($\sim 99\%$) in selecting the hadronic events but at the same time serve to be too strict for the continuum events, $e^+e^- \to q\overline{q}$, $q = u, d, s$ and c).

Heavy Jet Mass (HJM): To increase some efficiency of the continuum events, the Heavy Jet Mass (HJM) variable, which is an invariant mass of the τ particles,

is introduced as a conditional criterion. However, a study on the HJM variable revealed that its shift is deeply correlated to the shift in the visible energy, E_{vis}. So to cancel the correlations, their ratio HJM/E_{vis} along with the conditional selection criteria on HJM, is used to retrieve some of the lost continuum events which are not coming from τ-pair events. These conditions are as follows: $HJM/E_{vis} > 0.25$ or $HJM > 1.8$ GeV/c^2.

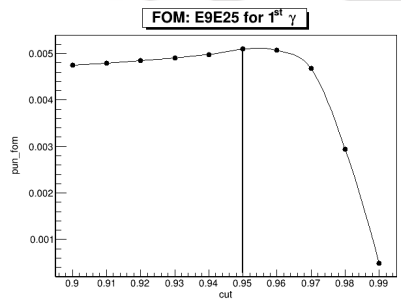
Calorimeter cluster multiplicity ($nECL$): Number of ECL clusters. The condition is $nECL > 1$.

Average energy per cluster ($E_{sum}/nECL$): If one of the electrons in a Bhabha event passes through a gap in the ECL, then the remaining detected energy will pass the selection criteria on E_{sum} described earlier. However, these events can still be removed by applying a condition on the average cluster energy ($E_{sum}/nECL$). We require that $E_{sum}/nECL < 1$ GeV.

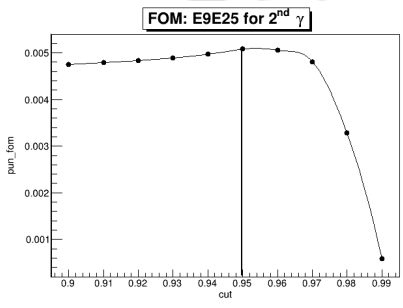
Appendix B

Figure of Merit Plots

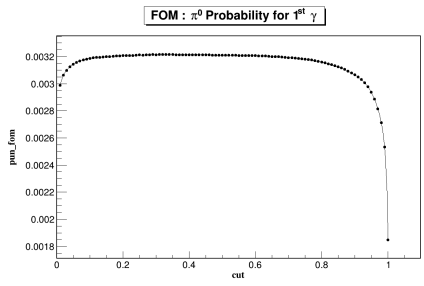
In this appendix, we present the plots of the figure of merit for the selection variables used in the $B^0 \to \gamma\gamma$ analyses.



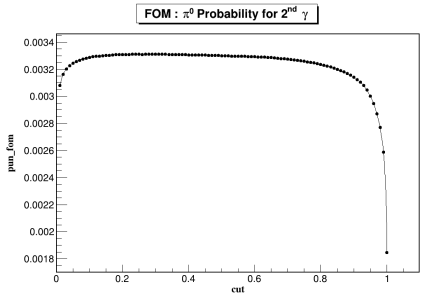
(a) E9E25 for 1^{st} γ



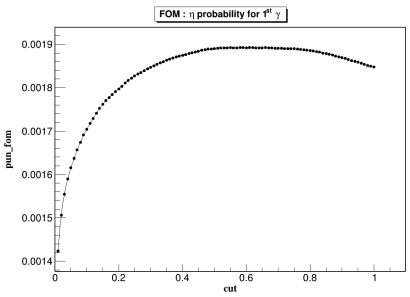
(b) E9E25 for 2^{nd} γ



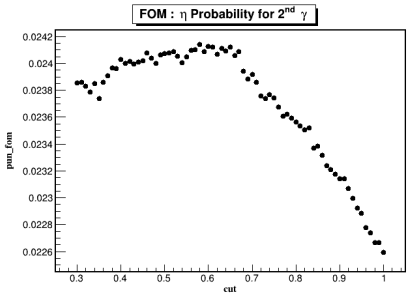
(c) π^0 probablity for 1^{st} γ



(d) π^0 probablity for 2^{nd} γ



(e) η probablity for 1^{st} γ



(f) η probablity for 2^{nd} γ

Figure B.1: FOM plots for the Belle selection variables.

Belle II Analysis

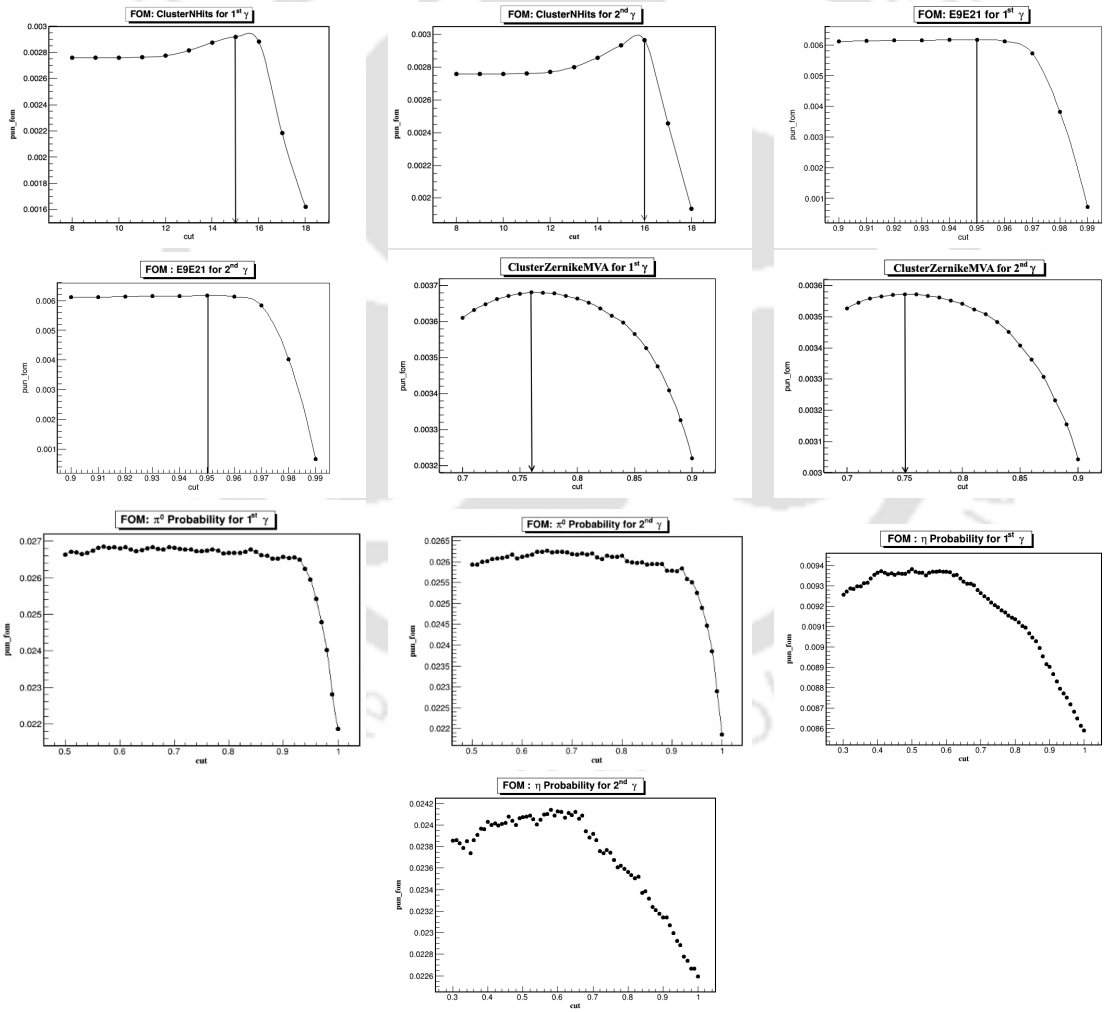


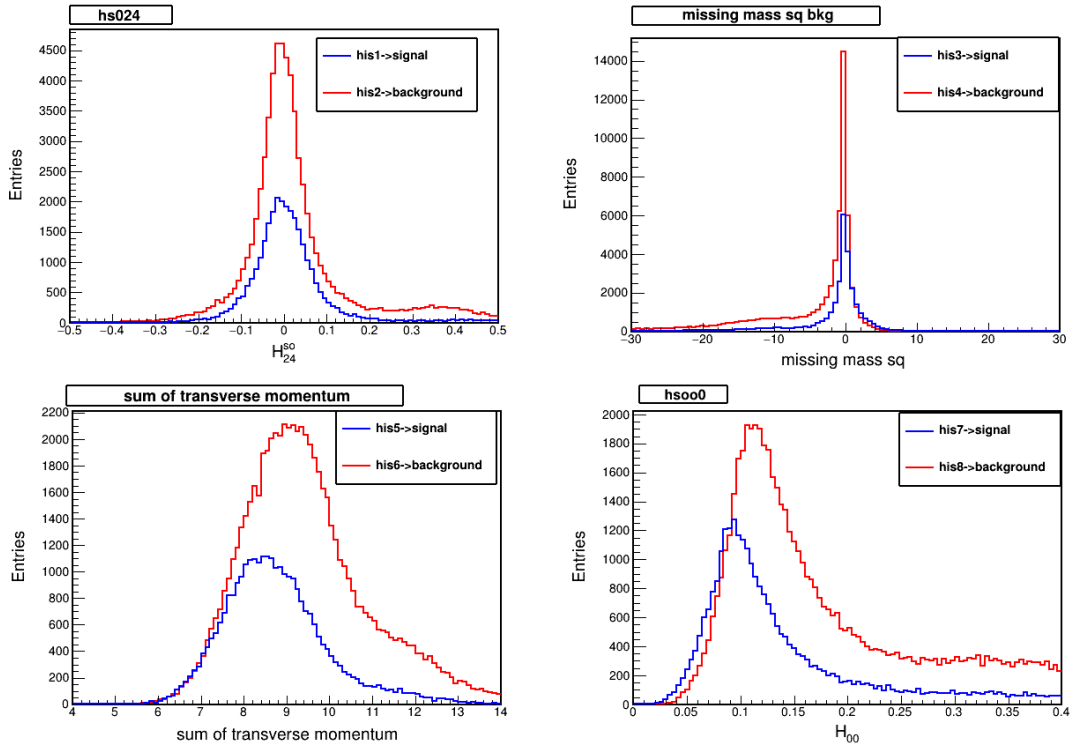
Figure B.2: FOM plots for the Belle II selection variables

Appendix C

Event Shape Variables

Belle Analysis

In this appendix, we present the plots of the event shape variables used in the event classifier (FastBDT) for Belle analysis are shown in Figures C.1 and C.2.



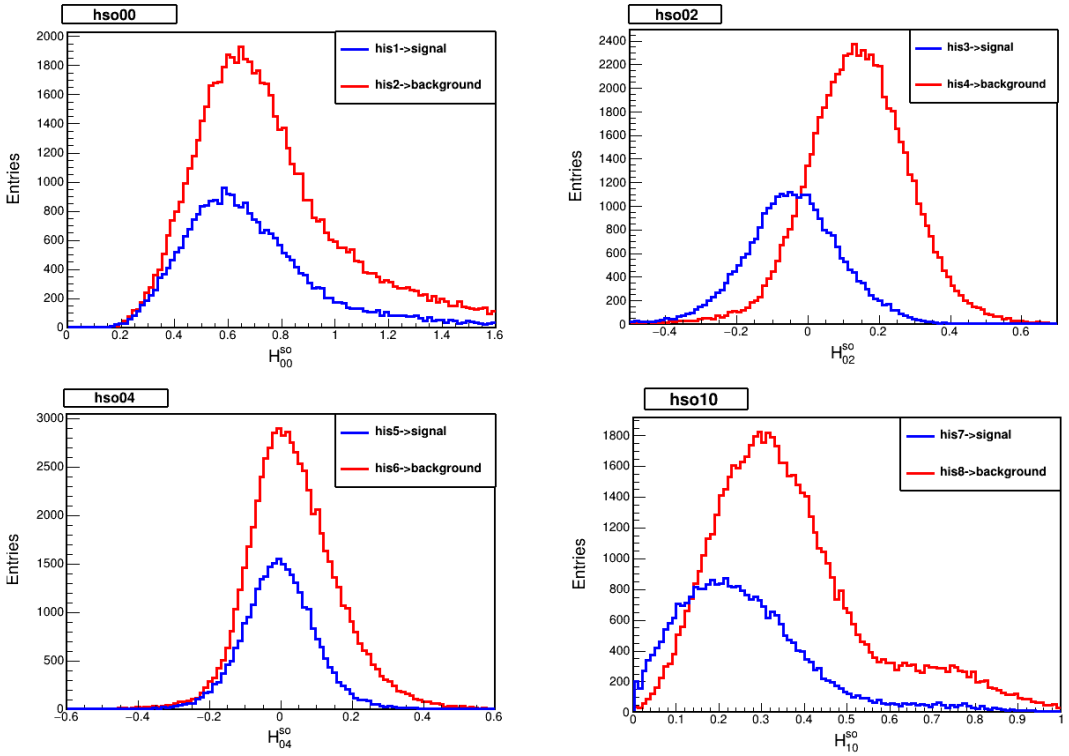
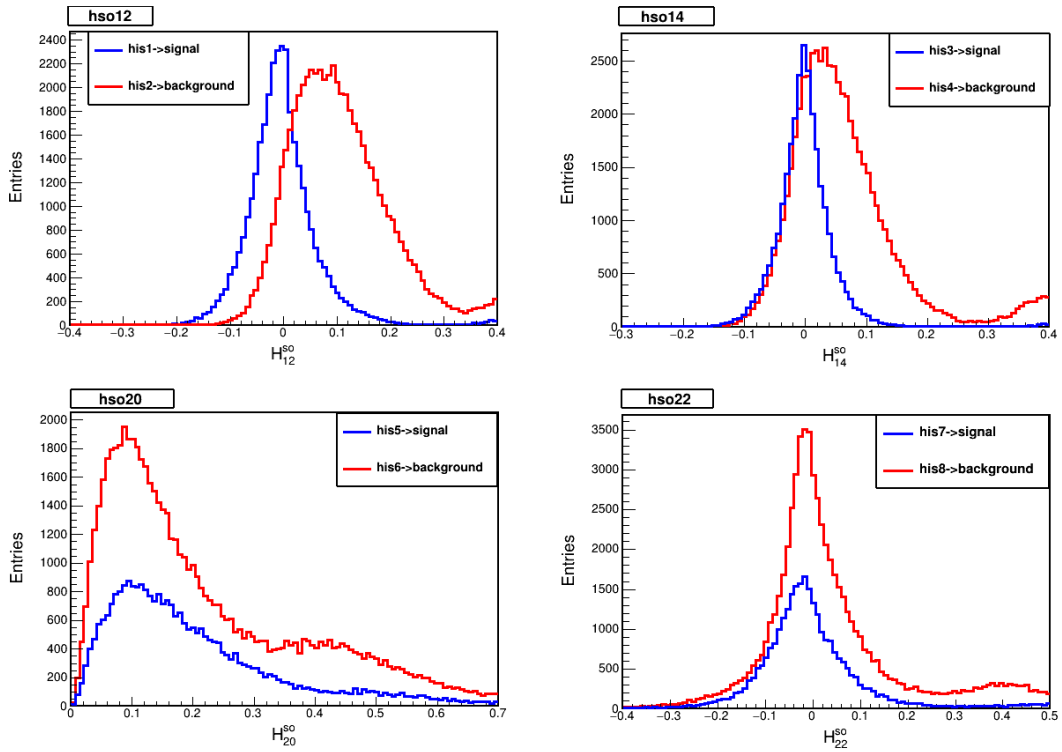


Figure C.1: FastBDT input variables(KSFW) for Belle analysis



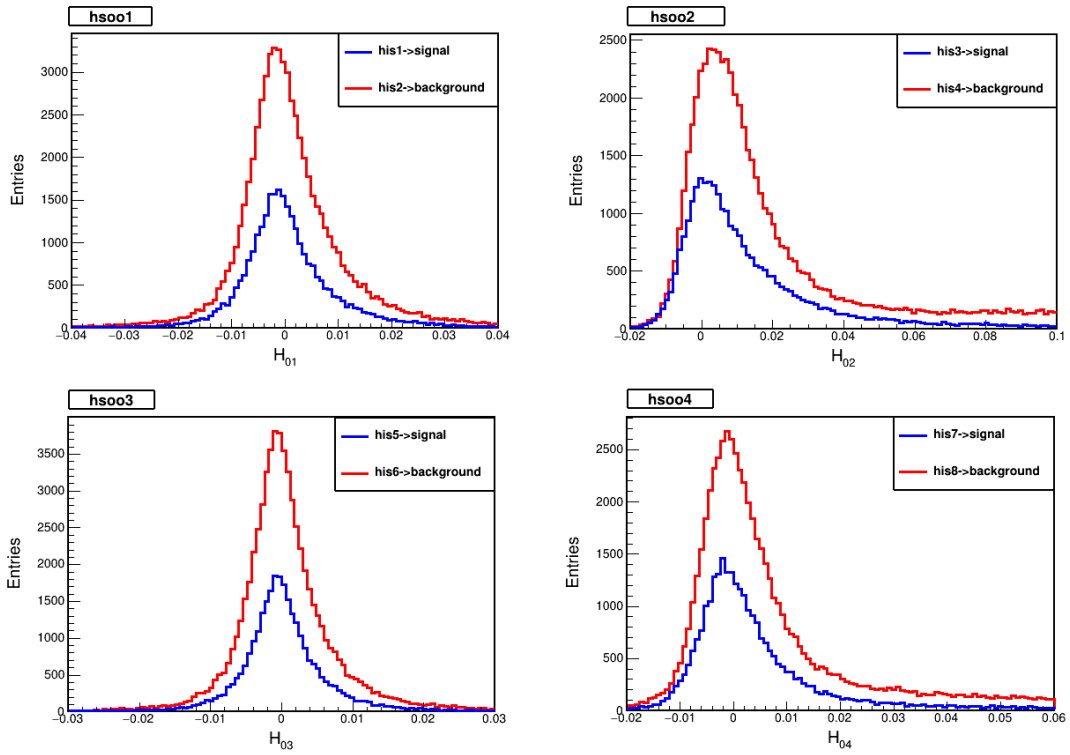
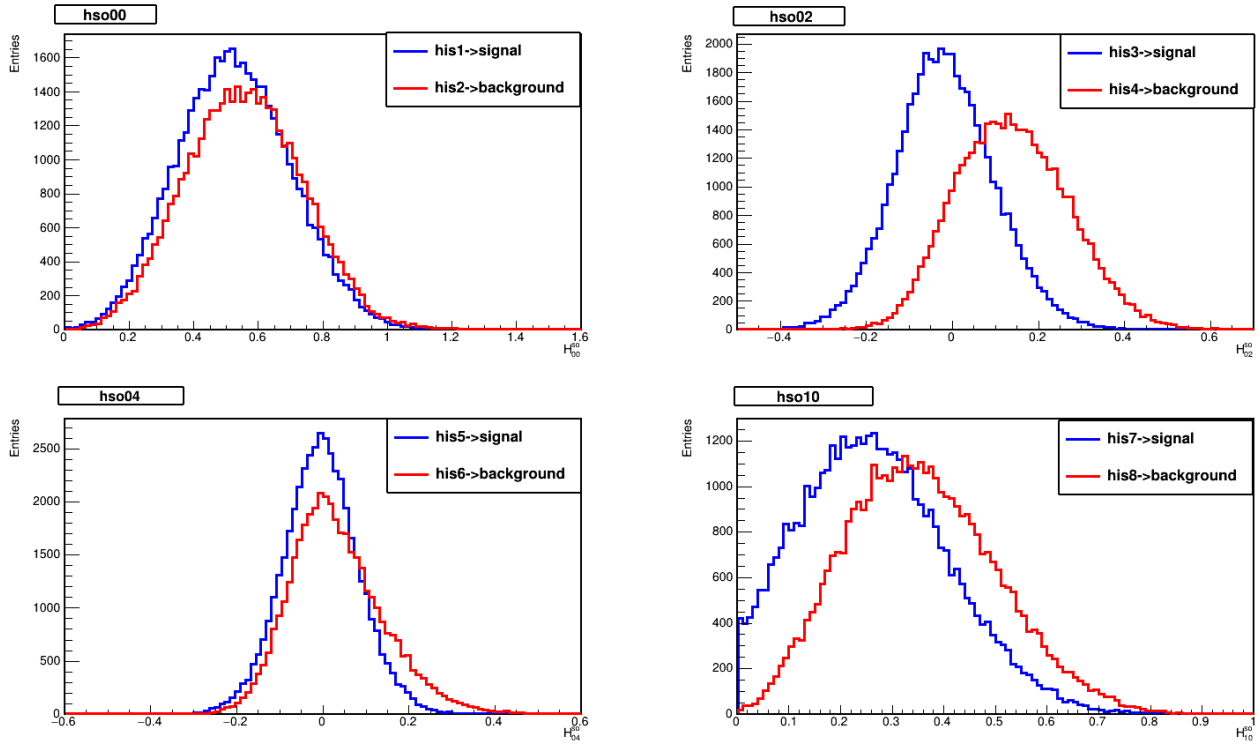


Figure C.2: FastBDT input variables(KSFW) for Belle analysis

Belle II Analysis

The plots of the event shape variables used in the event classifier (FastBDT) for $B^0 \to \gamma\gamma$ analyses, as shown in Figure C.3, C.4 and C.5.



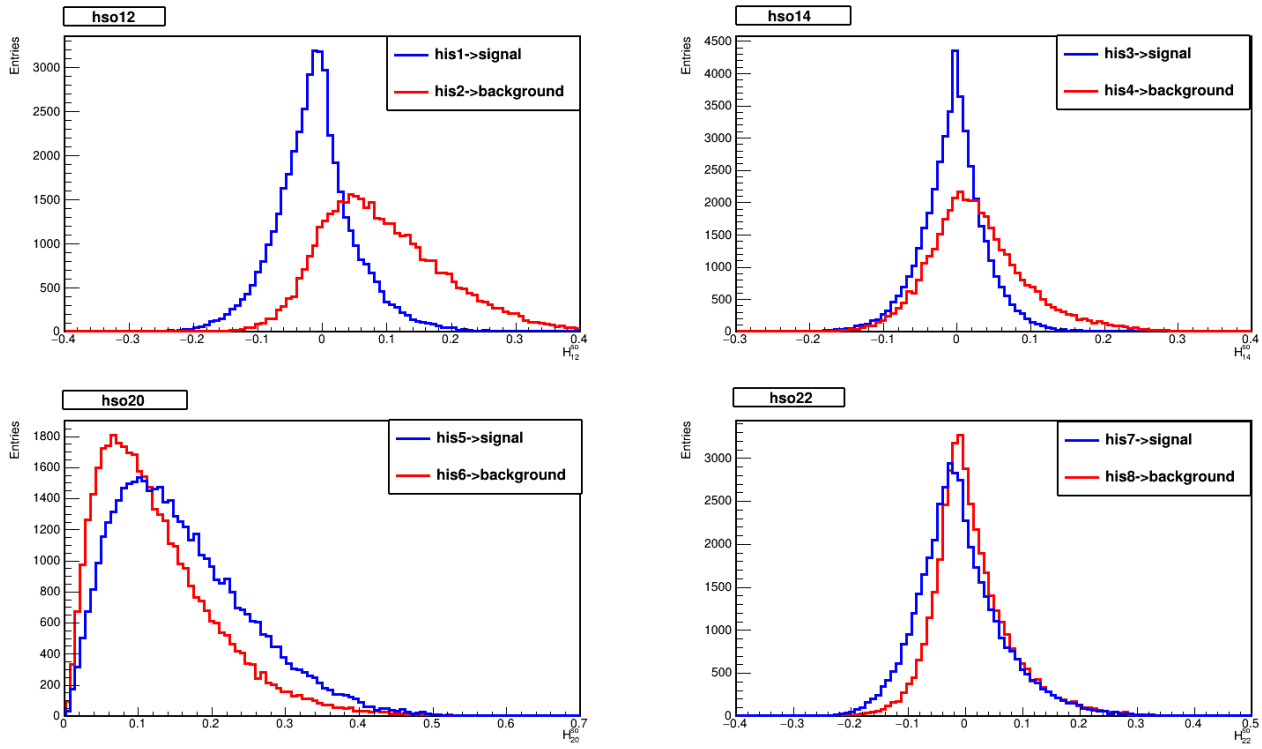
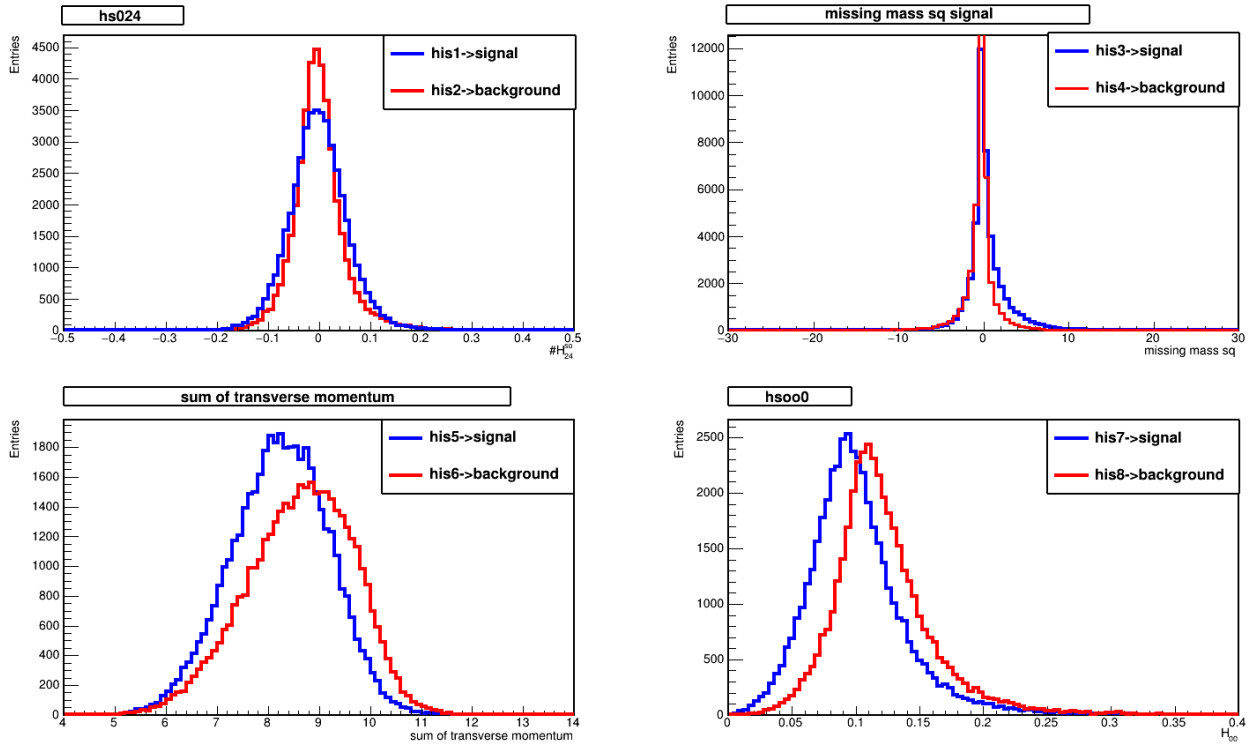


Figure C.3: FastBDT input variables(KSFW) for Belle II analysis



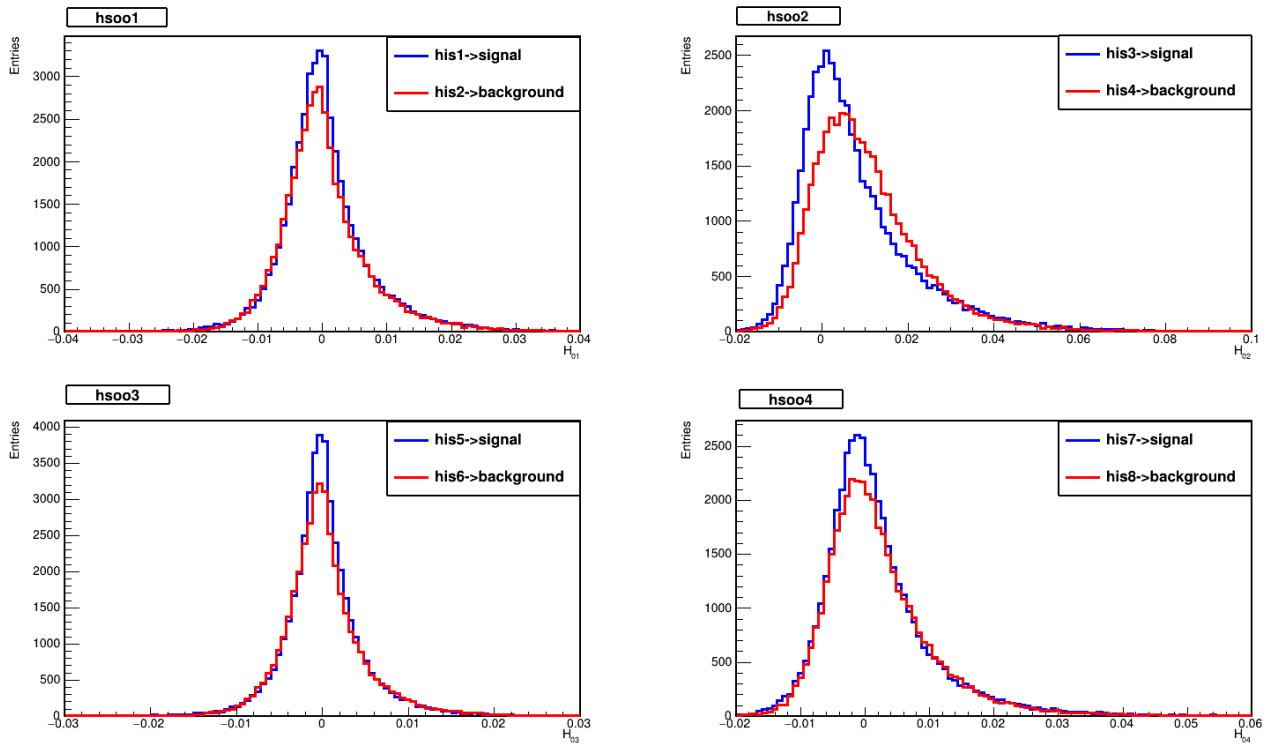


Figure C.4: FastBDT input variables(KSFW) for Belle II analysis

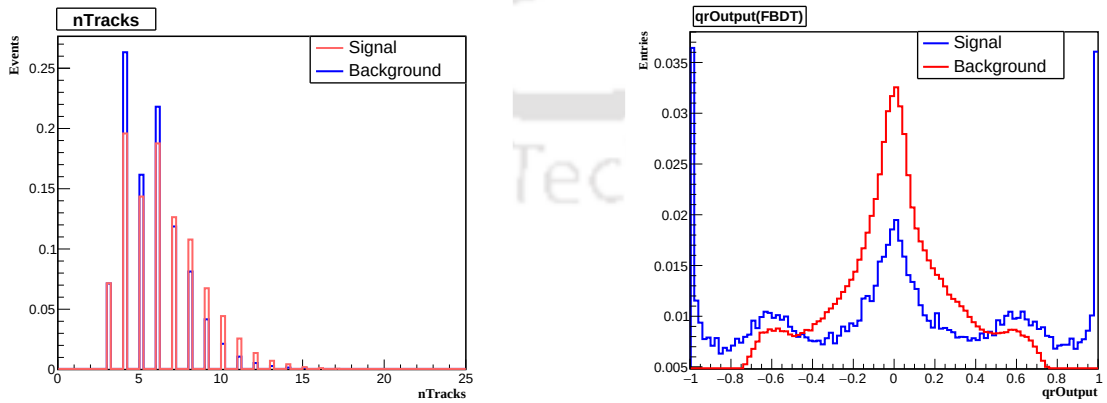


Figure C.5: FastBDT input variables(KSFW) for Belle II analysis

Appendix D

Variable Importance

In Table D.1-D.2, we have tabulated the MVA variables in decreasing order of their importance.

Rank	Variable	Importance
1	cosTBTO	0.396918565
2	KSFWVariables(hso12)	0.187304541
3	KSFWVariables(hso02)	0.131246164
4	KSFWVariables(hso14)	0.125659809
5	qrOutput(FBDT)	0.0565899014
6	KSFWVariables(hso04)	0.0305726454
7	useCMSFrame(cosTheta)	0.0235724282
8	KSFWVariables(mm2)	0.00982598681
9	KSFWVariables(hoo0)	0.00867537409
10	KSFWVariables(hoo2)	0.00620997557
11	KSFWVariables(hso10)	0.0053381715
12	KSFWVariables(hso22)	0.00458427286
13	cosTBz	0.00438240683
14	KSFWVariables(hso20)	0.00195238844
15	KSFWVariables(hso00)	0.00174377381
16	KSFWVariables(hoo1)	0.00153717073
17	KSFWVariables(hoo4)	0.00129995553
18	KSFWVariables(et)	0.00122002978
19	KSFWVariables(hso24)	0.0010119814
20	nTracks	0.000225382595
21	KSFWVariables(hoo3)	0.000129074193

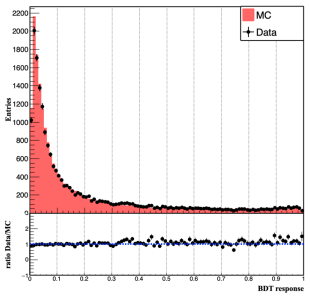
Table D.1: Importance of variables in Belle.

Rank	Variable	Importance
1	cosTBTO	0.46256277
2	KSFWVariables(hso12)	0.202244401
3	KSFWVariables(hso02)	0.115411982
4	KSFWVariables(hso14)	0.062394172
5	qrOutput(FBDT)	0.0498660319
6	useCMSFrame(cosTheta)	0.0241517834
7	KSFWVariables(hoo0)	0.0185289998
8	KSFWVariables(hoo2)	0.01609636628
9	cosTBz	0.0124249654
10	KSFWVariables(hso20)	0.0112800198
11	KSFWVariables(hso04)	0.0087535847
12	KSFWVariables(hso10)	0.00464894017
13	KSFWVariables(hso24)	0.00323372707
14	nTracks	0.00311524188
15	KSFWVariables(et)	0.00172015326
16	KSFWVariables(mm2)	0.00138313556
17	KSFWVariables(hoo3)	0.00126467319
18	KSFWVariables(hso00)	0.00124954036
19	KSFWVariables(hoo1)	0.00069062499
20	KSFWVariables(hoo4)	0.00044673820
21	KSFWVariables(hso22)	0.00036201856

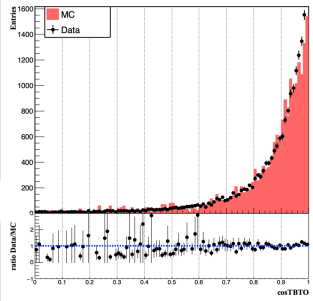
Table D.2: Importance of variables in Belle II

Appendix E

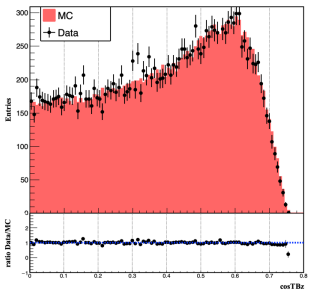
Data-MC Comparison for Belle Analysis



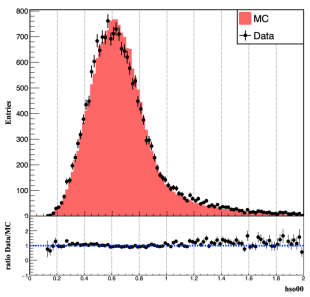
(a) BDT response



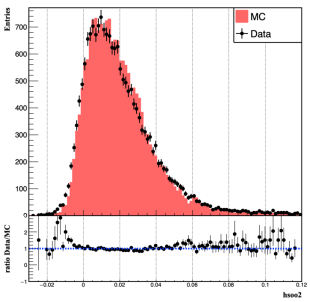
(b) cosTBTO



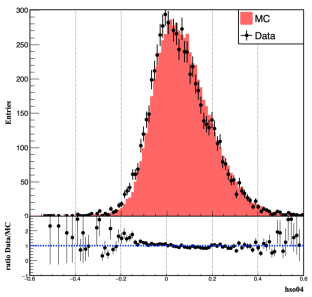
(c) cosTBz



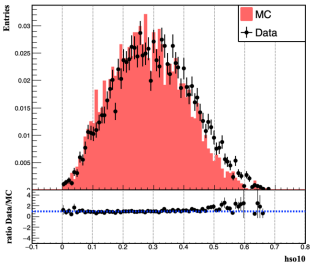
(d) hso00



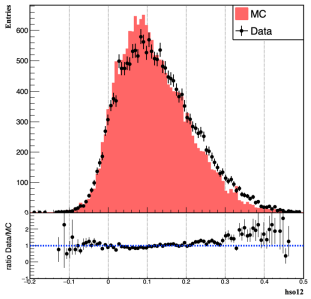
(e) hso02



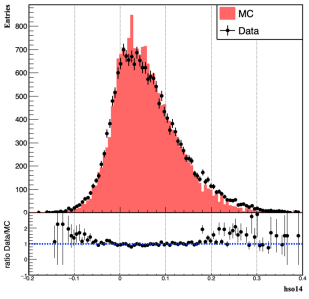
(f) hso04



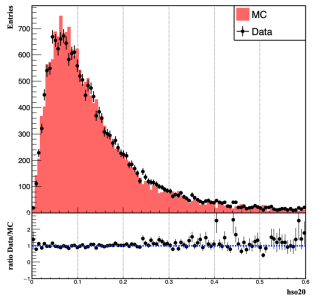
(g) hso10



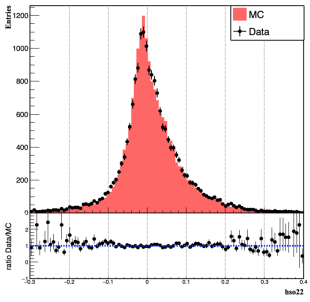
(h) hso12



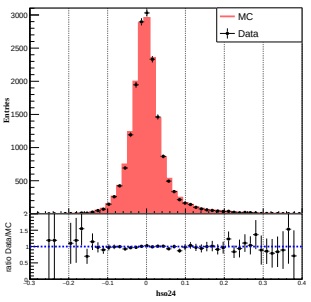
(i) hso14



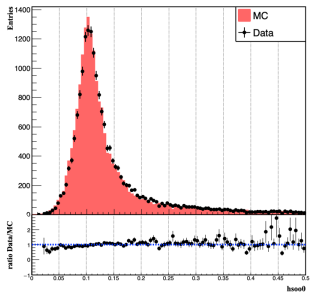
(a) hso20



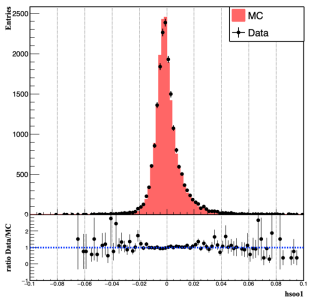
(b) hso22



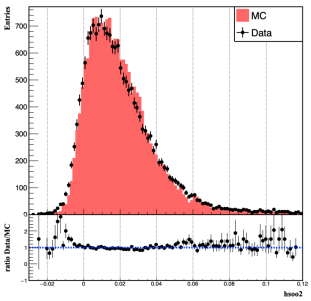
(c) hso24



(d) hsoo0

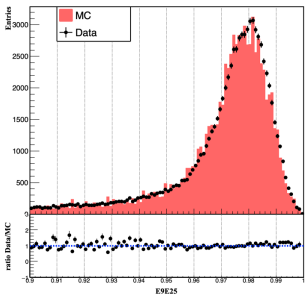


(e) hsoo1

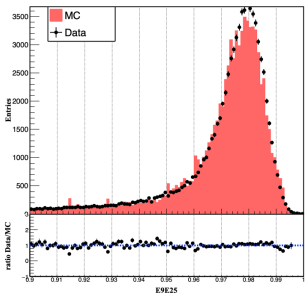


(f) hsoo2

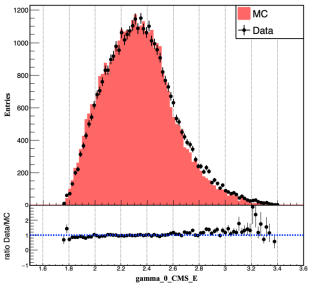
Figure E.2: BDT output and Eventshape variables for Belle study



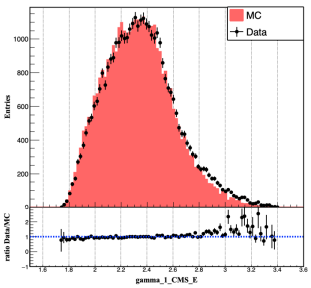
(a) E9E21 for 1^{st} γ



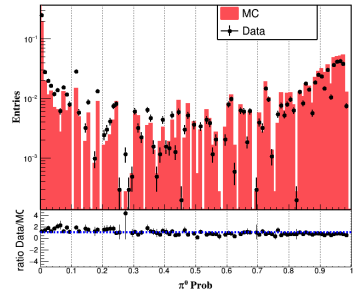
(b) E9E21 for 2^{nd} γ



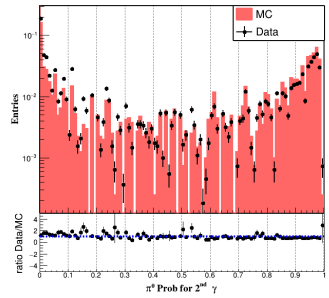
(c) 1^{st} γ energy



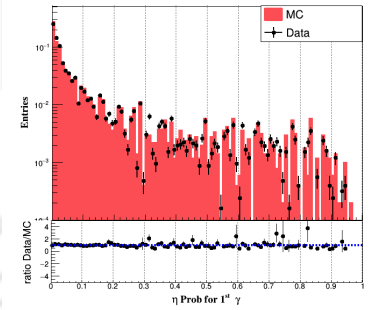
(d) 2^{nd} γ energy



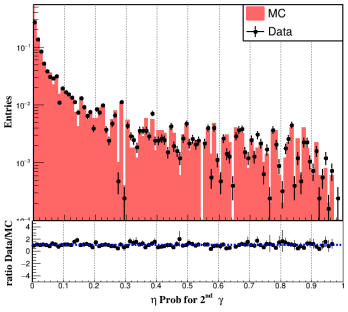
(e) π^0 probabilty for 1^{st} γ



(f) π^0 probabilty for 2^{nd} γ



(g) η probabilty for 1^{st} γ

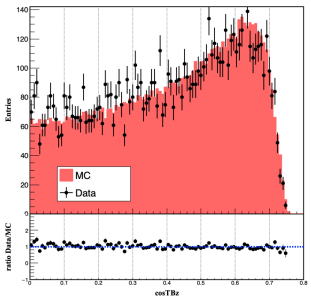


(h) η probabilty for 2^{nd} γ

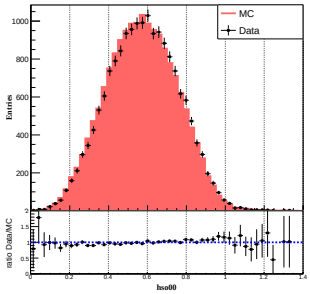
Figure E.3: Photon selection variables

Appendix F

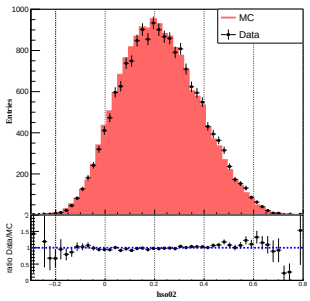
Data-MC Comparison for Belle II Analysis



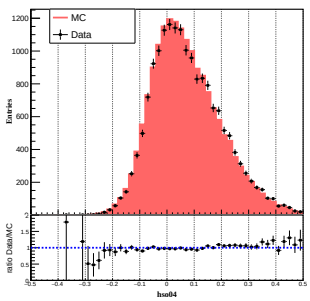
(a) cosTBz



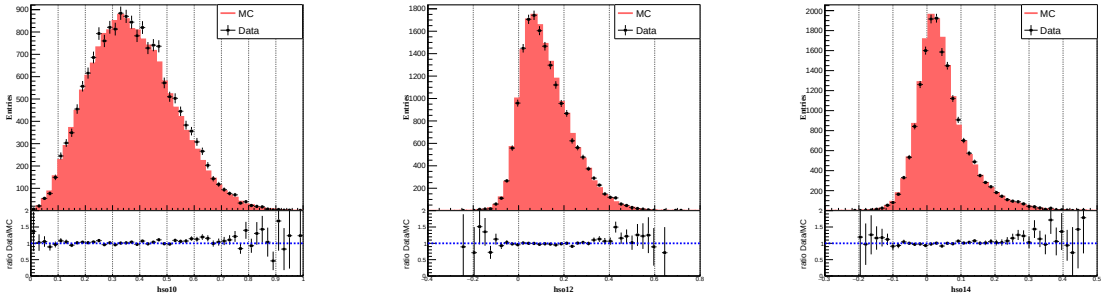
(b) hso00



(c) hso02



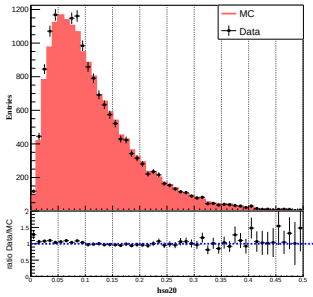
(d) hso04



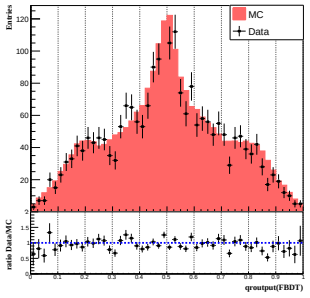
(e) hso10

(f) hso12

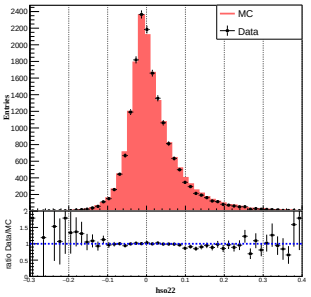
(g) hso14



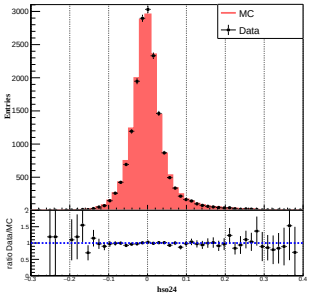
(h) hso20



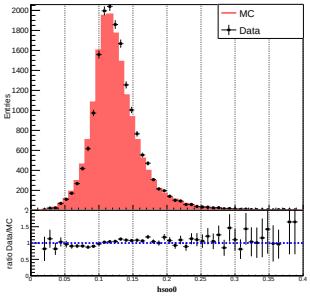
(i) qrOutput



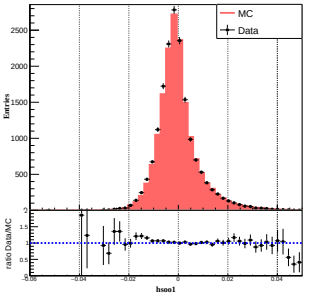
(j) hso22



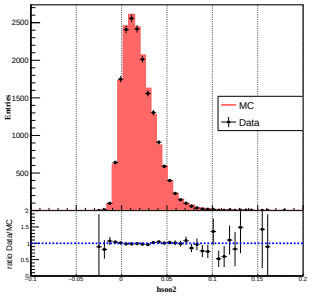
(k) hso24



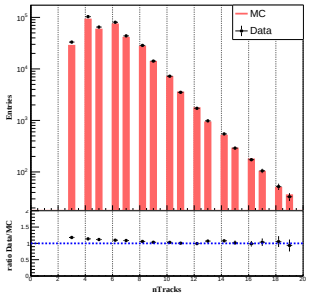
(l) hsoo0



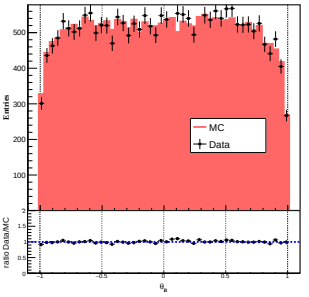
(m) hsoo1



(n) hsoo2

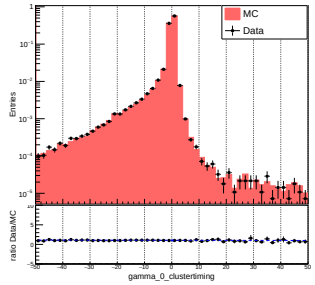


(o) nTracks

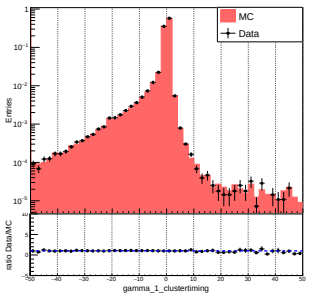


(p) θ_B

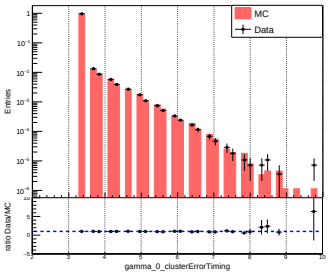
Figure F.1: BDT output and Eventshape variables for the Belle II study



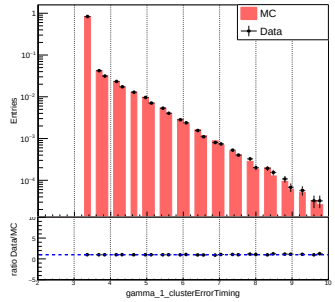
(a)
gamma_0_clusterTiming



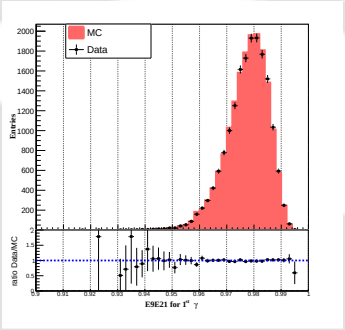
(b)
gamma_1_clusterTiming



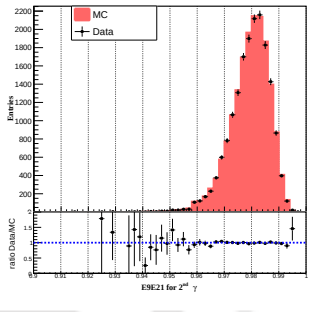
(c)
gamma_0_clusterErrorTiming



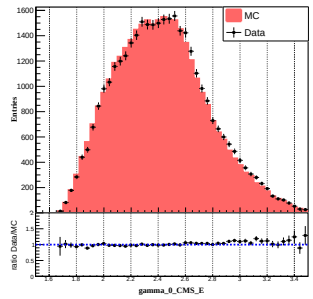
(d)
gamma_1_clusterErrorTiming



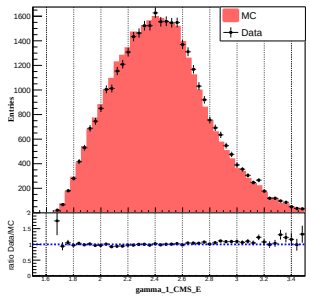
(e) E9E21 for 1^{st} γ



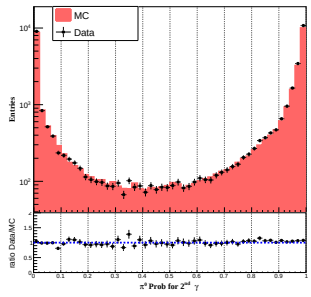
(f) E9E21 for 2^{nd} γ



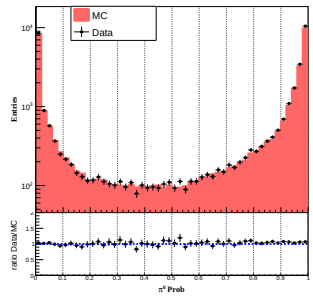
(g) 1^{st} γ energy



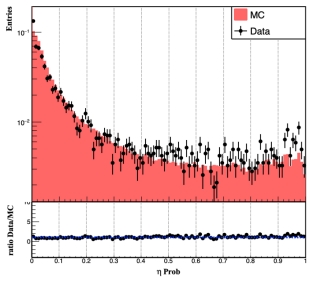
(h) 2^{nd} γ energy



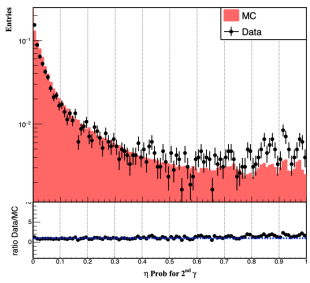
(i) π^0 probabilty for 1^{st} γ



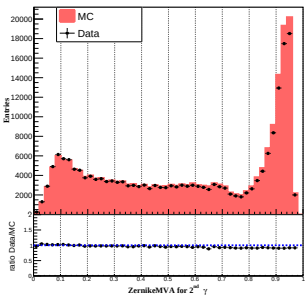
(j) π^0 probabilty for 2^{nd} γ



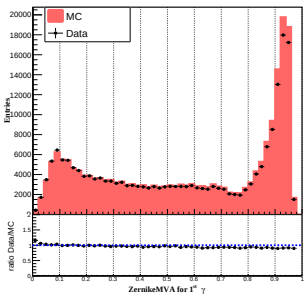
(k) η probabilty for 1^{st} γ



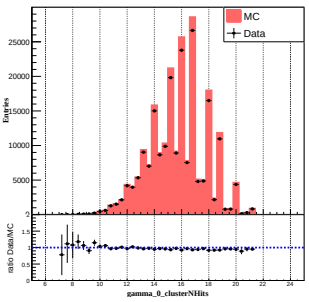
(l) η probabilty for 2^{nd} γ



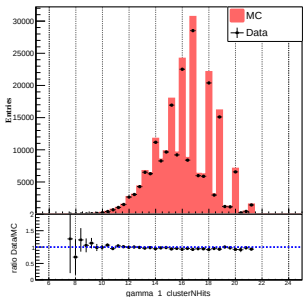
(a) ZernikeMVA for 1^{st} γ



(b) ZernikeMVA for 2^{nd} γ



(c) clusterNHits for 1^{st} γ



(d) clusterNHits for 2^{nd} γ

Figure F.3: Photon selection variables

Appendix G

Probability Distribution Functions

1.0 ARGUS function

The ARGUS function for a continuous random variable m is defined as

$$f(m; m_0, \xi) = \begin{cases} m\sqrt{1-\left(\frac{m}{m_0}\right)^2}\exp\left[\xi\left(1-\left(\frac{m}{m_0}\right)^2\right)\right], & \text{if } m < m_0 \\ 0, & \text{otherwise} \end{cases}, \tag{G.1}$$

where m_0 is the upper bound on the distribution of m and ξ is the curvature of the PDF. This function is defined only for $m < m_0$, beyond which it becomes zero. The function was first introduced by the ARGUS collaboration [102]. The ARGUS function is used significantly in the physics of B mesons to fit the invariant mass or $M_{\rm bc}$ distribution of the combinatorial background.

2.0 Chebychev Polynomial

The Chebychev polynomials [109] of the first kind ($T_n(x)$) are the solutions of the Chebychev differential equation given by

$$(1-x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + n^2 y = 0, \tag{G.2}$$

where $n = 0, 1, 2, \ldots$ are the integer numbers. The first few polynomials are as follows

$$T_0(x) = 1. \tag{G.3}$$
$$T_1(x) = x. \tag{G.4}$$
$$T_2(x) = 2x^2 - 1. \tag{G.5}$$
$$T_3(x) = 4x^3 - 3x. \tag{G.6}$$
$$T_4(x) = 8x^4 - 8x^2 + 1. \tag{G.7}$$
$$\vdots$$

All the polynomials can be determined recursively by

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \tag{G.8}$$

if any two consecutive polynomials are known. The Chebychev polynomials are orthogonal in the interval $x \in [-1, 1]$ and normalized that for all n, $T_n(1) = 1$. Any arbitrary polynomial of degree n can be written in terms of the Chebychev polynomials as

$$f(x; a_1, \ldots, a_n) = \frac{1}{N}\left[T_0(x) + \sum_{i=1}^{n} a_i T_i(x)\right], \tag{G.9}$$

where $T_i(x)$ is the i^{th} order Chebychev polynomial and a_i is the corresponding coefficient. Chebychev polynomials have better stability than ordinary polynomials due to the lower correlation between the coefficients [97]. Alternatively, the Chebyshev polynomials of the first kind can also be defined using the trigonometric function cosine. Specifically, they can be expressed using the following formula:

$$T_n(x) = \cos\big(n \cos^{-1}(x)\big) \tag{G.10}$$

Chebyshev polynomials have several important properties, including orthogonality, recurrence relations, and explicit formulas. They are particularly useful for approximating functions, solving differential equations, and performing numerical computations due to their advantageous properties and numerical stability.

3.0 Kernal Density Estimation

In this study, we incorporate the correlation between M_{bc} and ΔE by employing a two-dimensional kernel density estimation (KDE). KDE models the distribution of a given dataset by superimposing density functions, or kernels, at each data point. It is a non-parametric approach used to estimate the probability density function (PDF) of a random variable. When provided with independent and identically distributed samples $x_1, \ldots, x_n$ drawn from a distribution with an unknown density f. Given a kernel function K, the KDE estimator of the PDF f at a point x is defined as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{G.11}$$

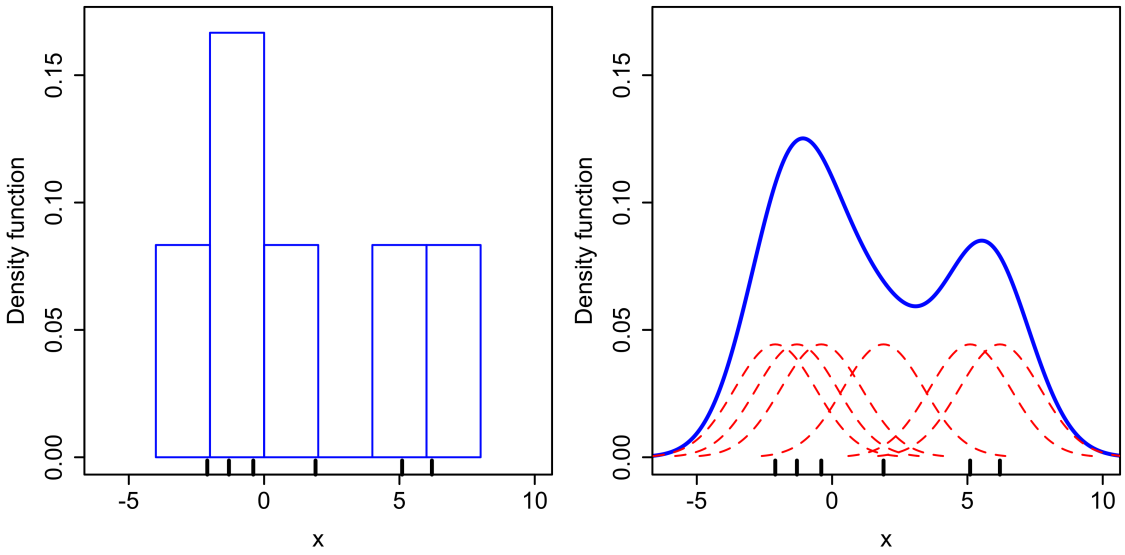


Figure G.1: Comparison of the histogram (left) and kernel density estimate (right) created using the same six data points, represented as lines on the horizontal axis. The kernels corresponding to each data point are the red dashed curves, and the kernel density estimate is the blue curves. Plots taken from [110].

Here, K represents the kernel, and h is the bandwidth parameter, which determines the "smoothness" of the PDF. A scaled kernel with subscript h is defined as $K_h(x) = 1/hK(x/h)$. Typically, the kernel is a symmetric, non-negative function that integrates to one. It assigns weights to x_i based on their distance from x. As illustrated in Figure G.1, the goal is to calculate the contribution of each data point to the estimated density at any given point in the feature space. A lower bandwidth implies that only points very close to the current position receive significant weight, leading to high sensitivity to the input data. Conversely, a higher bandwidth results in a wider kernel where distant points can contribute, yielding a smoother density estimate.

The two-dimensional M_{bc} and ΔE KDE generated by ROOT cannot be stored and exported for use in a fit due to the technical limitations of the software package. To overcome this challenge, we employ the following method to store the data:

- 2D kernel density estimation PDFs are generated from MC.
- Kernel density estimation PDFs are then converted to TH2 histograms with a very large number of bins (300×300) to preserve as much data as possible and save in a ROOT file.
- The fitting code reads in this ROOT file, the TH2 is loaded and converted into a RooDataHist, then a RooHistPdf, which is a format that the RooFit can use based fitter

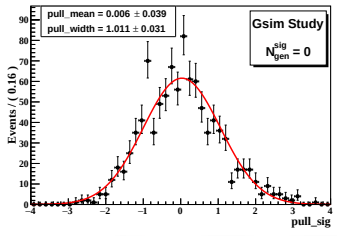
This technique allows the benefits of kernel density estimation PDFs with the speed and convenience of storage as a ROOT histogram.

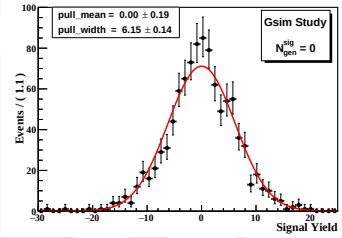
Appendix H

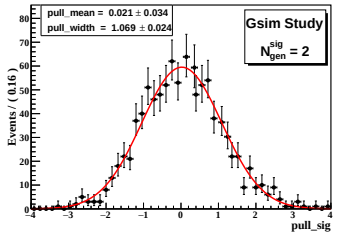
Ensemble Study: Signal Pull and Fitted Yield Distributions

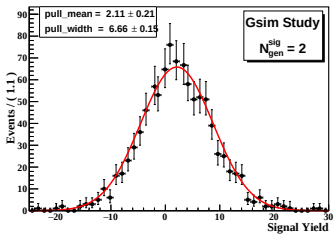
1.0 Gsim Study: Belle

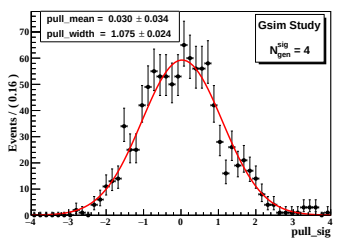
Figures H.10 shows the distributions of signal pull and fitted signal yield extracted from the toy ensembles for $N_{\text{gen}}^{\text{sig}} = (0, 2, 4, 6, 8, 10, 12, 14, 16, 18 \text{ and } 20)$

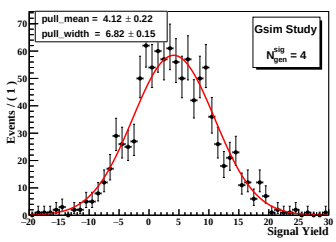


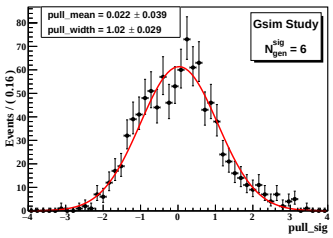


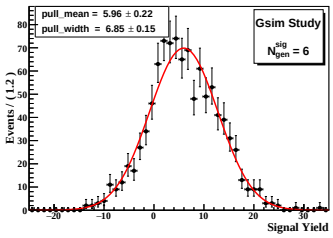


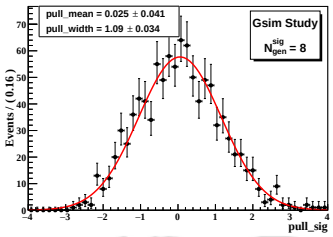


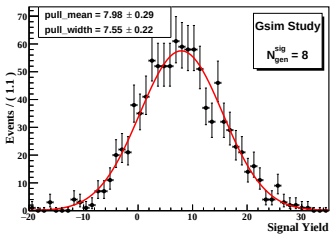


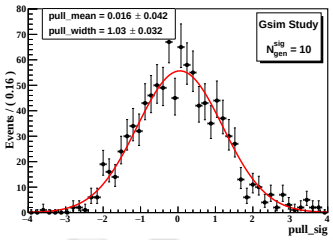


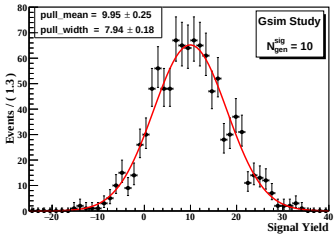

pull_mean = 0.022 ± 0.039
pull_width = 1.02 ± 0.029
Gsim Study
N_{gen}^{sig} = 6
Events / (0.16)
pull_sig

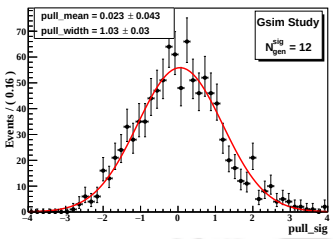

pull_mean = 5.96 ± 0.22
pull_width = 6.85 ± 0.15
Gsim Study
N_{gen}^{sig} = 6
Events / (1.2)
Signal Yield

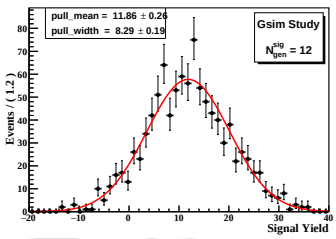

pull_mean = 0.025 ± 0.041
pull_width = 1.09 ± 0.034
Gsim Study
N_{gen}^{sig} = 8
Events / (0.16)
pull_sig

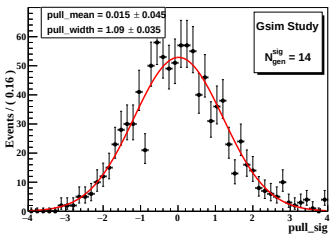

pull_mean = 7.98 ± 0.29
pull_width = 7.55 ± 0.22
Gsim Study
N_{gen}^{sig} = 8
Events / (1.1)
Signal Yield

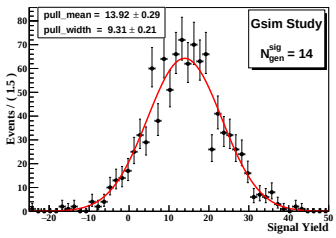

pull_mean = 0.016 ± 0.042
pull_width = 1.03 ± 0.032
Gsim Study
N_{gen}^{sig} = 10
Events / (0.16)
pull_sig

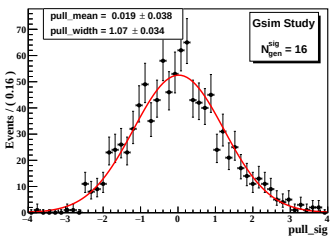

pull_mean = 9.95 ± 0.25
pull_width = 7.94 ± 0.18
Gsim Study
N_{gen}^{sig} = 10
Events / (1.3)
Signal Yield

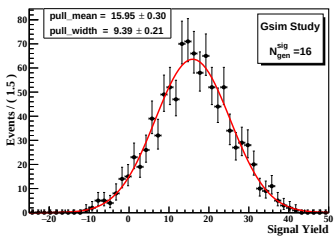

pull_mean = 0.023 ± 0.043
pull_width = 1.03 ± 0.03
Gsim Study
N_{gen}^{sig} = 12
Events / (0.16)
pull_sig


pull_mean = 11.86 ± 0.26
pull_width = 8.29 ± 0.19
Gsim Study
N_{gen}^{sig} = 12
Events / (1.2)
Signal Yield


pull_mean = 0.015 ± 0.045
pull_width = 1.09 ± 0.035
Gsim Study
N_{gen}^{sig} = 14
Events / (0.16)
pull_sig


pull_mean = 13.92 ± 0.29
pull_width = 9.31 ± 0.21
Gsim Study
N_{gen}^{sig} = 14
Events / (1.5)
Signal Yield


pull_mean = 0.019 ± 0.038
pull_width = 1.07 ± 0.034
Gsim Study
N_{gen}^{sig} = 16
Events / (0.16)
pull_sig


pull_mean = 15.95 ± 0.30
pull_width = 9.39 ± 0.21
Gsim Study
N_{gen}^{sig} =16
Events / (1.5)
Signal Yield

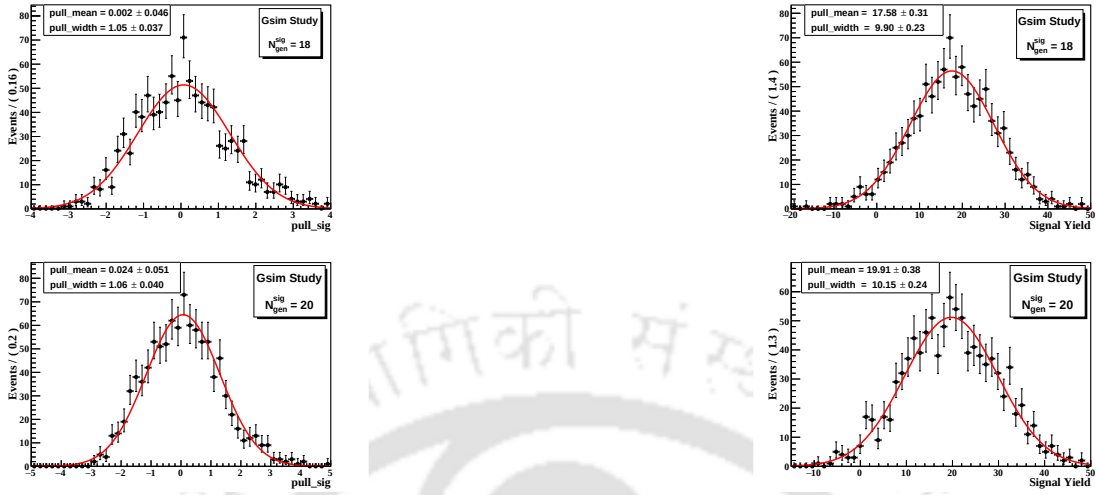
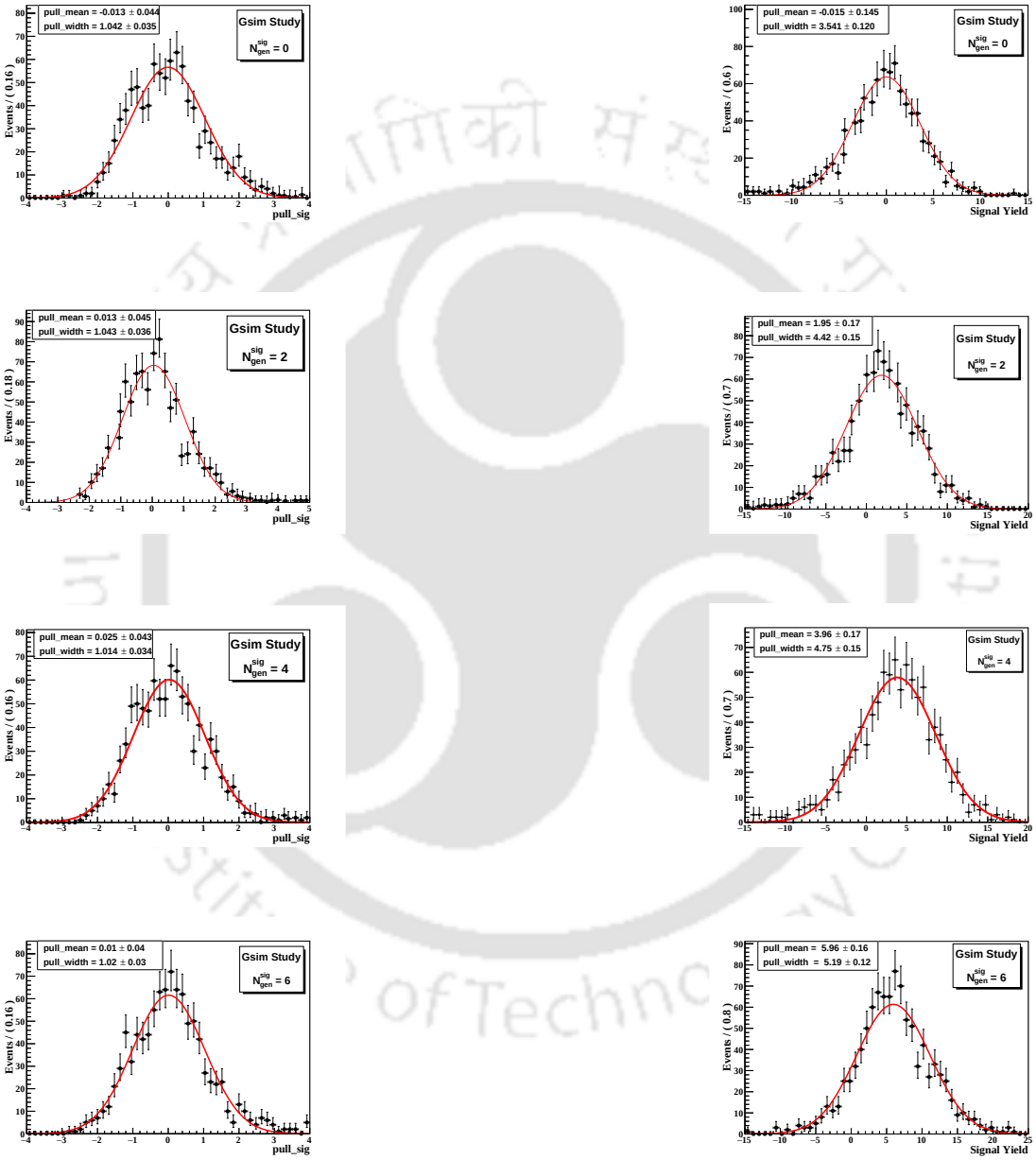
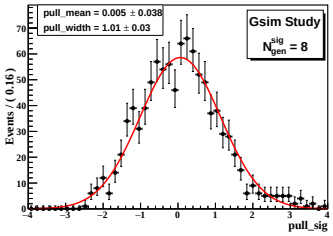


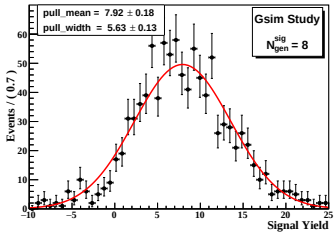
Figure H.10: Pull and yield distributions for $N_{gen}^{sig} = 0$, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 in Belle.

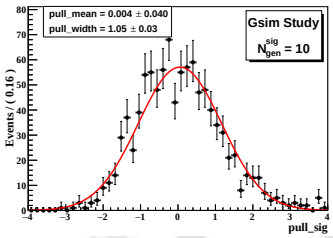
2.0 Gsim Study: Belle II

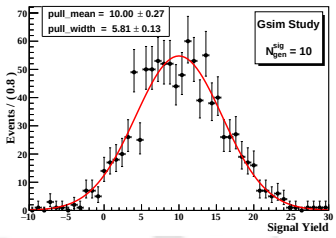
Figures H.18 shows the distributions of signal pull and fitted signal yield extracted from the toy ensembles for $N_{\text{gen}}^{\text{sig}}$ = (0, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20).

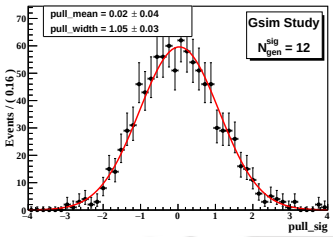


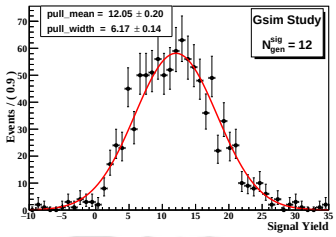

pull_mean = 0.005 ± 0.038
pull_width = 1.01 ± 0.03
Gsim Study
N$^{sig}_{gen}$ = 8
Events / (0.16)
pull_sig

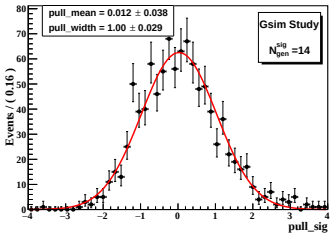

pull_mean = 7.92 ± 0.18
pull_width = 5.63 ± 0.13
Gsim Study
N$^{sig}_{gen}$ = 8
Events / (0.7)
Signal Yield

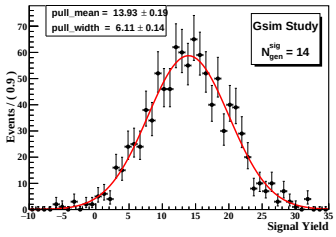

pull_mean = 0.004 ± 0.040
pull_width = 1.05 ± 0.03
Gsim Study
N$^{sig}_{gen}$ = 10
Events / (0.16)
pull_sig

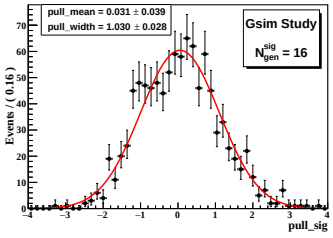

pull_mean = 10.00 ± 0.27
pull_width = 5.81 ± 0.13
Gsim Study
N$^{sig}_{gen}$ = 10
Events / (0.8)
Signal Yield

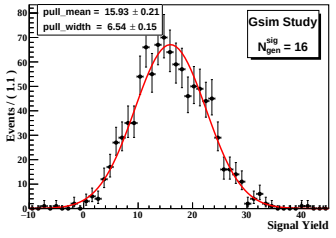

pull_mean = 0.02 ± 0.04
pull_width = 1.05 ± 0.03
Gsim Study
N$^{sig}_{gen}$ = 12
Events / (0.16)
pull_sig


pull_mean = 12.05 ± 0.20
pull_width = 6.17 ± 0.14
Gsim Study
N$^{sig}_{gen}$ = 12
Events / (0.9)
Signal Yield


pull_mean = 0.012 ± 0.038
pull_width = 1.00 ± 0.029
Gsim Study
N$^{sig}_{gen}$ =14
Events / (0.16)
pull_sig


pull_mean = 13.93 ± 0.19
pull_width = 6.11 ± 0.14
Gsim Study
N$^{sig}_{gen}$ = 14
Events / (0.9)
Signal Yield


pull_mean = 0.031 ± 0.039
pull_width = 1.030 ± 0.028
Gsim Study
N$^{sig}_{gen}$ = 16
Events / (0.16)
pull_sig


pull_mean = 15.93 ± 0.21
pull_width = 6.54 ± 0.15
Gsim Study
N$^{sig}_{gen}$ = 16
Events / (1.1)
Signal Yield

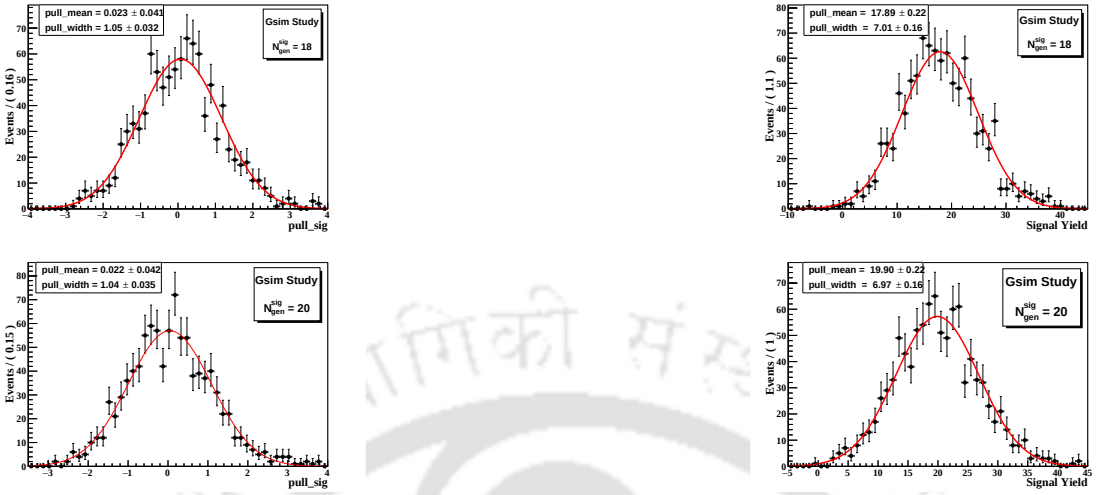


Figure H.18: Pull and yield distributions for $N_{gen}^{sig} = 0$, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20

Bibliography

[1] F. Sebastiani and F. Cordella, "Fermi toward quantum statistics (1923-1925)," *Enrico Fermi and the Universe of Physics*, p. 71, 2001.

[2] P. A. M. Dirac, "On the theory of quantum mechanics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 112, no. 762, pp. 661–677, 1926.

[3] Bose, "Plancks gesetz und lichtquantenhypothese," *Zeitschrift für Physik*, vol. 26, no. 1, pp. 178–181, 1924.

[4] C. Burgard and D. Galbraith, "Standard model of physics," *TEXample. net*, 2016.

[5] M. Thomson, *Modern particle physics*. New York: Cambridge University Press, 2013.

[6] G. L. Kane, *Modern Elementary Particle Physics*. Cambridge University Press, 2 2017.

[7] P. Kooijman and N. Tuning, "Lectures on cp violation," *no. January*, p. 112, 2015.

[8] L.-L. Chau and W.-Y. Keung, "Comments on the Parametrization of the Kobayashi-Maskawa Matrix," *Phys. Rev. Lett.*, vol. 53, p. 1802, 1984.

[9] R. L. Workman *et al.*, "Review of Particle Physics," *PTEP*, vol. 2022, p. 083C01, 2022.

[10] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa Matrix," *Phys. Rev. Lett.*, vol. 51, p. 1945, 1983.

[11] J. Charles, A. Hocker, H. Lacker, S. Laplace, F. R. Le Diberder, J. Malcles, J. Ocariz, M. Pivk, and L. Roos, "CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories," *Eur. Phys. J. C*, vol. 41, no. 1, pp. 1–131, 2005.

[12] E. Kou, P. Urquijo, W. Altmannshofer, F. Beaujean, G. Bell, M. Beneke, I. Bigi, F. Bishara, M. Blanke, C. Bobeth, *et al.*, "The Belle II physics book," *Progress of Theoretical and Experimental Physics*, vol. 2019, no. 12, p. 123C01, 2019.

[13] S. P. Del Amo, J. Lees, V. Poireau, E. Prencipe, V. Tisserand, T. J. Garra, E. Grauges, M. Martinelli, A. Palano, M. Pappagallo, *et al.*, "Search for the decay $B^0 \to \gamma\gamma$," *Physical Review D: Particles, Fields, Gravitation and Cosmology*, vol. 83, no. 3, pp. 032006–032006, 2011.

[14] Y.-L. Shen, Y.-M. Wang, and Y.-B. Wei, "Precision calculations of the double radiative bottom-meson decays in soft-collinear effective theory," *Journal of High Energy Physics*, vol. 2020, no. 12, pp. 1–55, 2020.

[15] Q. Qin, Y.-L. Shen, C. Wang, and Y.-M. Wang, "Deciphering the long-distance penguin contribution to $\bar{B}_{d,s} \to \gamma\gamma$ decays," *Physical Review Letters*, vol. 131, no. 9, p. 091902, 2023.

[16] J. Aranda, J. Montano, F. Ramirez-Zavaleta, J. Toscano, and E. Tututi, "Bounding the $B_s \to \gamma\gamma$ decay from higgs mediated flavor changing neutral current transitions," *Physical Review D*, vol. 82, no. 5, p. 054002, 2010.

[17] G. Devidze and G. Jibuti, "Double-photon decays of Bs-and Bd-mesons in the mssm," *Physics Letters B*, vol. 429, no. 1-2, pp. 48–50, 1998.

[18] G. Devidze, A. Liparteliani, and U.-G. Meißner, "$B_{s,d} \to \gamma\gamma$ decay in the model with one universal extra dimension," *Physics Letters B*, vol. 634, no. 1, pp. 59–62, 2006.

[19] M. Acciarri, A. Adam, O. Adriani, M. Aguilar-Benitez, S. Ahlen, B. Alpat, J. Alcaraz, J. Allaby, A. Aloisio, G. Alverson, *et al.*, "Search for the decays $B_d^0 \to \gamma\gamma$ and $B_s^0 \to \gamma\gamma$," *Physics Letters B*, vol. 363, no. 1-2, pp. 137–144, 1995.

[20] S. Villa, K. Abe, I. Adachi, H. Aihara, Y. Asano, T. Aushev, S. Bahinipati, A. Bakich, V. Balagura, E. Barberio, *et al.*, "Search for the decay $B^0 \to \gamma\gamma$," *Physical Review D*, vol. 73, no. 5, p. 051107, 2006.

[21] S. W. Herb and H. et al., "Observation of a dimuon resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions," *PhysRevLett*, vol. 39, pp. 252–255, Aug 1977.

[22] H. Albrecht and R. C. et al., "A determination of the total width of the $\Upsilon(9.46)$ meson," *Phys.Lett.B*, vol. 93, no. 4, pp. 500–504, 1980.

[23] C. Berger and H. G. et al., "First observation of hadron production in e^+e^- collisions at 13 and 17 GeV CMS energy with the Pluto detector at Petra: Pluto Collaboration," *Physics Letters B*, vol. 81, no. 3, pp. 410–415, 1979.

[24] C. Darden and H. et al., "Observation of a narrow resonance at 9.46 GeV in electron-positron annihilations," *Physics Letters B*, vol. 76, no. 2, pp. 246–248, 1978.

[25] D. Besson and T. Skwarnicki, "Upsilon spectroscopy: transitions in the bottomonium system," *Annual Review of Nuclear and Particle Science*, vol. 43, no. 1, pp. 333–378, 1993.

[26] B. Barish, S. Chadha, *et al.*, "Measurement of the B semileptonic branching fraction with lepton tags," *Physical review letters*, vol. 76, no. 10, p. 1570, 1996.

[27] E. B. et al. Heavy Flavor Averaging Group *et al.*, "Averages of b-hadron properties at the end of 2006," *arXiv preprint arXiv:0704.3575*, 2007.

[28] S. Choudhury, S. Sandilya, K. Trabelsi, Giri, *et al.*, "Measurement of the B^+/B^0 production ratio in e^+e^- collisions at the $\Upsilon(4S)$ resonance using $B \to J/\psi(\ell\ell)K$ decays at belle," *Phys. Rev. D*, vol. 107, p. L031102, Feb 2023.

[29] D. Besson, M. Green, N. Horwitz, *et al.*, "Observation of new structure in the e^+e^- cross section above the $\Upsilon(4S)$," *Physical review letters*, vol. 54, no. 5, p. 381, 1985.

[30] M. Artuso, C. Boulahouache, Blusk, *et al.*, "Evidence for $B_s^*\overline{B_s^*}$ production at the $\Upsilon(5S)$ resonance," *Physical review letters*, vol. 95, no. 26, p. 261801, 2005.

[31] H. Kichimi, Trabelsi, *et al.*, "Kekb beam collision stability at the picosecond timing and micron position resolution as observed with the belle detector," *Journal of Instrumentation*, vol. 5, no. 03, p. P03011, 2010.

[32] S. Kurokawa and E. Kikutani, "Overview of the KEKB accelerators," *Nucl. Instrum. Meth. A*, vol. 499, pp. 1–7, 2003.

[33] N. Toge and B. KEK, "Factory design report," tech. rep., Technical report, 1995.

[34] R. Palmer, "Slac technical report no," tech. rep., SLAC-PUB-4707, 1988.

[35] K. Oide and K. Yokoya, "Beam-beam collision scheme for storage-ring colliders," *Physical Review A*, vol. 40, no. 1, p. 315, 1989.

[36] "Belle public webpage."

[37] A. Abashian *et al.*, "The Belle Detector," *Nucl. Instrum. Meth. A*, vol. 479, pp. 117–232, 2002.

[38] J. Brodzicka, K. Browder, Kinoshita, *et al.*, "Physics achievements from the belle experiment," *Progress of Theoretical and Experimental Physics*, vol. 2012, no. 1, p. 04D001, 2012.

[39] E. Nygård, P. Aspell, P. Jarron, P. Weilhammer, and K. Yoshioka, "Cmos low noise amplifier for microstrip readout design and results," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 301, no. 3, pp. 506–516, 1991.

[40] O. Toker, S. Masciocchi, E. Nygård, A. Rudge, and P. Weilhammer, "Viking, a cmos low noise monolithic 128 channel frontend for si-strip detector readout," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 340, no. 3, pp. 572–579, 1994.

[41] G. Alimonti *et al.*, "The belle silicon vertex detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 453, no. 1-2, pp. 71–77, 2000.

[42] Y. Horii, "Phd thesis, tohoku university (2010),"

[43] Z. Natkaniec *et al.*, "Status of the belle silicon vertex detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 560, no. 1, pp. 1–4, 2006.

[44] H. Hirano *et al.*, "A high-resolution cylindrical drift chamber for the kek b-factory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 455, no. 2, pp. 294–304, 2000.

[45] H. Bethe, "Theory of the passage of fast corpuscular rays through matter," in *Selected Works Of Hans A Bethe: (With Commentary)*, pp. 77–154, World Scientific, 1997.

[46] H. Bethe and B. W. Heitler, "On the stopping of fast particles and on the creation of positive electrons," 1934.

[47] H. Kichimi *et al.*, "The belle tof system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 453, no. 1-2, pp. 315–320, 2000.

[48] I. Nakamura, "Implemantation of timing information of ecl clusters," *Belle Internal Note, BN 774*, 2004.

[49] S. Villa, "Search for the $B^0 \to \gamma\gamma$ decay in the presence of off-time QED background.," *Belle Internal Note, BN 811*, 2005.

[50] Y. Ushiroda, A. Mohapatra, H. Sakamoto, Y. Sakai, M. Nakao, Q. An, and Y. F. Wang, "Development of the central trigger system for the BELLE detector at the KEK B factory," *Nucl. Instrum. Meth. A*, vol. 438, pp. 460–471, 1999.

[51] K. Akai, K. Furukawa, and H. Koiso, "SuperKEKB Collider," *Nucl. Instrum. Meth. A*, vol. 907, pp. 188–199, 2018.

[52] T. Abe *et al.*, "Belle ii technical design report," *arXiv preprint arXiv:1011.0352*, 2010.

[53] https://www.bnl.gov/newsroom/news.php?a=117285.

[54] C. Bernardini, G. F. Corazza, G. Di Giugno, G. Ghigo, J. Haissinski, P. Marin, R. Querzoli, and B. Touschek, "Lifetime and beam size in a storage ring," *Phys. Rev. Lett.*, vol. 10, pp. 407–409, May 1963.

[55] L. Piilonen, "B-KLM Summary Talk. 11th B2GM, mar, 2012," *B2GM*, 2012.

[56] V. Aulchenko *et al.*, "Electromagnetic calorimeter for Belle II," *Journal of Physics: Conference Series 587 no. 1,(2015) 012045*, 2015.

[57] https://www2.kek.jp/proffice/archives/feature/2010/BelleIIBPID.html.

[58] https://docs.belle2.org/record/2895/files/Lepton_identification_Moriond_2022__v2.pdf.

[59] Y. Iwasaki, B. Cheon, E. Won, X. Gao, L. Macchiarulo, K. Nishimura, and G. Varner, "Level 1 trigger system for the Belle II experiment," *IEEE Trans. Nucl. Sci.*, vol. 58, pp. 1807–1815, 2011.

[60] S. Lee, R. Itoh, N. Katayama, and S. Mineo, "Development of high level trigger software for belle ii at superkekb," in *Journal of Physics: Conference Series*, vol. 331, p. 022015, IOP Publishing, 2011.

[61] P. M. Lewis *et al.*, "First Measurements of Beam Backgrounds at SuperKEKB," *Nucl. Instrum. Meth. A*, vol. 914, pp. 69–144, 2019.

[62] F. Forti, "Snowmass whitepaper: The Belle II Detector Upgrade Program," 2022.

[63] A. Moll, "The software framework of the belle ii experiment," in *Journal of Physics: Conference Series*, vol. 331, p. 032024, IOP Publishing, 2011.

[64] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, "The belle ii core software: Belle ii framework software group," *Computing and Software for Big Science*, vol. 3, pp. 1–12, 2019.

[65] R. Brun and F. Rademakers, "Root — an object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1, pp. 81–86, 1997. New Computing Techniques in Physics Research.

[66] N. Katayama, R. Itoh, A. Manabe, and T. Sasaki, "Belle computing model," *Computer physics communications*, vol. 110, no. 1-3, pp. 22–25, 1998.

[67] B. Casey, "HadronB," *Belle Note 390*, January 2001. http://belle.kek.jp/secured/nbb/nbb.html.

[68] I. Adachi, "Selection Criteria in Classifying Beam Data."." https://belle.kek.jp/secured/runinfo/dst/skimmed_file/SkimmedFile/e000007/selection.html, Belle Internal.

[69] M. Gelb, T. Keck, M. Prim, H. Atmacan, J. Gemmler, R. Itoh, B. Kronenbitter, T. Kuhr, M. Lubej, F. Metzner, *et al.*, "B2bii: Data conversion from belle to belle ii," *Computing and Software for Big Science*, vol. 2, pp. 1–7, 2018.

[70] D. J. Lange, "The evtgen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1-2, pp. 152–155, 2001.

[71] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. . Barrand, *et al.*, "Geant4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[72] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, "GEANT3," 9 1987.

[73] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to pythia 8.1," *Computer Physics Communications*, vol. 178, p. 852–867, June 2008.

[74] B. Ward, S. Jadach, and Z. Was, "Precision calculation for $e^+e^- \to 2f$: The kk mc project," *Nuclear Physics B-Proceedings Supplements*, vol. 116, pp. 73–77, 2003.

[75] S. Jadach, Z. Was, R. Decker, and J. Kühn, "The τ decay library tauola, version 2.4," *Computer Physics Communications*, vol. 76, no. 3, pp. 361–380, 1993.

[76] G. Balossini, C. M. C. Calame, G. Montagna, O. Nicrosini, and F. Piccinini, "Matching perturbative and parton shower corrections to bhabha process at flavour factories," *Nuclear Physics B*, vol. 758, no. 1-2, pp. 227–253, 2006.

[77] C. C. Calame, G. Montagna, O. Nicrosini, and F. Piccinini, "The babayaga event generator," *Nuclear Physics B-Proceedings Supplements*, vol. 131, pp. 48–55, 2004.

[78] A. Hershenhorn, T. Ferber, and C. Hearty, "Ecl shower shape variables based on zernike moments," *Internal Note (Jan, 2017)*, 2017.

[79] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[80] G. Punzi, "Sensitivity of searches for new signals and its optimization," *arXiv preprint physics/0308063*, 2003.

[81] Y. et. al., "The π^0 and η Veto Tool," *BELLE2-NOTE-PH-2021-013*.

[82] S. Villa, "Search for the B $\to$ $\gamma\gamma$ decay in the presence of the off-time QED background, Belle Note 811." https://belle.kek.jp/secured/belle_note/gn811/note-811.pdf, Belle Internal.

[83] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2014.

[84] T. Keck, "Fastbdt: a speed-optimized multivariate classification algorithm for the belle ii experiment," *Computing and Software for Big Science*, vol. 1, pp. 1–11, 2017.

[85] G. C. Fox and S. Wolfram, "Observables for the analysis of event shapes in e^+e^- annihilation and other processes," *Phys. Rev. Lett.*, vol. 41, pp. 1581–1585, Dec 1978.

[86] G. Fasano and A. Franceschini, "A multidimensional version of the kolmogorov–smirnov test," *Monthly Notices of the Royal Astronomical Society*, vol. 225, no. 1, pp. 155–170, 1987.

[87] A. Hocker, X. Prudent, J. Therhaag, Y. Mahalalel, M. Backes, R. Ospanov, M. Kruk, M. Jachowski, A. Voight, A. Robert, *et al.*, "Tmva-toolkit for multivariate data analysis with root: users guide," tech. rep., 2007.

[88] J. E. Angus, "The probability integral transform and related results," *SIAM review*, vol. 36, no. 4, pp. 652–654, 1994.

[89] X. Zhou, S. Du, G. Li, and C. Shen, "Topoana: A generic tool for the event type analysis of inclusive monte-carlo samples in high energy physics experiments," *Computer Physics Communications*, vol. 258, p. 107540, 2021.

[90] J. P. Lees, V. Poireau, V. Tisserand, J. Garra Tico, *et al.*, "Measurement of *cp* asymmetries and branching fractions in charmless two-body *b*-meson decays to pions and kaons," *Phys. Rev. D*, vol. 87, p. 052009, Mar 2013.

[91] B. Pal, A. Schwartz, A. Abdesselam, I. Adachi, H. Aihara, S. Al Said, K. Arinstein, D. M. Asner, V. Aulchenko, T. Aushev, *et al.*, "Evidence for the decay $B^0 \to \eta\pi^0$," *Physical Review D*, vol. 92, no. 1, p. 011101, 2015.

[92] B. Aubert, Y. Karyotakis, J. Lees, V. Poireau, E. Prencipe, X. Prudent, V. Tisserand, J. G. Tico, E. Grauges, M. Martinelli, *et al.*, "B meson decays to charmless meson pairs containing η or η' mesons," *Physical Review D*, vol. 80, no. 11, p. 112002, 2009.

[93] H. Cramér, "Mathematical methods of statistics, 1946," *Department of Mathematical SU*, 1946.

[94] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Reson. J. Sci. Educ*, vol. 20, pp. 78–90, 1945.

[95] L. Lista, *Statistical Methods for Data Analysis in Particle Physics*, vol. 909. Springer, 2016.

[96] G. Cowan, *Statistical data analysis*. 1998.

[97] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling," *eConf*, vol. C0303241, p. MOLT007, 2003.

[98] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth. A*, vol. 389, pp. 81–86, 1997.

[99] F. James, "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1," 1994.

[100] F. James and M. Winkler, "MINUIT User's Guide," 6 2004.

[101] K. Cranmer, "Kernel estimation in high-energy physics," *Computer Physics Communications*, vol. 136, no. 3, pp. 198–207, 2001.

[102] H. Albrecht *et al.*, "Search for Hadronic $b \to u$ Decays," *Phys. Lett. B*, vol. 241, pp. 278–282, 1990.

[103] T. Horiguchi, A. Ishikawa, H. Yamamoto, I. Adachi, H. Aihara, S. Al Said, D. Asner, V. Aulchenko, T. Aushev, R. Ayad, *et al.*, "Evidence for isospin violation and measurement of cp asymmetries in $B \to K^*(892)\gamma$," *Physical review letters*, vol. 119, no. 19, p. 191802, 2017.

[104] S. Choudhury, S. Sandilya, K. Trabelsi, A. Giri, I. Adachi, H. Aihara, D. Asner, H. Atmacan, T. Aushev, R. Ayad, *et al.*, "Measurement of the B^+/B^0 production ratio in e^+e^- collisions at the $\Upsilon(4S)$ resonance using $B \to J/\psi(ll)K$ decays at Belle," *Physical Review D*, vol. 107, no. 3, p. L031102, 2023.

[105] H. Svidras *et al.*, "Measurement of the data to MC ratio of photon reconstruction efficiency of the Belle II calorimeter using radiative muon pair event," *https://docs.belle2.org/record/2133/files/BELLE2-NOTE-TE-2020-026.pdf*.

[106] H. W. K. et al., "Study of high energy photon detection efficiency using radiative Bhabha," *Belle Note 499, (2002)*.

[107] "Number of B in HadronB(J)," *Belle internal, 1316*.

[108] "KEK High Energy Accelerator Research Organization. SuperKEKB Project: Projected Luminosity." https://public.belle2.org/downloads/Belle2_Luminosity_2022Mar.pdf.

[109] G. B. Arfken, F. E. Harris, and H. J. Weber, *Mathematical Methods for Physicists*. 2012.

[110] "Wikimedia commons. comparison of 1d histogram and kde, 2010. url:," *https://commons.wikimedia.org/wiki/File:Comparison_of_1D_histogram_and_KDE.png*.