Measurement of the branching fraction and CP asymmetry of $B^0 \to \pi^0\pi^0$ decays at Belle II

by

Francis Pham

ORCID: 0000-0003-0608-2302

A thesis submitted in total fulfillment for the
degree of Doctor of Philosophy

in the
School of Physics
THE UNIVERSITY OF MELBOURNE

August 2023

THE UNIVERSITY OF MELBOURNE

Abstract

School of Physics

Doctor of Philosophy

by Francis Pham
ORCID: 0000-0003-0608-2302

This thesis presents a measurement of the branching fraction and CP-violation asymmetry in $B^0 \to \pi^0\pi^0$ decays. The analysis uses a sample that corresponds to 198×10^6 $B\overline{B}$ pairs, collected by the Belle II experiment at the SuperKEKB accelerator in Tsukuba, Japan. Among collider experiments, only Belle II can efficiently record $B^0 \to \pi^0\pi^0$ events at rates enabling competitive measurements to previous results.

The large uncertainties of the branching fraction and CP-violation asymmetry of $B^0 \to \pi^0\pi^0$ decays are the greatest limitation in determining the least known angle of the unitarity triangle, ϕ_2. To enhance the precision of the $B^0 \to \pi^0\pi^0$ measurement, this analysis employs improved machine learning algorithms to suppress misreconstructed photons and continuum background. Simulated samples are used to optimise event selection criteria, compare observed data distributions with expectations, study background sources, and model distributions. The branching fraction and direct CP asymmetry are extracted from a three-dimensional unbinned extended maximum likelihood fit simultaneously to events divided into seven data sets.

The measured branching fractions and direct CP asymmetries are:

$$\mathcal{B}(B^0 \to \pi^0\pi^0) = (1.38 \pm 0.27 \pm 0.22) \times 10^{-6}$$
$$\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0) = 0.14 \pm 0.46 \pm 0.07$$

where the first uncertainty is statistical and the second uncertainty is systematic. These values are in agreement with previous results. The statistical and systematic uncertainty of the $\mathcal{B}$ measured in this work is similar in size to those obtained by Belle despite using a dataset almost a quarter in size. This demonstrates Belle II's potential for high-precision measurements of charmless hadronic B decays measurements, enabling the parameter space of new physics to be further constrained.

Declaration of Authorship

I, FRANCIS PHAM, declare that this thesis titled, 'Measurement of the branching fraction and *CP* asymmetry of $B^0 \to \pi^0\pi^0$ decays at Belle II' and the work presented in it are my own. I confirm that:

- The thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface;
- due acknowledgement has been made in the text to all other material used; and
- the thesis is fewer than the maximum word limit in length, exclusive of tables, maps, bibliographies and appendices as approved by the Research Higher Degrees Committee.

Signed:

Date: 15th August 2023

Preface

This thesis was written with the academic supervision of Professor Martin Sevior.

The Belle II experiment is an large international scientific collaborative effort and represents the collective work of hundreds of members from dozens of institutes around the world. Working groups, consisting of teams of theorists and experimentalists, are formed to work on a different aspects of the Belle II physics program. The research conducted for this thesis was carried out as part of the hadronic B working group.

Chapter 1 is a brief introduction to CP violation and the motivation behind this research. Chapter 2 provides the theoretical background necessary to understand the $B^0 \to \pi^0\pi^0$ decay mode from an experimental and theoretical perspective. Chapter 3 details the Belle II experiment, focusing on the working principle behind each subdetector and how the data flows from detector to storage.

Chapter 4 outlines the software, the dataset, and the preselection used throughout the analysis. Chapter 5 describes the development of a new boosted decision tree classifier that is used to discriminate between genuine and misreconstructed photons. Chapter 6 details the selection and reconstruction of the $B^0 \to \pi^0\pi^0$ candidates and describes the continuum and $B\overline{B}$ background. Chapter 7 describes the training and testing of another boosted decision tree classifier that is used to discriminate between signal and continuum background. Chapter 8 describes the probability distribution function used to model the signal and background components. Chapter 9 describes the $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ control mode used to validate the $B^0 \to \pi^0\pi^0$ analysis.

Chapter 10 applies the full analysis to experimental data the branching ratio and CP asymmetry parameters are extracted. Chapter 11 discusses the sources of systematic uncertainty. The research carried out led to the publication 'Measurement of the branching fraction and CP asymmetry of $B^0 \to \pi^0\pi^0$ decays using 198×10^6 $B\overline{B}$ pairs in Belle II data' published in Physics Review D in July 2023 [1]. The author was the point of contact for the peer review process, with comments from the internal review process providing some editorial assistance. Chapter 12 summarises all the results of this thesis.

This work was made possible by the Research Training Program scholarship provided by the Australian Government.

Acknowledgements

During my four years of study, I have met so many exceptional people who have had a profound impact on my academic journey and outlook on life. It has been my pleasure to work alongside so many brilliant and like-minded people. Foremost, my gratitude goes to my supervisor Professor Martin Sevior, whose constant guidance and unwavering support have been invaluable. His insights, mentorship, and encouragement have been the bedrock of both my academic and personal development. I also extend my thanks to the members of my supervisory panel, Professors Stuart Wyithe and Geoff Taylor, for their valuable input and constructive feedback. I wish to also express my sincere appreciation to the B hadronic working group. Their tireless dedication, insightful comments, and hours of discussions have significantly enhanced the quality of my research. Special recognition goes to Diego, Mirco, Alan, Jim, Bob, Taichiro, and Tom for their generous contributions to my work.

It has been a privilege to work alongside the entire particle physics group in Melbourne. Although I came from Sydney, everyone made me feel very welcome and I never felt like I didn't belong. A big thank you to Marcel, Michael, Mike, Daniel F, Daniel M, Neuton, Hans, Brian, and Cate for their friendship and most importantly for being sources of constant distraction. To Priykana and Tommy for being amazing people during my KEK visit. I couldn't imagine anyone else with whom I would rather visit every bar in Tsukuba with.

There are no words to describe how grateful I am for having such supportive parents. Whether its driving ten hours to help me move, or cooking my favourite meal every time I came home, their unwavering dedication and selflessness have given me every possible opportunity to succeed. I also couldn't forget my friends from Sydney; Andrew, John-Mark, Sebastian, Nick, Donna, Vu, and James. Although so much has changed since I've been gone, I know I can always rely on your friendship to brighten up my day, week, and even year.

I would like to express my heartfelt thanks to Jackie, Catherine, Ngan, and Linh for being some of my closest friends throughout my time in Melbourne. The late night zoom sessions got me through two years of lockdowns and I know that we will remain life-long friends. Last but not least, I am profoundly grateful to my partner, Yen Ha My, for her unwavering faith in me. Her boundless support, encouragement, and belief in my potential have driven me to surpass my own expectations. I hope one day I can find the right words to express my eternal gratitude to everyone who has supported me throughout my journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A fundamental and powerful concept in physics is symmetry. A system is said to exhibit symmetry if after an operation it remains indistinguishable from its original state. One such symmetry, known as CP, inverts the internal quantum numbers (C) and the spatial coordinates (P), transforming particles into antiparticles and vice versa. If CP were an exact symmetry of nature, it would be conserved in all particle interactions, and there would be an equal amount of matter and antimatter in the universe. However, cosmological measurements show that the observable universe is overwhelmingly dominated by matter [19]. In 1967, Sakharov [20] proposed three necessary conditions for a matter-dominated universe. One of these conditions is the violation of CP, in which processes involving particles are slightly different from those involving antiparticles.

In 1964, Cronin and Fitch [21] discovered CP violation in the decay of neutral kaons. They observed that for every five hundred CP conserving $K_L^0 \to \pi^+\pi^-\pi^0$ decays there was one CP violating $K_L^0 \to \pi^+\pi^-$ decay. This observation showed that the laws of physics treated matter and antimatter differently and that one could unambiguously differentiate them. In 1973, Kobayashi and Maskawa (KM) provided a mechanism for CP violation [22]. However, this CP violation is many orders of magnitude too small to explain the observed matter-antimatter asymmetry of the universe. Therefore, sources of violation of CP beyond the KM model are required to account for the preference of matter over antimatter. In 1981, Sanda, Carter *et al.* [23] showed that bottom-meson systems were an excellent area for studying CP violations, since many decay modes were expected to exhibit large CP violation.

To produce and measure B mesons decays, two ‘B-factories’, Belle and Babar, were constructed in the 1990s. In 2001, the Belle experiment in Japan [24], and the BaBar experiment in the United States [25] observed large CP asymmetries consistent with the KM mechanism in $B^0 \to J/\psi K_S^0$ decays. As a result, Kobayashi and Maskawa received the 2008 Nobel Prize in Physics for the confirmation of CP violation in B mesons. Belle’s achievements demonstrated substantial potential for B-physics and led to a second experiment named Belle II. This experiment, which featured significantly upgraded accelerator and detector, started collecting collision data in March 2019. Belle II is expected to achieve a peak luminosity that is forty times greater than the one achieved by its predecessor and to collect fifty times more data.

The Unitarity Triangle provides a visualisation of CP violation and is a geometric representation in the complex plane. This triangle is defined by three angles, ϕ_1, ϕ_2, and ϕ_3, which correspond to the amount of CP violation in different processes. If all CP violation is indeed described by the KM mechanism, then the angles should be internally consistent and sum up to 180 degrees. Any deviation would signal the presence of new physics. Among the angles of the unitarity triangle, ϕ_2 is the least known. The greatest limitation to precise measurements lies in the uncertainty of the branching fraction, $\mathcal{B}$, and the direct CP asymmetry parameter, $\mathcal{A}_{CP}$, of the $B^0 \to \pi^0\pi^0$ decay mode. The world average values of $\mathcal{B}(B^0 \to \pi^0\pi^0) = (1.59 \pm 0.26) \times 10^{-6}$ and $\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0) = 0.33 \pm 0.22$ [26] combine measurements reported by the BaBar [27] and Belle [28] collaborations.

The measurement of the decay properties of the $B^0 \to \pi^0\pi^0$ mode presents significant challenges. First, the decay is extremely rare and occurs only in approximately one in a million B meson decays. Second, the final state of this decay consists solely of photons, as well as a substantial background from $e^+e^- \to q\bar{q}$ $(q = u, d, s, c)$ events, which overwhelms the signal. Lastly, the reconstruction of the signal is susceptible to neutral backgrounds from beam interactions that mimic photons. To overcome these challenges, this analysis takes advantage of the upgraded Belle II detector, along with improved analysis techniques and machine learning algorithms. These developments enable a measurement of the branching fraction and $\mathcal{A}_{CP}$ of the $B^0 \to \pi^0\pi^0$ decay mode with statistical and systematic uncertainties that are comparable to those achieved by Belle with a data set that is approximately four times smaller.

Chapter 2

Physics of Belle II

The Belle experiment, which operated from 1999 to 2010, collected more than 1040 fb^{-1} of data at several Υ resonances. This resulted in more than six hundred published papers, including the first measurement of large CP violation in the $B^0 \to J/\psi K_S^0$ channel [24, 25] which led to the 2008 Nobel Prize in Physics. Building on Belle's success, a new experiment called Belle II began data collection in March 2019. Belle II features an upgraded detector and accelerator, allowing for a significantly higher data-taking rate and improved precision. The new detector incorporates state-of-the-art technologies, including silicon vertex detectors and particle identification systems, that will enable Belle II to perform high-precision measurements of B meson decays and other physics processes. Belle II aims to collect a dataset fifty times larger than Belle's, enabling a more extensive exploration of CP violation in the B meson sector and a greater sensitivity for possible New Physics contributions.

This chapter presents the theoretical background necessary to understand this thesis. It provides an introduction to the Standard Model (SM), CP violation in the B meson system, and the physics underlying the decay of $B^0 \to \pi^0\pi^0$ from an experimental and theoretical perspective. There is an emphasis on the unitarity angle, ϕ_2, and, in particular, its determination from $B \to \pi\pi$ decays, the current constraints on its value, and theoretical expectations.

2.1 The Standard Model of particle physics

The Standard Model is the theoretical framework that forms the foundation of modern particle physics. It describes the fundamental particles of nature and three of the four fundamental forces that govern their interactions: the strong force, the weak force, and the electromagnetic force. However, the model is not complete and may represent the 'low-energy limit' of a more fundamental theory of everything. The SM does not account for the gravitational force, dark matter, the large difference in the strength of the fundamental forces, and most relevant to this thesis, the matter-antimatter asymmetry in the universe. In addition, the SM is an 'effective' theory, containing free parameters that must be experimentally determined. There are 19 free parameters that require determination, and the model's inability to predict these parameters clearly indicates the need for a more fundamental theory. Despite its shortcomings, the SM has demonstrated remarkable precision in its predictions and remains the most comprehensive explanation of the nature of reality.

2.1.1 The Particles of the Standard Model

The Standard Model describes nature as being composed of two fundamentally different types of particles, fermions, and bosons. Fermions have half-integer spin and follow Fermi-Dirac statistics, while bosons have integer spin and follow Bose-Einstein statistics. The fundamental fermions are spin-1/2 particles and can be further classified into two types known as quarks and leptons. These categories are further split into three 'generations' based on the masses of the particles. Each generation consists of two flavours, each with a different 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 the up, charm, and top quarks is $+\frac{2}{3}$, while the down, strange, and bottom quarks carry an electric charge of $-\frac{1}{3}$. Similarly, lepton generations also consist of three flavours; electron (e) for the first generation, muon (μ) for the second, and tau (τ) for the third. Each of these particles has a corresponding neutrino: the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). Unlike quarks, the electron, muon, and tau leptons have an electric charge of -1, while neutrinos are electrically neutral.

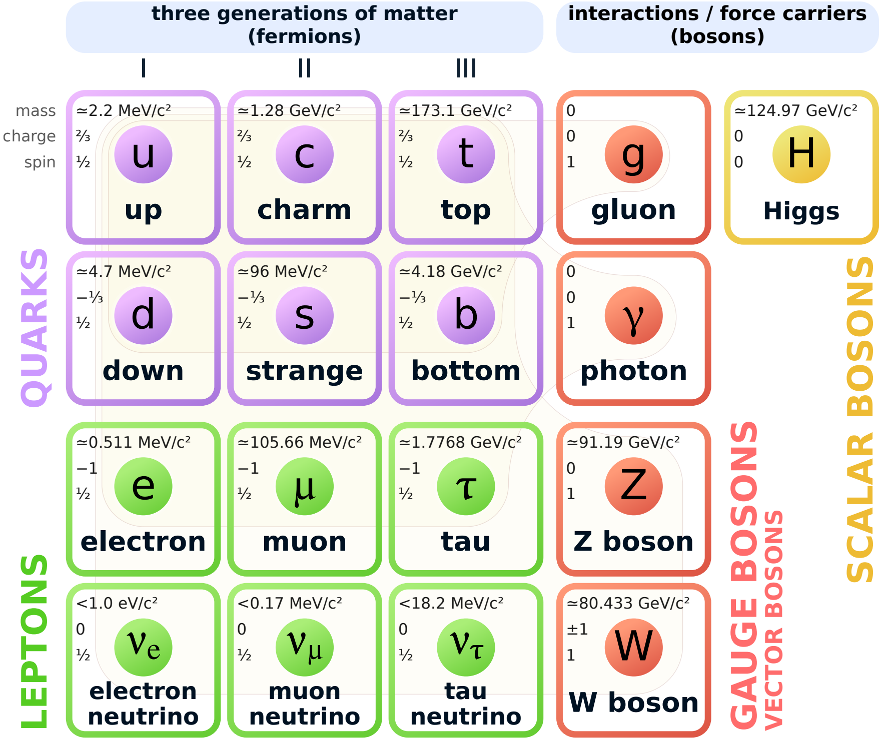


FIGURE 2.1: The Standard Model's elementary particles with associated mass, electric charge, and spin [2].

In the SM there are four spin-1 gauge bosons and one scalar boson. The gauge bosons are mediators of the fundamental forces that govern interactions among particles. Meanwhile, the only scalar boson is the Higgs boson (H^0), an excitation of the Higgs field which is responsible for giving fundamental particles their mass. The strong force is meditated by eight types of gluons (g), the electromagnetic force is meditated by the photon (γ), and the weak force is meditated by two charged W bosons ($W^{\pm}$) and one neutral Z boson (Z^0). The fundamental SM particles are summarised in Figure 2.1.

All other particles are composite particles called hadrons, which consist of quarks and/or antiquarks in a bound state. Hadrons are classified as mesons if they are composed of a quark and an antiquark, and baryons if they are composed of three quarks/antiquarks. Hadrons have excited states called resonances. These resonances have the same quark content but different quantum numbers. An excitation refers to the state of a particle which has more energy than its ground state. For example, $\Upsilon(3S)$ and $\Upsilon(4S)$ are spin-1 bound states consisting of an bottom and an anti-bottom quark ($b\bar{b}$) with zero orbital angular momentum (S) in the 3^{rd} and 4^{th} excited state, respectively. These excited states of the meson are analogous to the excited states of the hydrogen atom, except

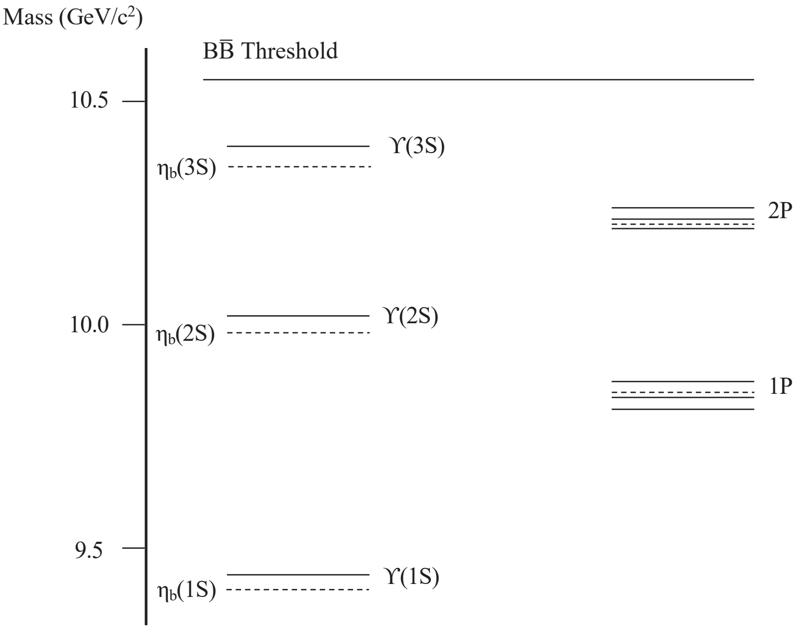


FIGURE 2.2: Energy-level diagrams for $b\bar{b}$ states consist of various states labeled as 1S, 2S, 3S, 1P, and 2P. The states labeled 1S, 2S, and 3S have orbital angular momentum $L = 0$, whereas the 1P and 2P states have $L = 1$. The intrinsic quark spins can couple to the $S = 0$ state, resulting in the creation of states with total angular momentum $J = L$, such as the η_b states. These states are represented by dashed lines. Additionally, the intrinsic quark spins can couple to give $S = 1$, leading to the formation of states such as the Υ states. States with $S = 1$ are indicated by solid line.

that it is the strong force, rather than the electromagnetic force, that binds the particles. As shown in Figure 2.2, different energy levels and quantum numbers are labelled as different particles. Spectroscopic notation is commonly used to classify these particles:

$$n^{2S+1}L_J$$

where n represents the principal quantum number of the excited state. S is the total spin quantum number, while L represents the orbital angular momentum of the particle, with values denoted by letters S, P, D, and F, which correspond to total angular momentum values of 0, 1, 2, and 3, respectively. Finally, J represents the total angular momentum quantum number, which is the vector sum of S and L. For example, $\eta_b(2S)$ is a spin-0 particle with zero angular momentum in the second excited state (2^1S_0) while $\Upsilon(3S)$ is spin-1 with zero angular momentum in the third excited state (3^3S_1).

2.1.2 The Fundamental Forces

In descending order of strength at the energy scales accessible with accelerators, the fundamental forces of the SM are the strong force, the electromagnetic force, and the weak force. Typical Feynman diagrams for each of the fundamental forces are shown in Figure 2.3. These fundamental forces are responsible for all particle interactions, including decay processes and scattering events. The gravitational force is not described by the SM, and is considered too weak to measurably affect particle interactions at accessible energies.

The interaction of all particles with an electric charge is governed by the electromagnetic force. The strong force is the interaction that takes place between gluons and quarks, binding quarks into hadrons, which then combine to form nuclei. Particles that experience the strong force carry an additional quantum number called 'colour'. For quarks, these colours are red, green, and blue, while for antiquarks, the corresponding colours are anti-red, anti-green, and anti-blue. This colour nomenclature is simply a label that reflects the empirical observation that quarks cannot freely exist and must be bound in 'colourless' combination. For example, if the quark in a meson has a red colour charge, then its partner antiquark must have an anti-red colour charge. The weak force affects both quarks and leptons and comes in two types: charged and neutral. The charged weak interactions, mediated by $W^{\pm}$ particles, can change the flavour of the quarks, whereas the neutral weak interactions, mediated by the Z^0 particle, cannot change the flavour of the quarks.

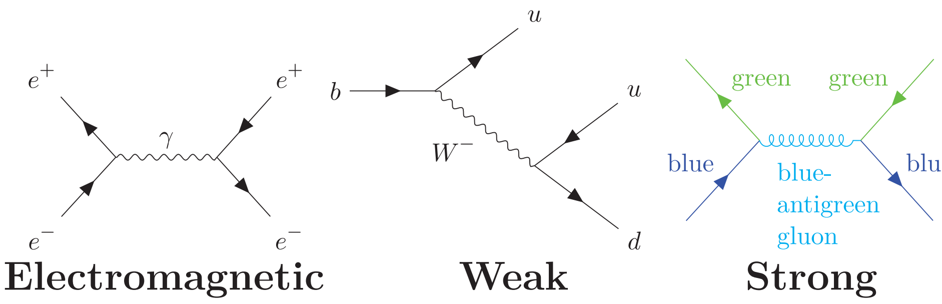


FIGURE 2.3: The fundamental forces of the Standard Model and their associated Feynman diagrams.

2.1.3 Particle production

In the SM, stable particles are limited to the electron, the neutrinos, the photon, the proton, and the neutron when bound in an atomic nucleus. Although unstable particles can be produced naturally, such as atmospheric muons from cosmic rays, particle accelerators that collide particles are needed to produce and study more massive particles with high precision. In these accelerators, two particles are collided at high energies, resulting in a wide range of interactions. Many decay products are possible, as long as they conserve energy, momentum, charge, colour, baryon number, and lepton number.

Predicting the specific products of any given collision is impossible, only their probabilities can be determined. The probability that two particles will collide and produce a particular particle is represented by the 'cross section', denoted as σ. For GeV-scale e^+e^- collisions, the cross section is typically measured in units of femtobarn (fb^{-1}), which is equivalent to $10^{-43}\,\mathrm{m}^2$. The luminosity, denoted as L, quantifies the rate at which a collider explores interaction cross sections. To be more precise, the total number of interactions for a specific process can be obtained by multiplying the luminosity by the cross section of that process, expressed as $L \times \sigma =$ number of interactions. The integral of the luminosity over time is called the integrated luminosity and is a measurement of the size of the collected data. The inverse femtobarn (fb^{-1}) is the unit that is typically used. For example, if a detector has accumulated 100 fb^{-1} of integrated luminosity, one expects to find 100 events per femtobarn of cross section within the dataset. The integrated luminosity of Belle and BaBar over time is shown in Figure 2.4.

2.1.4 Resonances and decays

In high-energy collisions, short-lived particles known as resonances can be produced. These particles have extremely short lifetimes, approximately 10^{-23} seconds, and consequently, they do not traverse an appreciable distance before decaying. This makes direct measurements impossible. Instead, their existence is inferred by analysing the sum of energies of their decay products. The energy sum tends to centre around specific values, creating peaks in the data that exhibit a Breit-Wigner distribution [29]. These peaks indicate the formation of an intermediate particle during the collision, which subsequently decays into longer-lived particles.

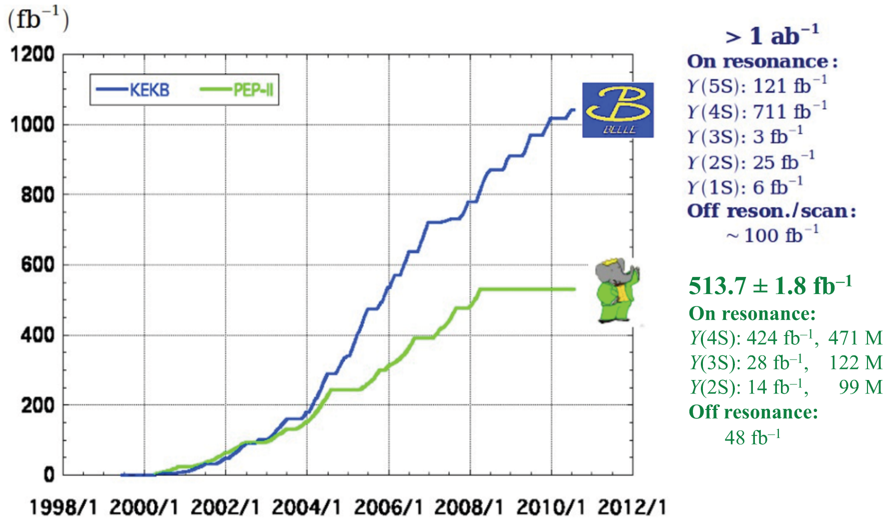


FIGURE 2.4: Integrated luminosity of Belle (1999-2010) and BaBar (1999-2008). Belle collected approximately 1000 fb^{-1} while BaBar collected approximately 500 fb^{-1} at the various Υ resonances [3].

Resonant states have a well-defined mass, width, and lifetime. The width at half-maximum of this peak, known as the decay width, is given by

$$\Gamma = \frac{\hbar}{\tau} \tag{2.1}$$

where $\hbar$ is the reduced Planck constant and τ is the characteristic mean lifetime. This equation implies that the decay width is inversely proportional to the mean lifetime, such that a rapidly decaying particle will exhibit a larger decay width. For example, the $\Upsilon(4S)$ resonance has a decay width of 20.5 ± 1.2 MeV, which corresponds to a lifetime of about 0.32×10^{-22} seconds. On the other hand, the B^0 meson, with a decay width of approximately 4.1×10^{-10} MeV, has a correspondingly longer lifetime of approximately 1.52×10^{-12} seconds. The Breit-Wigner distribution of the Higgs boson, which has a predicted decay width of 4.1 MeV and a corresponding lifetime of 1.6×10^{-22}, is shown in Figure 2.5.

Particle decay is a spontaneous process in which one particle transforms into two or more particles, provided conservation laws are followed. Depending on whether the interaction is strong, weak, or electromagnetic, different timescales are involved. Strong decays are exceptionally fast and often occur in approximately 10^{-23} seconds. On the other hand, electromagnetic decays, such as the emission of a photon from an excited atomic state, typically occur on a timescale of 10^{-16} to 10^{-20} seconds. Weak decays vary more widely and can be as fast as approximately 10^{-13} seconds for the decay of

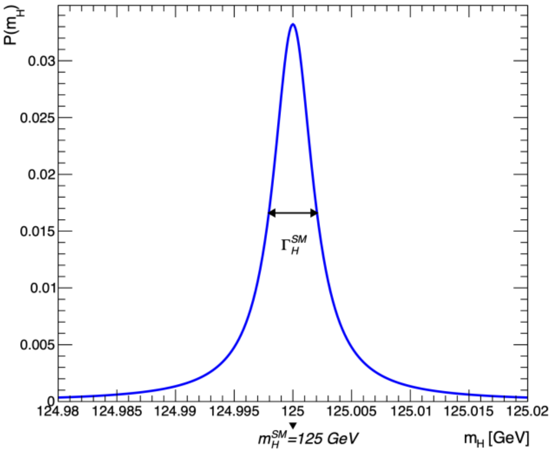


FIGURE 2.5: The Breit-Wigner distribution of the Higgs boson resonance with a width Γ_H^{SM} of 4.1 MeV [4].

the tau lepton, or as slow as about 15 minutes for the decay of a neutron. The different ways through which a particle can decay are known as decay modes or channels. The fraction of particles that undergo decay through a specific decay mode i, relative to the total number of particles, is defined as the branching fraction $\mathcal{B}_i$. In other words, the branching fraction $\mathcal{B}_i$ is the probability that a particle will decay through a particular decay mode i. It is given by the equation,

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma} \tag{2.2}$$

where Γ_i is the partial decay width for the decay mode i, and Γ is the total decay width of the particle. The partial decay width is the fraction of the total width of a resonance that is due to a specific decay mode, and the total width of the resonance is the sum of the partial widths for all potential decay channels. For example, consider a particle resonance with a total width of 10 MeV/c^2 that can only decay into two modes. If this particle decays into one mode and the partial width is 3 MeV/c^2, then the partial width of the other mode must be 7 MeV/c^2.

2.1.5 Feynman Diagrams

Any particle interaction, such as production and decay, can be represented pictorially through Feynman diagrams. Each diagram represents one possible way the interaction could have occurred. In Feynman diagrams, particles are depicted as arrows pointing from left to right, while antiparticles are represented by arrows pointing from right to left. The interactions of particles are represented as a vertex: a point where the lines representing the different particles come together. Only the incoming and outgoing lines are real particles and obey Einstein's energy-momentum relation, while everything in between are virtual particles and violate energy-momentum conservation. This violation only occurs for an exceedingly short amount of time, as allowed by Heisenberg's uncertainty principle.

Each vertex within a diagram suppresses the overall contribution of that particular diagram. The probability of a transition rapidly decreases with the number of vertices, so that diagrams with many vertices contribute very little to the overall interaction. As shown in Figure 2.6, diagrams with no closed loops are known as 'tree diagrams' and represents the leading-order process of a particle interaction. The term 'tree' refers to the branching structure of these diagrams, where the mediator (W, Z, γ, g) serves as a branch connecting the incoming and outgoing particles. Diagrams with loops are typically called 'loop diagrams' and represent higher-order corrections to the tree-level processes. These involve virtual particles circulating in the loops. In general, the heavier the virtual particles involved, the more suppressed the decay. 'Penguin diagrams' are a specific type of loop diagram that involves the weak interaction between quarks.

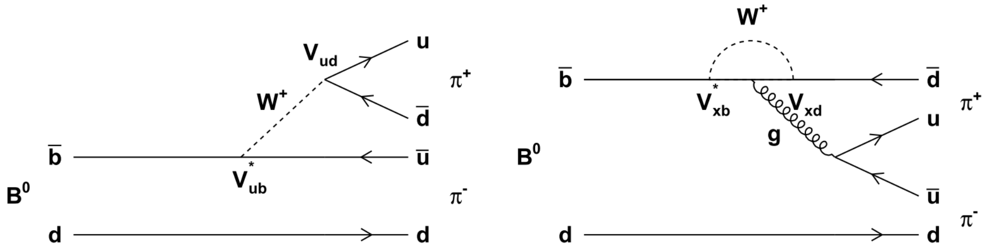


FIGURE 2.6: Leading order Feynman diagrams for $B^0 \to \pi^+\pi^-$ decays. First order colour-allowed tree diagram (left) and second-order colour-suppressed penguin diagram (right) where $x = u, c, t$.

Decays can also be colour-allowed or colour-suppressed. Recall that all combination of quarks and antiquarks must be colourless, meaning that the total color charge equals zero. In decays involving the weak or strong interaction, a virtual particle, either a W or a gluon, can be emitted. This virtual particle can then decay into two quarks $(q\bar{q})$. In colour-allowed decays, the two quarks that form a charged hadron may be in any colourless combination. However, in colour-suppressed decays, the quark from the virtual W or gluon is combined with an antiquark other than its partner. In this scenario, only one out of the three possible colours for the quark can form a colourless combination with the antiquark. Consequently, a colourless combination is only possible one-third of the time. For example, in the right-hand diagram of Figure 2.6, if the d 'spectator' quark possesses a green colour charge, then the $\bar{u}$ quark must possess an anti-green colour charge to form a colourless π^0. Similarly, to form two π^0 mesons, the quark (antiquark) from the W boson decay must pair with the spectator antiquark (quark), and hence the $B^0 \to \pi^0\pi^0$ decay is also colour-suppressed.

2.2 The Cabibbo-Kobayashi-Maskawa Matrix

In 1963, only the u, d, and s quarks were known. The rates of weak transitions involving the s quark were found to be different from those involving the u quark. To preserve the universality of the weak interaction, the idea that quarks couple to the $W^{\pm}, Z^0$ bosons with the same strength, Nicola Cabbibo [30] proposed that the weak eigenstates d' and s' are actually mixtures of flavour eigenstates d and s,

$$d' = d\cos\theta_c + s\sin\theta_c, \quad s' = -d\sin\theta_c + s\cos\theta_c \tag{2.3}$$

or, in matrix form;

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \tag{2.4}$$

where V_{ij} represents elements of the Cabibbo matrix, with i being either u or c, and j being either d or s. The specifics of these elements will be explained shortly. The Cabibbo angle has been experimentally determined to be approximately $\theta_c = 13^\circ$ [26]. The non-zero value of this angle indicates that the weak interaction does not respect quark generations. Nonetheless, the universality of the weak interactions is respected since, by construction, the sum of all couplings of any one of the up-type quarks to all the down-type quarks is the same for all generations,

$$|V_{ud}|^2 + |V_{us}|^2 = |V_{cd}|^2 + |V_{cs}|^2 = 1 \tag{2.5}$$

$$\cos\theta_c{}^2 + \sin\theta_c{}^2 = (-\sin\theta_c)^2 + \cos\theta_c{}^2 = 1 \tag{2.6}$$

In other words, the strength of the weak interaction remains the same when changing from one quark flavour to another.

To incorporate the CP violation into the SM, Makoto Kobayaski and Toshihide Maskawa extended the number of quark generations from two to three. They proposed that the weak eigenstates, d', s', b' are related to the flavour eigenstates d, s, b via the 3×3 unitarity Cabibbo-Kobayaski-Maskawa (CKM) matrix. This relationship is shown below with the magnitudes of the elements specifying the amount of quark flavour mixing in the weak interaction.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.7}$$

The number of free parameters for the n generations of quarks is $(n-1)^2$. The CKM matrix with three generations ($n = 3$) of quarks can be represented by four free parameters; commonly, three angles and one irreducible phase that represents SM CP violation. It can also be seen that with only two generations ($n = 2$) of quarks, there is only one free parameter and hence no CP violation. The current best estimates of the magnitudes of the CKM matrix elements are [26]:

$$V_{\text{CKM}} = \begin{pmatrix} 0.97373 \pm 0.00031 & 0.2243 \pm 0.0008 & 0.00382 \pm 0.0002 \\ 0.221 \pm 0.004 & 0.975 \pm 0.006 & 0.0408 \pm 0.0014 \\ 0.0086 \pm 0.0002 & 0.0415 \pm 0.0009 & 1.014 \pm 0.029 \end{pmatrix} \tag{2.8}$$

The diagonal elements are close to one while the off diagonal elements are significantly smaller. This corresponds to the experimental observation that quark flavour transitions ‘prefer’ to be in the same generation. Consequently, $b \to u$ transitions, such as in $B^0 \to \pi^0\pi^0$ decays, are suppressed since they are proportional to the smallest CKM element $|V_{ub}|$.

In the ‘standard’ parameterisation of the CKM matrix, there exist three Euler angles $(\theta_{12}, \theta_{23}, \theta_{13})$ and one CP-violating phase δ_{13}. Cosine and sines of the angles are denoted c_{jk} and s_{jk} respectively.

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -s_{23}c_{12} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{2.9}$$

If $\theta_{23} = \theta_{13} = 0$, the third generation does not mix, and Equation 2.4 can be recovered and $\theta_{12} = \theta_c$. An alternative parameterisation is the Wolfenstein parameterisation [31] that has the following relations:

$$\lambda^2 = \frac{|V_{us}|^2}{|V_{ud}|^2 + |V_{us}|^2} = s_{12}^2, \tag{2.10}$$

$$A^2\lambda^4 = \frac{|V_{cb}|^2}{|V_{ud}|^2 + |V_{us}|^2} = s_{23}^2, \tag{2.11}$$

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \tag{2.12}$$

where $\bar{\rho} = \rho(1 - \lambda^2/2 + ...)$ and $\bar{\eta} = \eta(1 - \lambda^2/2 + ...)$. Using this parameterisation, the CKM matrix V_{CKM} can be expressed to an order of λ^3 as follows:

$$V_{\text{CKM}} \approx \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} \tag{2.13}$$

where $\lambda \approx \sin(\theta_c)$. A, ρ and η are real numbers of order unity. The accuracy of the λ^3 approximation is $\mathcal{O}(0.1\%)$, which is sufficient given current experimental and theoretical uncertainties. The parameter η amounts to an irreducible phase that is responsible for all CP-violating processes since the CKM matrix and its complex conjugate, which describe the rotation between weak and mass eigenstates for antiquarks, would be different. The fits for the Wolfenstein parameters using a global fit to all available measurements and imposing the SM constraints such as unitarity of the three generation CKM matrix by the CKMFitter Group [9] are shown below,

$$\begin{aligned} \lambda &= 0.22500 \pm 0.00067, & A &= 0.826^{+0.018}_{-0.015}, \\ \bar{\rho} &= 0.159 \pm 0.010, & \bar{\eta} &= 0.348 \pm 0.010 \end{aligned} \tag{2.14}$$

2.3 The Unitarity Triangle

The unitarity of the CKM matrix imposes the conditions $\sum_k V_{ik}V_{jk}^* = 0$. Although this condition yields six equations, or equivalently, six unitarity triangles of equal area, most have large differences between the lengths of their sides. This makes them difficult to explore experimentally. It is only when the unitarity condition is applied to the first and third columns of Equation 2.7 that an equation with terms of approximately equal magnitudes is obtained. This equation in the complex plane is shown in Figure 2.7.

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{2.15}$$

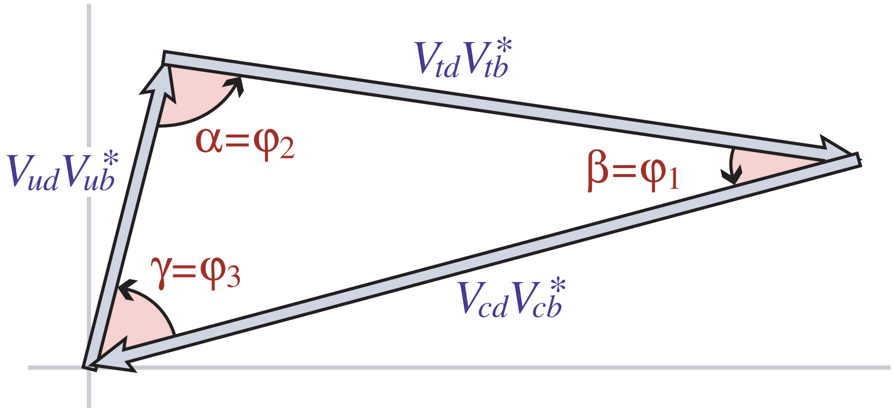


FIGURE 2.7: Graphical representation of the unitarity constraint $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^*$ as a triangle in the complex plane [5].

Dividing each side by the most precisely known of the three terms $V_{cd}V_{cb}^*$

$$1 + z_1 + z_2 = 0, \quad \text{where} \quad z_1 = \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}, \quad z_2 = \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \tag{2.16}$$

This can be represented in the complex plane as a closed loop, and is commonly referred to as 'the Unitarity Triangle', as shown in Figure 2.8. Using the Wolfenstein parametrisation the vertices are exactly (0,0), (1,0) and $(\bar{\rho}, \bar{\eta})$. The three internal angles of the triangles are given by

$$\phi_1 = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{cd}V_{cb}^*}\right), \quad \phi_2 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{td}V_{tb}^*}\right), \quad \phi_3 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right) \tag{2.17}$$

or in terms of ρ and η,

$$\tan\phi_1 = \frac{\bar{\eta}}{1-\bar{\rho}}, \quad \tan\phi_2 = \frac{\bar{\eta}}{\bar{\eta}^2 - \bar{\rho}(1-\bar{\rho})}, \quad \tan\phi_3 = \frac{\bar{\eta}}{\bar{\rho}} \tag{2.18}$$

If the CKM mechanism describes all the quark mixing processes, then all of the measurements should agree, and they should converge on a single apex of the triangle. An indication of CP violation beyond the SM would be if the angles do not sum to 180 degrees or the sides fail to converge to their vertex. Therefore, the precise measurement of the parameters of the unitarity triangle is one of the primary goals of the Belle II experiment. Current data shows that the sum of the angles is $173 \pm 6^\circ$ [26], which consistent with SM expectations.

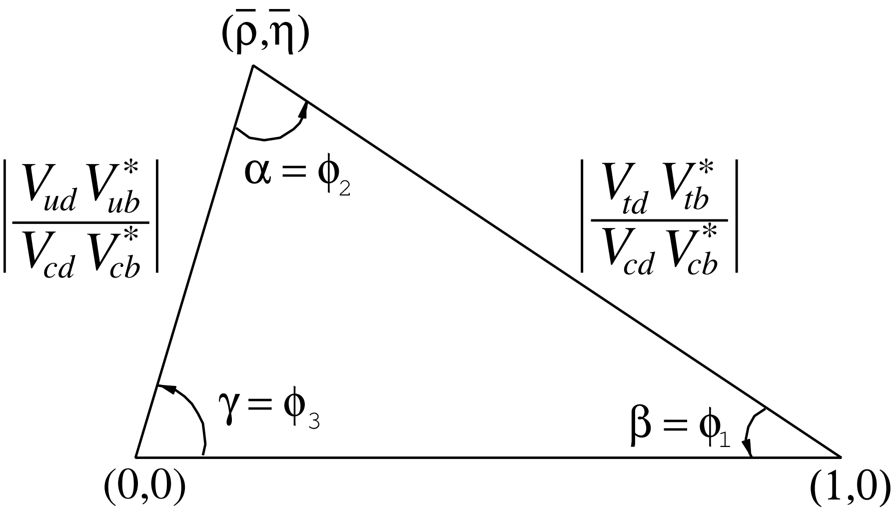


FIGURE 2.8: Rescaled unitarity triangle obtained by dividing all sides by $V_{cd}V_{cb}^*$ [6].

2.4 *CP* Symmetry

Parity transformation (P) and charge conjugation (C) are two key discrete symmetry operations that can be performed on a physical system. Parity transformation and charge conjugation are two discrete operations that can be performed on a physical system. Parity transformation (P) can be thought of as a reflection through the origin, since it reverses the spatial coordinates of a system:

$$\begin{pmatrix} x \rightarrow -x \\ y \rightarrow -y \\ z \rightarrow -z \end{pmatrix} \tag{2.19}$$

Applying this operation to the wavefunction $\psi(r)$ of a given system yields

$$\hat{P}\psi(r) = \psi(-r) \tag{2.20}$$

which implies

$$\hat{P}\psi(r) = \psi(-r) = p\psi(r) \qquad \hat{P}\hat{P}\psi(r) = p^2\psi(r) = \psi(r)$$

Hence, the possible eigenvalues are ±1. Mesons have an intrinsic parity of -1, as particles and antiparticles have opposite intrinsic parities, and parity is a multiplicative quantum number. The overall parity of a meson is a product of the intrinsic parities of the particles (-1) and the parity associated with the orbital motion (-1^L):

$$P = (-1)^{L+1} \tag{2.21}$$

where L is the angular momentum.

In contrast, charge conjugation flips internal quantum numbers such as electric charge, strangeness, charm, beauty, and so on. This operation transforms particles into their corresponding antiparticles and vice versa. Just like a parity transformation, applying C twice must leave the particle's state unchanged, meaning that the only allowed eigenvalues are ±1. However, unlike P, most particles are not eigenstates of C. This fact can be understood by considering that if $|p\rangle$ is an eigenstate of C, then

$$C\,|p\rangle = \pm\,|p\rangle = |\bar{p}\rangle\,. \tag{2.22}$$

Therefore, only particles that are their own antiparticles, such as photons and neutral mesons, can be eigenstates of C. Photons, for example, have a charge conjugation number of -1 because the electric and magnetic fields switch signs under C. The conservation of C-symmetry is observed in strong and electromagnetic interactions, which means that the electromagnetic decay of a π^0 into two photons must also respect C-symmetry. Therefore, π^0 has $C = (-1)(-1) = +1$. Hadrons can be classified using J^{PC} where J, P, and C stand for total angular momentum, P-symmetry, and C-symmetry, respectively. Therefore, π^0 is denoted by $J^{PC} = 0^{-+}$.

2.4.1 Discovery of CP Violation in the Kaon System

Until 1964, physicists believed that the combined operation CP was conserved in nature. At this time, Cronin, Fitch, Christenson *et al.* discovered evidence of CP violation in the neutral-kaon system. The neutral K meson $K^0 = (\bar{s}d)$ and $\bar{K}^0 = (s\bar{d})$ are flavour eigenstates, and can spontaneously transform to its counterpart through a second-order weak interaction known as a 'box' diagram as shown in Figure 2.9. A state which was a K^0 at $t = 0$ will evolve over time into a mixture of K^0 and $\bar{K}^0$. This mixing can occur for other particles if the following is true:

1. It is a meson since baryon number is conserved in the weak interaction
2. It is a neutral particle since charge is conserved in all interactions

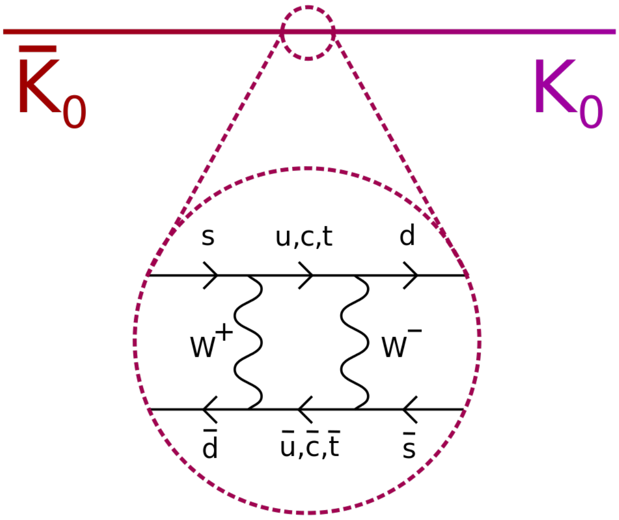


FIGURE 2.9: Kaon mixing via a box diagram

3. The antiparticles are distinct from the particles, meaning it must carry a differentiating tag such as net charm or strangeness.

For example, the K^0 and its antiparticle $\bar{K}^0$ are neutral mesons with net strangeness 1 and -1, respectively. These particles can be produced in strong interactions, such as in the following reactions:

$$\pi^+ + p \to \Sigma_c^{++} + K^0$$
$$\pi^+ + p \to K^+ + \bar{K}^0 + p$$

These neutral kaon states are antiparticles of each other with exactly the same mass eigenstate under the strong-interaction Hamiltonian. However, they do not have well-defined lifetimes because of the weak interaction, which allows for oscillation between the two states.

This means that the particle observed in the Cronin-Fitch experiment is some linear combination of K^0 and $\bar{K}^0$. Eigenstates of CP can be formed by noting that from Equation 2.21

$$\hat{P}\,|K^0\rangle = -\,|\bar{K}^0\rangle\,, \quad \hat{P}\,|\bar{K}^0\rangle = -\,|K^0\rangle \tag{2.23}$$

and, from Equation 2.22

$$\hat{C}\,|K^0\rangle = |\bar{K}^0\rangle\,, \quad \hat{C}\,|\bar{K}^0\rangle = |K^0\rangle \tag{2.24}$$

By combining these two operations, one obtains:

$$\hat{C}\hat{P}\,|K^0\rangle = -\,|\bar{K}^0\rangle\,, \quad \hat{C}\hat{P}\,|\bar{K}^0\rangle = -\,|K^0\rangle \tag{2.25}$$

This leads to the formation of two new particles, denoted as K_1 and K_2. These particles are the normalised eigenstates of CP. The mass and CP eigenstates are two different bases to describe the same set of particles, related by a unitarity rotation.

$$|K_1\rangle = \frac{1}{\sqrt{2}}(\,|K^0\rangle - \,|\bar{K}^0\rangle) \tag{2.26}$$

$$|K_2\rangle = \frac{1}{\sqrt{2}}(\,|K^0\rangle + \,|\bar{K}^0\rangle) \tag{2.27}$$

with

$$\hat{C}\hat{P}\,|K_1\rangle = +\,|K_1\rangle \tag{2.28}$$

$$\hat{C}\hat{P}\,|K_2\rangle = -\,|K_2\rangle \tag{2.29}$$

K_1 and K_2 are not antiparticles of each other and therefore do not need to have the same mass. Since K_1 is the eigenstate with $CP = +1$, if CP is conserved, it can only decay into a state with the same CP. Since the kaon is a spin-0 particle, there cannot be any orbital momentum between the two pions. Thus, from Equation 2.21, the parity of the two pions must equal $(-1)^2 = +1$. The charge conjugation number for two and three pions must be $C = +1$. Hence, the CP of the two-pion decay is +1. It is concluded that if CP is conserved, K_1 must always decay into two pions, and by similar reasoning, K_2 must decay into three pions.

$$K_1 \to 2\pi, \quad K_2 \to 3\pi \tag{2.30}$$

However, Cronin and Fitch found evidence for $K_2 \to 2\pi$ by producing K^0 of approximately 1 GeV through the strong interaction, as illustrated in Equation 2.23. By rearranging Equation 2.26, one can see that

$$|K^0\rangle = \frac{1}{\sqrt{2}}(|K_1\rangle + |K_2\rangle)$$

The K_1 component will rapidly decay away, as the $K_1 \to 2\pi$ decay is much faster due to greater energy release, with lifetimes [26]

$$\tau_1 = 0.89 \times 10^{-10}\ \text{s} \quad \text{and} \quad \tau_2 = 5.29 \times 10^{-8}\ \text{s}$$

At the end of the 18 m spectrometer, only $K_2 \to 3\pi$ decays was expected. However, they observed that K_2 decayed to 2 pions about 1 in 500 times. This was a small effect but definitive proof of CP violation. It was concluded that long-lived neutral kaon, denoted as K_L, is not a perfect eigenstate of CP, but contains a small K_1 component:

$$|K_L\rangle = \frac{1}{\sqrt{1+|\epsilon|^2}}(|K_2\rangle + \epsilon\,|K_1\rangle) \tag{2.31}$$

where ϵ measures the amount of mixing-induced CP violation. Experimentally, $\epsilon = 1.596 \pm 0.013 \times 10^{-3}$ with a mass difference $m_L - m_S = 3.483 \pm 0.006 \times 10^{-15}$ GeV/c^2 [26]. These different eigenstates are all representations of the same underlying particle, and sometimes it is more convenient to use one set over the other.

2.4.2 Mixing-induced CP violation

A more general approach can be adopted by denoting the flavour eigenstate meson as X^0. The mesons, X_1 and X_2, can then be defined as the eigenstates of the $\hat{C}\hat{P}$ operator:

$$|X_1\rangle = \frac{1}{\sqrt{2}}\left(|X^0\rangle + |\bar{X}^0\rangle\right), \quad |X_2\rangle = \frac{1}{\sqrt{2}}\left(|X^0\rangle - |\bar{X}^0\rangle\right) \tag{2.32}$$

where X_1 has eigenvalue $+1$ while X_2 has eigenvalue -1. Hence, X_1 is a CP even state, which means that it remains invariant under a CP transformation, whereas X_2 is a CP-odd state and changes sign. X_1 and X_2 are not antiparticles of each other and therefore can have different masses and can decay in different ways.

To examine how the X_1 and X_2 decays, first consider

$$\hat{H}\psi = E\psi \tag{2.33}$$

where $\hat{H} = i\partial/\partial t$ is the energy Hamiltonian operator in quantum mechanics. For a free and stable particle at rest ($E = m$), this has a solution

$$\psi = \psi_0 e^{-imt} \tag{2.34}$$

If the particle decays, it will follow an exponential decay distribution:

$$|\psi|^2 \propto e^{-\frac{t}{\tau}}$$
$$\psi \propto e^{-\frac{\Gamma}{2}t}$$

where τ is the mean lifetime, $\Gamma = 1/\tau$ is known as the width and t is the time measured in the rest frame of the particle. Applying this leads to:

$$\psi = \psi_0 e^{-imt} e^{-\frac{\Gamma}{2}t} = \psi_0 e^{-i(m-i\frac{\Gamma}{2})t} \tag{2.35}$$

The neutral X^0 and $\bar{X}^0$ system can be can written as

$$|\Psi(t)\rangle = a(t)\,|X^0\rangle + b(t)\,|\bar{X}^0\rangle \tag{2.36}$$

where X^0 and $\bar{X}^0$ are the flavour eigenstates. The evolution of this system is described by the time-dependent Schrodinger equation.

$$i\frac{d}{dt}\Psi(t) = \mathbf{H}\Psi(t) \tag{2.37}$$

where $\mathbf{H}$ is the Hamiltonian and takes the form of the solution derived in Equation 2.35

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} - \frac{i}{2}\begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \tag{2.38}$$

where $\mathbf{M}$ is the mass matrix and $\mathbf{\Gamma}$ is the decay matrix. States that evolve over time are derived by diagonalising the Hamiltonian. This process results in eigenstates and eigenvalues that correspond to freely propagating particles with well-defined masses, $|X^0\rangle$ and $|\bar{X}^0\rangle$. Note that if only the strong interaction was present, there would be no mixing and the off-diagonal terms would be zero. $\mathbf{M}$ and $\mathbf{\Gamma}$ must be Hermitian, and so $m_{21} = m_{12}^*$ and $\Gamma_{21} = \Gamma_{12}^*$. Invariance under CPT transformation requires $m_{11} = m_{22} = m$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$ as particles and antiparticles have equal masses and lifetimes in the absence of the weak interactions. This leads to

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{pmatrix} m & m_{12} \\ m_{12}^* & m \end{pmatrix} - \frac{i}{2}\begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix} \tag{2.39}$$

In the case of CP symmetry, $\mathbf{M}$ and $\mathbf{\Gamma}$ must be real, because if any component of $\mathbf{H}$ were imaginary, the probabilities of oscillation between the two states would vary over time. That is, the rate at which $X_0 \to \bar{X}_0$ might not match the reverse process, $\bar{X}_0 \to X_0$. Therefore, $m_{12} = m_{12}^* = \Delta m$ and $\Gamma_{12} = \Gamma_{12}^* = \Delta\Gamma$, where the choice of symbols will become apparent shortly. Diagonalising Equation 2.39 gives eigenvalues

$$\lambda_{L,H} = m \mp \Delta m - \frac{i(\Gamma \pm \Delta\Gamma)}{2} \tag{2.40}$$

and the corresponding eigenvectors,

$$|X_L\rangle = \frac{1}{\sqrt{2}}\left(|X^0\rangle + |\bar{X}^0\rangle\right) = |X_1\rangle\,, \quad |X_H\rangle = \frac{1}{\sqrt{2}}\left(|X^0\rangle - |\bar{X}^0\rangle\right) = |X_2\rangle \tag{2.41}$$

where X_L and X_H is the light and heavy mass eigenstates, respectively, which correspond to the CP eigenstates X_1 and X_2. The X_L has mass $(m - \Delta m)$ and lifetime $(\Gamma + \Delta\Gamma)$, while the X_H has mass $(m + \Delta m)$ and lifetime $(\Gamma - \Delta\Gamma)$. The sign convention for the mass difference, Δm, is chosen such that it is always positive, whereas $\Delta\Gamma$ can have any sign. This can be rewritten as

$$m = \frac{m_L + m_H}{2} \qquad \Delta m = \frac{m_H - m_L}{2} \tag{2.42}$$

$$\Gamma = \frac{\Gamma_L + \Gamma_H}{2} \qquad \Delta\Gamma = \frac{\Gamma_H - \Gamma_L}{2} \tag{2.43}$$

where m and Γ are the average mass and lifetime of the two mass eigenstates, respectively, while Δm and $\Delta\Gamma$ denote the differences in mass and width, respectively. The weak force splits the mass and lifetimes, as qualitatively shown in Figure 2.10.

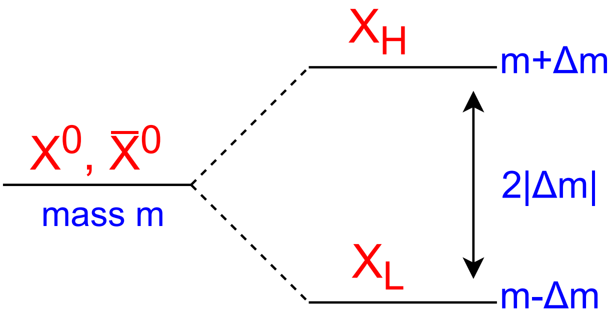


FIGURE 2.10: The weak force splits the mass and lifetimes in an analogous way to the splitting of levels in atomic physics due to a magnetic field.

In the absence of CP symmetry, diagonalising Equation 2.39 gives eigenvalues

$$\lambda_{L,H} = m - \frac{i}{2}\Gamma \pm \sqrt{(m_{12} - \frac{i}{2}\Gamma_{12})(m_{12}^* - \frac{i}{2}\Gamma_{12}^*)} \tag{2.44}$$

and the corresponding eigenstates

$$|X_L\rangle = p\,|X^0\rangle + q\,|\bar{X}^0\rangle\,, \quad |X_H\rangle = p\,|X^0\rangle - q\,|\bar{X}^0\rangle \tag{2.45}$$

where

$$p^2 = m_{12} - \frac{i}{2}\Gamma_{12}, \quad q^2 = m_{12}^* - \frac{i}{2}\Gamma_{12}^* \tag{2.46}$$

The masses, m_L and m_H, can be found from $\text{Re}(\lambda_{\pm})$ while the lifetimes, Γ_L and Γ_H, are found from -2lm($\lambda_{\pm}$) as

$$m_{L,H} = m \pm \text{Re}(pq), \quad \Gamma_{L,H} = m \mp 2\text{Im}(pq). \tag{2.47}$$

The splitting is given by $\Delta m = m_H - m_L = 2\text{Re}(pq)$ and $|\Delta\Gamma| = \Gamma_L - \Gamma_H = -4\text{Im}(pq)$. In the presence of CP-violation, the CP-even and CP-odd states are no longer pure. There is the presence of a small admixture of the other CP-eigenstate:

$$|X_L\rangle = \frac{1}{\sqrt{1+\epsilon^2}}\left(|X_1\rangle + \epsilon\,|X_2\rangle\right), \quad |X_H\rangle = \frac{1}{\sqrt{1+\epsilon^2}}\left(|X_2\rangle + \epsilon\,|X_1\rangle\right) \tag{2.48}$$

$$p = \frac{1}{\sqrt{2}}\frac{1+\epsilon}{\sqrt{1+\epsilon^2}}, \quad q = \frac{1}{\sqrt{2}}\frac{1-\epsilon}{\sqrt{1+\epsilon^2}} \tag{2.49}$$

$$\frac{q}{p} = \frac{1-\epsilon}{1+\epsilon}, \quad \epsilon = \frac{p-q}{p+q} \tag{2.50}$$

Hence, CP is violated in mixing if $\epsilon \neq 0$, which occurs if $|\frac{q}{p}| \neq 1$. Physically this is because the rate of $X^0 \to \bar{X}^0$ is different from $\bar{X}^0 \to X^0$. The ratio $\frac{q}{p}$ characterises the degree of CP violation in the $X^0\bar{X}^0$ system. More specifically, the phase of $\frac{q}{p}$ is a measure of CP violation in the mixing, and the magnitude of $\frac{q}{p}$ deviating from unity indicates direct CP violation. The relation stated in Equation 2.31 is a special case where X is the kaon system. However, the primary focus of this thesis is the B meson system, and so $X = B$. It is also interesting to note that, unlike the $K^0\bar{K}^0$ system, the lifetimes of the two B meson mass eigenstates are extremely close and difficult to measure experimentally. The quantity that is usually measured is

$$\frac{\Delta\Gamma_B}{\Gamma_B} = (0.1 \pm 1.0) \times 10^{-2}$$

where $\Delta\Gamma_B$ and Γ_B are the decay rate difference and average between the B_H and B_L states, respectively. It is evident from the relative uncertainty that the exact lifetime is difficult to resolve. On the other hand, the mass difference is $\Delta m = m_H - m_L = 3.334 \pm 0.013 \times 10^{-15}$ GeV/c^2 [26] and is much larger than the lifetime difference. Therefore, it is more appropriate to distinguish the neutral B mesons by their masses rather than by their lifetimes. This mass difference is still a tiny amount compared to the rest mass of the neutral B meson, which is approximately 5.28 GeV/c^2.

From the eigenvalues found in Equation 2.44, the time-dependent Schrodinger equation can be rewritten as

$$i\frac{d}{dt}\left|B_{L,H}(t)\right\rangle = \lambda_{L,H}\left|B_{L,H}(0)\right\rangle \tag{2.51}$$

where t is the decay time in the B meson rest frame. Hence the time dependence of the mass eigenstate is:

$$\begin{aligned} |B_{L,H}(t)\rangle &= e^{-i\lambda_{L,H}t}\,|B_{L,H}(0)\rangle \\ &= e^{-i\lambda_{L,H}t}(p\,|B^0(0)\rangle \pm q\,|\bar{B}^0(0)\rangle) \end{aligned} \tag{2.52}$$

It also follows from Equation 2.45 that

$$|B_{L,H}(t)\rangle = p\,|B^0(t)\rangle \pm q\,|\bar{B^0}(t)\rangle \tag{2.53}$$

Equating Equation 2.53 and 2.52 and solving for $B^0(t)$ and $\bar{B}^0(t)$, leads to the time evolution of the flavour eigenstate as

$$|B^0(t)\rangle = \frac{1}{2}\left[(e^{-i\lambda_L t} + e^{-i\lambda_H t})\,|B^0(0)\rangle + \frac{q}{p}(e^{-i\lambda_L t} + e^{-i\lambda_H t})\,|\bar{B}^0(0)\rangle\right] \tag{2.54}$$

$$|\bar{B}^0(t)\rangle = \frac{1}{2}\left[(e^{-i\lambda_L t} + e^{-i\lambda_H t})\,|\bar{B}^0(0)\rangle + \frac{p}{q}(e^{-i\lambda_L t} + e^{-i\lambda_H t})\,|B^0(0)\rangle\right] \tag{2.55}$$

This expression can be simplified by using the definitions in Equation 2.42 and 2.43 which leads to

$$|B^0(t)\rangle = g_+(t)\,|B^0(0)\rangle + \frac{q}{p}g_-(t)\,|\bar{B}^0(0)\rangle \tag{2.56}$$

$$|\bar{B}^0(t)\rangle = g_+(t)\,|\bar{B}^0(0)\rangle + \frac{p}{q}g_-(t)\,|B^0(0)\rangle \tag{2.57}$$

where

$$\begin{aligned} g_+(t) &= e^{-imt}e^{-\frac{\Gamma t}{2}}\left(+\cosh\frac{\Delta\Gamma t}{4}\cos\frac{\Delta m t}{2} - \sinh\frac{\Delta\Gamma t}{4}\sin\frac{\Delta m t}{2}\right) \\ g_-(t) &= e^{-imt}e^{-\frac{\Gamma t}{2}}\left(-\sinh\frac{\Delta\Gamma t}{4}\cos\frac{\Delta m t}{2} + i\cosh\frac{\Delta\Gamma t}{4}\sin\frac{\Delta m t}{2}\right) \end{aligned} \tag{2.58}$$

The flavour states either remain unchanged $(+)$ or oscillate into each other $(-)$ with time-dependent probabilities proportional to the magnitude of the coefficient of the $|B^0(0)\rangle$ and $|\bar{B}^0(0)\rangle$ squared.

$$|g_{\pm}(t)|^2 = \pm\frac{e^{-\Gamma t}}{2}\left(\cosh\frac{\Delta\Gamma t}{2} \pm \cos\Delta m t\right) \tag{2.59}$$

The time-integrated mixing probability χ_d measures the probability for an initial B^0 meson to be a $\bar{B}^0$ meson when it decays. In the absence of CP violation, it is given by

$$\chi_d = \frac{\int |g_-(t)|^2 dt}{\int |g_-(t)|^2 dt + \int |g_+(t)|^2 dt} \tag{2.60}$$

$$= \frac{x^2+y^2}{2(1+x^2)} \tag{2.61}$$

where $x = \frac{\Delta m}{\Gamma}$ and $y = \frac{\Delta\Gamma}{2\Gamma}$. Assuming that $\Delta\Gamma = 0$, no CP violation in mixing, and a B^0 lifetime of 1.591 ± 0.004 ps [32], the current world average is [26],

$$\chi_d = 0.1858 \pm 0.0011$$

2.4.3 CP violation via interference of mixing and decay

If a final state, f, is accessible to both B^0 and $\bar{B}^0$ decay then there will be interference between the direct decay ($B^0 \to f$) and indirect decay through mixing ($B^0 \to \bar{B}^0 \to f$), as illustrated in Figure 2.11.

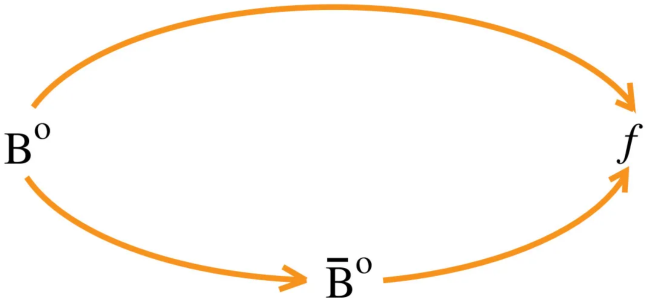


Figure 2.11: Interference via mixing and decay [7]

This interference between the direct and indirect decays leads to observable differences in the rates of B^0 and $\bar{B}^0$ decays into the final state f. Over time, the probability of $B^0 \to \bar{B}^0$ transformation can be observed to fluctuate, causing the decay amplitudes to become time-dependent.

Let the final state to be a CP eigenstate, denoted as f_{CP} and define the decay amplitudes for B^0 and $\bar{B}^0$ as:

$$A = \langle f_{CP}| \mathcal{H} |B^0\rangle , \quad \bar{A} = \langle f_{CP}| \mathcal{H} |\bar{B}^0\rangle \tag{2.62}$$

where $\mathcal{H}$ represents the weak Hamiltonian, given that the decay occurs through the weak interaction. To simplify the calculation, assume that there is no difference in the decay rates of the B_L and B_H states ($\Delta\Gamma = 0$), and that there is no CP violation in the mixing ($|\frac{q}{p}| = 1$). These are reasonable approximations, since the current world average measurement indicates that $\Delta\Gamma/\Gamma = 0.001 \pm 0.010$ and $|\frac{p}{q}| = 1.001 \pm 0.008$ [26]. Both of these determinations are compatible with zero. Substituting Equation 2.56 and 2.57 into Equation 2.62 leads to the time-dependent amplitudes

$$A(t) = \langle f_{CP}| \mathcal{H} |B^0(t)\rangle = Ae^{-\frac{\Gamma t}{2}} \left(\cos\frac{\Delta mt}{2} + i\lambda \sin\frac{\Delta mt}{2} \right) \tag{2.63}$$

$$\bar{A}(t) = \langle f_{CP}| \mathcal{H} |\bar{B}^0(t)\rangle = \bar{A}e^{-\frac{\Gamma t}{2}} \left(\cos\frac{\Delta mt}{2} + \frac{i}{\lambda} \sin\frac{\Delta mt}{2} \right) \tag{2.64}$$

where

$$\lambda = \frac{q}{p}\frac{\bar{A}}{A} \tag{2.65}$$

The parameter λ characterises the interference between decay via mixing and direct decay, represented by $\frac{q}{p}$ and $\frac{\bar{A}}{A}$, respectively. The e^{-imt} term has been eliminated since, in quantum mechanics, one can always multiply a $|B^0\rangle$ by an arbitrary phase factor without changing any physical quantities. The time-dependent decay rate is

$$\begin{aligned}\Gamma(B^0 \to f_{CP}) &= |\langle f_{CP}| \mathcal{H} |B^0(t)\rangle|^2 \\ &= |A|^2 e^{-\Gamma t} \left[(1 + |\lambda|^2) + \frac{1-|\lambda|^2}{2} \cos\Delta mt - 2\mathrm{Im}(\lambda) \sin\Delta mt \right]\end{aligned} \tag{2.66}$$

$$\begin{aligned}\Gamma(\bar{B}^0 \to f_{CP}) &= |\langle f_{CP}| \mathcal{H} |\bar{B}^0(t)\rangle|^2 \\ &= |\bar{A}|^2 e^{-\Gamma t} \left[(1 + |\lambda|^2) - \frac{1-|\lambda|^2}{2} \cos\Delta mt + 2\mathrm{Im}(\lambda) \sin\Delta mt \right]\end{aligned} \tag{2.67}$$

The term $\frac{q}{p}$ originates from the $B^0 - \bar{B}^0$ mixing via the box diagrams, as shown in Figure 2.12. In this diagram, the b and d quarks couple to the W boson and the internal u, c, t quarks. Given that $V_{tb} \gg V_{cb} \gg V_{ub}$, the top quark is dominant, the contribution of the other quarks can be neglected to a good approximation. Therefore, with negligible

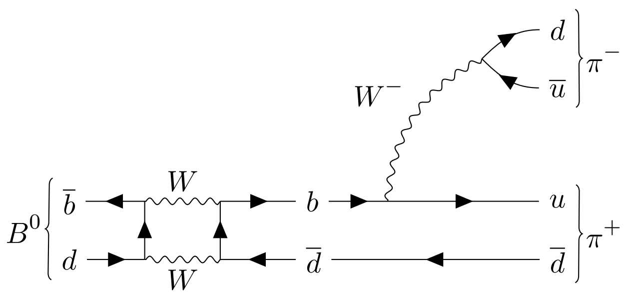


FIGURE 2.12: Feynman diagram for CP violation via interference of mixing and decay in $B^0 \to \pi^+\pi^-$ decays.

corrections, the mixing phase is

$$\frac{q}{p} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} = \frac{(V_{tb}^* V_{td})^2}{|V_{tb} V_{td}|^2} = e^{-2i\phi_1} \tag{2.68}$$

$$\mathrm{Im}(\frac{q}{p}) = \sin 2\phi_1 \tag{2.69}$$

where ϕ_1 is one of the angles of the unitarity triangle. To quantify the CP violation in the interference between decays via mixing ($B^0 \to \bar{B}^0 \to f_{CP}$) and direct decays ($B^0 \to f$), the time-dependent asymmetry is defined as

$$\begin{aligned} a_{f_{CP}} &= \frac{\Gamma(B^0 \to f_{CP}) - \Gamma(\bar{B}^0 \to f_{CP})}{\Gamma(B^0 \to f_{CP}) + \Gamma(\bar{B}^0 \to f_{CP})} && (2.70)\\ &= \frac{|\lambda|^2 - 1}{1 + |\lambda|^2} \cos \Delta mt + \frac{2\mathrm{Im}(\lambda)}{1 + |\lambda|^2} \sin \Delta mt && (2.71)\\ &= \mathcal{A}_{CP} \cos \Delta mt + \mathcal{S}_{CP} \sin \Delta mt && (2.72) \end{aligned}$$

where the coefficient of the $\sin \Delta mt$ and $\cos \Delta mt$ terms correspond to the mixing-induced CP violation parameter $\mathcal{S}_{CP}$ and direct CP violation parameter $\mathcal{A}_{CP}$, respectively. Hence,

$$\mathcal{A}_{CP} = \frac{|\lambda|^2 - 1}{1 + |\lambda|^2} = \frac{|\bar{A}|^2 - |A|^2}{|\bar{A}|^2 + |A|^2}, \quad \mathcal{S}_{CP} = \frac{2\mathrm{Im}(\lambda)}{1 + |\lambda|^2} \tag{2.73}$$

where $|\lambda|^2 = \frac{|\bar{A}|^2}{|A|^2}$ assuming CP violation in mixing ($|\frac{q}{p}| = 1$). Note that $\mathcal{S}_{CP}$ changes sign depending on the CP of the final state, while $\mathcal{A}_{CP}$ does not. Experimentally, by measuring the time dependence of $a_{f_{CP}}$ one can extract the coefficient of the $\sin \Delta mt$ and $\cos \Delta mt$ terms, and thereby determine $|\lambda|$ and $\mathrm{Im}(\lambda)$. It can also be seen that CP violation can occur as long as $\lambda = \frac{q}{p}\frac{\bar{A}}{A} \neq 1$. This holds true even if $|\lambda| = 1$, as λ can acquire a non-zero phase, resulting in a non-vanishing $\mathrm{Im}(\lambda)$. Hence, $\mathrm{Im}(\lambda)$ signifies the amount of mixing-induced CP violations. Moreover, if $|\lambda| = 1$, then there is only one decay amplitude, and $|A| = |\bar{A}|$, leading to no direct CP violation.

2.4.4 Direct CP violation

Direct CP violation occurs when the magnitude of a decay differs from its CP-conjugate. This condition can be expressed as:

$$|\frac{\bar{A}}{A}| \neq 1 \tag{2.74}$$

In other words, the decay rate of $B \to f$ is different to $\bar{B} \to \bar{f}$ where f is some final state. Furthermore, direct CP violation is the only source of CP violation in the decay of charged mesons.

Without CP violation, any process and its corresponding antiparticle process can be represented by the same complex number, separable into magnitude and phase. When CP violation is included, one can introduce a weak phase term originating from the complex CKM matrix elements. This results in the conjugate process acquiring a negative phase term.

$$A(B^0 \to f) = |A|e^{i\phi}e^{i\theta}, \quad \bar{A}(\bar{B}^0 \to \bar{f}) = |A|e^{i\phi}e^{-i\theta} \tag{2.75}$$

where $|A|$ is the decay amplitude, ϕ is the strong phase, and θ is the weak phase. The sign of the strong phase remains unchanged since CP is conserved in strong interactions. Meanwhile, the weak phase has opposite signs in $A(B^0 \to f)$ and $\bar{A}(\bar{B}^0 \to \bar{f})$.

If there are two different routes, $B \to 1 \to f$ and $B \to 2 \to f$, the total amplitude is given by

$$A(B \to f) = |A_1|e^{i\phi_1}e^{i\theta_1} + |A_2|e^{i\phi_2}e^{i\theta_2} \tag{2.76}$$

$$\bar{A}(\bar{B} \to \bar{f}) = |A_1|e^{i\phi_1}e^{-i\theta_1} + |A_2|e^{i\phi_2}e^{-i\theta_2} \tag{2.77}$$

Consequently, the amplitude squared becomes

$$|A(B \to f)|^2 = |A_1|^2 + |A_2|^2 + 2|A_1||A_2| \cos{(\Delta\phi + \Delta\theta)} \tag{2.78}$$

$$|\bar{A}(\bar{B} \to \bar{f})|^2 = |A_1|^2 + |A_2|^2 + 2|A_1||A_2| \cos{(\Delta\phi - \Delta\theta)} \tag{2.79}$$

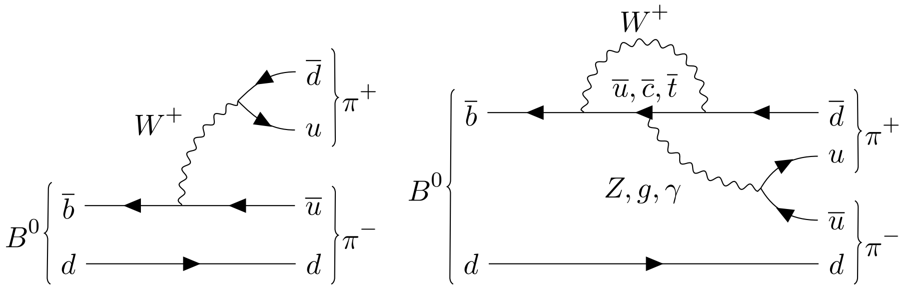


FIGURE 2.13: Leading order Feynman diagrams for Dominant first-order tree amplitude (left) and second-order loop penguin (right) processes for $B^0 \to \pi^+\pi^-$ decay.

From Equation 2.73, the direct violation asymmetry is

$$\begin{aligned}\mathcal{A}_{CP} &= \frac{|\bar{A}(\bar{B} \to \bar{f})|^2 - |A(B \to f)|^2}{|\bar{A}(\bar{B} \to \bar{f})|^2 + |A(B \to f)|^2} \\ &= \frac{2|A_1||A_2| \sin\Delta\phi \sin\Delta\theta}{|A_1|^2 + |A_2|^2 + 2|A_1||A_2| \cos\Delta\phi \cos\Delta\theta}\end{aligned} \tag{2.80}$$

where $\Delta\phi = \phi_1 - \phi_2$ and $\Delta\theta = \theta_1 - \theta_2$. Therefore, for direct violation of CP to occur, there must be at least two contributing decay amplitudes with different weak phases and strong phases. These could come from two tree diagrams, two penguin diagrams, or one tree and one penguin. Furthermore, to obtain large direct CP violation, A_1 and A_2 should have similar magnitudes while $\Delta\phi$ and $\Delta\theta$ should be close to $\frac{\pi}{2}$. Nevertheless, extraction of the weak phase from data is often complicated by the large hadronic uncertainties in the amplitudes and strong phases, which are typically induced by mixing and penguin contributions.

2.5 Analysis of $B \to \pi\pi$

The decay of B mesons into two pions is described by the weak transition $b \to \bar{u}u\bar{d}$ followed by a hadronisation process involving a spectator quark. Hadronisation refers to the formation of hadrons from quarks and gluons. Quantitative predictions are difficult and have large theoretical uncertainties due to the large coupling constant of the strong interaction between quarks and gluons. The weak process receives dominant contributions from both the tree-level charged transition and the flavour-changing neutral current penguin transition, as shown in Figure 2.13.

Despite the complex interactions, it remains possible to describe the charmless $B^0 \to (\bar{u}u\bar{d}, q) \to \pi\pi$ transition amplitudes by using the CKM matrix elements associated with the quark transitions involved in tree and penguin diagrams. The penguin diagrams has three different up-type quark flavors (u, c, t) within the W loop. Both tree and penguin-level topologies mediate transitions involving the u quark, while transitions that include c and t quarks are solely mediated by penguin topologies.

$$\begin{aligned}
\mathcal{A}^{ij} &= \left\langle h_i^i h_2^j \right| \mathcal{H}_{\text{eff}} \left| B_d \right\rangle &(2.81)\\
&= V_{ud}V_{ub}^*(\mathcal{T}_u^{ij} + \mathcal{P}_u^{ij}) + V_{cd}V_{cb}^*(\mathcal{P}_c^{ij}) + V_{td}V_{tb}^*(\mathcal{P}_t^{ij}) &(2.82)
\end{aligned}$$

where $\mathcal{H}_{\text{eff}}$ is the effective Hamiltonian describing the transition. The amplitudes $\mathcal{T}_u^{ij}$ and $\mathcal{P}_{u,c,t}^{ij}$ represents the tree-level and (u, c, t)-loop mediated topologies, respectively. These CKM matrix terms are not independent, the unitarity of the CKM matrix can be used to eliminate one of the three terms. Using the constraint $V_{td}V_{tb}^* + V_{ud}V_{ub}^* + V_{cd}V_{cb}^* = 0$,

$$\begin{aligned}
\mathcal{A}^{ij} &= V_{ud}V_{ub}^*(\mathcal{T}_u^{ij} + \mathcal{P}_u^{ij}) - (V_{ud}V_{ub}^* + V_{ud}V_{ub}^*)(\mathcal{P}_c^{ij}) + V_{td}V_{tb}^*(\mathcal{P}_t^{ij}) &(2.83)\\
&= V_{ud}V_{ub}^*(\mathcal{T}_u^{ij} + \mathcal{P}_u^{ij} - \mathcal{P}_c^{ij}) + V_{td}V_{tb}^*(\mathcal{P}_t^{ij} - \mathcal{P}_c^{ij}) &(2.84)
\end{aligned}$$

The amplitudes can be rewritten as

$$\mathcal{A}^{ij} = V_{ud}V_{ub}^*\tilde{\mathcal{T}}^{ij} + V_{td}V_{tb}^*\tilde{\mathcal{P}}^{ij} \tag{2.85}$$

where $\tilde{\mathcal{T}}^{ij} = \mathcal{T}_u^{ij} + \mathcal{P}_u^{ij} - \mathcal{P}_c^{ij}$ and $\tilde{\mathcal{P}}^{ij} = \mathcal{P}_t^{ij} - \mathcal{P}_c^{ij}$ will be referred to as 'tree' and 'penguin' amplitudes. This choice is arbitrary and has no physical implication on the determination of the weak phase affecting the transition. Pulling out the weak phases $\phi_3 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right)$ and $\phi_1 = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right)$ from the first and second term, respectively.

$$\mathcal{A}^{ij} = -e^{i\phi_3}T^{ij} + e^{-i\phi_1}P^{ij}, \tag{2.86}$$

where the magnitude of the CKM products $R_u = |V_{ud}V_{ub}^*|$ and $R_t = |V_{td}V_{tb}^*|$ is included in the redefined amplitudes $T^{ij} = R_u\tilde{\mathcal{T}}^{ij}$ and $P^{ij} = -R_t\tilde{\mathcal{P}}^{ij}$.

Similarly, the decay amplitudes of the CP-conjugate isodoublet $(\overline{B}^0, B^-)$ can be expressed as

$$\bar{\mathcal{A}}^{ij} = \frac{p}{q}\tilde{\mathcal{A}}^{ij} = -e^{-i\phi_3}T^{ij} + e^{i\phi_1}P^{ij}, \tag{2.87}$$

where $\tilde{\mathcal{A}}^{ij}$ is the complex conjugation of $\mathcal{A}^{ij}$, and $\bar{\mathcal{A}}^{ij}$ is obtained by consistently rotating the $\tilde{\mathcal{A}}^{ij}$ to absorb the mixing phase, i.e. $\bar{\mathcal{A}}^{ij} = e^{-2i\phi_1}\tilde{\mathcal{A}}^{ij}$. Here, the factor $p/q = e^{2i\phi_1}$ is included to take into account the $B^0 - \overline{B}^0$ mixing phase of neutral B mesons. For the purposes of consistency, this phase is also applied when defining the amplitudes of the charged B meson. The CP invariance of the strong interaction ensures that the hadronic amplitudes T^{ij} and P^{ij} remain the same between a process and its CP-conjugate. However, the weak phases undergo complex conjugation. A phase redefinition is performed by rotating the amplitudes by the weak phase ϕ_1 so that $A^{ij} = e^{i\phi_1}\mathcal{A}^{ij}$ and $\bar{A}^{ij} = e^{i\phi_1}\bar{\mathcal{A}}^{ij}$.

$$\begin{aligned} A^{ij} &= -e^{i(\phi_3+\phi_1)}T^{ij} + P^{ij} && (2.88)\\ &= -e^{i(\pi-\phi_2)}T^{ij} + P^{ij} && (2.89)\\ &= -(e^{i\pi}e^{-i\phi_2})T^{ij} + P^{ij} && (2.90)\\ &= e^{-i\phi_2}T^{ij} + P^{ij} && (2.91) \end{aligned}$$

Similarly, it is found that $\bar{A}^{ij} = e^{i\phi_2}T^{ij} + P^{ij}$ where $2\phi_2$ represents the phase difference between the tree contributions of the CP-conjugate amplitudes:

$$e^{2i\phi_2} = \frac{\bar{A}^{ij} - P^{ij}}{A^{ij} - P^{ij}} \tag{2.92}$$

In the absence of a penguin contribution ($P^{ij} = 0$), ϕ_2 can be related to the relative phase of CP-conjugate amplitudes describing the B^0 and $\overline{B}^0$ mesons decaying into the same final state. In particular, if only the tree amplitude contributes to $B^0/\bar{B}^0 \to \pi^+\pi^-$, there would be no direct CP violation. This implies that $|\lambda_{+-}| = |A^{+-}|/|\bar{A}^{+-}| = 1$, and thus, $\mathcal{A}_{CP} = 0$. Consequently, the time-dependent decay rate asymmetry ($a_{\mathrm{CP}} = \mathcal{S}^{+-}_{CP}$)), as described by Equation 2.72, would yield $\sin(2\phi_2)$:

$$\mathcal{S}^{+-}_{CP} = \frac{2\mathrm{Im}(\lambda_{+-})}{1 + |\lambda_{+-}|^2} = \frac{2\mathrm{Im}(\frac{\bar{A}^{+-}}{A^{+-}})}{1 + |\lambda_{+-}|^2} = \frac{2\mathrm{Im}(e^{2i\phi_2})}{1 + |e^{2i\phi_2}|^2} = \sin(2\phi_2) \tag{2.93}$$

Hence, in the limit where only the tree ($b \to u$) amplitude contributes, one finds from Equation 2.73 that $\mathcal{S}_{CP} = \sin 2\phi_2$, and that Equation 2.95 simplifies to $\lambda_{+-} = e^{2i\phi_2}$. However, in reality there are significant penguin ($b \to d$) contributions. This occurs because, even though the penguin process is loop-suppressed, the tree-level process is also suppressed due to the small magnitude of $|V_{ub}|$ in the CKM matrix. Consequently, both processes may have similar amplitudes and the interference can be significant. Unless explicitly stated otherwise, electroweak penguins are assumed to be negligible. The penguin amplitudes introduce a different CKM phase, causing the value of ϕ_2 to shift by an amount $\Delta\phi_2$. Moreover, with two amplitudes contributing, $|\lambda| \neq 1$ and therefore, $\mathcal{A}_{CP}$ is not zero. Thus, the observable mixing-induced CP parameter, as defined by Equation 2.73, becomes

$$\begin{aligned}\mathcal{S}_{CP}^{+-} &= \frac{2\mathrm{Im}(\lambda_{+-})}{1+|\lambda_{+-}|^2} \\ &= \frac{2|\lambda_{+-}|}{1+|\lambda_{+-}|^2}\sin\left(2[\phi_2+\Delta\phi_2]\right) \\ &= \sqrt{1-\mathcal{A}_{CP}^2}\sin(2\phi_2+2\Delta\phi_2)\end{aligned} \tag{2.94}$$

and λ_{+-} is given by

$$\lambda_{+-} = \frac{q}{p}\frac{\bar{A}^{+-}}{A^{+-}} = e^{2i\phi_2}\frac{-e^{i\phi_2}T^{+-}+P^{+-}}{-e^{-i\phi_2}T^{+-}+P^{+-}} = e^{2i\phi_2}\frac{1-\frac{P^{+-}}{T^{+-}}e^{-i\phi_2}}{1-\frac{P^{+-}}{T^{+-}}e^{+i\phi_2}} = |\lambda|e^{2i\phi_{2,\mathrm{eff}}} \tag{2.95}$$

Experimentally, one can only measure the effective (or penguin polluted) angle, $2\phi_{2,\mathrm{eff}} = 2\phi_2 + 2\Delta\phi_2$, and hence $\Delta\phi$ must be determined to measure ϕ. Fortunately, as will be demonstrated shortly, penguin and tree contributions can be disentangled using the $B \to \pi\pi$ isospin relations. This allows ϕ_2 to be extracted with only small theoretical uncertainty.

2.6 $B \to \pi\pi$ Isospin Analysis

2.6.1 Isospin

In 1932, Heisenberg proposed that if electric charge could be 'switched off', the proton and neutron would be indistinguishable due to the symmetry of the strong force experienced by both particles. Heisenberg further proposed that the neutron and proton could be considered as the two states of a single particle, termed the nucleon. A rotation through 180° converts protons into neutrons and vice versa. This is directly analogous to the spin-up and spin-down state of spin-1/2 particle, and as such it is referred to as 'isospin' (I). Isospin is a vector in abstract 'isospin space', with the components (I_1, I_2, I_3). The proton and neutron form an isospin doublet with total isospin $I = 1/2$ and third component $I_3 = \pm 1/2$:

$$p = \left|\frac{1}{2}, +\frac{1}{2}\right\rangle, \quad n = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \tag{2.96}$$

The families of similar particles are known as isospin multiplets. A two-particle families are called doublets, three-particle families are called triplets, and so on. Physics involving the strong interactions, such as hadronisation, is invariant under rotations in 'isospin space'. The isospin numbers for the proton and neutron arise from the isospin assignment for the u, d quark, and $\bar{u}$, $\bar{d}$ anti-quark, each of which form an isospin doublet. All other flavors carry isospin zero.

$$u = \left|\frac{1}{2}, +\frac{1}{2}\right\rangle, \quad d = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \tag{2.97}$$

$$\bar{u} = \overline{\left|\frac{1}{2}, -\frac{1}{2}\right\rangle}, \quad \bar{d} = -\overline{\left|\frac{1}{2}, +\frac{1}{2}\right\rangle} \tag{2.98}$$

where the bar over the ket indicates isospin representation of an anti-quark. The ordering and minus sign in the anti-quark doublet ensures the anti-quark transforms the same way as quark ($u \leftrightarrow d$, $\bar{u} \leftrightarrow \bar{d}$).

As with protons and neutrons, the strong interaction does not differentiate among quark flavors. Therefore, no changes occur if all up quarks are replaced by down quarks and vice versa. By combining two isospin-1/2 states representing a quark and an anti-quark, and following the usual rules for adding spins, one obtains a triplet of isospin-1 states and a singlet isospin-0 state:

Triplet

$$\begin{aligned}
|1,1\rangle &= \left|\frac{1}{2}, +\frac{1}{2}\right\rangle \overline{\left|\frac{1}{2}, \frac{1}{2}\right\rangle} && = -u\bar{d} \\
|1,0\rangle &= \frac{1}{\sqrt{2}}\left[\left|\frac{1}{2}, +\frac{1}{2}\right\rangle \overline{\left|\frac{1}{2}, -\frac{1}{2}\right\rangle} + \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \overline{\left|\frac{1}{2}, +\frac{1}{2}\right\rangle}\right] && = \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}) \qquad (2.99) \\
|1,-1\rangle &= \left|\frac{1}{2}, +\frac{1}{2}\right\rangle \overline{\left|\frac{1}{2}, -\frac{1}{2}\right\rangle} && = d\bar{u}
\end{aligned}$$

Singlet

$$|0,0\rangle = \frac{1}{\sqrt{2}}\left[\left|\frac{1}{2}, +\frac{1}{2}\right\rangle \overline{\left|\frac{1}{2}, -\frac{1}{2}\right\rangle} - \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \overline{\left|\frac{1}{2}, +\frac{1}{2}\right\rangle}\right] = \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}) \tag{2.100}$$

The π mesons are the lightest mesons that form a SU(2) isospin triplet ($I = 1$):

$$\pi^{+}(u\bar{d}), \quad \pi^{0}(\frac{u\bar{u} - d\bar{d}}{\sqrt{2}}), \quad \pi^{-}(\bar{u}d) \quad \text{with } I_3 = -1, 0, +1 \tag{2.101}$$

Interestingly, the π^0 is a linear combination of $u\bar{u}$ and $d\bar{d}$, implying that if somehow a π^0 could be split, half the time one would find only up quarks and the other half of the time one would find only down quarks. The charged pions $\pi^{\pm}$ have a mass of $m_{\pi^{\pm}} = 139.6\,\mathrm{MeV}/c^2$ and possess quantum numbers $J^P = 0^-$. The neutral pions π^0

have a mass $m_{\pi^0} = 135\,\mathrm{MeV}/c^2$ and possess quantum numbers $J^{PC} = 0^{-+}$. Similarly, B^0 and B^+, form an SU(2) isospin doublet ($I = 1/2$)

$$B^0(d\bar{b}), B^+(u\bar{b}) \text{ with } I_3 = -\frac{1}{2}, +\frac{1}{2} \tag{2.102}$$

$$\overline{B}^0(\bar{u}b), B^-(\bar{d}b) \text{ with } I_3 = -\frac{1}{2}, +\frac{1}{2} \tag{2.103}$$

2.6.2 The Eightfold way

When the strange quark ($I = 0$) is included, the three light quarks (u, d, s) form a three-dimensional representation of SU(3). This breaks down into an isodoublet (u, d) and an isosinglet (s) under SU(2). From the direct sum decomposition of the approximate SU(3) flavour symmetry one finds that nine mesons can be created ($3 \otimes \bar{3} = 8 \oplus 1$) and arranged into geometrical patterns, based on their charge and strangeness. Depending on the spin, total orbital angular momentum, and total angular momentum, many meson nonets are possible. For instance, pions ($S = 0, L = 0, J = 0$) are denoted $1S_0$, while ρ particles ($S = 1, L = 0, J = 1$) are denoted $1S_3$. Although it has the same quark content, $1S_3$ nonet has a significantly larger mass than the $1S_0$ nonet. This difference in their masses arises from spin-spin interactions, which can be understood as the QCD equivalent of hyperfine splitting in the ground state of the hydrogen atom.

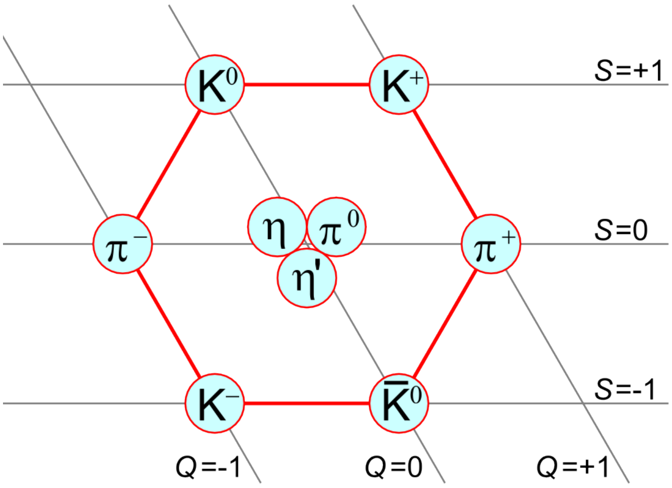


FIGURE 2.14: The psuedoscalar meson nonet for $1S_0$ states. Diametrically opposite particles are anti-particles of one-another while particles in the center are their own anti-particle.

As shown in Figure 2.14, every meson nonet can be organised by arranging the nine lightest baryons into a hexagonal pattern, with three particles positioned at the center. The charge of each particle is determined by the diagonal lines, whereas the strangeness is determined by the horizontal lines. For the $1S_0$ nonet, the flavorless η particles have zero total isospin and are defined as:

$$\begin{aligned}\eta_1 &= \frac{1}{\sqrt{3}}\left(\mathrm{u\bar{u}} + \mathrm{d\bar{d}} + \mathrm{s\bar{s}}\right)\\ \eta_8 &= \frac{1}{\sqrt{6}}\left(\mathrm{u\bar{u}} + \mathrm{d\bar{d}} - 2\mathrm{s\bar{s}}\right)\end{aligned} \tag{2.104}$$

The η_1 belongs to a singlet, and η_8 is part of an octet. The π^0, η_1, and η_8 are three mutually orthogonal linear combinations of the quark pairs $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$. Analogous to the previously discussed Cabibbo matrix, the electroweak interaction induces a mixing of the η_1 and η_8 eigenstates, so that the physical states are η and η',

$$\begin{pmatrix}\eta\\ \eta'\end{pmatrix} = \begin{pmatrix}\cos\theta_\mathrm{P} & -\sin\theta_\mathrm{P}\\ \sin\theta_\mathrm{P} & \cos\theta_\mathrm{P}\end{pmatrix}\begin{pmatrix}\eta_8\\ \eta_1\end{pmatrix} \tag{2.105}$$

where the mixing angle $\theta_\mathrm{P} = -11.5^\circ$. If SU(3) flavour symmetry were ‘exact’, the up, down, and strange quarks would be identical in all respects, and all nine mesons would share the same mass. However, in reality, the masses of the lights quarks are not identical, and the SU(3) flavour symmetry is only an approximation. This approximation is still reasonable since the mass difference between the light quarks, $m_s - m_{u,d} \approx 90\,\mathrm{MeV}/c^2$, is smaller than the strong interaction scale $\Lambda_\mathrm{QCD} \approx 200\,\mathrm{MeV}$. SU(3) flavour symmetry is also broken by electromagnetism, since the light quarks do not have the same electric charges. Electromagnetic corrections are expected to be small, approximately 2-3%, due to their small interaction strength. Indeed, the $\pi^\pm$ is slightly heavier than the π^0 due to corrections to their self-energy from electromagnetism[1].

2.6.3 $B \to \pi\pi$ isospin relations

Gronau and London [8] demonstrated that an SU(2) isospin analysis can be used to disentangle the penguin and tree contributions in the amplitudes of $B \to \pi\pi$ decays. This is possible due to the conservation of isospin throughout the hadronisation process. Recall that the pion carries an isospin of $I = 1$, and the third component $I_3 = -1, 0, 1$ corresponds to π^-, π^0, π^+, respectively. Hence, in the two-pion ($\pi\pi$) final state, the isospin of each pion is added together to give a total isospin of $I = 0, 1$, or 2.

---

[1] Self-energy is the additional contribution to the particle's energy due to the emission and absorption of virtual photons. Because charged pions carry an electric charge, they can interact with the electromagnetic field, leading to a higher self-energy than for the electrically neutral π^0.

However, not all total isospins are allowed since pions are bosons and their overall wavefunction must be symmetric due to the requirement of Bose-Einstein statistics. The wavefunction is a product of the spatial part, the spin part, and the isospin part. The spatial part is symmetric because the initial B meson has no spin and hence the total angular momentum of the two-pion system is zero. In addition, pions are spin-0 particles, and so the spin part is always symmetric. Hence, for the total wavefunction to be symmetric, the isospin part also has to be symmetric. The isospin $I = 1$ state is obtained by combining two isospin-1 states and is given by:

$$|1,1\rangle = \frac{1}{\sqrt{2}}|1;0\rangle - \frac{1}{\sqrt{2}}|0;1\rangle = \frac{1}{\sqrt{2}}(\pi^+\pi^0 - \pi^0\pi^+) \tag{2.106}$$

$$|1,0\rangle = \frac{1}{\sqrt{2}}|1;-1\rangle - \frac{1}{\sqrt{2}}|-1;1\rangle = \frac{1}{\sqrt{2}}(\pi^+\pi^- - \pi^-\pi^+) \tag{2.107}$$

$$|1,-1\rangle = \frac{1}{\sqrt{2}}|0;-1\rangle - \frac{1}{\sqrt{2}}|-1;0\rangle = \frac{1}{\sqrt{2}}(\pi^0\pi^- - \pi^-\pi^0) \tag{2.108}$$

where $|m;n\rangle = |1,m\rangle\,|1,n\rangle$ are the triplet of isospin-1 states from Equation 2.99. The $I = 1$ state is forbidden because it would lead to an antisymmetric wavefunction under exchange of the two pions. A similar analysis shows that the $I = 0, 2$ states are symmetric. Hence, for a two-pion system, the possible isospins are $I = 0, 2$.

Furthermore, it turns out that the tree diagram can give rise to states with $I = 0$, $I = 1$ or $I = 2$, whereas for the penguin diagram, only $I = 0$ or $I = 1$ are possible. This is because the B meson decay process via a penguin diagram involves the creation of intermediate states, as shown in the right-hand diagram of Figure 2.13. The strong force conserves isospin at each vertex of the Feynman diagram, and therefore the isospin of the intermediate state should be conserved. For an $I = 2$ state, the intermediate state would also need to have $I = 2$ to conserve isospin. However, the gluon that mediates the strong interaction between quarks is an isospin singlet that carries $I = 0$. Therefore, the total isospin is due solely to the two quarks ($I = \frac{1}{2}$) at the quark-gluon vertex. Thus, when combined, they can only create states of isospin 0, 1 or $\frac{1}{2}$, but not 2. As such, an $I = 2$ intermediate state is not possible and hence only isospin states $I = 0$ or $I = 1$ for the penguin diagram are allowed. This, coupled with the restriction imposed by Bose-Einstein statistics on $I = 1$ states, means that tree diagrams contribute to the two-pion final state with either $I = 0$ or 2, whereas penguin diagrams only contribute with $I = 0$. Similarly, the difference between the final and initial isospin (ΔI), is different for tree and penguin diagrams. B mesons possess an isospin of $I = \frac{1}{2}$ and hence $\Delta I = \frac{3}{2}$ is exclusive to tree-level interactions, while $\Delta I = \frac{1}{2}$ can occur at both the tree and penguin levels.

To find the amplitudes of the $B \to \pi\pi$ decays, let the two pion system be represented by $|\pi\pi\rangle = |I, I_3\rangle$, then $\pi^+\pi^- = (\pi_1^+\pi_2^- + \pi_1^-\pi_2^+)/\sqrt{2}$ and $\pi^+\pi^0 = (\pi_1^+\pi_2^0 + \pi_1^0\pi_2^+)/\sqrt{2}$. The final states are found using the Clebsh-Gordon coefficients:

$$|\pi^0\pi^0\rangle = |1,0\rangle\,|1,0\rangle = \sqrt{\frac{2}{3}}\,|2,0\rangle - \sqrt{\frac{1}{3}}\,|0,0\rangle \tag{2.109}$$

$$|\pi^0\pi^+\rangle = \frac{1}{\sqrt{2}}\left(|1,0\rangle\,|1,+1\rangle + |1,+1\rangle\,|1,0\rangle\right) = |2,1\rangle \tag{2.110}$$

$$|\pi^+\pi^-\rangle = \frac{1}{\sqrt{2}}\left(|1,+1\rangle\,|1,-1\rangle + |1,-1\rangle\,|1,+1\rangle\right) = \sqrt{\frac{1}{3}}\,|2,0\rangle + \sqrt{\frac{2}{3}}\,|0,0\rangle \tag{2.111}$$

As expected, the two-pion final state can only possess a total isospin of 0 or 2. Define the amplitudes as $A_{\Delta I, I_f}$, where ΔI is between the final and initial isospin and I_f is the isospin of the final state. The $B^0 \to \pi^0\pi^0$ decays at the quark level involve an anti-bottom quark forming an anti-down, up, and anti-up quarks. The final quark state of this process is given by

$$\left|\bar{b} \to \bar{u}u\bar{d}\right\rangle = A_{\frac{3}{2},2}\left|\frac{3}{2},+\frac{1}{2}\right\rangle + A_{\frac{1}{2},0}\left|\frac{1}{2},+\frac{1}{2}\right\rangle \tag{2.112}$$

where $A_{\frac{3}{2},2}$ and $A_{\frac{1}{2},0}$ are the unknown amplitudes to be extracted for the $\Delta I = \frac{3}{2}$ and $\Delta I = \frac{1}{2}$ processes, respectively. Finally, with the down quark spectator, it follows:

$$\begin{aligned}|B^0\rangle = \left|\bar{b} \to \bar{u}u\bar{d}\right\rangle |d \to d\rangle &= \left(A_{\frac{3}{2},2}\left|\frac{3}{2},+\frac{1}{2}\right\rangle + A_{\frac{1}{2},0}\left|\frac{1}{2},+\frac{1}{2}\right\rangle\right)\left|\frac{1}{2},-\frac{1}{2}\right\rangle \\ &= \sqrt{\frac{1}{2}}A_{\frac{3}{2},2}\,|2,0\rangle + \sqrt{\frac{1}{2}}(A_{\frac{1}{2},0} + A_{\frac{3}{2},2})\,|1,0\rangle + \sqrt{\frac{1}{2}}A_{\frac{1}{2},0}\,|0,0\rangle\end{aligned} \tag{2.113}$$

where $|B^0\rangle$ represents the B^0 meson initial state. To find the amplitude of the $B^0 \to \pi^0\pi^0$ decay

$$\begin{aligned}A^{00} &= \langle\pi^0\pi^0|B^0\rangle \\ &= \left[\sqrt{\frac{2}{3}}\,\langle 2,0| - \sqrt{\frac{1}{3}}\,\langle 0,0|\right]\left[\sqrt{\frac{1}{2}}A_{\frac{3}{2},2}\,|2,0\rangle + \sqrt{\frac{1}{2}}(A_{\frac{1}{2},0} + A_{\frac{3}{2},2})\,|1,0\rangle + \sqrt{\frac{1}{2}}A_{\frac{1}{2},0}\,|0,0\rangle\right] \\ &= \sqrt{\frac{1}{3}}A_{\frac{3}{2},2} - \sqrt{\frac{1}{6}}A_{\frac{1}{2},0}\end{aligned}$$

where A^{ij} are the amplitudes of the $B \to \pi^i \pi^j$ decay. Applying a similar calculation on all three final states, one finds

$$\begin{aligned} A^{00} &= \frac{1}{\sqrt{3}} A_{\frac{3}{2},2} - \frac{1}{\sqrt{6}} A_{\frac{1}{2},0} \\ A^{+0} &= \frac{\sqrt{3}}{2} A_{\frac{3}{2},2} \\ A^{+-} &= \frac{1}{\sqrt{6}} A_{\frac{3}{2},2} + \frac{1}{\sqrt{3}} A_{\frac{1}{2},0} \end{aligned} \tag{2.114}$$

For convenience, two variables are defined to absorb some Clebsh-Gordon coefficients.

$$A_0 = -\frac{1}{\sqrt{6}} A_{\frac{1}{2},0}, \quad A_2 = \frac{1}{2\sqrt{3}} A_{\frac{3}{2},2} \tag{2.115}$$

where A_0 and A_2 are the amplitudes for a B to decay into a $\pi\pi$ with $I = 0$ and $I = 2$ respectively. A_0 has tree and penguin contributions, while A_2 is purely tree. Here, the $\Delta I = 3/2$ contribution from electroweak penguins is assumed to be negligible, and only the $\Delta I = 1/2$ gluonic and electroweak penguins can contribute. This simplifies Equation 2.114 to

$$\begin{aligned} A^{00} &= 2A_2 + A_0 \\ A^{+0} &= 3A_2 \\ \frac{1}{\sqrt{2}} A^{+-} &= A_2 - A_0 \end{aligned} \tag{2.116}$$

This yields the complex triangle relation along with its charge-conjugated counterpart.

$$A^{+0} = \frac{1}{\sqrt{2}} A^{+-} + A^{00}, \quad \bar{A}^{-0} = \frac{1}{\sqrt{2}} \bar{A}^{+-} + \bar{A}^{00} \tag{2.117}$$

where A^{ij} are the amplitudes of the $\bar{B} \to \pi^i \pi^j$ decays. These two triangles are related by observing that the $B \to \pi^+ \pi^0$ decay is purely tree, as indicated in Equation 2.116. This can also be shown by noting that since the third component of isospin of the final state is $I_3 = +1$, the total isospin must be $I = 1$ or $I = 2$. However, as previously discussed, Bose-Einstein statistics forbid $I = 1$, meaning that $\pi^+\pi^0$ decays can only have $I = 2$, i.e., only the tree diagram contributes. Consequently, $|A^{+0}|$ and its conjugate $|A^{-0}|$ are equal. The weak phase of A^{+0} is ϕ_3, while that of $\bar{A}^{-0}$ is $-\phi_3$. Thus, ϕ_2 is half of the phase difference between the CP-conjugate amplitudes of the charged modes. This can also be seen in Equation 2.92 when there are no penguin contributions. By multiplying all $\bar{B}$ amplitudes by $e^{2i\phi_3}$, $\tilde{A} = e^{2i\phi_3}\bar{A}$ is redefined so that $\tilde{A}^{+0}$ and A^{+0} share the

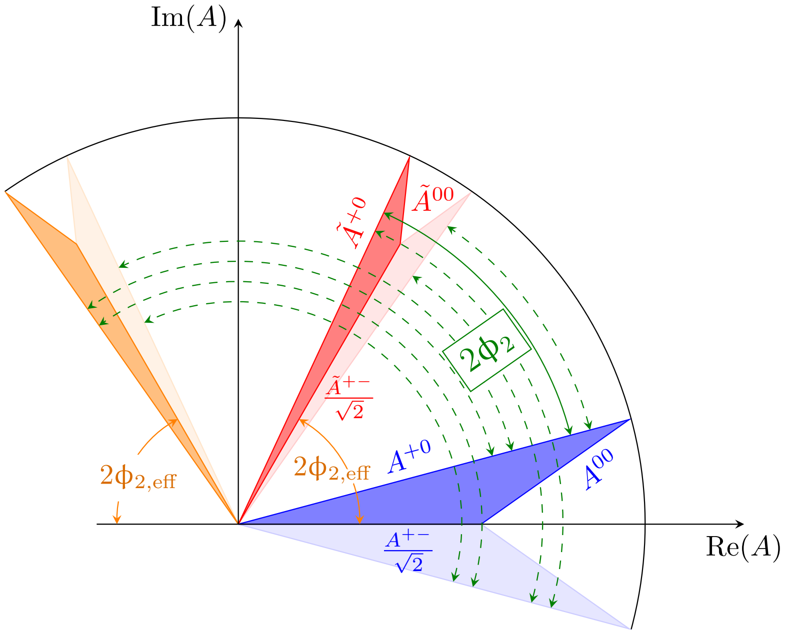


FIGURE 2.15: Geometric representation of the isospin triangular relations in the complex plane of $B^{i+j} \to \pi^i \pi^j$ amplitudes. The blue and the red shaded areas correspond to the isospin triangles. The angle between the *CP* conjugate charged amplitudes A^{+-} and $\bar{A}^{+-}$ corresponds to twice the weak phase $\phi_{2,\text{eff}}$ (orange solid arcs). The angle between the *CP* conjugate charged amplitudes A^{+0} and $\bar{A}^{+0}$ corresponds to twice the CKM angle ϕ_2 (green solid arc). The other triangles with lighter shading represent the mirror solutions allowed by the discrete ambiguities in the isospin relationships, with the corresponding values for ϕ_2 represented by the green dashed curves.

same base. This results in the triangle depicted in Figure 2.16. Geometrically, $\Delta\phi_2$ is represented by the angle between $\text{Arg}(A^{+-}/A^{+0})$ and $\text{Arg}(\bar{A}^{+-}/\bar{A}^{+0})$.

2.6.4 Extracting ϕ_2

As demonstrated in Equation 2.91, the three $B \to \pi\pi$ decays depend on their tree and penguin amplitudes. This amounts to a total of 12 hadronic parameters, along with the common weak phase ϕ_2. However, these 12 parameters can be reduced to 6. This reduction can be achieved using the complex isospin relations in Equation 2.117 to provide 4 real constraints, and by setting the penguin contribution in the purely tree $B^+ \to \pi^+\pi^0$ decay to zero to provide 2 real constraints. Therefore, the isospin-related $B \to \pi\pi$ decays can be represented with six real independent parameters along with the common weak phase.

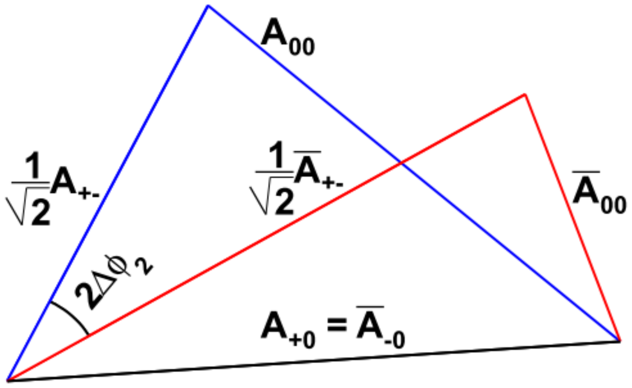


FIGURE 2.16: Complex isospin triangles from which $\Delta\phi_2$ can be determined

The amplitudes $|A^{+0}| = |\bar{A}^{+0}|$, $|A^{+-}|$, $|\bar{A}^{+-}|$, $|A^{00}|$, and $|\bar{A}^{00}|$ can be experimentally determined using Equation 2.72, through the measurement of the CP violating parameters and the branching ratio for each $B \to \pi\pi$ decay. The branching ratio, $\mathcal{B}(B \to f)$, is defined as the square of the amplitude and is related to $\mathcal{B}(\bar{B} \to f)$ through Equation 2.72, where B represents either a neutral or charged B meson, and f represents any final state. This can be expressed as

$$\mathcal{B}(\bar{B} \to f) = \frac{1 + \mathcal{A}_{CP}}{1 - \mathcal{A}_{CP}}\mathcal{B}(B \to f) \tag{2.118}$$

The branching fraction and CP asymmetries are connected to the decay amplitudes as given by the following equations:

$$\frac{1}{\tau_{B^{i+j}}}\mathcal{B}^{ij} = \frac{|A^{ij}|^2 + |\bar{A}^{ij}|^2}{2} \tag{2.119}$$

$$\mathcal{A}^{ij} = \frac{|A^{ij}|^2 - |\bar{A}^{ij}|^2}{|A^{ij}|^2 + |\bar{A}^{ij}|^2} \tag{2.120}$$

$$\mathcal{S}^{ij} = \frac{2\mathrm{Im}(\bar{A}^{ij}A^{ij*})}{|A^{ij}|^2 + |\bar{A}^{ij}|^2}, \tag{2.121}$$

where $\tau_{B^{i+j}}$ is the measured lifetime of the charged $(i + j = 1)$ or neutral $(i + j = 0)$ B meson. The six observables that can experimentally constrain ϕ_2 are $\mathcal{B}(\pi^+\pi^-)$, $\mathcal{B}(\pi^+\pi^0)$, $\mathcal{B}(\pi^0\pi^0)$, $\mathcal{A}_{CP}(\pi^0\pi^0)$, $\mathcal{A}_{CP}(\pi^+\pi^-)$, and $\mathcal{S}_{CP}(\pi^+\pi^-)$.

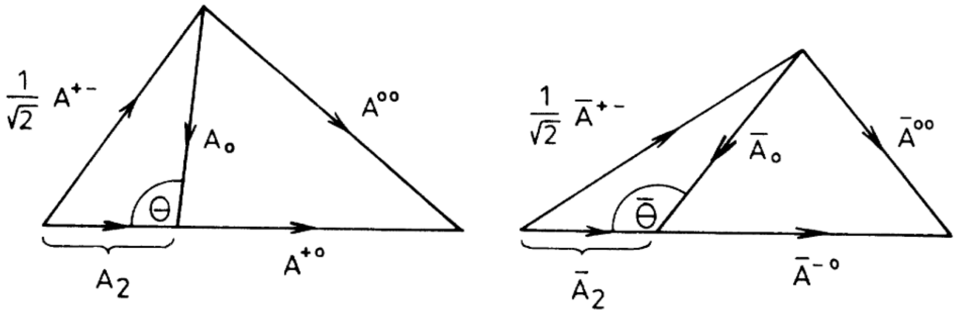


FIGURE 2.17: Complex triangles of Equation 2.117 [8].

To determine the $\Delta\phi_2$ shift, Equation 2.69 is modified to include penguin contributions using the isospin relation from Equation 2.116 for the $B^0 \to \pi^+\pi^-$ decay.

$$\begin{aligned}\mathrm{Im}(\lambda_{+-}) = \mathrm{Im}(\frac{q}{p}\frac{\bar{A}^{+-}}{A^{+-}}) &= \mathrm{Im}\left(e^{-2i\phi_1}\frac{\bar{A}_2 - \bar{A}_0}{A_2 - A_0}\right) \\ &= \mathrm{Im}\left(e^{-2i\phi_1}\frac{\bar{A}_2}{A_2}\frac{1-\bar{A}_0/\bar{A}_2}{1-A_0/A_2}\right) \\ &= \mathrm{Im}\left(e^{-2i\phi_1}e^{-2i\phi_3}\frac{1-\bar{z}}{1-z}\right) \\ &= \mathrm{Im}\left(e^{2i\phi_2}\left[\frac{1-\bar{z}}{1-z}\right]\right) \qquad (2.122)\end{aligned}$$

Here, $z \equiv A_0/A_2$, $\bar{z} \equiv \bar{A}_0/A_2$, and $\bar{A}_2$ is purely tree. It can be seen that in the limit that penguin contributions are neglected, $z = \bar{z}$ and the terms inside the square bracket equal one and Equation 2.122 reduces to 2.69. Hence, the phase of the terms within the square brackets represents the penguin contributions, $\Delta\phi_2$.

If the magnitudes of the decay amplitudes are known, the magnitudes of z and $\bar{z}$ can be geometrically obtained from the complex triangle relation in Figure 2.17, using Equation 2.116. However, only $\cos\theta$, where θ is the angle between A_0 and A_2, can be found. This leaves an ambiguity in the sign of θ, i.e. whether the triangle is pointing up or down. Consequently, z and $\bar{z}$ each have a two-fold ambiguity in their phase, which means that a measurement of $\phi_{2,\mathrm{eff}}$ will yield four possible values of ϕ_2. The ambiguity can be shown more explicitly by factoring out the magnitude and phase in Equation 2.122 so that

$$\mathrm{Im}(\lambda_{+-}) = \mathrm{Im}\left(e^{2i\phi_2}\left[\frac{1-|\bar{z}|e^{\pm i\bar{\theta}}}{1-|z|e^{\pm i\theta}}\right]\right) \qquad (2.123)$$

where θ is the angle between A_0 and A_2 and $\bar{\theta}$ is the angle between $\bar{A}_0$ and $\bar{A}_2$. This is shown in Figure 2.17. If the sides of the complex triangle of Equation 2.116 are known, then θ, $\bar{\theta}$, $|z|$, and $|\bar{z}|$ can all be determined. Here, the four potential ϕ_2 phase shifts are denoted as $\pm\epsilon_{+-}$ and $\pm\eta_{+-}$. The magnitude of the terms within the square brackets are represented as m_{+-}, which gives rise to four equations:

$$\begin{aligned} \mathrm{Im}(\lambda_{+-}) &= m_{+-}\sin\left(2\phi_2 \pm \epsilon_{+-}\right) \\ \mathrm{Im}(\lambda_{+-}) &= m_{+-}\sin\left(2\phi_2 \pm \eta_{+-}\right) \end{aligned} \tag{2.124}$$

These equations, along with Equation 2.94, yield four possible values for ϕ_2. To resolve this ambiguity, the $B \to \pi\pi$ decay which also includes penguin contributions are considered. Using a similar process, one finds that

$$\begin{aligned} \mathrm{Im}(\lambda_{00}) = \mathrm{Im}(\frac{q}{p}\frac{\bar{A}^{00}}{A^{00}}) &= \mathrm{Im}\left(e^{-2i\phi_1}\frac{2\bar{A}_2 - \bar{A}_0}{2A_2 - A_0}\right) \\ &= \mathrm{Im}\left(e^{2i\phi_2}\left[\frac{1+\frac{1}{2}\bar{z}}{1+\frac{1}{2}z}\right]\right) \\ &= \mathrm{Im}\left(e^{2i\phi_2}\left[\frac{1+\frac{1}{2}|\bar{z}|e^{\pm i\bar{\theta}}}{1+\frac{1}{2}|z|e^{\pm i\theta}}\right]\right) \end{aligned} \tag{2.125}$$

Then the four possible ϕ_2 phase shifts are denoted as $\pm\epsilon_{00}$ and $\pm\eta_{00}$, and the magnitude of the terms within the square bracket as m_{00}.

$$\begin{aligned} \mathrm{Im}(\lambda_{00}) &= m_{00}\sin\left(2\phi_2 \pm \epsilon_{00}\right) \\ \mathrm{Im}(\lambda_{00}) &= m_{00}\sin\left(2\phi_2 \pm \eta_{00}\right) \end{aligned} \tag{2.126}$$

Equation 2.124 and 2.126 each have four solutions for ϕ_2. The overlapping solution between these two equations unambiguously determines ϕ_2.

In summary, the complex isospin triangles given by Equation 2.116 and ϕ_2 can be fully determined from the branching fractions $\mathcal{B}(B^0 \to \pi^+\pi^-)$, $\mathcal{B}(B^0 \to \pi^0\pi^0)$ and $\mathcal{B}(B^+ \to \pi^+\pi^0)$, and the CP violation parameters $\mathcal{A}_{CP}(B^0 \to \pi^-\pi^+)$, $\mathcal{S}_{CP}(B^0 \to \pi^+\pi^-)$ and $\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0)$. The isospin analysis of charmless B meson decays is one of the most theoretically precise ways to determine ϕ_2, since it is almost free of hadronic uncertainties due to QCD effects in hadrons, and the isospin breaking effects are small compared to experimental uncertainty.

2.7 Current constraints

Before reviewing the current status of $B^0 \to \pi^0\pi^0$ and ϕ_2, it is useful to define how the constraints are defined. Constraints are determined through an exploration of the N-dimensional parameter space, utilising a frequentist statistical approach. The set of experimental observable, denoted O_{exp} is measured in terms of likelihoods that can be used to create a χ^2-like test statistic:

$$\chi^2(p) = -2\log\mathcal{L}(O_{\text{exp}} - O(p)), \tag{2.127}$$

where $O(p)$ represents the theoretical value of the observable for a fixed set of parameters p. The test statistic χ^2 is initially minimised over the entire parameter space, allowing all N parameters p to vary freely. To convert χ^2 into a p-value, $\chi^2(p)$ is interpreted as a random variable distributed according to the χ^2 law with a certain number of degrees of freedom. The degrees of freedom are given by $N_{\text{dof}} = M - N$, where M is the number of independent variables and N is the quantity of free parameters. This is valid in the limit of large samples under the condition of Wilks' theorem which states that as the number of events approaches infinity, the distribution of the test statistic $-2\log\big(O_{\text{exp}} - O(p)\big)$ asymptotically approaches the χ^2 distribution.

To constrain ϕ_2, the independent hadronic parameters were considered as 'nuisance parameters'. These parameters, while necessary for a complete model, are not of immediate interest. This approach allowed for the adjustment of these parameters to their optimal values in order to minimise the χ^2 statistic. The test statistic was then defined as the difference between χ^2 values:

$$\Delta\chi^2(\phi_2) = \min[\chi^2(\phi_2)] - \chi^2_{\min} \tag{2.128}$$

where $\min[\chi^2(\phi_2)]$ refers to the value of χ^2 minimised with respect to the nuisance parameters for a fixed ϕ_2 value, while $\chi^2_{\min}$ represents the absolute minimum. This serves to quantify the degree to which any hypothesised ϕ_2 agrees with the data, independent of the value of the nuisance parameters. Confidence intervals on ϕ_2 can be calculated assuming that $\Delta\chi^2(\phi_2)$ follows a χ^2 distribution with one degree of freedom:

$$p(\phi_2) = \text{Prob}(\Delta\chi^2(\phi_2), N_{\text{dof}} = 1) = \frac{\Gamma(N_{\text{dof}}, \Delta\chi^2/2)}{\Gamma(N_{\text{dof}}/2)} \tag{2.129}$$

The confidence intervals at a specific confidence level (CL) are obtained by identifying the values of ϕ_2 with a p value that exceeds 1− CL.

2.8 Current constraints on $B \to \pi\pi$ isospin

The current values of $\mathcal{B}(B^0 \to \pi^0\pi^0)$ and $\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0)$ are shown in Table 2.1.

	Belle ($\times 10^{-6}$)	**BaBar ($\times 10^{-6}$)**	**Average ($\times 10^{-6}$)**
$\mathcal{B}(B^0 \to \pi^0\pi^0)$	$1.31 \pm 0.19 \pm 0.19$	$1.83 \pm 0.21 \pm 0.13$	1.59 ± 0.26
$\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0)$	$0.14 \pm 0.36 \pm 0.10$	$0.43 \pm 0.26 \pm 0.05$	0.33 ± 0.22

TABLE 2.1: The branching fraction and A_{CP} of $B^0 \to \pi^0\pi^0$ as measured at Belle and BaBar along with their current value [5].

The consistency of the experiment $B \to \pi\pi$ data with the isospin relation can be assessed with the two-side sum:

$$t = \frac{|a^{+-}|}{\sqrt{2}} + |a^{00}| \tag{2.130}$$

where a^{ij} denotes the normalised amplitude $a^{ij} = A^{ij}/A^{+0}$. The sum of the two lengths, t, must be greater than one. The values [33]

$$t = 1.05 \pm 0.09 \quad \text{and} \quad \bar{t} = 1.45 \pm 0.08 \tag{2.131}$$

are consistent with an almost flat and open triangle for $B \to \pi\pi$ and $\bar{B} \to \pi\pi$, respectively. This is shown in Figure 2.18.

2.8.1 Current constraints on ϕ_2

The experimental extraction of these parameters using $B \to \pi\pi$ decays has been performed by BaBar, Belle, and LHCb. The isospin analysis uses the global averages of these measurements $\mathcal{S}_{CP}(\pi^+\pi^-) = -0.68 \pm 0.04$, $\mathcal{A}_{CP}(\pi^+\pi^-) = -0.27 \pm 0.04$, the branching fractions of all three modes, and the direct CP asymmetry $\mathcal{A}_{CP}(\pi^0\pi^0) = -0.33 \pm 0.22$. Due to experimental uncertainties, some of these solutions are not well separated, as demonstrated in Figure 2.19. Only the following constraints [5] can be obtained at

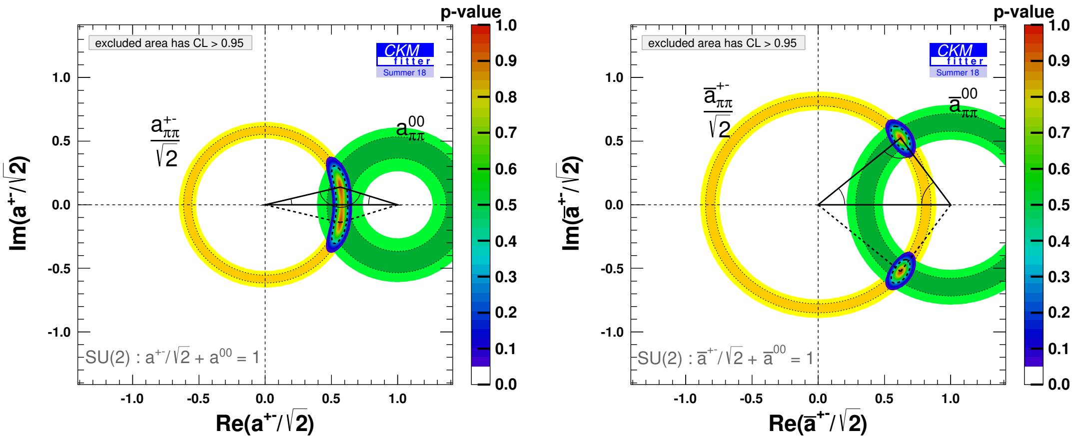


FIGURE 2.18: Constraint on the reduced isospin amplitude $a^{+-} = A^{+-}/A^{+0}$ (left) and $\bar{a}^{+-} = \bar{A}^{+-}/\bar{A}^{+0}$ (right) in the complex plane for the $B \to \pi\pi$ system. The individual constraint from the $B^0 \to \pi^+\pi^-$ (left) and $\bar{B}^0 \to \pi^+\pi^-$ (right) system and from the $B^0 \to \pi^0\pi^0$ (left) and $\bar{B}^0 \to \pi^0\pi^0$ (right) observables are indicated by the yellow and green circular areas, respectively. The corresponding isospin triangular relation is represented by the black triangle [9]

the 68% confidence level (CL): $-13.5^\circ < \phi_2 < 15.7^\circ$, $74.3^\circ < \phi_2 < 105.6^\circ$, and $118.5^\circ < \phi_2 < 151.5^\circ$.

The solution centred at 90° agrees with the indirect determination of $\phi_{2,\text{indirect}} = (91.9^{+1.6}_{-1.2})^\circ$ by a global fit of the flavour data performed by the CKMFitter group in Spring 2021 [9]. The largest source of uncertainty is in the measurement of $B^0 \to \pi^0\pi^0$ due to its low branching ratio and the high experimental background.

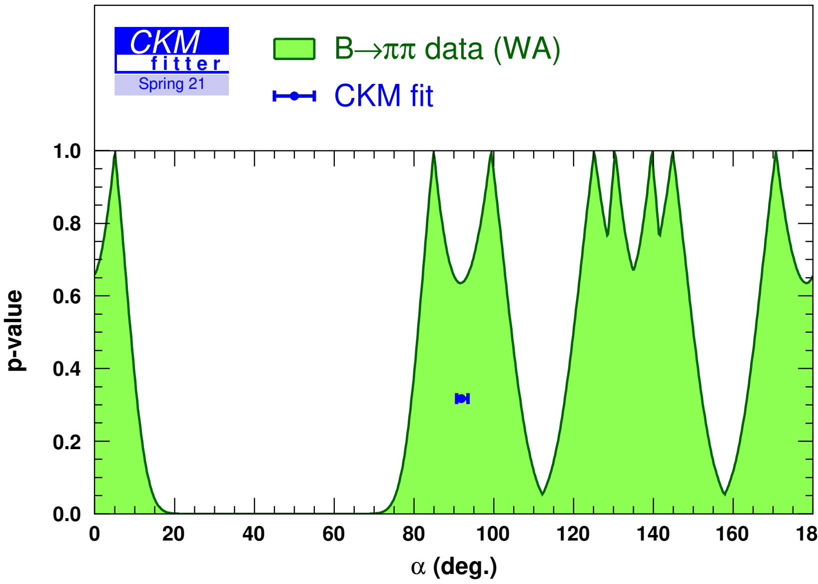


FIGURE 2.19: Constraints on ϕ_2 from $B \to \pi\pi$ isospin and the global CKM fit [9].

2.8.2 Constraints on $\bar{\rho}$ and $\bar{\eta}$

The constraint on ϕ_2 can be recast as a constraint on the $(\bar{\rho}, \bar{\eta})$ Wolfenstein parameters of the CKM matrix representing the apex of the B meson unitarity triangle. The following relation can be derived:

$$\left(\bar{\eta} - \frac{\cotan(\phi_2)}{2}\right)^2 + \left(\bar{\rho} - \frac{1}{2}\right)^2 = \frac{1}{4\sin^2(\phi_2)} \tag{2.132}$$

The curves at fixed ϕ_2 consist in circles centred on the point $(\bar{\rho}, \bar{\eta}) = (1/2, \cotan(\phi_2)/2)$. All curves meet at the points (0,0) and (1,0). Figure 2.20 shows the constraint result from the direct determination of ϕ_2 in the $(\bar{\rho}, \bar{\eta})$ plane. The constraint of the unitarity triangle from all measurements as determined by the CKMFitter Group [9] is shown in Figure 2.21.

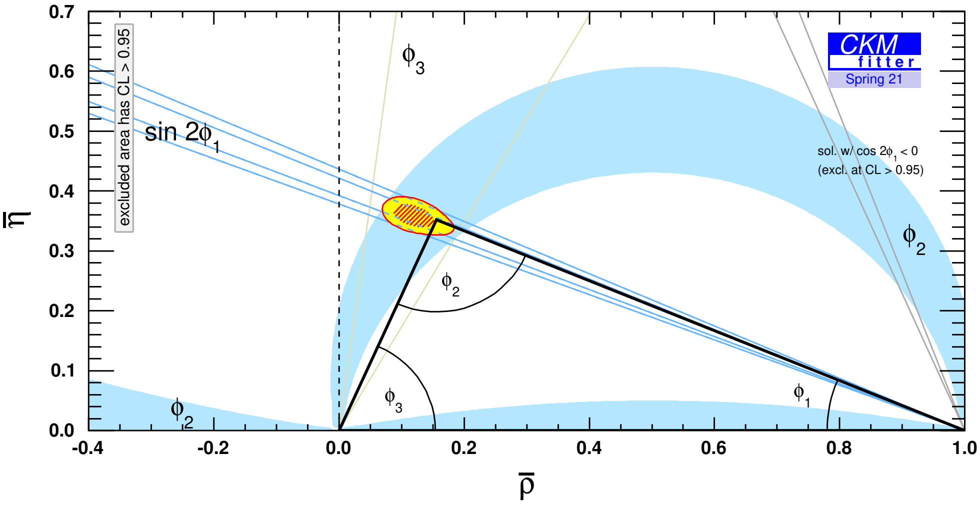


FIGURE 2.20: Constraints in the $(\bar{\rho}, \bar{\eta})$ plane including only the angle measurements [9].

2.9 $B^0 \to \pi^0\pi^0$ predictions

Beyond the importance of the $B^0 \to \pi^0\pi^0$ for improving ϕ_2 measurements, this decay mode is interesting to study in its own right. The quantitative prediction of the $B^0 \to \pi^0\pi^0$ decay requires the computation of challenging hadronic matrix elements, which involve both short-distance perturbative and long-distance non-perturbative QCD. Hadronic matrix elements describe the transition between the initial B meson and the two-pion final-state. This involves understanding how the quarks and gluons within the B meson rearrange themselves to form the pions. Hence, theoretical predictions of $\mathcal{B}(\to \pi^0\pi^0)$

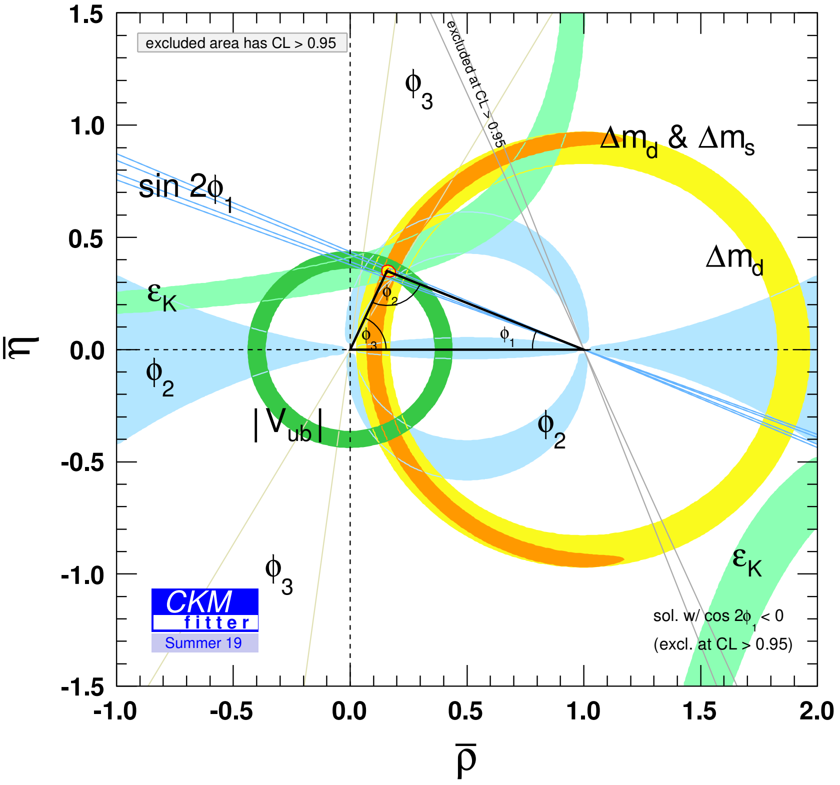


FIGURE 2.21: Summary of the measurements for the unitarity triangle [9]. *CP* violation is proportional to the height $\bar{\eta}$.

serve to test our theoretical understanding of charmless hadronic B decays. In particular, the consistency of commonly employed tools such as QCD factorisation and perturbative QCD.

Prior to the initial measurement of $\mathcal{B}(B^0 \to \pi^0\pi^0)$ in 2003 by Belle and Babar [34, 35], the branching fraction was expected to be below 1×10^{-6} [36, 37]. This can be seen most easily in the topological amplitude parameterisation [38], which is independent of the theoretical models used for the calculation of hadronic matrix elements.

$$\begin{aligned} \sqrt{2}A(B^+ \to \pi^+\pi^0) &= -T[1 + \frac{C}{T} + \frac{P_{ew}}{T}e^{i\phi_2}], \\ A(B^0 \to \pi^+\pi^-) &= -T(1 + \frac{P}{T}e^{i\phi_2}), \\ \sqrt{2}A(B^0 \to \pi^0\pi^0) &= T[(\frac{P}{T} - \frac{P_{ew}}{T})e^{i\phi_2} - \frac{C}{T}], \end{aligned} \tag{2.133}$$

where, T, C, P, and P_{ew} stand for the colour-allowed tree, colour-suppressed tree, gluonic penguin, and electroweak penguin amplitudes, respectively. The colour-suppressed electroweak penguin, exchange, and penguin-annihilation amplitudes have been neglected.

The amplitudes obey the hierarchy of the amplitudes of different processes in the SM [39],

$$\frac{P}{T} \approx \lambda, \quad \frac{C}{T} \approx \lambda, \quad \frac{P_{ew}}{T} \approx \lambda^2, \tag{2.134}$$

Hence, the main contributions to the $B^0 \to \pi^0\pi^0$ decays are expected to come from colour-suppressed tree and gluonic penguin amplitudes. Furthermore, the P_{ew} contribution to $B^+ \to \pi^+\pi^0$ and $B^0 \to \pi^+\pi^-$ is expected to be very small, especially in the latter case due to colour suppression. The branching ratio for $B^0 \to \pi^0\pi^0$ decays is predicted to be of the order $O(\lambda^2)$, whereas the branching ratio for $B^0 \to \pi^+\pi^-$ decays is of the order $O(\lambda)$. That is, $\mathcal{B}(B^0 \to \pi^+\pi^-) \gg \mathcal{B}(B^0 \to \pi^0\pi^0)$. However, the observed measurements contradict this prediction. This inconsistency is also reflected in theoretical predictions based on QCD factorization [36, 40–42] and perturbative QCD [37, 43], which are approximately five times smaller than the world average value.

Furthermore, the ratio of color-suppressed to color-allowed tree amplitudes, as inferred from other charmless two-body decay modes, does not agree well with expectations [44]. This might indicate large electroweak-penguin contributions, which are difficult to explain in the SM [45, 46]. Various approaches, which predict a wide range of values for $\mathcal{B}$ and $\mathcal{A}_{CP}$, have been proposed as possible solutions to this disagreement [47–50]. Obtaining more precise measurements of these quantities would aid in distinguishing among the various proposed solutions aimed at addressing this disagreement.

Chapter 3

The Belle II Experiment

The Belle II experiment is an international particle physics experiment focused on analysing data gathered by the Belle II detector situated at the interaction point (IP) of the electron-positron collider SuperKEKB. Belle II is located at the High Energy Accelerator Research Organisation (KEK) in Tsukuba, Ibaraki Prefecture, Japan and is the successor to the Belle Experiment that operated from 1999 to 2010. The hardware upgrade of the Belle II detector and SuperKEKB began at the conclusion of Belle's collision data-taking and has taken almost a decade to complete. After a long period of commissioning, on the 25th of March 2019, the first electron-positron collision with the almost complete[1] Belle II detector began. The Belle II target data set is 50 ab^{-1}, which is approximately fifty times more data than Belle.

Belle II is known as a 'B-factory', as it is designed to produce and detect large numbers of B meson pairs. The decay of B mesons is an excellent way to probe new physics due to the relatively large mass of the bottom quark, which can enhance the contributions from new virtual particles. Furthermore, B mesons have hundreds of measurable decay modes, providing greater opportunities to study potential new physics. Belle II will use this data set to perform a wide range of extremely high-precision and fundamental measurements in flavour physics. This precision is required to probe the Standard Model, which has withstood almost every experimental test at current levels of precision, and constrain the parameter space for new physics. Belle II is also the only operating experiment that

---

[1] Due to technical and budget constraints a partial VXD system was used

can competitively study the $B^0 \to \pi^0\pi^0$ decay mode due to the high levels of background that other experiments would struggle to disentangle from the signal.

3.1 Studying $B\overline{B}$ Meson Pairs

Experiments in the search for new physics are typically categorised into two frontiers: the energy frontier and the intensity frontier. At the energy frontier, hadronic particles, such as protons, collide at high energies, leading to the production of high-mass particles. These collisions result in the formation of narrow cones, known as ‘jets’, which can comprise dozens of hadrons and other particles. Such jets are the result of protons being strongly bound composite particles with constituent quarks and gluons that interact in complex ways. In contrast, electrons and positrons are fundamental particles and leave no remnants from their collisions to form ‘jets’.

The centre-of-mass (c.m.) energy of electron-positron collisions is constrained by the energy loss from synchrotron radiation. This limitation arises because the energy radiated by a charged particle in a magnetic field per turn is inversely proportional to the fourth power of the particle’s mass. As a result of the difference in mass, electrons/positrons radiate energy at a rate approximately 10^{13} times higher than that of protons. Consequently, electron-positron colliders are designed to collide at very specific c.m. energies. At Belle II, the SuperKEKB accelerator is used to produce $B\overline{B}$ pairs by colliding 7 GeV electrons (E_{e^-}) and 4 GeV positrons (E_{e^+}) with a c.m. energy ($E_{\text{c.m.}}$) slightly greater than the mass of the $\Upsilon(4S)$,

$$E_{\text{c.m.}} = \sqrt{4E_{e^-}E_{e^+}} = 10.58 \text{ GeV} \approx M_{\Upsilon(4S)} \tag{3.1}$$

In electron-positron collisions at this energy, the most probable process is Bhabha scattering $e^-e^+ \to e^-e^+$ with a cross section of approximately 300 nb. In contrast, the cross section for $e^-e^+ \to \Upsilon(4S)$ processes is only around 1.110 ± 0.008 nb [51]. If a $\Upsilon(4S)$ is created, it will rapidly decay through the strong force, producing a quantum mechanically entangled pair of short-lived B mesons. Due to the conservation of angular momentum, the resultant spin-0 B mesons from the spin-1 $\Upsilon(4S)$ decay are in a coherent $L = 1$ state. In the case where neutral B mesons are produced, each of the mesons evolves according to Equation 2.54. The two neutral B mesons evolve in phase, so that at any given time, there is always precisely one B^0 and one $\overline{B}^0$. This is true until one of the two neutral B mesons decays. If one of the mesons decays into a flavour-specific decay mode, the other meson in the pair, at that same instant, must have the opposite flavour. For example, consider a flavour-specific decay mode such as $B^0 \to D^+\pi^-$. The D^+ meson contains a c quark and the π^- contains a $\bar{u}$ and d quark, so the decay must originate from a B^0 ($b \to c\bar{u}d$) and not from a $\bar{B}^0$. Once this decay occurs, the other meson in the pair is now known to be a $\bar{B}^0$. Similarly, if one of the mesons decays into a CP eigenstate, the other meson in the pair must have the opposite CP. Once one of the B mesons decays, the other B meson is able to evolve independently. Consequently, it is possible to observe events with two B^0 or two $\bar{B}^0$. Therefore, it is the difference in decay times between the two B mesons that is important, rather than the time from creation to decay.

In the c.m. frame, the two B mesons are produced almost at rest without additional particles, since the mass of the $\Upsilon(4S)$ is only slightly greater than the B mesons. In the laboratory frame, the two B meson pairs get a relativistic Lorentz boost in the direction of the electron beam due to the asymmetric beam energies. The Lorentz boost factor,

$$\beta\gamma = \frac{E_e^- - E_e^+}{E_{\text{c.m.}}} = 0.28$$

of the e^-e^- system increases the vertex difference from 2 μm to 200 μm, so that the B mesons travel an appreciable distance before decaying. The decay modes can be efficiently reconstructed as the initial energy state is very well defined. By extrapolating the tracks of the decay products of the B mesons, the decay vertex of the B meson can be determined from where all tracks coincide. From the two vertices, a precise determination of the time between the B meson decays can be made, allowing precise measurement of the lifetimes, mixing parameters, and CP violation.

3.2 SuperKEKB Collider

SuperKEKB is an electron-positron collider, located in Ibaraki Prefecture, Tsukuba, Japan. The collision region employs a beryllium double wall beam pipe and two storage rings with a circumference of 3 km, aligned side by side. The High-Energy Ring (HER) circulates the electron beam and the Low-Energy Ring (LER) circulates the positron beam in opposite directions, as shown in Figure 3.1. The electrons are produced by irradiating a cold cathode with photons from a laser, while positrons are produced by capturing positrons created by energetic electron collisions on a tungsten target.

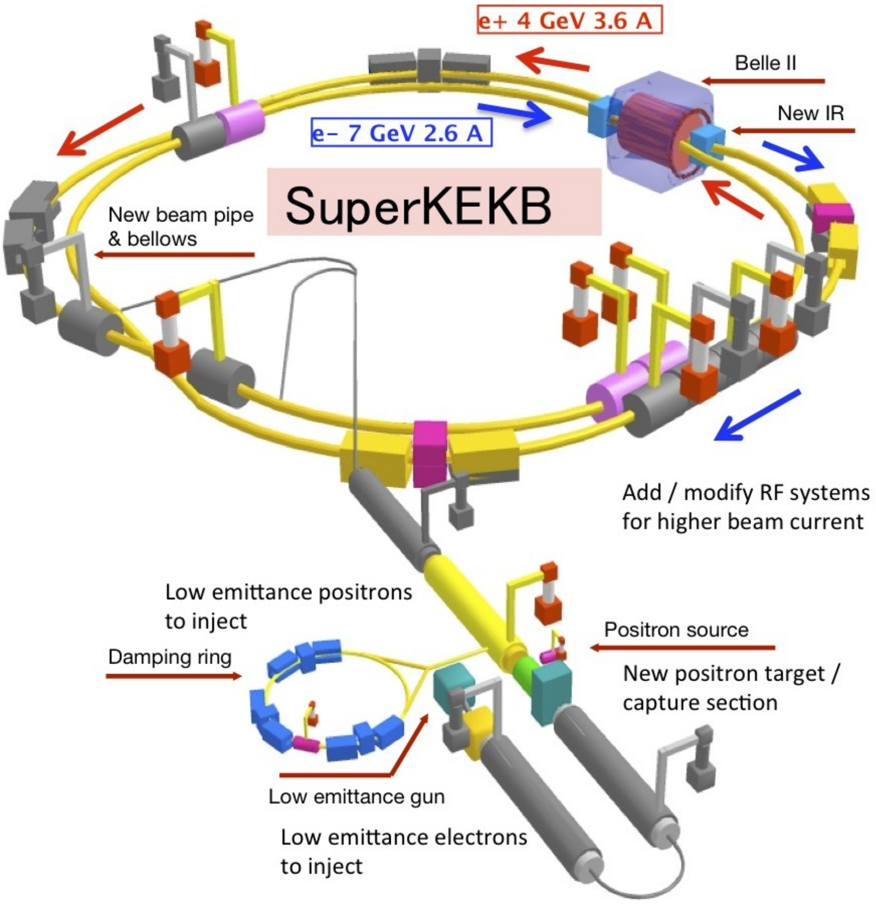


FIGURE 3.1: Layout of the SuperKEKB collider [10].

The electrons and positrons are accelerated to 7 GeV and 4 GeV, respectively, by repeatedly traversing several RF cavities. Each RF cavity is designed to generate an oscillating electromagnetic field at a specific frequency that resonates with the particles as they pass through. This results in the acceleration of the particles as a result of the transfer of energy from the electromagnetic field to the particles. By passing the particles through multiple RF cavities, each tuned to a specific frequency, the particles gain energy in a stepwise fashion until the desired level of acceleration is achieved. The particles in the beam will form densely packed structures known as ‘bunches’. This bunching effect is largely attributed to the cosine nature of the electric field generated within the RF cavities. Specifically, the strength of the electric field peaks in the centre of the cavity and tapering off towards the edges. A particle positioned ahead of the bunch experiences a weaker field strength, reducing the energy gain through the cavity and causing it to decelerate. Conversely, a particle lagging behind the bunch finds itself in a region of higher field strength, which imparts a larger gain in energy, causing the particle to speed up. This effect, over time and over many passes through multiple RF cavities, causes the beam to naturally form into bunches. The particles are constrained in a circular path by an array of dipole 'bending magnets'. Quadrupole magnets are used to constrain the transverse beam size and to provide the final focus at the IP.

To achieve a target luminosity that is forty times greater than the peak of KEKB, the 'nano-beam' scheme is used [52]. Relative to KEKB, the beam size at the IP is reduced by a factor of 20, resulting in a beam size of approximately 10 μm × 50 nm (horizontal × vertical). As shown in Figure 3.2, the current is increased by a factor of 2, and the two beams collide at 83 mrad instead of 22 mrad. The Belle beam pipe is reused with the gold foil shielding covering the inner wall, rather than the outer wall. The shielding is thinned to a thickness of only 10 μm. This helps reduce the deviation of a charged particle from its original trajectory as it travels through matter as a result of successive interactions with multiple atoms, an effect known as 'multiple scattering'. Positioning it on the inner surface helps improve the impact parameter resolution, which refers to the precision with which the distance from the closest approach of a track to the collision point can be measured.

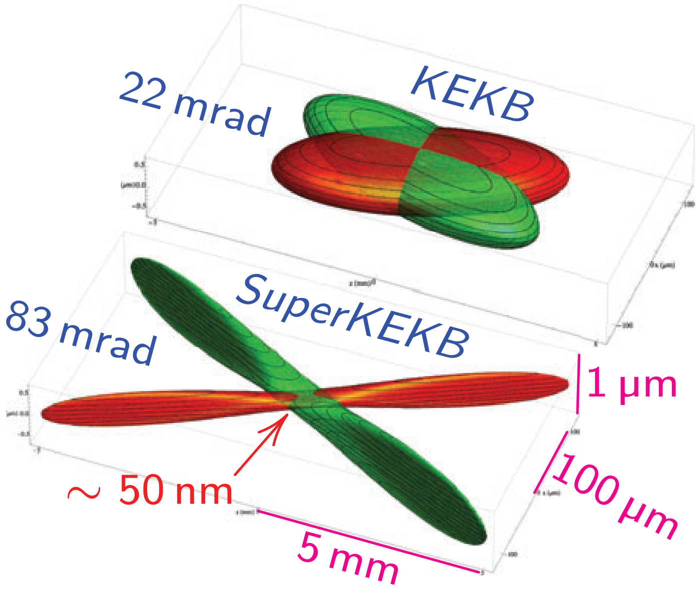


FIGURE 3.2: Comparison of beam size and angle at Belle and Belle II

3.3 The Belle II Detector

The Belle II detector is a general-purpose hermetic detector located around the interaction region of the asymmetric energy e^+e^- collider, SuperKEKB. A full description of the Belle II detector is given in Ref. [53]. The detector consists of several subdetectors arranged in a concentric structure around the beam pipe. The z-axis of the lab frame is defined as the symmetry axis of a superconducting solenoid, which generates a 1.5 T uniform field along the beam direction. The positive direction is given by the direction of the electron beam and the polar angle, θ, is defined with respect to the positive z-axis. The detector is divided into three regions, named in order of increasing θ as the forward endcap, the barrel, and the backward endcap. The detector material is minimised to reduce multiple scattering, enabling high efficiency and excellent resolution for low-energy photons. The coverage of the polar angle is asymmetric (17° to 150°) to match the boost from the asymmetric energy collisions. The complete Belle II detector is 8 metres wide, 8 metres high, and weighs about 1,400 tons.

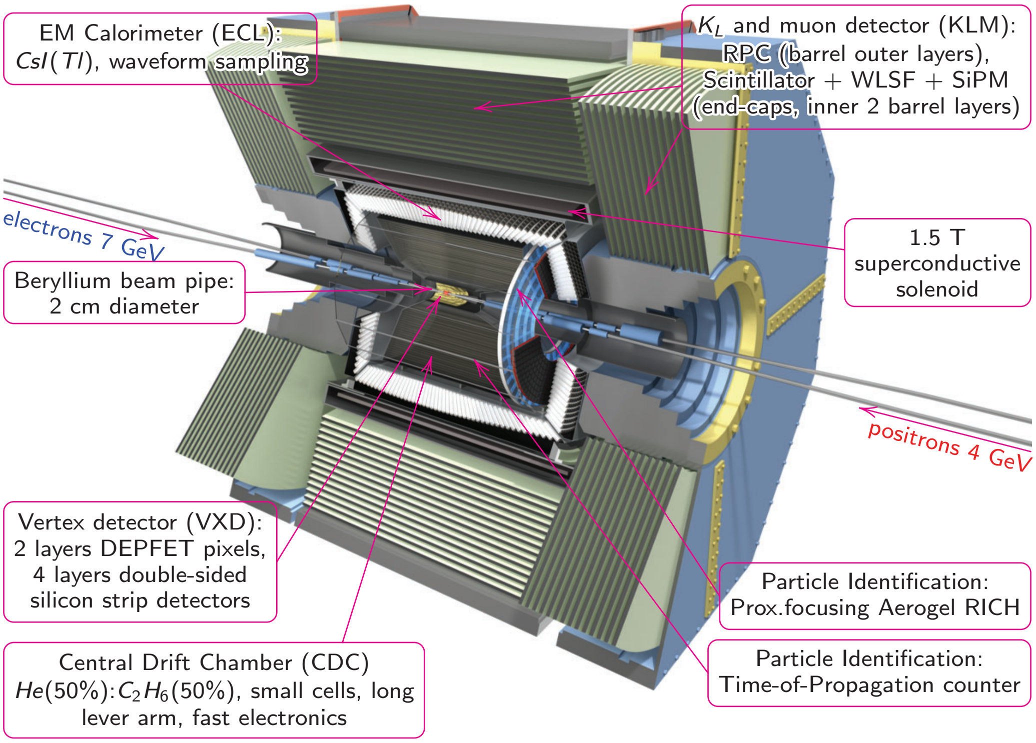


FIGURE 3.3: Comparison of beam size and angle at Belle and Belle II.

The inner subdetectors consist of a silicon pixel detector (PXD) surrounded by a four-layer, double-sided silicon strip detector (SVD), and a central drift chamber (CDC).

These subdetectors are used to reconstruct charged particles and measure their momentum. A time-of-propagation counter (TOP) and an aerogel ring-imaging Cherenkov detector (ARICH) cover the barrel and forward endcap regions, respectively, and are used to identify charged particles. The electromagnetic calorimeter (ECL) is a segmented array of 8736 thallium-doped caesium iodide [CsI(Tl)] crystals arranged in a projective geometry toward the IP and covering about 90% of the solid angle in the c.m. frame. The ECL identifies electrons and photons in an energy range of 20 MeV to 4 GeV and occupies the remaining volume inside the superconducting solenoid. The K_L and muon detector (KLM) is responsible for detecting particles of the same name. It is composed of resistive plate chambers and scintillating fibers, which are interspersed between the iron plates of the magnetic flux return. The Belle II subdetectors are shown in Figure 3.3. The analog signals from the various subdetectors are digitised in either internal front-end electronics (FEE) cards (CDC, TOP, and ARICH) or near the detector (SVD, ECL, and KLM) so that they can be read and processed by the data acquisition system (DAQ). However, before describing each subdetector in more detail, it is important to review how different particles interact with various materials.

3.3.1 Particle interaction in the detector material

In particle physics experiments, only the electron, muon, photon, charged pion, charged and neutral kaon, proton, and neutron are stable enough to travel through the detector to be measured. These particles interact with the detector material, and their effects are measured using specialised instruments to determine properties about the particle, such as energy/momentum, electric charge, etc. This requires an extensive theoretical understanding of how different particles interact with different materials under different conditions. The charged track transverse momentum spectrum of a typical Belle II event ranges between tens of MeV/c and a few GeV/c, with an average of 11 charged tracks. Here, the primary ways in which particles interact with the detector material at energies accessible at Belle II, i.e. above tens of MeV, are discussed.

The primary ways charged particles lose energy are through ionisation, radiation, and pair production. Ionisation occurs when charged particles interact with the atomic electrons of a medium through the electromagnetic force, resulting in the transfer of energy and momentum. This process is known as Coulomb scattering and causes the incoming particles to lose energy and the atom to be excited or ionised. All charged particles impart energy to the medium through which they traverse via this mechanism. Ionisation is the dominant form of energy loss for charged particles other than electrons and positrons. The rate of ionisation energy loss per distance travelled can be described by the Bethe-Bloch formula,

$$-\frac{dE}{dx} = K\frac{Z}{A}\frac{Z}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_ec^2\beta^2\gamma^2T_{max}}{I^2}\right) - \beta^2 - \frac{1}{2}\frac{1}{\gamma^2}\right], \tag{3.2}$$

where K is a constant, Z and A are the atomic number and atomic mass of the medium, respectively, m_e is the mass of the electron, c is the speed of light, β is the velocity of the particle divided by c, γ is the Lorentz factor, T_{max} is the maximum transfer of kinetic energy to an electron, and I is the mean excitation energy of the medium. From this equation, it can be seen that the ionisation energy loss depends only on β, with a weak dependence on the medium itself, since $\frac{Z}{A} \approx 0.5$ for most materials. The Bethe-Bloch formula for positive muons in copper as a function of muon momentum is shown in Figure 3.4. Radiation energy loss occurs when the electric field of a nucleus accelerates and decelerates the particles as they pass through, causing the particles to radiate photons and hence lose energy. This process is called ‘bremsstrahlung’ and is an additional energy loss mechanism for electrons and positrons. This interaction can be characterised by a parameter known as the ‘radiation length’, X_0, which is the average distance over which an electron loses all but $1/e$ of its energy by bremsstrahlung. It is a property of the material of a detector through which the particles are passing and is typically given in units of centimetres. Pair production is the process by which a photon loses energy by converting into an electron-positron pair. This rate of this process is proportional to the square of the charge of the atomic nuclei of the material.

Electrons lose energy through bremsstrahlung and pair production as they pass through the high Z-material in the ECL, creating electromagnetic showers. These electromagnetic showers are a cascade of secondary particles, typically composed of a large number of electrons, positrons, and photons. The shower progresses until the energy of the shower particles falls below a critical energy. This causes the dominant energy-loss mechanism to switch from bremsstrahlung to ionisation, leading to the gradual dissipation of the shower. Photons lose energy through an analogous mechanism via pair production. They can also lose energy, to a lesser extent, through Compton scattering with atomic electrons. As shown in Figure 3.5, the ECL absorbs almost all electrons and

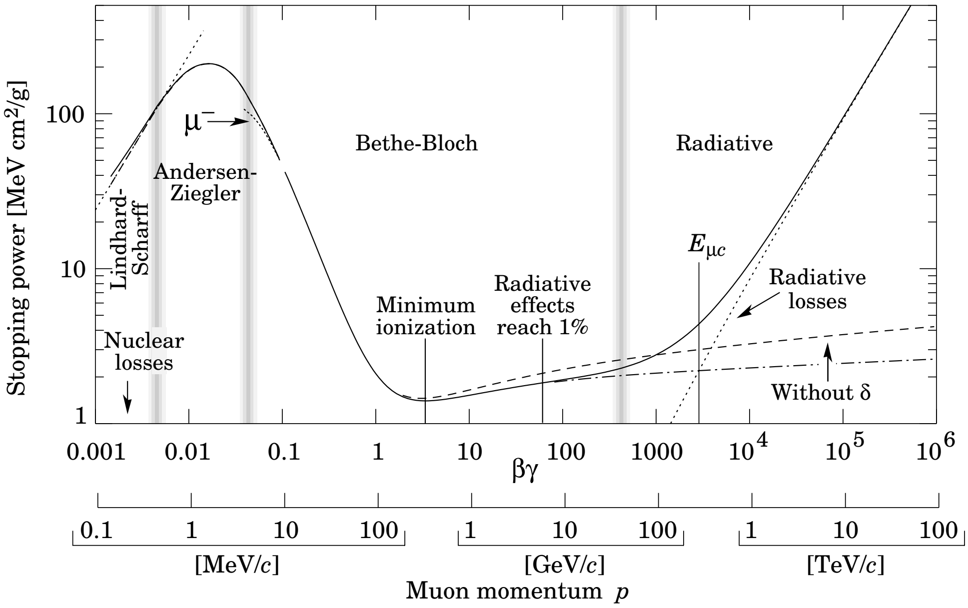


FIGURE 3.4: The energy loss, dE/dx vs muon momentum according to the Bethe-Bloch formula [11]. The energy loss decreases as the particle energy increases, reaching a minimum, after which there is a 'relativistic rise'.

photons. This is because CsI(Tl) crystals have a radiation length of 16.2 X_0. Therefore, the probability that a photon that enters a crystal from the front does not interact is $e^{(7/9)(16.2)} = 3.37 \times 10^{-6}$.

For hadrons such as pions, kaons, protons, and neutrons, it is important to consider strong interactions with the atomic nucleus. These interactions can be categorised into two types: elastic scattering ($A + B \rightarrow A + B$) and inelastic scattering ($A + B \rightarrow C + D$). Inelastic collisions at high energies involve the production of several particles in the final state, leading to what are referred to as 'hadronic showers'. These showers tend to be complex due to the inelastic production of secondary hadrons. The charged particles within these showers lose energy through ionisation and are absorbed by the large amount of material in the ECL and KLM, as shown in Figure 3.5. This interaction can be characterised by a parameter known as the 'absorption length', λ_0, which is the average distance over which the intensity of a particle beam decays to $1/e$. It is a property of the material of a detector through which the particles are passing and is typically given in units of centimetres.

Muons, on the other hand, do not participate in the strong interaction and hence rarely undergo collisions with atomic nuclei, resulting in a very long absorption length. In addition, muons also have a very long radiation length since they are 200 times heavier than electrons, and consequently only radiate a small fraction of their energy compared to electrons. Muons only lose energy via ionisation as they pass through the tracking detector, ECL and KLM. As a result, muons are highly penetrating charged particles that are distinguished from other charged particles by the depth with which they penetrate the KLM, as shown in Figure 3.5.

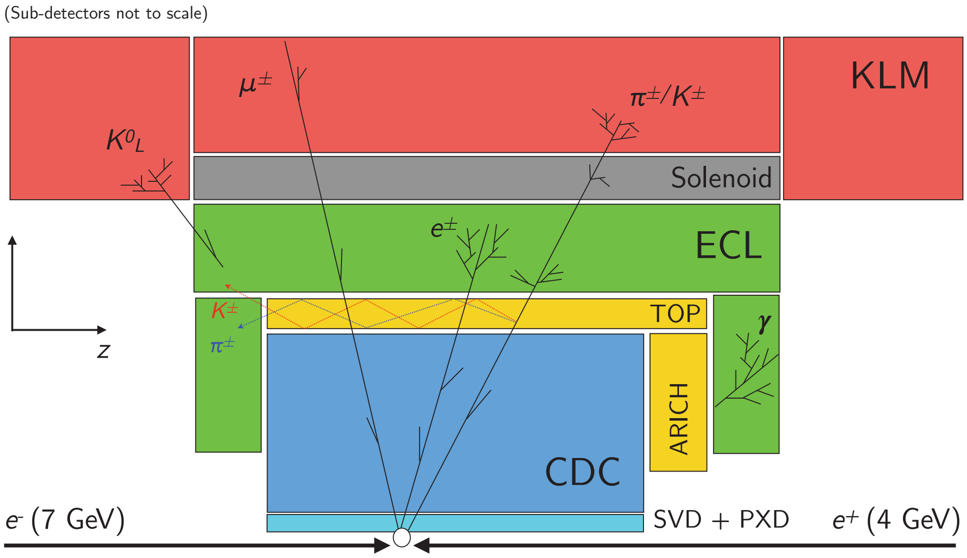


FIGURE 3.5: Interactions of different particles in the Belle II subdetectors.

3.3.2 Pixel Detector (PXD)

The PiXel Detector (PXD) is in the innermost subdetector and is designed to precisely measure the tracks of charged particles and reconstruct their decay vertex. As shown in the left-hand diagram of Figure 3.6, the full configuration of the PXD is composed of two concentric layers, with 8 ladders in the inner layer (L1) and 12 ladders in the outer layer (L2), located 14 and 22 mm from the IP, respectively. These ladders contain two-dimensional arrays of sensors based on Depleted P Channel Field Effect Transistor (DEPFET) technology, which are engineered to be extremely thin (75 micrometres on average or 0.2% X_0 per layer) to minimise multiple scattering that would degrade the precision of the vertex. The read-out electronics, which is located outside the detector acceptance, use application-specific integrated circuits (ASICs) directly attached to the sensor substrate at the ends of the sensor. Flexible electrical circuits are attached to the module's ends to carry data and power. To form a ladder, two modules are attached together with the readout circuits and other passive elements at both ends to support the entire structure. To cool the read-out electronics, liquid CO_2 circulates through stainless steel blocks on which the ladders are mounted.

The sensors themselves use very little power and require only air cooling. There are a total of 40 sensors (12.5×(L1 44.8, L2 61.44) mm^2) with pixels that are 50 by 55-85 μm. The sensors use a rolling shutter readout, where the image capture process proceeds from the top to the bottom, scanning one row of pixels at a time. After some exposure time, consecutive lines are read out and reset one after the other. The duration of a single image acquisition or one readout cycle is 20 μs. The full PXD configuration will contain a total of 3.072×10^6 pixels in the inner layer and 4.608×10^6 pixels in the outer layer, for a total of 7.7M readout channels. During the work conducted for this thesis, the inner layer was fully populated, but only two ladders in the outer layer were present.

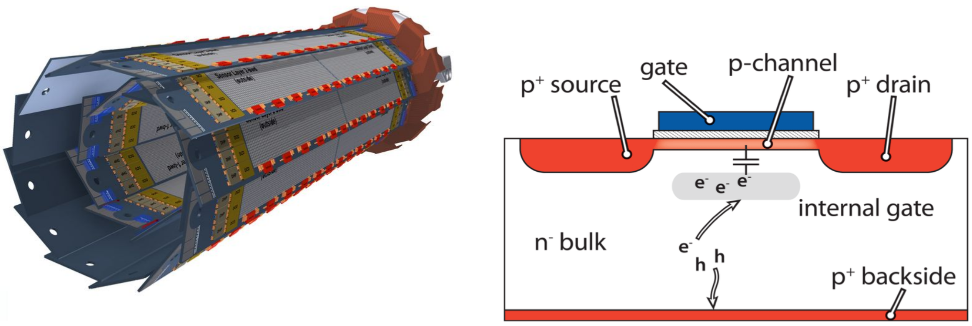


Figure 3.6: Model of the PXD (left) and diagram of a DEPFET pixel showing the principle of detection (right).

Compared to Belle, the beam pipe and the first two detector layers are closer to the IP and must handle a significantly higher dose of potentially damaging radiation. For this reason, pixel sensors are used instead of the silicon strips used in Belle. These pixel sensors have a larger number of channels, which results in the fraction of channels hit in each triggered event, or the occupancy, being smaller. In DEPFET-based sensors, the signal from a single particle hitting the pixel is amplified in the pixel itself. As shown in the right-hand diagram in Figure 3.6, DEPFET-based sensors are based on field-effect transistors (FETs), which are integrated directly into the n-type depleted silicon bulk of the pixel and serve as the readout electronics for the sensor. When a charged particle passes through the region with no free space charges (depleted region), it creates electron-hole pairs. The electrons will drift towards the ‘internal gate’, a phosphorus implantation that creates a minimum potential for the electrons and is located about 1 μm below the FET, while the holes will drift to the p-doped backside. The internal gate is capacitive coupled to the FET, which is biased with a constant voltage. Any accumulated charge at the internal gate modulates the current flowing through the FET, resulting in a voltage signal across the pixel that is used as a read-out signal. The output signal is amplified by the intrinsic gain of the transistor. Therefore, DEPFET pixels perform the amplification process directly at the particle interaction site, rather than through external readout electronics. The internal gain or current per electron collected is around 400 pA/e^-. This internal amplification is what allows the DEPFET sensors to be exceptionally thin while maintaining low noise, high sensitivity, and fast readout times. A matrix of DEPFET pixels forms an image sensor that can determine the tracks of a particle with an accuracy of about 10 μm. In addition, a n^+ contact is placed on the fringe of each pixel that resets the pixel by applying a sufficiently high voltage to force electrons from the internal gate to the contact.

3.3.3 Silicon Vertex Detector (SVD)

The silicon vertex detector (SVD) is the second sub-detector and is designed to work alongside the PXD to precisely measure and reconstruct the tracks and decay vertex of particles. It extrapolates tracks with high precision to the PXD, allowing for the correct identification of hits and improvement in the particle identification in the low-momentum region. The SVD is composed of four concentric layers of ladders (38 mm, 80 mm, 115 mm, and 140 mm from the IP) that contain Double-sided Silicon Strip Detectors (DSSD), as shown in the left-hand diagram of Figure 3.7. The strips are orientated perpendicular to each other, with the p-strip on top and the n-strip on bottom, to provide two-dimensional position information. The p-strips provide $r - \phi$ information, while the n-strips provide z information. The innermost layer of the SVD consists of seven ladders, each equipped with two small rectangular silicon sensors. As shown in the right-hand diagram of Figure 3.7, the second through fourth layer (layers 4, 5, and 6) are composed of two to four large rectangular sensors with trapezoidal sensors in the forward region angled towards the beam line to account for the forward boost. The rectangular and trapezoidal sensors have a thickness of 320 μm and 300 μm, respectively. The average material budget is 0.7% X_0 per layer. Strips measuring the longitudinal direction (z) have a pitch of 160 to 240 μm, while for the measurement in the transverse plane the strip pitch is between 50 and 75 μm, depending on the layer. The second, third, and fourth layers consist of 10, 12 and 16 ladders, each with 3, 4, and 5 sensors per ladder, respectively. There are a total of 224,000 strips contained in 172 sensors.

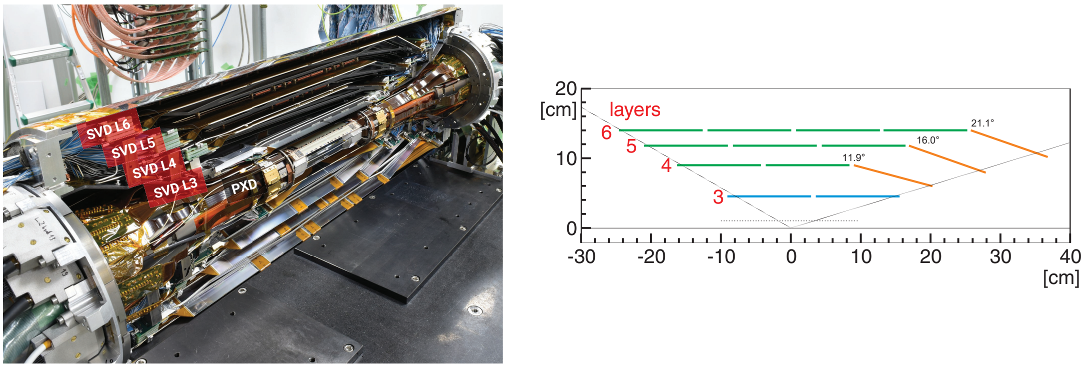


FIGURE 3.7: Model of the four layers of the SVD (left) and diagram of SVD layout (right).

The DSSD are based on reverse-biased p-n junction technology. In the depleted region of a p-n junction, a bias voltage is applied, creating an electric field that causes electrons and holes created by the passage of charged particles to be attracted to the oppositely charged strips. These strips are shown in the left-hand diagram of Figure 3.8. As the electron-hole pairs move in the electric field, their drift motion generates a current in the sensor. The closest strips collect the most current, while strips farther away collect progressively less. Charged particle tracks are reconstructed by measuring the small ionisation current that occurs as they pass through multiple layers of DSSDs. As shown in the right-hand diagram of Figure 3.8, the DSSD strip sensors are connected to APV25 ASICs which perform read-out and suppress the background with a fast shaping time of around 50 ns to keep occupancy low ($< 10\%$). The 'shaping time' is the time taken for the signal to be processed and amplified before being read out. The APV25 chips are partially thinned to 100 μm, folded around and bonded to the sensor with 10-12 readout chips per sensor. This allows the SVD to have a resolution of approximately 10-20 μm, depending on the layer. The analog data are guided by flex circuits, read out at the end of the ladder, and transmitted to flash analog/digital converter (FADC) boards, where they are digitised using FPGAs. The data are then sent from the FADC boards to optical converter boards that convert digital electrical signals into optical signals. Finally, the data are sent by optical fibres to the DAQ.

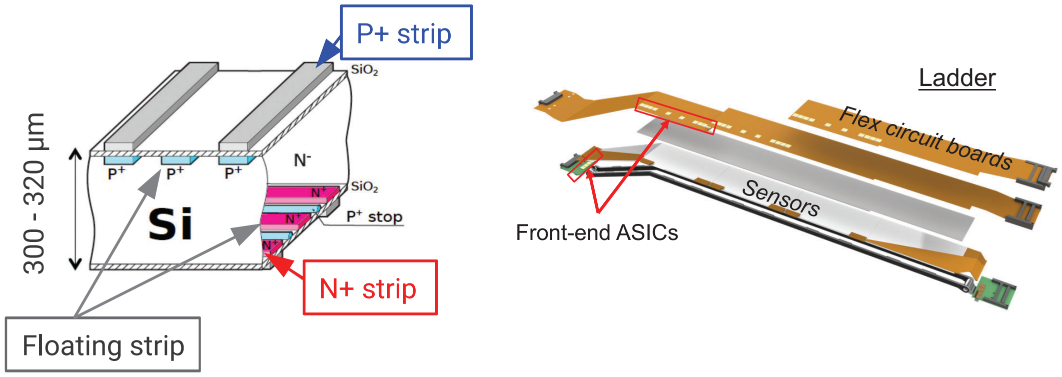


FIGURE 3.8: Model of the SVD strips (left) and ladders (right).

3.3.4 Central Drift Chamber (CDC)

The Central Drift Chamber (CDC) is the third subdetector and is designed to reconstruct charged tracks with precise momentum measurements, provide additional particle identification capabilities via ionisation energy-loss measurements, and serve as a reliable trigger signal for charged particles. The CDC comprises 56 concentric layers of 14,336 sense wires, extending from 160 to 1130 mm, and operates within a He(50%):C_2H_6(50%) gas mixture. The sense wires are thin metal wires that run parallel to the beam axis and are held at a high positive voltage (minimum of 2.2 kV) to attract electrons. When a charged particle passes through the gas, its gives up a small amount of kinetic energy (few keV/cm) creating a trail of gaseous ionisation (ions and electrons). These charged particles create an avalanche of electrons that drift towards the sense wires. The electronics attached to the sense wires measure the arrival times and positions of the drifting electrons. The drift times of the electrons are employed to interpolate between the locations of the sense wires, which enables a typical position resolution of approximately 100 microns for each sense wire. This enables the trajectory of the particles to be reconstructed. It also provides particle identification information by measuring the track ionisation energy loss energy (dE/dx) in the He(50%):C_2H_6(50%) gas which differs for different meson species and can be used to identify particles, as shown in Figure 3.9.

Surrounding the sense wires are the field wires, which are thicker metal wires that also run parallel to the beam axis. They are held at a lower negative voltage and are used to shape the electric field, which helps to guide the drifting electrons towards the sense wires. One sense wire surrounded by eight field wires forms a cell, as shown in the left-hand diagram of Figure 3.10. The cells are grouped every 6 or 8 layers into superlayers

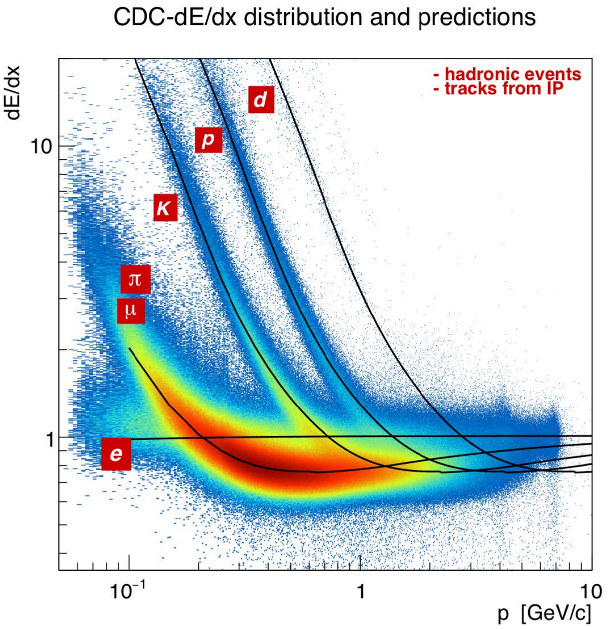


FIGURE 3.9: Energy loss for different particles as a function of momentum.

that alternate between axial and stereo angles, as shown in the right-hand diagram of Figure 3.10. The axial angles are aligned with the 1.5 T magnetic field generated by superconducting magnets, whereas the stereo angles are slightly skewed with respect to the axial wires to provide resolution along the z-axis. Information from the axial and stereo layers is combined to reconstruct the full three-dimensional helix track. The momentum of the particle can be determined based on the curvature of the track.

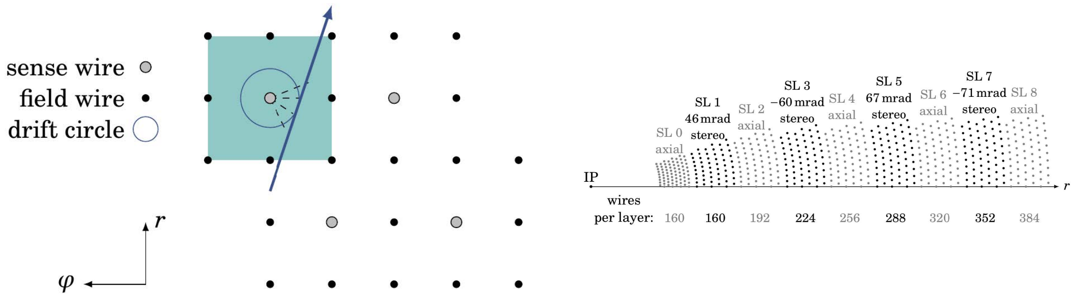


FIGURE 3.10: Layout of the CDC wires (left) and diagram of the working principle behind particle detection (right). The average drift velocity of the electrons is approximately 3.3 cm/μs with a maximum drift time of approximately 350 ns for a cell size of 17 mm.

3.3.5 Particle Identification (PID)

Two sub-detectors are primarily used for the identification of particles: the time-of-propagation counter (TOP) and the Aerogel Ring-Imaging CHerenkov counter (ARICH).

The TOP is located on the outer wall of the CDC in the barrel region ($32^\circ < \theta < 120^\circ$) while the ARICH is located in the front end cap ($17^\circ < \theta < 35^\circ$). The TOP detector is used in the barrel region, as there is not enough room to build an ARICH-like detector. Both sub-detectors measure the optical photons, known as Cherenkov radiation, that are emitted when charged particles traverse through materials at velocities faster than the speed of light in that particular medium. The electromagnetic 'shockwave' forms a cone around the particle's track with a characteristic angle that is inversely proportional to its velocity. Cherenkov radiation is emitted at an angle θ_C with respect to the direction of motion of the particle. The angle depends on the mass of the incoming particle and is given by

$$\cos(\theta_c) = \frac{1}{\beta n} \tag{3.3}$$

where n is the refractive index and β is the ratio of the velocity of the particle to the speed of light. Hence, particles with higher velocities have smaller angles, i.e. narrower cones. By measuring the Cherenkov angle and combining it with the momentum measurements in the CDC, a likelihood for each final-state particle hypothesis can be defined.

The TOP detector is made up of sixteen 45 cm wide, 2.6 m long and 2 cm thick quartz bars with a small expansion volume about 10 cm long at the sensor end of the bar. A TOP quartz bar is shown in the left-hand diagram of Figure 3.11. Quartz is chosen as the radiator material because of its high refractive index. The generated photons propagate through the bars as a result of total internal reflection, as shown in the right-hand diagram in Figure 3.11. A spherical mirror is glued to the forward end of the bar to reflect any photons back into the bar and towards the sensor plane. The other end of the bar has a quartz prism attached to it, which acts as an expansion volume for the Cherenkov photons that are captured, widening their geometric distribution on the sensor plane. At the exit window of the wedge, two rows of sixteen Micro-Channel Plate (MPC) Photomultiplier Tube (PMT) are attached to the prism to record both the position and arrival times of the photons with high precision.

To avoid overwhelming the DAQ system, the recorded waveforms are processed using fast waveform sampling FEE, and only the timing and pulse parameters of the observed photons are transferred out of the detector. 'Waveform sampling' refers to the process of sampling an electrical signal at regular intervals and converting each sample into a digital value. The signal from all 8192 MPC-PMT channels is readout using Subdetector

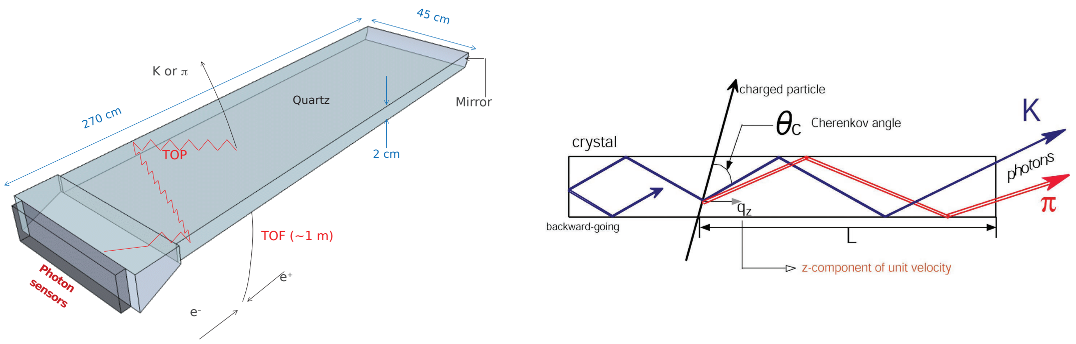


FIGURE 3.11: Model of one of the TOP quartz bars (left) and diagram of the photon path as it experiences total internal reflection inside the quartz bar (right).

Readout Modules (SRM). SRMs consist of four carrier boards and one Control Read-Out Data (SCROD) data aggregator board. There are four SRMs on every quartz bar, each of which reads eight MCP-PMTs. Each carrier board is equipped with four custom-designed 8-channel waveform sampling ASICs. These ASICs use a switched capacitor array, which is a series of capacitors that are selectively switched to sample signals at discrete time intervals and convert the acquired analog waveforms to a digital form. The resulting digitised waveform segments from all four carrier boards are then transferred to the attached SCROD board for processing. Here, the timing of photon pulses and other important pulse parameters are extracted from the waveforms, and the data are transferred to the DAQ system. The use of switched-capacitor array ASICs allows for high-speed and precise waveform sampling of the scintillation signals produced by the TOP detectors. The TOP has a single-photon timing resolution of less than 100 picoseconds and a spatial resolution of a few millimetres on the sensor plane.

The ARICH detector directly measures the circular pattern formed by emitted photons and provides good separation of pions and kaons from 0.4 to 4 GeV. The ARICH uses a non-homogeneous aerogel radiator with two different refractive indices for the 2 layers of aerogel: $n_1 = 1.045$ and $n_2 = 1.056$. The use of two layers increases the photon yield by focusing the Cherenkov rings onto the photon detector. The radius of this ring depends on the angle at which the Cerenkov radiation is emitted and thus on the velocity of the particle. A diagram of the ARICH detector is shown in the left-hand diagram of Figure 3.12, while the right-hand diagram illustrates the difference in photon rings formed by kaons and pions.

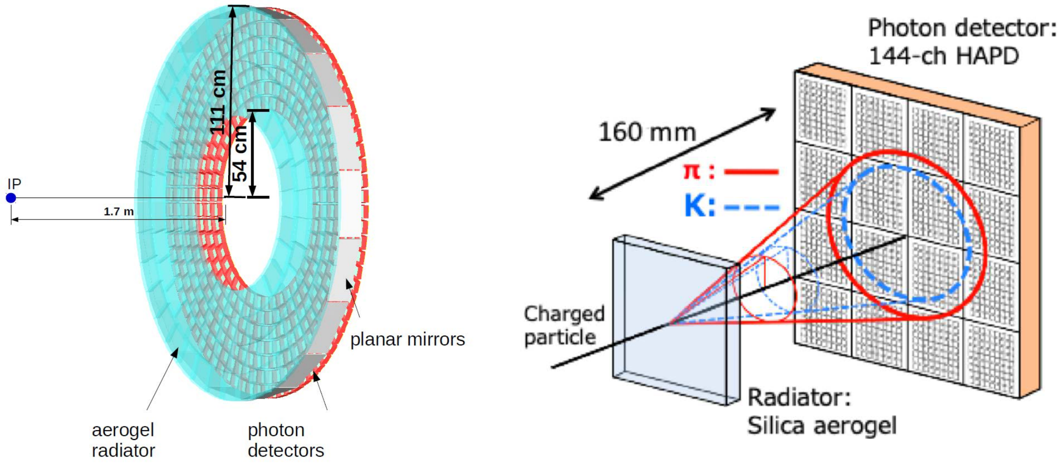


FIGURE 3.12: Model of the ARICH detector with its main components (left) and diagram of the difference in the photon path for Cherenkov photons from kaons and pions (right).

3.3.6 Electromagnetic Calorimeter (ECL)

The Electromagnetic Calorimeter (ECL) primary function is to detect and measure electromagnetic showers that are produced by high-energy photons, electrons and positrons. It is also used to assist in distinguishing electrons from hadrons such as pions. The ECL is the most relevant subdetector for this thesis, as it is the only one capable of detecting photons from π^0 decays. In fact, a third of the decay products from B mesons are neutral particles that decay into photons. The ECL starts at a radius of 1.25 m and extends to 1.62 m, with a 3 m long barrel section. The back and forward end caps are located at -1.02 and 1.96 m relative to the IP. The ECL is a segmented array of thallium-doped caesium iodide CsI(Tl) crystals. As shown in Figure 3.13, the crystals are a truncated pyramid with a front face area of 4.5×4.5 cm^2, a rear face area of 5×5 cm^2, and 30 cm long. Each crystal is wrapped with a layer of 200 μm thick Gore-Tex porous teflon and covered by a laminated sheet of 25 μm thick aluminium and 2μm thick mylar. There are 8736 crystals in total that cover all three detector regions (about 90% of the solid angle in the c.m. system) and point towards the IP. The ECL can precisely measure the photon energy and direction from a few tens of MeV to ≈ 7 GeV, and with an angular coverage that is summarised in Table 3.1.

When an electron or photon passes through one of the CsI(Tl) crystals, it initiates a particle shower in which Bremsstrahlung and pair production create a shower of electrons,

	θ Coverage	Number of crystals
Forward end-cap	12.4° - 31.4°	1152
Barrel	32.2° - 128.7°	6624
Backward end-cap	130.7° - 155.1°	960

TABLE 3.1: Description of the geometry of the ECL.

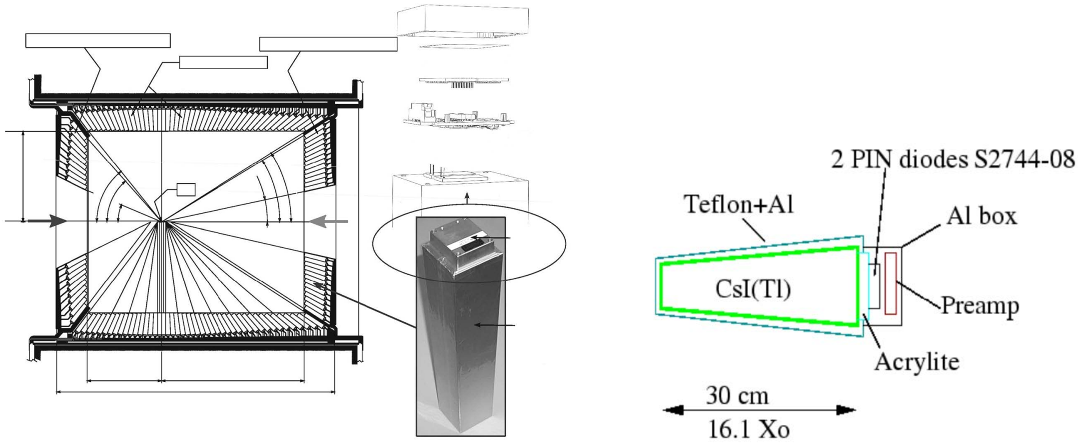


FIGURE 3.13: Schematic design of the electromagnetic calorimeter (left) and CsI(Tl) crystals with attached electronics (right).

positrons, and photons. The energy of these particles decreases exponentially until almost all of them stop as a result of ionisation loss. This energy excites atoms in the crystal, which then emit photons when they return to the ground state. The resolution of the calorimeter improves as the energy of the particle increases, making it possible to measure particles with energies up to several tens of GeV. In fact, CsI(Tl) is one of the brightest known scintillators, emitting approximately 50,000 photons per MeV and has a short radiation length, ($X_0 = 1.860$ cm), which ensures that the incoming photon is completely absorbed. Furthermore, the Molière radius, which is the radius of a cylinder that contains 90% of the energy deposited by the shower, is only 3.53 cm. This allows the crystals to be made smaller laterally, leading to improved spatial resolution and better separation of the particle showers. Lastly, the photon emission spectrum peaks at around 550 nm, which is ideal for photodiode readout. However, the time it takes for the light in the crystals to decay is relatively long, approximately 1 μs, which increases the overlapping of pulses from neighbouring (background) events. This means that scintillation light may be present when a charged particle from a later event arrives. The intrinsic energy resolution is approximately 7.7% at 0.1 GeV and 2.25% at 1 GeV [54].

The CsI(Tl) crystals, preamplifiers, and support structure were reused from Belle, while the readout electronics and reconstruction software were upgraded. The upgraded readout electronics produce full waveform information for each recorded hit, resulting in improvement to timing extraction, hit energy information, and subtraction of the off-time background. Additionally, the shaping time is decreased from 1.0 μs to 0.5 μs. To read out the signals produced by the scintillating crystals, the Belle II ECL detector uses two photodiodes attached at the back surface of the crystal. The signals from the photodiode are sent to two pre-amplifiers mounted on the rear of the crystal to integrate the signal emitted by each photodiode, as shown in Figure 3.14. The two raw signals after amplification from each crystal are sent to a readout board, called the 'ShaperDSP board' [55] (DSP stands for Digital Signal Processing), located outside the Belle II detector. The ShaperDSP contains 16 analog shaper circuits: slow (Shaper) and fast (Fast Shaper), 16 analog-to-digital converter (ADC) circuits, and an FPGA. These 16 channels of analog shaping and signal digitisation receive signals from up to 16 CsI(Tl) crystal. A total of 576 ShaperDSP modules are needed to process signals from all 8736 crystals.

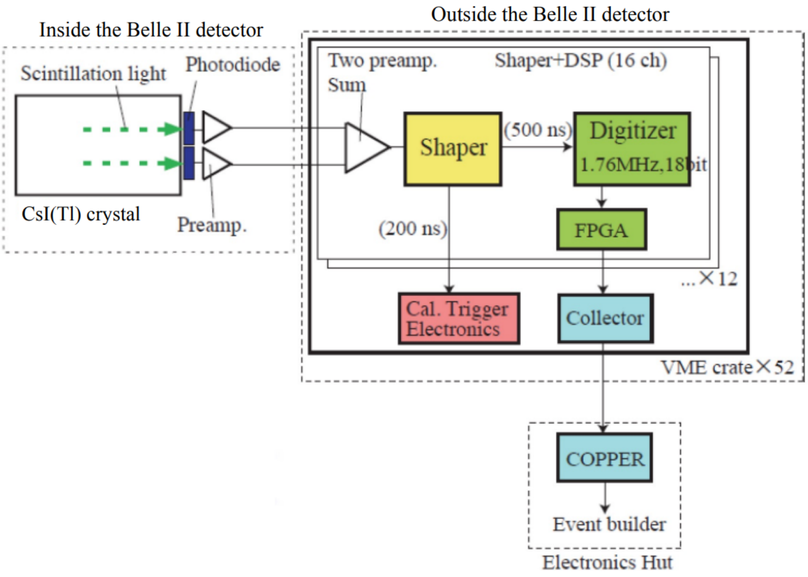


FIGURE 3.14: Schematic diagram of the data flow from scintillation light to the data acquisition system for a single crystal.

Shaper circuits amplifies and shapes the analog signal to provide a clean and well-defined pulse while removing noise, pile-up noise due to overlapping pulses from neighbouring events, and other unwanted features from the signal. The 'Fast Shaper' sends a signal to the trigger system with a frequency of about 200 ns. Meanwhile, the signals from the slow shaper are continuously sampled by an 18-bit precision ADC (digitiser) with a sampling frequency of 1.76 MHz, which corresponds to an interval between measurements of $T_{SR} = 567$ ns, where SR stands for Sampling Rate. After collecting 31 ADC samples, the waveform is processed by an FPGA using a photon template fit to compute the amplitude of the signal, the time relative to the trigger signal, and the χ^2 fit quality. An example of a fit is shown in Figure 3.15. This takes the form of a 32-bit word known as ECLDigit for each crystal that contains the parameters of the FPGA fit result: the 18-bit amplitude, 12-bit timing information, and 2-bit status information. If the ECLDigit amplitude is high enough, the waveforms are also stored for pulse-shape discrimination. The waveforms will also be randomly saved to form the beam background overlay in simulated data. The output data from 8-12 ShaperDSP boards are sent to a Collector module, which groups, packs, and transfers the data to the DAQ system.

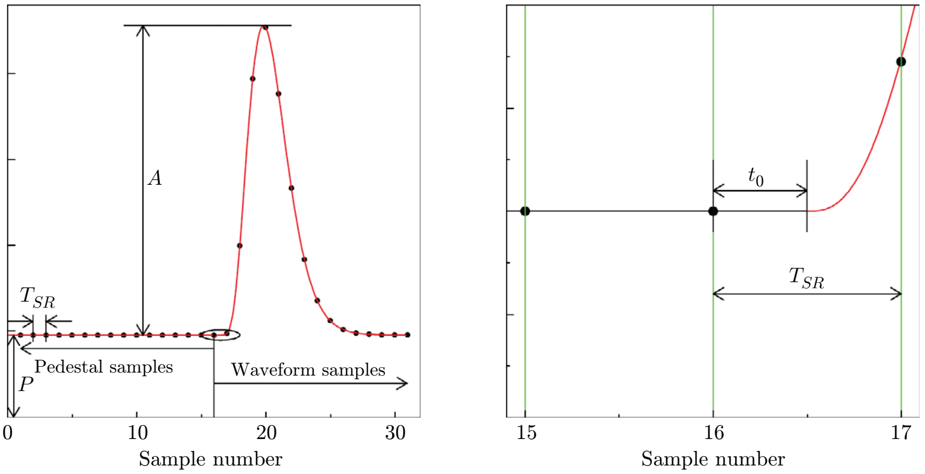


FIGURE 3.15: The first 16 samples (pedestal) contain information about the baseline value, while the remaining 15 samples contain the signal waveform (left). The signal starts the 16th sample and this time is referred to as t_0 (right) [12].

The signal waveforms can be analysed using pulse-shape discrimination (PSD) to improve particle identification, since the decay time of CsI(Tl) consists of more than one component. The fast component is around 0.6 μs while the slow component is around 3.5 μs. The fast component is associated with the scintillation response to electromagnetic (e.g. gamma ray) interactions, while the slow component is associated with the scintillation response to hadronic (e.g., proton or neutron) interactions. This is because the scintillation process in CsI(Tl) for electromagnetic interactions involves the excitation and de-excitation of Tl atoms, while the scintillation process for hadronic interactions involves the excitation and de-excitation of both Tl and Cs atoms, which leads to a longer decay time. The ratio between the intensity of these two decay components varies as a function of the ionising power of the absorbed particle. When analysing the shape of the scintillation pulse, PSD can help to distinguish between these different types of particles.

3.3.7 K_L and Muon Detector (KLM)

The K_L and muon detector (KLM) is positioned at the very edge of the experiment, outside the superconducting solenoid. Here, only the muons, K_L, and charged pions are likely to register a signal. The KLM is designed to detect these particles and assist in their identification on the basis of their interactions. K_L mesons are neutral strongly interacting particles that have a significant probability of leaving hadronic showers in the ECL (0.8 λ_0), and being completely absorbed in the thick iron plates of the KLM (3.9 λ_0), while leaving no tracks in the CDC. This characteristic allows K_L mesons to be identified. Muons, on the other hand, interact only electromagnetically and lose energy mostly via ionisation. Therefore, muons are identified as tracks that penetrate the detector layers and reach the KLM, leaving tracks in the CDC, depositing a small fraction of their energy in the ECL, and producing no hadronic showers.

The KLM consists of alternating layers of active particle parts and 4.7 cm thick iron plates, which serve as the magnetic flux return of the solenoid. The KLM is divided into two parts, the barrel (BKLM) aligned parallel to the beam and covers the polar range from 45° to 125°, and the end cap (EKLM), aligned normal to the beam and covers 20° to 155°. The barrel region has 15 detector layers and 14 iron layers, while each of the forward and backward end caps has 14 detector layers. A diagram of the KLM is shown in Figure 3.16.

The detection of charged particles in the BKLM is performed with 208 glass electrode resistive plate chambers (RPCs) and 32 scintillators panels with wavelength shifting fibers. To tolerate higher background rates such as by neutrons that are mostly produced in electromagnetic showers from background processes, the first two layers of the Belle II KLM are fitted with scintillators, positioned directly adjacent to the magnet, without any iron layer in front. The remaining layers are equipped with RPCs. RPCs are two high bulk resistivity parallel plate electrodes held at positive high voltage and separated by a gas-filled gap, as shown in the left-hand diagram of Figure 3.17. When a charged particle passes through the gas-filled gap, it ionises the gas and initiates a streamer, as shown in the right-hand diagram of Figure 3.17. A streamer is a cascade of free ions

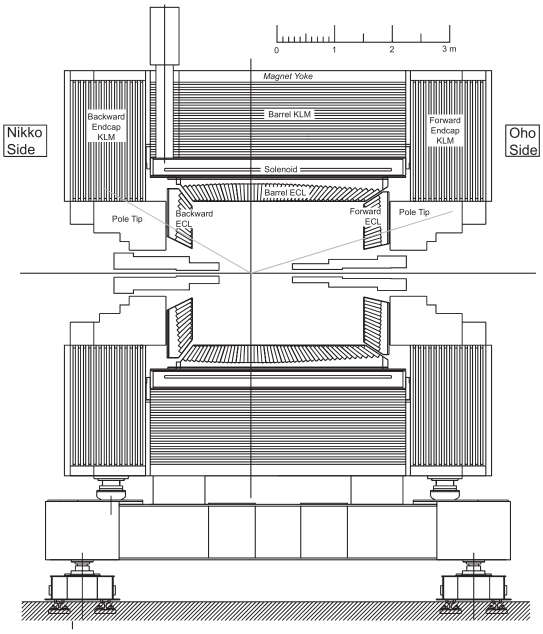


FIGURE 3.16: Side view of the KLM. The grey lines denote the polar angular acceptance of Belle II.

and electrons that are accelerated by an electric potential and further ionise the gas, leading to a large avalanche of charge carriers. The charge carriers then move towards the electrodes, resulting in a local discharge of the plates and inducing a signal on the read-out strips. This signal is strong enough to be measured without amplification. The read-out process involves the use of strips at both ends of the RPC. A signal initiated in one chamber produces a mirror charge in the other chamber, which increases the detection efficiency to 99% and provides redundancy.

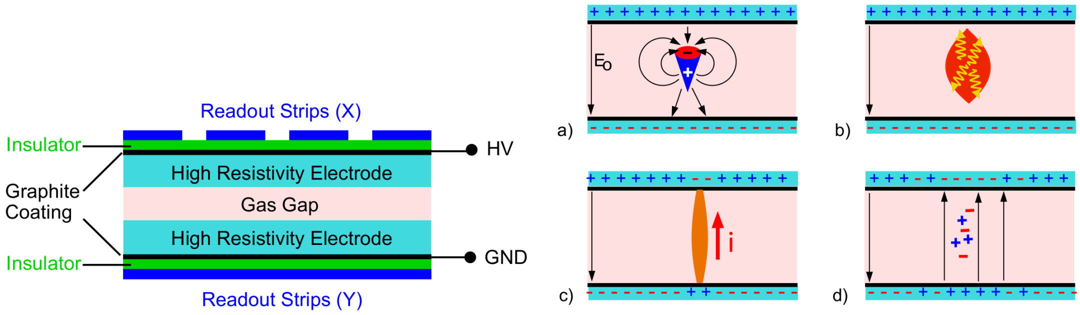


FIGURE 3.17: Schematic sketch of an RPC (left) and the development of a streamer inside an RPC, from ionisation in gas to detection by the readout strips (right).

In the EKLM regions, 104 scintillator panels are used instead of RPCs due to the high level of background expected at design luminosity. The strips are between 885 mm and 2820 mm in length and have a groove in the center to accommodate a fiber, as shown in the left-hand diagram of Figure 3.18. The fiber's purpose is to carry the scintillation light to the photodetectors and prevent re-absorption of the generated photons in the material of the detector. It is also mirrored at one end to increase the total light yield. The scintillation light is created by charged particles exciting electrons in the crystal into higher energy bands, as shown in the right-hand diagram of Figure 3.18. The electrons then relocate themselves into impurities, which undergo quenching and emit scintillation light. The scintillation light is collected by a fiber and brought to silicon photomultipliers (SiPM). SiPMs are used instead of PMTs because of limited space and the strong magnetic field in the KLM endcap region. The fibers are made of a material that absorbs scintillation light and shifts the light from blue to green wavelengths to improve the sensitivity and efficiency of the SiPMs.

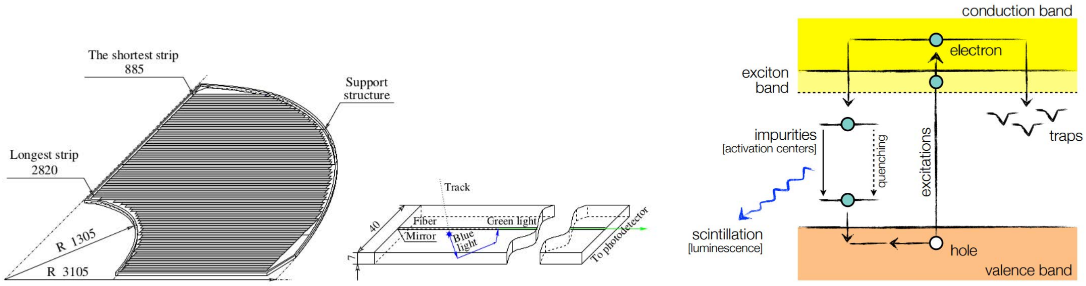


Figure 3.18: Schematic sketch of one EKLM strip layer and one scintillation strip (left) and a sketch of the scintillation process (right).

3.3.8 Track reconstruction

Track reconstruction is the process of determining the trajectories and properties of charged particles produced in e^+e^- collisions. The track reconstruction process involves two main steps: track finding in the Vertex Detectors (VXD) and the Central Drift Chamber (CDC), and track fitting [56]. In the track finding step, sets of hits that are close to each other in space and time are grouped together based on the assumption that they originated from the same particle. This process is challenging due to the presence of a large amount of background hits from other particles and beam background. Moreover, the charged particles may undergo multiple scattering, energy loss, and can even loop several times in the CDC producing hundreds of hits, making it difficult to accurately identify the hits that belong to the same track. Pattern recognition algorithms are employed to identify the hits that belong to a particular track and to reject hits that do not belong to any track. For example, the VXD Track finder algorithm uses a cellular automation model, where each hit is assigned to a track segment, which is defined as a straight line between two adjacent hits. The algorithm then searches for possible extensions of the track segment, and appends additional hits if specific geometric criteria are satisfied. These requirements include a maximum distance and angle between hits, as well as a maximum curvature for the track. The process is repeated until no more hits can be added to the track, or until a maximum number of hits is reached. In the track fitting step, a curve is fitted to these hits to estimate the set of parameters that describe the track, such as its momentum and trajectory.

3.3.9 Luminosity Monitor

Measurements of the instantaneous and integrated luminosity are essential to track the performance of the accelerator and to determine the data-taking efficiency of the detector. One method to determine the luminosity is to measure the rate of a specific process with a well-known cross section. At Belle II, this is achieved using large-angle $e^+e^- \to e^+e^-$ Bhabha scattering and $e^+e^- \to \gamma\gamma$ two-photon annihilation processes [57]. The procedure involves searching for large isolated energy depositions in the forward ($E > 4.5\,\text{GeV}$) and backward ($E > 1.5\,\text{GeV}$) endcaps of the ECL with a back-to-back topology. The number of events fulfilling this requirement is determined and then compared to the expected number based on MC simulation to estimate the luminosity. The luminosity, L, is estimated using the following formula.

$$L = \frac{\text{Count rate}}{\sigma_{\text{vis}}} \tag{3.4}$$

where σ_{vis} is the visible cross section for both the $e^+e^- \to e^+e^-$ and $e^+e^- \to \gamma\gamma$ processes. σ_{vis} is determined from MC simulations as,

$$\sigma_{\text{vis}} = \frac{N_{\text{det}}}{N_{\text{gen}}}\sigma \tag{3.5}$$

where N_{gen} is the number of generated events, N_{det} is the number of events counted as Bhabha events, and σ is the total cross section for a particular process within a given detector geometry. The computed visible cross sections are $\sigma_{ee}^{vis} = 28.095 \pm 0.044$ nb for the $e^+e^- \to e^+e^-$ events and $\sigma_{\gamma\gamma}^{vis} = 0.926 \pm 0.002$ nb for the $e^+e^- \to \gamma\gamma$ events.

3.4 Data Flow from Detector to Storage

The Belle II Data Acquisition (DAQ) system [58] is used to acquire and process data from subdetectors. It is a synchronous system based on a pipelined trigger flow control, meaning that it operates in a coordinated and synchronised manner to ensure that data are collected, passed through a series of processing elements in a predetermined order, and processed in real-time. The pipeline structure enables the DAQ system to efficiently process large amounts of data and identify interesting events in real-time, while also maintaining synchronisation and coordination between the various stages of the processing pipeline. The DAQ system consists of several components, including:

- Front-end electronics: responsible for converting the signals from the detectors into digital signals that can be processed.
- Trigger system: responsible for identifying and selecting events of interest.
- Data acquisition software: responsible for collecting, processing, and storing the data from the detectors and for transmitting the data to the event builder for further processing.
- Event builder: responsible for merging the data from the various subdetectors into a single event and preparing the data for storage.
- Data storage system: responsible for storing raw data and processed events.

The front-end electronic (FEE) readout systems installed near each subdetector are custom designed to process and digitise the distinct raw data of each subdetector. These FEEs are necessary because the raw data rate would overwhelm the transfer capabilities of the DAQ system. As a result, the data must first be processed within the front-end of the detector, so only the relevant information is transmitted from the detector.

The Belle II DAQ is composed of two levels, a hardware-based Level 1 (L1) trigger and a software-based High Level Trigger (HLT), that are used to differentiate between background events and events of interest to the Belle II physics program. These triggers are necessary due to high background rates and to satisfy the limitations of the data transfer, computing, and storage systems.

The L1 Trigger [59] gathers information from the detectors by receiving data from various front-end electronics that are directly connected to the CDC, ECL, TOP, and KLM subdetectors. These front-end electronics digitise and send the signals from the detectors to the L1 Trigger system. The L1 Trigger uses specialised algorithms running in FPGAs to quickly combine raw and unprocessed data from the CDC, ECL, TOP, and KLM subdetectors. The algorithm uses the two- and three- dimensional charged track information from the CDC, the total energy and cluster information based on trigger towers (4x4 set of crystals) from the ECL, hit topology and precise timing from the TOP, and muons identified in KLM. The information from each sub-detector is then sent to the Global Reconstruction Logic (GRL) where a low level reconstruction is performed, e.g. matching between tracks found in the CDC and clusters found by the ECL. Finally, all trigger information is sent to the Global Decision Logic (GDL) which determines if an event is to be accepted. An overview of the L1 Trigger is summarised is shown in Figure 3.19. The time it takes for an event to pass through the L1 trigger is less than 4.4 μs, and the trigger output rate is 20 kHz at the design luminosity. The L1 trigger provides an efficiency larger than 99% for B-meson events, and significantly reduces two-photon and Bhabha events. If a Bhabha event is detected by the Extreme Forward Calorimeter, it is stored for luminosity measurements.

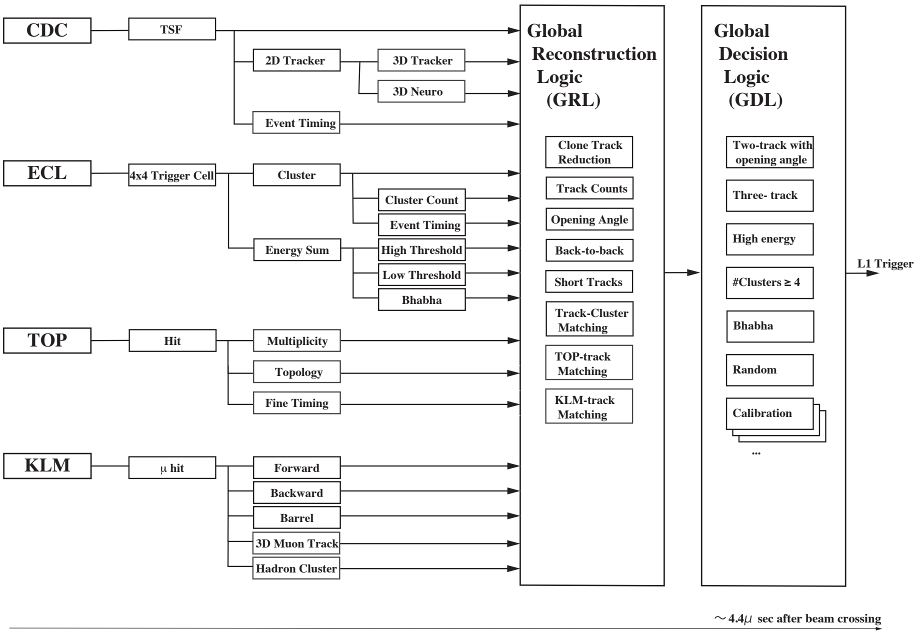


FIGURE 3.19: Overview of the Belle II L1 Trigger system. The CDC, ECL, TOP, and KLM sub-trigger system outputs are sent to the Global Decision Logic, where the final L1 Trigger decision is made.

Once the GDL accepts an event, the DAQ system is engaged. Data from all subdetectors, except for the PXD, are digitised by front-end electronics, and then collected, merged, and transferred through a unified readout and high-speed data transmission system using custom high-speed optical fibres (Belle2LINK) [60] to high-speed data receiver boards on COPPER modules (COmmon Pipeline Platform for Electronics Readout). Belle II's high trigger rate is transformed into a massive amount of data, about 1 MB per event, through this link. Each COPPER board receives and collects data from the various subdetectors and forwards it to readout PCs via 1 GBit Ethernet. Each readout PC can receive data from up to 16 COPPER boards. At the readout PC, the data fragments are combined by 'event builder 0'. This data is then sent via 10 Gbit connections to an input server, called 'event builder 1', where a more comprehensive event building process is performed, i.e. all the event fragments from the readout PCs are combined together. This process involves data from all subdetectors except the PXD. The combined data is forwarded into the High Level Trigger (HLT) to perform event reconstruction and generate software triggers. The HLT will suppress the event rate to 15 kHz using information from the CDC track finding and ECL reconstruction, i.e. rejecting beam background. It does this by analysing the features of each event, such as the properties and trajectories of the detected particles and selecting the events that are most likely to result from interesting physics processes. The event rate is then further reduced to 10 kHz using full reconstruction information. To process the PXD data, an FPGA-based readout system (ONSEN) is used.

Being close to the IP, the PXD is exposed to many background events, primarily two-photon processes. This results in about 100,000 fired pixels per read-out cycle, with only a small fraction being signal hits from events of interest. To extract these signals from the background noise, it is necessary to filter the PXD data, which otherwise exceeds 20 GB/s at 3% occupancy, more than 10 times the combined output rate of all other subdetectors. If all the PXD information was kept, the data output rate would quickly overwhelm the Belle II event builder and data transfer system. Hence, the Online Selection Nodes (ONSEN) [61], a FPGA-based online data reduction system, was designed to filter events in real time and reduce the data-taking rate. It takes the digitised data of PXD collected by the Data Handling Hybrid (DHH) and DHH Concentrator (DHHC) modules and keeps the entire output data from the pixel detector in local buffer for up to 5 seconds. During this time, the HLT computing farm (6400 cores) performs an

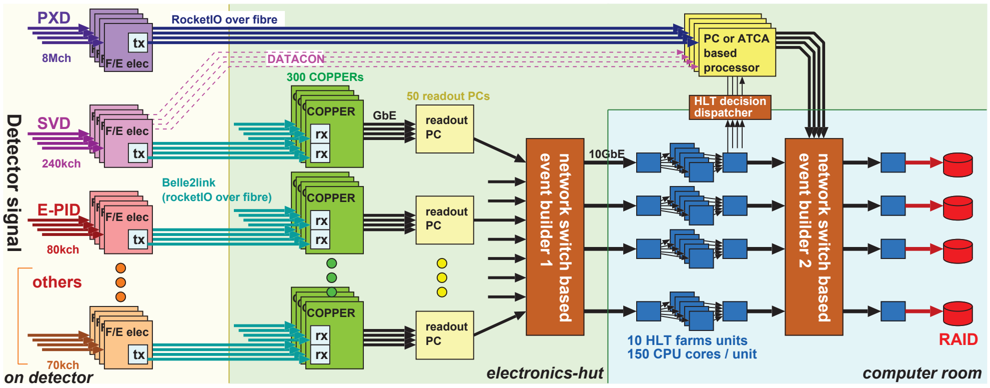


FIGURE 3.20: Overview of the Belle II Data Acquisition system from Front-End Electronic detector readout to data storage.

online event reconstruction and data reduction using data from the other Belle II subdetectors. It exploits expected hit positions of charged tracks reconstructed by the HLT, extrapolates reconstructed tracks to the layers of the pixel detector, and defines regions of interest (ROI). In addition, SVD-based online track reconstruction is also performed by an FPGA-based Data Acquisition Tracking and Concentrator Online Node (DATCON) [62] to determine the ROI. Using two redundant ROIs, ONSEN perform an overall PXD data reduction and only transfers pixels inside the two ROIs (data reduction by factor of 10) to event builder 2 which merges the PXD data with the HLT output. Finally, the full-event data go into storage using a Redundant Array of Inexpensive Disks (RAID), which are hard drives connected together, so that the same data is stored in multiple places to prevent data loss in the event of drive failure. Despite the reduction in events by ONSEN the PXD still accounts for approximately 75% of total Belle II raw data size. The movement of data from the front-end electronics to storage is summarised in Figure 3.20.

Chapter 4

Analysis of $B^0 \to \pi^0\pi^0$

In order to measure the branching fraction and $\mathcal{A}_{CP}$ of any B decay, a standard Belle II analysis begins by employing simulated signal and background data produced by Monte Carlo (MC) event generators to determine optimal selections for experimental data. MC refers to the use of repeated random sampling to obtain numerical results that model the probabilistic nature of particle decays. They are used when a problem is too complex to solve directly. Instead, it is modelled as a random process and the algorithm repeatedly samples this process to estimate the solution. Event generators use computational algorithms that use theoretical predictions and prior experimental results to model the distributions of important quantities such as energy, momentum, event shape, etc. for the signal and background. Two key software packages, EvtGen [63] and PYTHIA [64], serve as MC event generators.

EvtGen simulates the decays of heavy-flavour particles, primarily B and D mesons. It encompasses a variety of decay models for intermediate and final states that include scalar, vector, and tensor mesons or resonances, as well as leptons, photons, and baryons. Meanwhile, PYTHIA is used for generating high-energy physics collision events between particles such as electrons, protons, photons, and heavy nuclei. Finally, GEANT4 [65] is used to simulate the passage of particles through the detector and the interaction of particles with the material.

The simulated data are processed using the same digitisation algorithms and detector calibration constants as those applied to real data. Reconstructed particle tracks, particle identification (PID) likelihoods, and MC information are stored in a mini data summary table (mDST), which is then skimmed using standard selections to create a user data summary table which includes pre-built composite particles like π^0 mesons and K_S^0 (μDST). Selection cuts can be applied to the μDST and the variables of interest are created. The code used in this analysis chain is then applied to experimental data from the Belle II detector and the results are compared. This general scheme is visualised in Figure 4.1 and Figure 4.2.

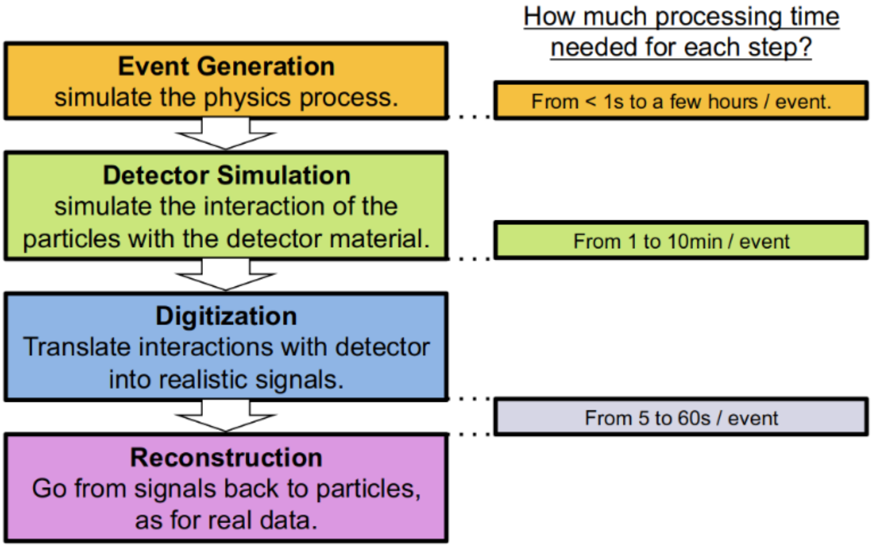


FIGURE 4.1: Production of simulated data, from event generation to reconstruction, along with the typical time requirement for each step.

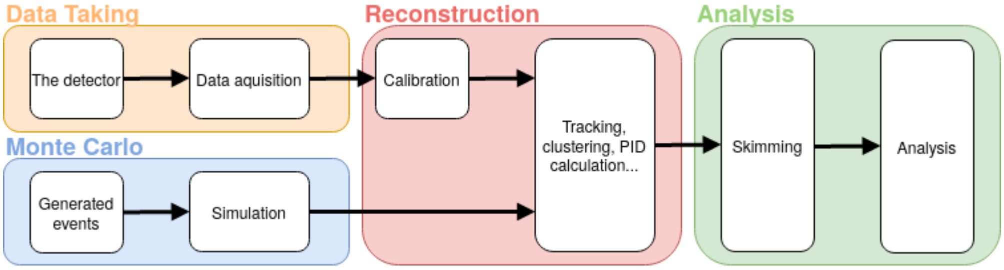


FIGURE 4.2: The overall scheme of a typical Belle II analysis.

4.1 Belle II Analysis Software Framework

Belle II Analysis Software Framework (basf2) [66] is a comprehensive data handling and analysis framework written in `C++` and `Python3`. The framework consists of packages that contain libraries, modules, and data objects used for data processing, event reconstruction, simulation, and analysis. Modules, which are constructed on the foundations of libraries, leverage the functionality that these libraries provide. Each module performs self-contained tasks on the data and operates on an event-by-event basis. A sequence of these modules is known as a 'path'. Data objects that contain shared information between modules are stored in a common DataStore, alongside their relationships. The user provides a Python steering file that, when executed, constructs a path populated with modules. This chain of modules is then applied to all specified events, as depicted in Figure 4.3. The design of basf2 emphasises flexibility and modularity, enabling users to incorporate their own modules and libraries to develop and customise their analysis codes and workflows.

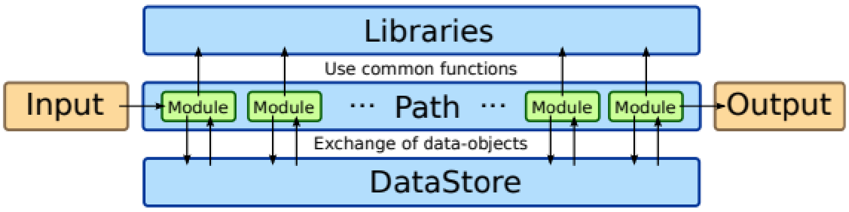


FIGURE 4.3: Schematic diagram of the general execution of modules chained into a path via a Python steering file. The modules operate event-by-event and use functionality provided by libraries and exchange data-objects via the common DataStore [13]

4.2 The Challenge

The $B^0 \to \pi^0\pi^0$ decay mode has an extremely low branching fraction, approximately on the order of 10^{-6}. In fact, even if every signal decay is reconstructed, one expects only approximately 1.72 ± 0.28 signal events per fb^{-1}. In addition, the $\pi^0 \to \gamma\gamma$ mode has a branching fraction of $98.823 \pm 0.034\%$ [26], meaning that the final state consists almost exclusively of photons. This has several implications:

- The final state is electrically neutral, and so there are no signal-side charged tracks and hence no vertex reconstruction, making this analysis time-independent.
- As many false photons as possible must be excluded, which will be difficult given that Belle II has a significantly higher beam background compared to Belle.
- The background from continuum will be substantial, necessitating excellent background rejection and signal retention for a competitive result.
- A comprehensive understanding of the performance of the electromagnetic calorimeter is required for high-precision measurement of the π^0.

Belle II is currently the only active experiment capable of competitively studying this decay mode due to the aforementioned challenges. The data sample employed for this analysis corresponds to an integrated luminosity of 189.9 fb^{-1} at the $\Upsilon(4S)$ resonance and represents all the data collected until July 1^{st}, 2021. The analysis is performed in a blinded manner, meaning that prior to applying the reconstruction procedure to experimental data, it is validated using simulated data and the $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ control mode. In the control mode, experimental and simulated data are compared in terms of signal yields, background levels, and relevant distributions. Any discrepancies found are taken into account or corrected. Only once the reconstruction and analysis procedures are validated can the analysis be applied to experimental data.

It should be noted that although there are three primary ways in which the $B^0 \to \pi^0\pi^0$ can decay:

1. $B^0 \to \pi^0(\to \gamma\gamma)\,\pi^0(\to \gamma\gamma)$
2. $B^0 \to \pi^0(\to e^+e^-\gamma)\,\pi^0(\to \gamma\gamma)$
3. $B^0 \to \pi^0(\to \gamma(\to e^+e^-)\gamma)\,\pi^0(\to \gamma\gamma)$

In this analysis, only the first decay is considered. The second decay occurs with a branching fraction of only $1.174 \pm 0.035\%$ [26] while the third decay mostly occurs when a photon interacts with an element of the detector and converts into an electron-positron pair. Both decays are not used in this analysis, as they are difficult to reconstruct and contribute a large amount of background for relatively small signal yield. Furthermore, the expected purity for the second and third decays is approximately 60%, while the first decay has a purity of approximately 95%. In subsequent discussion, all mention of the signal decays refers to the $B^0 \to \pi^0(\to \gamma\gamma)\,\pi^0(\to \gamma\gamma)$ mode.

4.2.1 Source of background

The main experimental challenge when studying the $B^0 \to \pi^0\pi^0$ is the reduction of the significant background from continuum events, beam background, and cross-feed. The 'continuum background' occurs when a pair of light quarks $q\bar{q}$ ($c\bar{c}$, $u\bar{u}$, $d\bar{d}$, $s\bar{s}$) is produced in place of an Υ(4S). These quarks are produced with high momentum and are initially moving almost directly away from each other due to the conservation of momentum. However, as the quark and antiquark from the $q\bar{q}$ pair move apart, the strong force between them increases. Eventually, the energy in the field becomes so great that it becomes energetically favourable to create a new quark-antiquark pair from the vacuum. This process is known as hadronisation and results in a spray or 'jet' of hadrons (protons, neutrons, pions, etc.) moving in roughly the same direction as the original quark. At energies close to the mass Υ(4S), the cross section of $e^-e^+ \to q\bar{q}$ is three times larger than $e^-e^+ \to \Upsilon$(4S). Fortunately, the continuum background can be suppressed by using event shape variables in the c.m. frame. Event shape variables quantify the geometric properties of the final state in a particle collision. They can be used to help distinguish between spherically symmetric $B\bar{B}$ and jet-like continuum events.

Another source of background is the electromagnetic interaction of the beam with the detector material. This 'beam background' is primarily the result of beam particles being lost due to the Coloumb and bremsstrahlung scattering with residual gas molecules in the beam pipe, and scattering with other beam particles in the same bunch. The latter is known as the 'Touschek scattering' and occurs when two electrons in a bunch elastically scatter and transfer momentum from the transverse to longitudinal direction, causing them to be lost. Touschek-scattered particles are subsequently lost at the beam-pipe inner wall after propagating further around the ring. If their loss position is close to the detector, they can generate showers which reach the detector, generating false hits. Heavy-metal shields in VXD volume, superconducting final focus cryostat, and horizontal and vertical movable collimators are used to prevent shower particles from entering the Belle II detector's acceptance volume. The beam background is particularly important for this analysis since the ECL energy resolution is expected to degrade with increasing beam background, as shown in Figure 4.4.

The last background to consider is known as cross-feed. The B meson that decays into two π^0 mesons in a $\Upsilon(4S) \to B^0\bar{B}^0$ decay is called the 'signal'. Cross feed occurs when a photon from the non-signal B meson is mistakenly included in the signal reconstruction. Fortunately, this is a relatively rare occurrence, as only about 1% of B^0 decays result in a π^0 as a direct daughter particle.

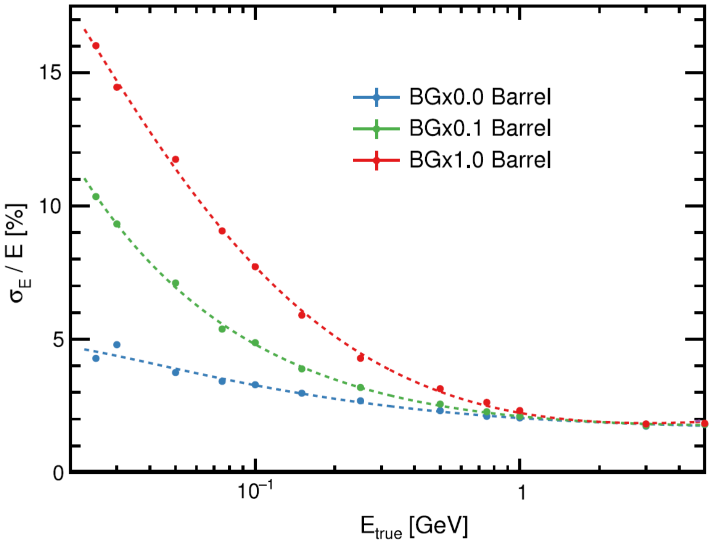


FIGURE 4.4: Belle II ECL energy resolution, simulated data with nominal (red), 0.1 times (green), and without (blue) beam background.

4.3 Data sets

Simulated samples were used to optimise event selection criteria, compare the distributions observed in the data with expectations, determine fit models, calculate signal efficiencies, and study background sources. For signal studies, 1×10^7 $\Upsilon(4S) \to B^0\bar{B}^0$ decays were generated using EvtGen [63]. In these decays, one B meson decayed into $B^0 \to \pi^0\pi^0$, while the other meson decayed generically. Here, generically means that the specific decay mode of the other meson was not specified. Instead, the probability of possible decay modes is determined by a decay table. This table lists all possible decay processes, their corresponding branching fractions, and their associated amplitude models.

Final-state radiation is a type of radiative emission that occurs when particles emit electromagnetic radiation as they fly apart from their originating B meson. This process is simulated with PHOTOS [67]. All event generators use the same beam parameters supplied by a central database. The interactions of particles with the detector material and the magnetic field, according to the final detector geometry, are simulated with GEANT4 [65]. The responses from the simulated sub-detectors are then stored in the same format as real experimental data, with additional 'truth' information provided by EvtGen. This truth information is the true properties (energy, momentum, direction, decay chain, etc.) of the generated particle. The beam background is simulated by overlaying random 'hits' in the detector on top of the events.

The simulation and processing of continuum, tau pairs ($\tau^+\tau^-$), and generic $B\overline{B}$ decays involve handling enormously large files, typically in the range of hundreds of terabytes. These tasks are performed centrally, and the data are subsequently distributed to users for access. The MC samples used for these decays are updated annually with the latest `basf2` software, calibration constants, and beam backgrounds. A simulated sample consisting of $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ processes and continuum $e^+e^- \to q\bar{q}$ background was generated with EvtGen and PYTHIA [64], where q denotes a u, d, s, or c quark. The $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ samples can be subdivided into B^+B^- (charged) and $B^0\bar{B}^0$ (mixed) decays. A sample of $e^+e^- \to \tau^+\tau^-$ events, generated with KKMC [68] and TAUOLA [69] and having the same size as the continuum sample was also used. This sample was employed to account for a large observed $\tau^+\tau^-$ production, which, while very interesting in its own right, is a background for this analysis. These samples are approximately five times larger than the data sample and correspond to 1 ab^{-1}. These samples are produced independently of any specific data-taking (or run) period, using background and calibration constants that represent typical experimental conditions. This sample is referred to in the analysis as MC14ri_a. In addition, 189.9 fb^{-1} of run-dependent MC are also employed. This MC sample, referred to in this analysis as MC14rd_a, uses the same calibration constants as those applied to the actual data and incorporates real background events that were randomly collected. This data set is used to provide a more accurate representation of the experimental conditions.

To validate the analysis, $B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$ decay is used as a control mode. It is useful as a control mode since it contains two π^0 particles in the final state and has an order of magnitude higher yield. The control mode is used to validate the procedure for event reconstruction, selection, and fitting. Moreover, it provides a means to estimate systematic uncertainties and check for any potential data-MC discrepancies. For this purpose, a simulated sample of 5×10^6 control mode events was generated with EvtGen.

To study photon selections and potential discrepancies between data and MC, the $D^+ \to D^0(\to K^0_S(\to \pi^+\pi^-)\pi^0)\pi^+$ calibration mode was used. The D^+ originates from the hadroniation of charm quarks and provides a large experimental data set suitable for validating photon selections and calibrating photon energies. Moreover, the π^0 covers a broad momentum range, enabling the study of photons across all energy scales. For this purpose, a simulated sample of 1×10^7 $D^+ \to \bar{D}^0(\to K_S^0(\to \pi^+\pi^-)\pi^0)\pi^+$ was generated using EvtGen. The π^0 momentum distribution of the signal, control and calibration mode is shown in Figure 4.5.

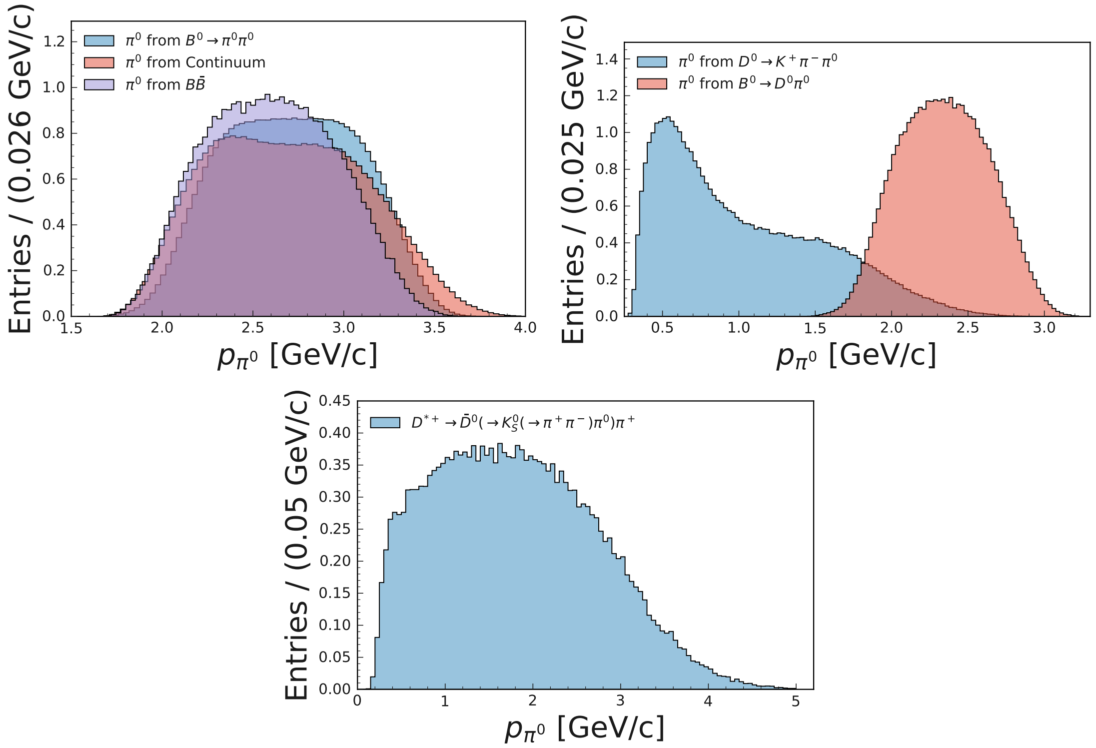


FIGURE 4.5: π^0 momentum of the $B^0 \to \pi^0\pi^0$ signal, continuum and $B\bar{B}$ components (top left), $B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$ control mode (top right) and $D^+ \to (\to K_S^0(\to \pi^+\pi^-)\pi^0)\pi^+$ calibration mode (bottom).

4.4 Skim Selections

The data sets described above are massive, with sizes typically ranging from tens to hundreds of terabytes. However, the vast majority of events are not relevant to the $B^0 \to \pi^0\pi^0$ decay. To reduce the volume of data requiring processing and storage, a procedure known as 'skimming' is employed. It is the process of selecting a subset of events from a large data set, based on certain criteria or requirements, such as having specific particles or triggers. These selections are chosen to maximise the signal-to-background ratio and are generally broad to retain as many signal events as possible, while minimising the amount of background events.

For an event to be included in the skim, it must pass at least one of the HLT triggers (see section 3.4) which are designed to remove uninteresting physics events. To remove the high number of low-energy photon hits in the ECL detector, a minimum energy requirement for photons is imposed. In the backward and barrel region, the energy is required to be greater than 20 MeV ($E > 20$ MeV). Meanwhile, in the forward region, the energy is required to be greater than 22.5 MeV ($E > 22.5$ MeV). The slight asymmetry in the forward region is due to a higher occurrence of forward-directed beam background from the electron beam.

Photons that meet these selection criteria are paired to form π^0 candidates. The diphoton mass is required to be between 0.105 GeV/c^2 and 0.150 GeV/c^2 ($M \in (105, 150)$ MeV/c^2), which corresponds to a range of approximately $+2.0\sigma$ and -2.5σ about the known π^0 mass. The mass requirement is asymmetric, as the reconstructed π^0 mass has a slight negative skew due to energy leakage from the ECL calorimeter. B^0 candidates are reconstructed by combining two π^0 candidates.

To select signal B^0 candidates, two kinematic variables are defined,

$$M_{bc} = \sqrt{E_{\text{beam}}^2 - |\vec{p}_B|^2} \quad \text{and} \quad \Delta E = E_B - E_{\text{beam}}, \tag{4.1}$$

where E_{beam} is the beam energy and $(E_B, \vec{p}_B)$ is the reconstructed four-momentum of the B^0 candidate. All quantities are calculated in the c.m. frame of the $\Upsilon(4S)$ resonance. M_{bc} is known as the beam-constrained mass. It is the mass-energy relation where the energy of the B meson has been replaced with the beam energy, which is well defined by the experimental setup. ΔE is the difference between the measured and experimental beam energy. The M_{bc} and ΔE distributions of the signal decay peaks at the B^0 mass and zero, respectively. Candidate B^0 mesons must have $M_{bc} > 5.20$ GeV/c^2 and $|\Delta E| < 0.5$ GeV. The M_{bc} and ΔE distribution from correctly reconstructed and misreconstructed B^0 candidates originating from $B^0 \to \pi^0\pi^0$ decays are shown in Figure 4.6. The M_{bc} and ΔE skim selections are purposely loose in order to preserve nearly all signal events, while also keeping enough experimental data for background study.

To investigate the impact of the selection criteria, each is individually applied to candidates matched with truth, ensuring that no other selections are in place during each

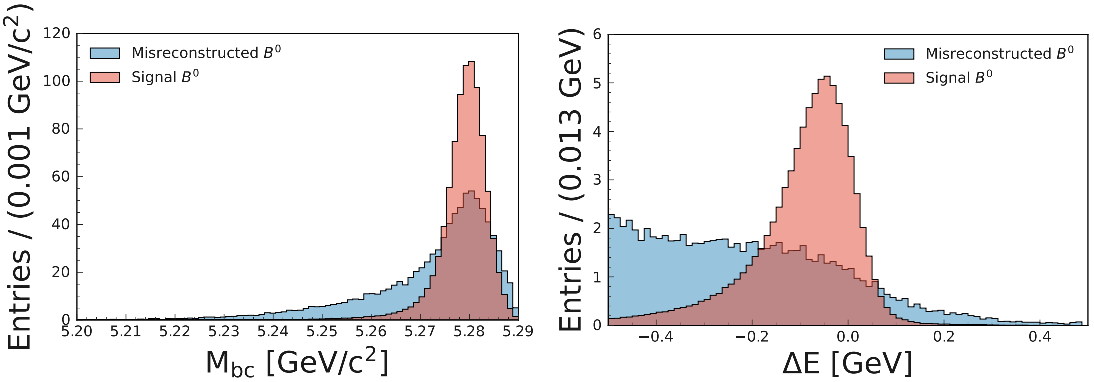


FIGURE 4.6: M_{bc} (left) and ΔE (right) distribution from correctly reconstructed and misreconstructed B^0 candidates originating from $B^0 \to \pi^0\pi^0$ decays following skimming. Misreconstruction of B^0 candidates occurs when a photon not from the signal decay (e.g. beam background, the partner B meson, etc.) is included in the signal reconstruction.

evaluation. Here, 'truth-matched' refers to events that are correctly reconstructed according to the MC truth information. The effect of each individual skim selection is shown in Table 4.1. The photon truth-matching algorithm is described in Section 5.2.

Particle	**Selection**	**Selection efficiency (%)**
γ	$E > 0.020$ GeV in barrel/backwards, $E > 0.0225$ GeV in forwards	99.4
π^0	$0.105 < M < 0.150$ GeV/c^2	99.9
	`massKFit` $\chi^2 > 0$	99.9
B^0	$\|\Delta E\| < 0.5$	99.8
	$M_{\rm bc} > 5.20$ GeV/c^2	100

TABLE 4.1: $B^0 \to \pi^0\pi^0$ skim selections and efficiency. There is significant overlap between selections, and the skim efficiency shown only serves to give a general idea of how much signal is lost during each part of the skim.

In other hadronic analyses, a standard hadronic skim is typically applied, requiring more than three tracks in the event. However, the final-state particles of the signal decay consist only of photons, resulting in no tracks. A comparison of the number of signal events, with and without the hadronic skim, showed that these additional selection criteria would cause a 23.1% loss in signal events. Therefore, the hadronic skim is not applied in this analysis. However, this exclusion significantly increases the $e^+e^- \to \tau^+\tau^-$ background, an issue that will be discussed in subsection 6.2.1.

4.4.1 Skim Results

The overall efficiency of the skim selections can be determined with two efficiencies: the skim efficiency ($\varepsilon_{\text{skim}}$) and the total efficiency ($\varepsilon_{\text{total}}$).

$$\varepsilon_{\text{skim}} = \frac{N_{\text{skim}}}{N_{\text{all}}} = 98.8\% \quad \text{and} \quad \varepsilon_{\text{total}} = \frac{N_{\text{skim}}}{N_{\text{gen}}} = 58.4\% \tag{4.2}$$

where N_{skim} is the number of truth-matched candidates passing the skim criteria, N_{all} is the total number of truth-matched candidates reconstructed without any selection criteria, and N_{gen} is the total number of generated signal events. N_{all} corresponds to the maximum number of $B^0 \to \pi^0\pi^0$ decays that could potentially be reconstructed. Hence, $\varepsilon_{\text{skim}}$ is the retention rate of reconstructable $B^0 \to \pi^0\pi^0$ events following skim selections. It is close to 99%, indicating that the skim selections retain the vast majority of useful signal events.

After skim selections, the file size of the continuum, tau pair, and generic $B\overline{B}$ data sets was determined to be less than 0.1% in both MC and data. This significant reduction provides massive improvements in terms of computing time and storage. From MC, it was found that the L1 trigger efficiency is 99.98% and therefore has a negligible effect on the signal yield.

The number of continuum and generic events after the skim is summarised in Table 4.2. The total number of continuum and generic $B\bar{B}$ events is 3,223,938 and 3,566, respectively. Notably, 55.1% of the events in the 'mixed' samples originate from signal events. This means 44.9% of the reconstructed $B^0\bar{B}^0$ events originate from other rare B^0 decays that can mimic the signal decay. Furthermore, there are a significant number of events from the 'charged' samples, indicating that there are B^+B^- decays that can similarly mirror the signal decay. These rare decays are collectively referred to as $B\overline{B}$ background, and are discussed in more detail in subsection 6.2.2.

MC Event Type	Events after skim
$u\overline{u}$	2,191,548
$d\overline{d}$	588,754
$s\overline{s}$	167,342
$c\overline{c}$	276,294
`charged`	1561
`mixed`	2005

TABLE 4.2: Number of events in 1 ab^{-1} of `MC14ri_a` after $B^0 \to \pi^0\pi^0$ skim.

Chapter 5

Suppression of misreconstructed photons

Bhabha scattering events ($e^+e^- \to e^+e^-$) can deposit substantial amounts of energy into the CsI(Tl) crystals of the electromagnetic calorimeter (ECL). This energy excites the crystals, causing them to emit scintillation light. However, the decay time of this scintillation light is relatively long when compared to the time interval between beam crossings. Consequently, the energy deposits from these Bhabha events can persist even when a subsequent hadronic event occurs. This residual energy, referred to as a misreconstructed photon, can result in a measurement in the ECL crystals that mimics the signature of a photon.

In the case of $B^0 \to \pi^0\pi^0$ decays, a random photon from a hadronic event can be combined with a misreconstructed photon to form a misreconstructed π^0. This misreconstructed π^0, along with a genuine π^0 from the event, can then combine to form a misreconstructed B^0 candidate. Such a process leads to the creation of additional B^0 candidates in the event, introducing an additional source of background. These background events, often induced by beam-background interactions, are expected to significantly increase as SuperKEKB approaches its design luminosity. Therefore, it is important to prepare the tools to suppress this background.

In the Belle analysis of the $B^0 \to \pi^0\pi^0$ decay [28], non-signal photons were suppressed by requiring the ECL signals to be in-time with the rest of the event. In this analysis, the approach is extended by not only including timing requirements but also by integrating ten variables into a binary boosted decision tree classifier, known as `photonMVA`. These variables, hereafter referred to as photon variables, provide enhanced discrimination between genuine and misreconstructed photons. These photon variables characterise the signal in the ECL crystal, i.e. photon candidates. The process of creating these photon candidate from signals in the ECL is detailed in Section 5.1. The process of identifying a photon candidate as genuine or misreconstructed is described in Section 5.2. Background information about binary boosted decision tree classifiers is provided in Section 5.3, while the photon variables themselves are described in Section 5.4. The training and validation of the `photonMVA` is described in Section 5.5.

5.1 From ECL signals to photon variables

When photons deposit their energy into the ECL, this energy can be spread across multiple ECL crystals and might overlap with energy deposits from other nearby photons. The process of transforming the raw information from the 8736 ECL crystals into photon candidates is known as ECL reconstruction. The raw information from a single crystal is the ECLDigit. It is the most basic data format and must be converted to ECLCalDigits, a higher-level representation of the data in which waveforms are fit, as described in subsection 3.3.6, to obtain integrated amplitudes, times and fit qualities which are then calibrated to obtain variables such as crystal energy and crystal time of measurement.

These ECLCalDigits are then grouped into connected regions known as ECLShower objects which are objects likely to originate from the same particle. The shower objects are further corrected and calibrated, and used to calculate physical attributes such as shower-shape quantities which quantify the energy deposition patterns. Finally, the ECLShower object is converted into a ECLCluster object, which is the highest level ECL reconstruction object. An ECLCluster that is not matched to any track is considered a photon candidate.

The initial step in forming an ECLShower object is to create connected regions (CR) known as ECLConnectedRegion. These are collections of ECLCalDigits that are likely to correspond to a single photon. The process of creating CR is illustrated in Figure 5.1. First, ECLCalDigits with an energy greater than 10 MeV are identified to serve as the initial ‘seed’ for CR. The CR expands by merging with adjacent ECLCalDigits with an

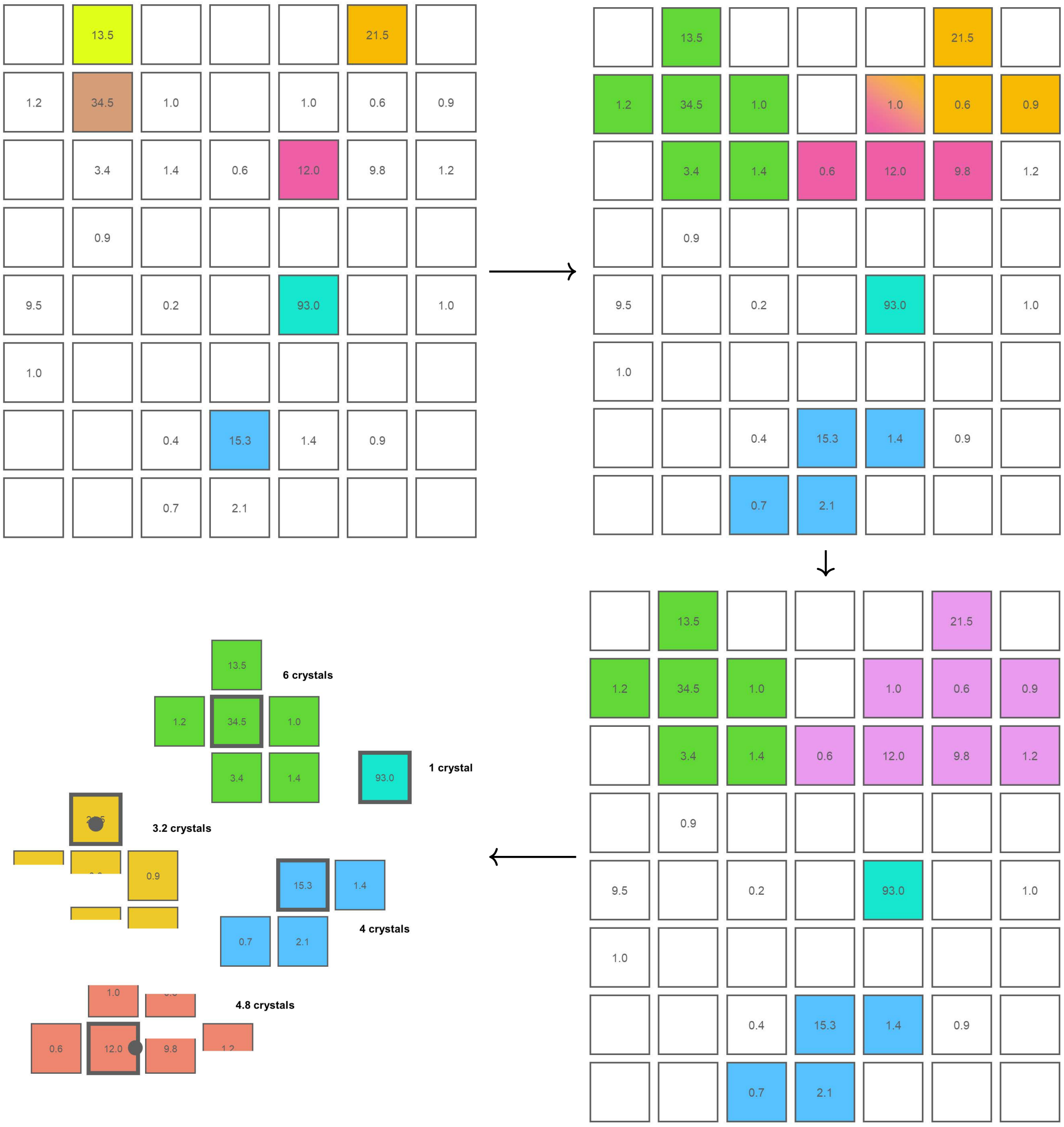


FIGURE 5.1: Steps for creation from ECLShowers from ECLCalDigits. Local maxima are identified (top left), then cells are formed by iteratively attaching adjacent ECLCalDigits with energy above 0.5 MeV (top right), overlapping cells are merged creating connected regions (bottom right), and finally if a connected regions has more than one local maxima, split the connected regions by iteratively finding a stable center of gravity to form ECLShowers (bottom left). ECLShowers can contain a non-integer number of crystals.

energy greater than 0.5 MeV. Then additional cells are included if the energy of the neighboring cells exceeds 1.5 MeV. This process is repeated until no more neighbouring ECLCalDigits can be included. Any CR that overlap are merged together. On average, an event contains around 100 CRs, with each consisting of 2-6 ECLCalDigits.

Local maxima are found by identifying ECLCalDigits in each CR. If a CR has multiple local maxima, it is split by iteratively finding stable centers of gravity (COG). The

COGs initially correspond to local maxima, and their positions are updated based on the following formula:

$$\vec{x} = \frac{\sum_i w_{1i} w_{2i} \vec{x}_i}{\sum_i w_{2i} w_{1i}}$$

$$w_{1i} = \frac{E_i e^{-Cd_i}}{\sum_k E_k e^{-Cd_x}}$$

$$w_{2i} = 4.0 + \ln\left(\frac{E_i}{E_{\text{cluster}}}\right)$$

where $\vec{x}$ is the cell position, C is a scaling factor to adjust the impact of distance, d is the distance between the current COG and another cell, E is the energy, i and k run over all cells. The constant 4.0 is included to ensure that the weight is always positive. The weights w_{1i} and w_{2i} ensure that cells farther away with less energy have an exponentially smaller impact on the final position of the COG. Employing two weights instead of one enables finer control over complex situations that a single weight cannot adequately capture.

During this process, each ECLCalDigit in the ECLConnectedRegion is assigned a set of weights that sum up to 1, with each weight corresponding to one of the local maxima in the region. This relationship is visualised in Figure 5.2. For instance, the weight of the second ECLCalDigit sums up to one and is split between the first and second ECLShower.

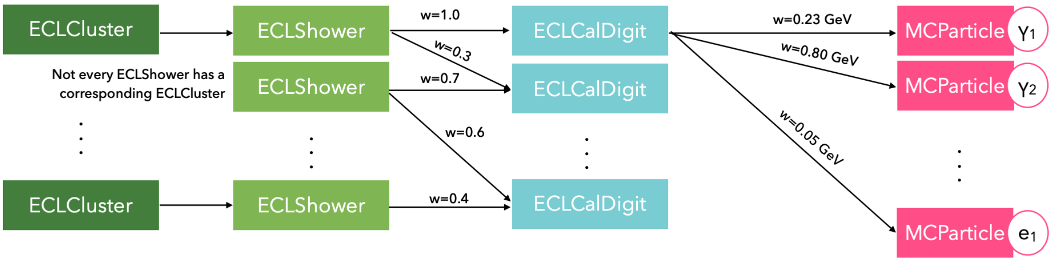


FIGURE 5.2: Diagram showing the weighted relations between ECL reconstruction objects and simulated particles.

Once the COG becomes stable across multiple iterations, the energies are split on the basis of these weights to create an ECLShower for each local maximum. ECLShowers can comprise up to 21 ECLCalDigits (a 5×5 grid that excludes corners). The weight of each of ECLCalDigits to each ECLShower is equal to the fraction of energy that it contributes to the shower. For example, in Figure 5.2, the second ECLCalDigit contributes 30% of its energy to the first ECLShower and hence the weight of the ECLCalDigit to ECLShower is 0.3.

To determine the energy of an ECLShower, the n most energetic crystals within it are summed up. The value of n is between 1 and 21, depending on the initial energy estimation and the level of the local beam background at the crystal position with the highest energy. This approach is intended to improve the resolution of true photons. However, photon energy distributions inevitably have a low-energy tail due to energy leakages in the longitudinal and lateral directions. To correct for this, the peak positions of the photon energy distributions are adjusted using large MC samples of mono-energetic single photons. The correction factor is determined by taking the ratio between the reconstructed peak and the true energy.

Meanwhile, the timing and position information of the ECLShower is determined solely by the timing and position of the local maximum within it. Using the crystal with the highest energy avoids using other crystals within a cluster, which could have different times as a result of time delays from shower propagation. Additionally, the maximum energy crystal typically contains more than 50%, of the total energy within the cluster. Consequently, the maximum energy crystal will often have a much better time resolution compared to the time resolutions for the other crystals within the cluster.

ECLClusters are formed by grouping ECLShowers with $E > 20$ MeV that are likely to originate from the same particle. They contain information such as the complete energy deposited in the ECL within the cluster, the energy distribution among its crystals, the shower-shape variables defining the cluster's shape, and the calibrated timing of the ECL shower. Each cluster usually corresponds to a single particle interaction, but there may be overlaps of multiple showers from different particles or instances where a single particle generates more than one shower outside of a connected region, resulting in more than one cluster. From the ECLClusters that do not match with a CDC track, variables known as 'photon variables' can be calculated. These variables are constructed to describe useful quantities such as the total momentum, how the energy is distributed between the ECL crystals, the distance to the nearest charged track, etc.

5.2 Photon matching

A reconstructed particle is considered ‘correctly reconstructed’ if it is matched with the underlying generator-level particle (MCParticle) that is responsible for it. In the case of a photon, an ECLCluster is correctly matched to an MCParticle if a certain fraction of its energy is attributable to the generated particle [70].

The relationship between ECLClusters and MCParticles is illustrated in Figure 5.2. It’s important to note that each ECLCluster corresponds to one ECLShower, and each ECLShower can have weighted relations with up to 21 ECLCalDigits. Each ECLCalDigit, in turn, can have a weighted relation with none, one, or multiple MCParticles. The overall weight between an ECLCluster object and an MCParticle is simply given by

$$\begin{aligned}&(\text{weight between the corresponding ECLShower and ECLCalDigit})\\&\times(\text{weight between the ECLCalDigit and MCParticle})\end{aligned}$$

The weight between the ECLShower and the ECLCalDigit, as described above, is just the fraction of energy the ECLCalDigit contributes to the ECLShower. The weight between the ECLCalDigit and the MCParticle was calculated using the total energy deposited by the MCParticle in each ECLCalDigit. For example, the weight of the relation between the first ECLCluster and MCParticle γ_2 in Figure 5.2 is given by $1.0 \times 0.8 = 0.8$ GeV. If multiple relations exist between a given ECLCluster and MCParticles, only the relation with the largest weight will be used for photon matching. The MC-truth matching is set between the ECLCluster and the MCParticles if the following conditions are met:

1. The MCParticle is actually a photon
2. $\text{weight}/E_{\text{rec}} > 0.2\,\text{GeV}$
3. $\text{weight}/E_{\text{true}} > 0.3\,\text{GeV}$

where E_{rec} is the recorded energy and E_{true} is the true energy as determined by MC. If the MCParticle turns out to be a particle other than a photon, such as an electron, no match will be made. This applies even if one of the other lower-weighted relations for the particle is correct.

5.3 Classification using machine learning

The 'photon variables' describe the ECLCluster and can be employed to differentiate between genuine and misreconstructed photons. The challenge lies in utilising these variables to minimise the number of misreconstructed photons. One strategy is to examine each variable (known as features) and apply binary selections that maximise some function (known as a 'figure of merit') which is a measure of the classifier's performance. However, if the variables are correlated, then the distribution of events (also known as samples or examples) in the feature space can have irregular shapes. Binary selections are insufficient to capture this dependency, leading to sub-optimal separation of genuine and misreconstructed photons. For example, in Figure 5.3, a substantial non-linear correlation is evident between variables var1 and var2. The rectangular selection, represented by the green lines, does not account for this correlation. Consequently, it rejects a significant portion of the signal and includes a large amount of background. In addition, the complexity of identifying and accounting for correlations between variables increases exponentially with the number of variables. This makes it increasingly impractical to use traditional statistical approaches. Therefore, a machine learning approach is necessary.

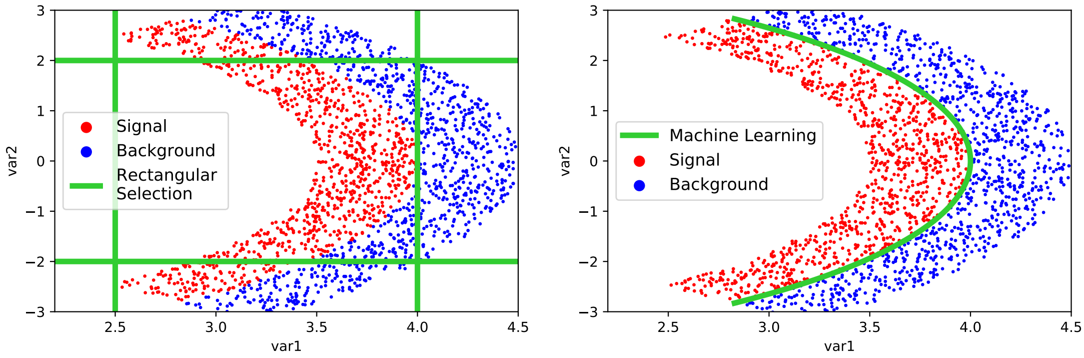


Figure 5.3: Separation of signal and background with two highly correlated variable with binary selections (left) and machine learning (right)

Machine learning (ML) approaches fall into three categories: supervised, unsupervised, and reinforcement learning. Supervised learning is used for classification problems such as distinguishing between signal and background. In this approach, the machine learns to predicts a category based on labelled training data. The labelled data consist of a set of features (the ECL variables), and a class label y (genuine $y = 1$ and misreconstructed $y = -1$ photons). A model with adjustable parameters is created along with a loss function which assesses the performance of the model. Training data are used to find the parameters which minimise the loss function by comparing the predictions with the true class label and adjusting the model parameters accordingly. This process allows the feature vectors to be mapped to a single variable known as a ‘test statistic’. A selection on the ‘test statistic’ corresponds to a selection on a hyper-surface in feature space. This allows for better separation between signal and background events, as the correlations between the features are taken into account.

Boosted decision trees (BDTs) are one example of supervised learning which can be used to combine multiple discriminating variables into a single final discriminator. They are generally easier to configure than other ML approaches, such as neural networks, as they usually contain fewer hyperparameters to tune. Hyperparameters are parameters whose values are set before the beginning of the training process. They play a pivotal role in adjusting the training process and must be carefully chosen, as the model’s overall performance can significantly depend on these values. The default hyperparameters work well for most scenarios, and only a small degree of adjustment for optimal results is necessary. In this thesis ‘FastBDT’ [71], a stochastic gradient-boosted set of decision trees, is employed. This multivariate classification algorithm is optimised for execution time and exhibits good out-of-the-box performance.

5.3.1 Decision Trees

Boosted decision trees are built from an ensemble of decision trees. A decision tree (DT) is a model composed of ‘questions’ organised hierarchically in a tree-like structure of nodes, branches, and leaves to model decisions and their outcomes, as shown in Figure 5.4. It performs classification by applying a consecutive series of selections that eventually results in a final decision. Each decision is known as a ‘decision node’, ‘branches’ represent the possible answers, and ‘leaves’ corresponding to the final classification value. The maximum number of consecutive selections is a hyperparameter and is called the depth of the tree.

The construction of a DT begins with the ’root node’ which contains the entire dataset. This root node, like all subsequent nodes, is designed to split the data into subsets that are as homogeneous as possible with respect to some target variable. This means selecting the feature that locally maximises the separation between signal and background events at each node of the tree. For example, the root node might pose a question such

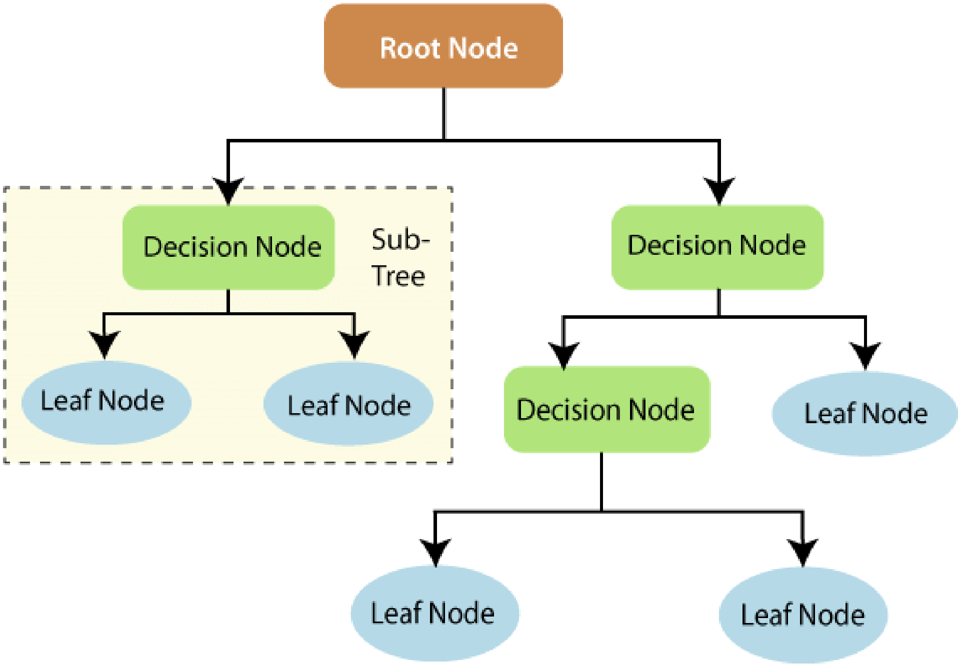


FIGURE 5.4: Diagram of a typical decision tree [14]

as ‘Is the total energy greater than 3 MeV?’. The possible answers to this question form the branches that connect the root node to the child nodes. Each child node can then ask another question about a different feature. This process continues recursively and independently, creating new child nodes until a stopping criterion is met, such as the child nodes providing less separation power or reaching a maximum depth. At this point, the decision node becomes a ‘leaf’.

Once the DT is built, it can be used to make predictions for new data points by traversing the tree from the root node to a leaf node. At each node, the algorithm checks the value of the corresponding feature for the new data point and follows the appropriate branch until it reaches a leaf node. The leaf prediction is usually determined by the majority vote of the data points. For example, if during training most of the events in a leaf node were signal events, then the leaf node will predict signal. The probability of any new data point to be signal is the fraction of signal training data points in the leaf.

In the case of FastBDT, at each node, a cumulative histogram (CH) for each feature is calculated for the signal and background. To create a CH, training samples are sorted by increasing feature value. Then the data is divided into suitable intervals or bins. The probabilities are calculated by counting the fraction of signal and background events with feature values less than or equal to the upper bin boundary. For example, consider

10 events where the total energy is:

$$2.0(S), 2.1(S), 2.5(S), 2.7(B), 3.0(B), 3.2(S), 4.0(S), 4.5(S), 5.0(B), 6.0(B).$$

Of the 10 events, 6 are labelled as signal (S) and 4 as background (B). The CH with a bin width of 1.0 is shown in Table 5.1.

Bin	Upper bin boundary	S	B
1	2.5	3	0
2	3.0	3	2
3	3.5	4	2
4	4.0	5	2
5	4.5	6	2
6	5.0	6	3
7	5.5	6	3
8	6.0	6	4

TABLE 5.1: Example cumulative histogram.

The separation gain is defined as the loss function (LF) at the decision node subtracted by the loss function of both leaves. A loss function quantifies the split based on the purity of the resulting nodes. The loss function for each node is the Gini index (I_G) [72] multiplied by the total number of events in that node. The Gini index is given by

$$I_G = 1 - \sum_{k=1}^{K} p_k^2 \tag{5.1}$$

where K is the number of classes and p_k is the proportion of samples in the node that belong to class k. A Gini index of 0 represents perfect classification, while a Gini index close to 1 indicates that the events are randomly distributed across various classes. The loss function is then given by:

$$\text{Loss Function} = \left[1 - \left(\frac{S}{S+B}\right)^2 - \left(\frac{B}{S+B}\right)^2\right] \times (S+B) \tag{5.2}$$

$$= 2\frac{SB}{S+B} \tag{5.3}$$

Here, S and B represent the sum of the weights of the signal and background events that belong to that node, respectively. The weight of an event signifies its contribution to the overall learning process. By default, this weight is set to one. For simplicity, the factor of two in Equation 5.3 is usually dropped. For all features and cuts, the information gain is determined by subtracting the loss function for both leaf nodes subtracted from the current loss.

$$\text{Separation Gain} = \text{Current LF} - \text{LF of left node} - \text{LF of right node}$$

This separation gain quantifies the separation between signal and background. In the given example, with 6 signals and 4 background events, the current loss function is 2.4. If one considers a split at 3.0 ($E \leq 3$), the loss function for the left node would be calculated as $(4 \times 1)/(4+1) = 0.8$, while for the right node it would be $(2 \times 3)/(2+3) = 1.2$. The separation gain for this cut can be determined as $2.4 - 0.8 - 1.2 = 0.4$. Performing a similar calculation at 4.0, one finds that the information gain is 0.3. Thus, a split at 3.0 provides a better separation between the signal and background.

5.3.2 Overfitting

The split must be carefully considered to prevent overfitting. Overfitting occurs when distinguishing features are not generalised, but instead statistical fluctuations are learnt in the training data, as shown in the left-hand diagram of Figure 5.5. In this case, the model learns features that are not actually present and is therefore more likely to have a higher error rate on unseen data. To avoid overfitting, regularisation is usually applied. Regularisation is the process of adding constraints or penalties to the model parameters to prevent them from becoming too large or too specific. For DTs, pruning is one form of regularisation that removes some of the branches that do not contribute much to the model's predictive power, as shown in the right-hand diagram of Figure 5.5. To perform pruning, a validation dataset is used. This dataset is a subset of the data that is not used for training, but is instead used to evaluate the performance of the model. The primary purpose of pruning is to compare the validation accuracy of the original tree with the validation accuracy of a pruned tree, where some decision nodes are replaced by leaves. If pruning a decision node enhances the validation accuracy, it indicates that the node is not useful for generalising to new data and can be removed.

5.3.3 Boosted Decision Trees

Another and more robust method to use DTs for classification is known as 'boosting'. In this ensemble method, a sequence of low-depth or shallow decision trees are trained one after another, with each subsequent tree being influenced by the performance of the

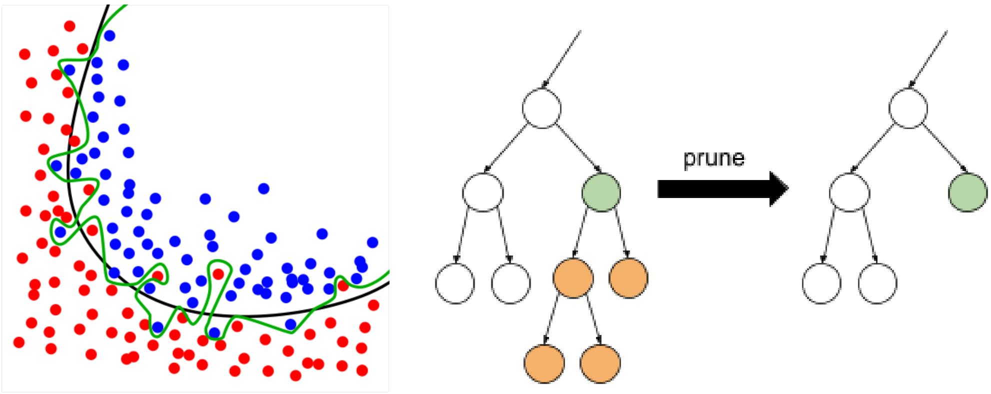


Figure 5.5: The green line represents an overfitted model and the black line represents a well trained model (left). Example of pruning a decision tree, thereby turning a decision node into a leaf (right).

previous trees. In each iteration of the boosting process, a new tree is constructed that is trained to correct the errors or misclassifications of the previous tree. This is done by assigning more weight to events that have not been correctly classified by previous models. Initially, all the weights are set to the same value, so that each event is of equal importance. The next classifier will concentrate on these misclassified events and reduce the weight of correctly classified ones after each training. Now, in the next iteration of training, when the algorithm selects a feature and a threshold value to split the events, it is more likely to choose a splitting point that correctly separates misclassified events. This has the effect of making the subsequent weak classifiers more sensitive to events that were misclassified by the previous classifiers and less sensitive to events that were correctly classified.

In FastBDT, each shallow DT is constructed such that the expectation value of a negative binomial log-likelihood loss-function is minimised. The negative binomial log-likelihood loss-function is defined as

$$\log(\mathcal{L}) = 1 + e^{-2yF} \tag{5.4}$$

where F is the prediction of the model, $y \in -1, 1$ is the actual value of the target variable. It measures the difference between the predicted values of a target variable and the actual values, assuming that the distribution of the target variable $(y_1, y_2, ..., y_n)$ follows a negative binomial distribution. A negative binomial log-likelihood of 0 indicates a perfect fit between the predicted and actual values of the target variable. The higher the value, the greater the difference between the predicted and actual values.

Models with lower expected negative binomial log-likelihoods are generally considered to be better fits to the data. The process iterates until a stopping criterion is met, and the final weight is the sum of all classifiers weighted by their error. Using many shallow DTs, known as weak learners, the classifier is able to have much larger separation power than a single large decision tree.

The number of boosting steps, i.e. the number of trees and the learning rate, are additional hyperparameters. The learning rate controls the contribution of each tree that is added to the model during training, and hence the magnitude of the update made to a model's parameters (or step size). Mathematically, it is a scaling factor that multiplies the contribution of each tree to the final prediction. A smaller learning rate will result in a smaller step size and a slower convergence of the boosting algorithm, but it may also improve the generalisation performance of the model by reducing overfitting. A higher learning rate will result in a larger step size and faster convergence of the boosting algorithm, but it may also increase the risk of overfitting the training data.

5.3.4 Gradient Boosted Decision Tree

Gradient Boosted Decision Trees (GBDT) [73] is an iterative algorithm that combines weak models $f_m(x)$, typically decision trees, into a strong model $F_m(x)$. Here, m refers to the m-th step of the training process, and so f_m is the m-th tree. In GBDT, the term 'gradient' refers to the use of the gradient of a loss function with respect to the prediction model. The loss function measures the discrepancy between the predicted output of the model and the actual output of the training data. It can be any function as long as it is differentiable, and its value increases as the performance of the classifier decreases. The negative gradient indicates the direction of steepest descent and hence minimises the loss function. Like boosted decision trees, newly created decision trees in GBDT correct the errors of the previous ones. But unlike BDTs which change the weights of each event, the newly created decision trees in the GBDT algorithm are trained on the pseudo-residuals. The pseudo-residual is defined as the negative gradient of the loss function with respect to the predicted output value. It provides a measure of the direction and magnitude of the change needed to reduce the error for each event in the training data.

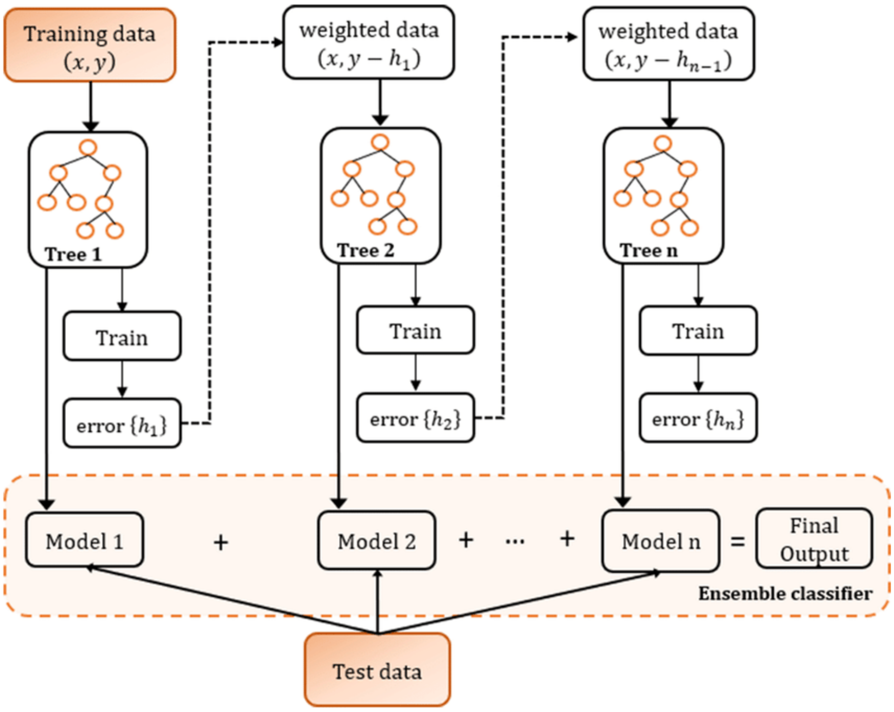


FIGURE 5.6: Diagram of the gradient boosted decision tree fitting algorithm [15]

As shown in Figure 5.6, a GBDT starts with a single tree that makes an initial naive prediction based on the training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where i is the i-th event in the training dataset, $\mathbf{x}_i \in \mathbb{R}^d$ are the training variables (energy, momentum, etc.), $y_i \in \mathbb{R}$ is the actual label (signal or background), and $L(y, F(x))$ is a differentiable loss function. The naive prediction is a constant value:

$$F_0(x) = \text{argmin}_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$$

It is the value of γ in which the sum $\sum_{i=1}^{n} L(y_i, \gamma)$ is minimised. In the case of least-square regression, the loss function is the mean square error and the naive prediction is just the average. Then the negative gradient of the loss function with respect to the output of the previous model is computed:

$$r_{im} = -\left[\frac{\partial L(y_i, F_{m-1}(\mathbf{x}_i))}{\partial F_{m-1}(\mathbf{x}_i)}\right]_{F(x)=F_{m-1}(x)} \quad \text{for} \quad i = 1, .., n$$

where r_{im} is known as the ‘pseudo-residual’, and y is the actual label. Here, the previous predictions $F_{m-1}(x)$ are used as the current prediction model. For the first iteration, F_1, the pseudo-residual for each event is calculated based on the initial naive prediction. Then a decision tree, $f_m(x)$, is constructed using the training data that has the target variable for each event replaced with their pseudo-residual, i.e., using the training dataset

$\{(\mathbf{x}_i, r_{im})\}_{i=1}^{n}$. In the case of least-square regression, the pseudo-residual are simply the difference between the actual and predicted values, i.e. the negative residuals.

Traditionally, the signal-fraction of the leaf node is used as the output of the DT. In contrast, a 'boost-weight' is calculated for each leaf, R_{jm}, for this newly created tree, where $j = 1, .., J$ runs over all the leaves. The output of each leaf node is this 'boost-weight':

$$\gamma_{jm} = \text{argmin}_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$$

In the case of least-squares regression, this is again just the average. This boost-weight is a measure of the importance of that leaf in terms of improving the overall prediction accuracy of the ensemble model. The final output value of the tree is the γ_{jm} value of the leaf node where any test event ends up.

The strong model is updated by adding the current tree along the learning rate, ν, to the ensemble model:

$$F_m(x) = F_{m-1}(x) + \nu f_m(x)$$

The addition of a learning rate is an important regularisation step which controls how quickly the overall model is learning. ν varies from 0 to 1 and is multiplied to each newly decision tree to prevent overfitting, as the contribution of each tree is decreased by an amount ν. Small learning rates ($\nu = 0.1$) generally improve the generalisability of the models to unseen data. New pseudo-residuals are computed and the process is iterated ($m = 0...M$). The pseudo-residuals become smaller at each tree. The pseudo-residuals of events that are easy to classify very quickly become small, and hence 'ignored' while constructing subsequent decision trees. Events that are difficult to classify, such as those near the decision boundary, continue to have large pseudo-residuals, which subsequent decisions tree focus on. As a result, the predictions for these types of events improve.

In this way, the GBDT is able to correct the errors of previous trees. The model repeats this process until a predefined number of trees is reached or no further improvement is possible, as determined on a validation data set. Ideally the final pseudo-residual for each event will be very small. The final model is an ensemble of these trees, where each tree is weighted according to its contribution to the overall prediction. Conceptually, this process can be thought of as training a decision tree to predict the pseudo-residuals,

r_{im}, of the previous ensemble model, F_{m-1}, given $\mathbf{x}_i$. Then using the pseudo-residual to 'correct' the prediction of F_{m-1}.

In FastBDT, the loss function is the negative binominal log-likelihood

$$L(y, F) = \log(1 + \exp\{-2yF\}), \quad y \in -1, 1$$

where

$$F(x) = \frac{1}{2} \log \left[\frac{\Pr(y = 1|x)}{\Pr(y = -1|x)} \right]$$

is the prediction of the model, $y \in -1, 1$ is the actual value of the target variable. It measures the difference between the predicted values of a target variable and the actual values, assuming that the distribution of the target variable $(y_1, y_2, ..., y_n)$ follows a negative binomial distribution. A negative binomial log-likelihood of 0 indicates a perfect fit between the predicted and actual values of the target variable. The higher the value, the greater the difference between the predicted and actual values. Hence, each shallow DT is constructed such that the expectation value of a negative binomial log-likelihood loss-function is minimised.

The psuedo-residuals is given by:

$$r_{im} = 2y_i/(1 + \exp(2y_i F_{m-1}(\mathbf{x}_i))$$

The boost-weight calculated for each leaf, R_{jm}, can be approximated by

$$\gamma_{jm} = \sum_{\mathbf{x}_i \in R_{jm}} r_{im} / \sum_{\mathbf{x}_i \in R_{jm}} |r_{im}|(2 - |r_{im}|)$$

When the model is applied to a testing dataset with unknown labels, the probability of a test data point belonging to a specific class is calculated by summing the boost-weights for all leaf nodes where that test data point ends up. With all this, the model update rule becomes:

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} \mathbf{1}(x \in R_{jm})$$

Here, $\mathbf{1}$ is the indicator function which maps elements of the subset $(x \in R_{jm})$ to one, and all other elements to zero. The weighted predictions of all the trees in the ensemble are summed, and a sigmoid function is applied to produce a probability. This predicted probability can then be compared to a threshold value to make the final prediction. The sigmoid function maps any input value to a value between 0 and 1 and is defined as:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

where x is the input value. When the input value is large and positive, the sigmoid function approaches 1. When the input value is large and negative, the sigmoid function approaches 0. When the input value is 0, the sigmoid function is 0.5.

5.3.5 Stochastic Gradient Boosted Decision Tree

Stochastic Gradient Boosted Decision Tree (SGBDT) combines the strength of gradient boosting and random subsampling to improve the accuracy of decision trees. Random subsampling is the process of randomly selecting, without replacement, a subset of events (a sample) from the entire training data set. In each boosting step, this subset is used to train the decision tree. By doing so, the algorithm effectively averages out the statistical fluctuations in the training data, which can improve the model's generalisability. The subsampling rate, denoted α, is a hyperparameter that controls the fraction of samples used in each boosting step. A smaller α value corresponds to a larger fraction of the training data being used, and vice versa. Furthermore, SGBDT can be significantly faster to train compared to GBDT since the dataset is smaller and requires less memory.

5.3.6 FastBDT

FastBDT [71] is a speed-optimised and cache-friendly stochastic gradient-boosted decision tree algorithm used in this thesis for multivariate classification. The training time is highly optimised due to an equal-frequency binning on the input data and a cache-friendly linear access pattern to the input data. It also employs CHs to calculate the best-cut at each node of the tree.

The equal-frequency binning on the input data discretises the continuous input variables so that each bin has roughly the same number of data-points, and maps them to integers (see Figure 5.7). Hence, the FastBDT algorithm only compares the bin values with each other and does not use the feature values themselves. This integer operation is much faster than computationally expensive floating-point operations. Furthermore, the integers can be used as indices of the CHs during the calculation of the best feature

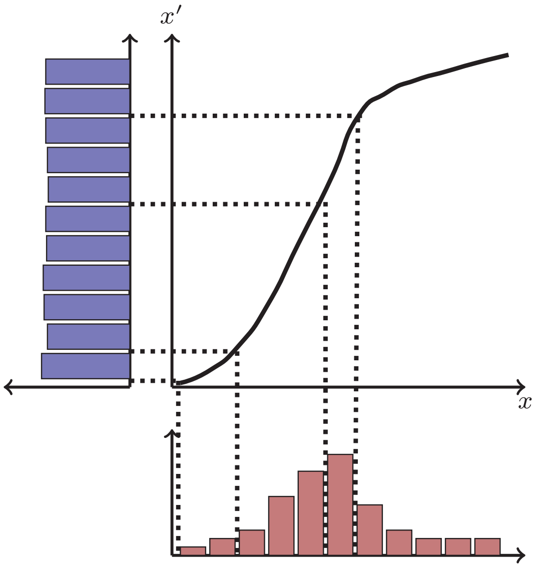


FIGURE 5.7: Equal-frequency binning. The bin boundaries are indicated by the dotted lines

and cut at each decision node. This allows the FastBDT algorithm to directly access the corresponding entry in the CHs without the need for any additional calculations or conversions. Finally, since any irregular shape (such as sharp peaks) in the input feature distribution are mapped to a uniform distribution, the separation can be improved. This is because irregular shapes can create challenges for classification algorithms as they can cause certain regions of the feature space to be overrepresented or underrepresented, leading to biased classification.

As such, FastBDT is used for multivariate classification in this thesis because of its high-speed processing, efficient signal and background separation, and ease of use.

5.4 Photon variables

FastBDT is used to combine 10 photon variables to create a single discriminating variable. The variables, whose meaning will be described shortly, are known as 'clusterHighestE, clusterZernikeMVA, minC2TDist, clusterSecondMoment, pt, E, clusterE1E9, clusterE9E21, clusterAbsZernikeMoment40, and clusterAbsZernikeMoment51'.

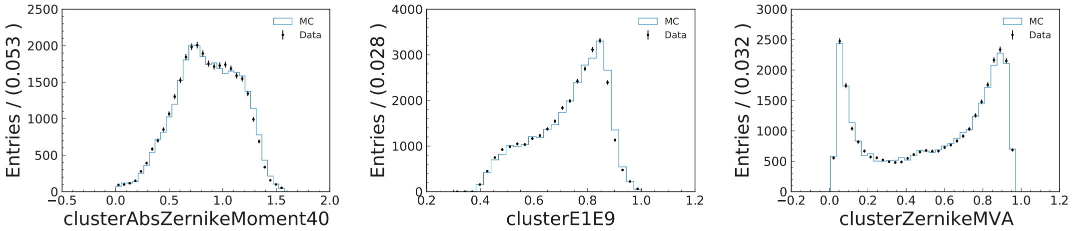


FIGURE 5.8: Comparison between `MC14ri_a` (blocks) and data (dots) photons for clusterAbsZernikeMoment40 (left), clusterE1E9 (middle) and clusterZernikeMVA (right).

5.4.1 E, pt, and clusterHighestE

The energy of an ECLShower, denoted as ‘E’, is determined by summing up the energies of the n most energetic crystals within the ECLShower. The value of n ranges from 1 to 21, and is dependent on an initial rough energy estimate and a background estimate. The energy distributions of photons typically exhibit a low-energy tail due to both longitudinal and transverse leakage in the ECL crystals. This tail can be further influenced by beam backgrounds and the clustering algorithm, which is described in subsection 5.1.

The variable ‘pt’ refers to the momentum transverse to the beam line. This is a significant parameter because the momentum perpendicular to the beam line is always associated with the physics processes occurring at the vertex.

The variable ‘clusterHighestE’ corresponds to the energy of the most energetic crystal in the ECL cluster. Generally, events where the photon deposits almost all of its energy into a single crystal are more likely to be signal events rather than beam background or electron events, which tend to distribute their energy more broadly.

Figure 5.9 illustrates the distribution of ‘E’, ‘pt’, and ‘clusterHighestE’ for misreconstructed and genuine photons. Any ECL hits not originating from photons are classified as misreconstructed, whereas ECL hits that do originate from photons are classified as genuine.

5.4.2 clusterE1E9 and clusterE9E25

The shower shape variables, ‘clusterE1E9’ and ‘clusterE9E21’, are used to calculate the ratio of energy contained in the central crystal and the surrounding crystal grids of an ECLCluster. The variable ‘clusterE1E9’ is the ratio of energy contained in the

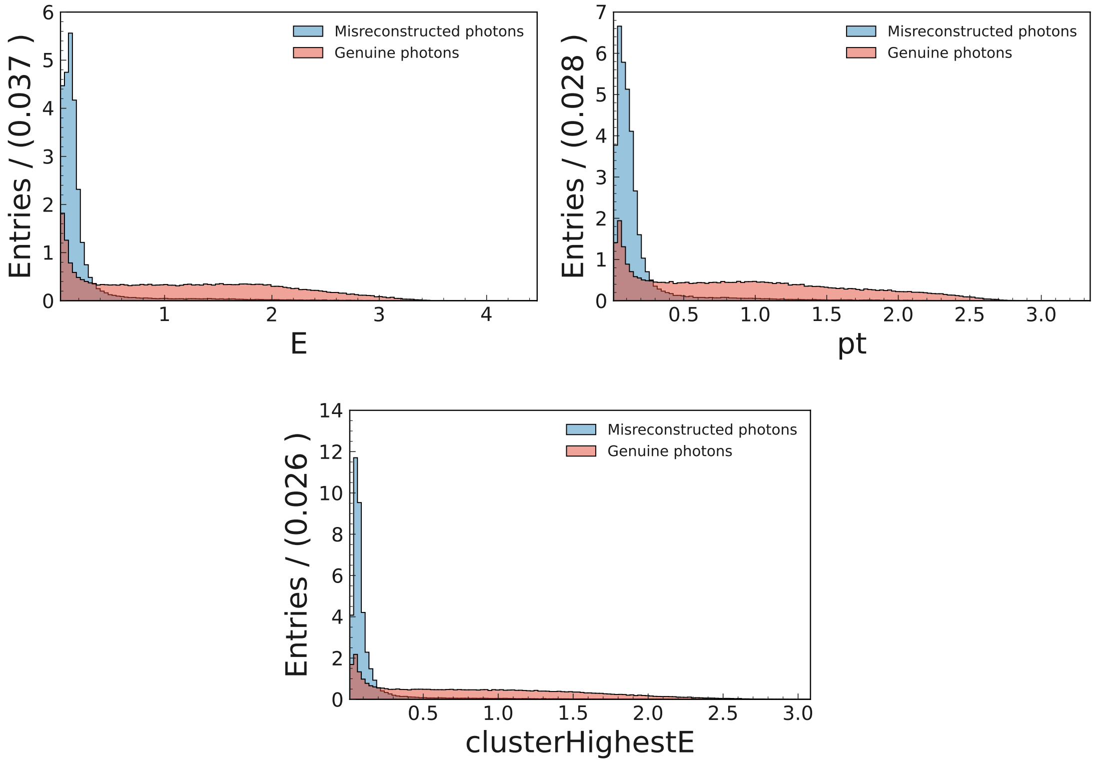


FIGURE 5.9: Comparison between misreconstructed (blue) and genuine (red) photons for E (left), pt (middle) and clusterHighestE (right).

central crystal of a cluster to the energy contained in a 3×3 grid of crystals surrounding the central crystal,as shown in the left-hand diagram of Figure 5.10. The term 'E1' represents the energy in the central crystal and 'E9' represents the total energy in the 3×3 crystal grid. For clusterE1E9, this ratio is between 0 and 1. Photon deposits, which concentrate their energy in a single crystal, tend toward unity, while hadronic clusters, which distribute their energy across multiple crystals, tend toward smaller values.

The 'clusterE9E21' variable represents the ratio of the energy deposited in the inner 9 crystal and the outer energy deposited in a 5×5 crystal grid surrounding the inner 9 crystals, with the four corners of the grid excluded, as shown in the right-hand diagram of Figure 5.10. Similar to 'clusterE1E9', 'clusterE9E21' also lies between 0 and 1. It tends to unity for photons and to smaller values for hadronic clusters. However, 'clusterE9E21' peaks closer to unity than 'clusterE1E9'. This is because energy deposits tend to be circular and focus on the central crystal. Figure 5.11 illustrates the distribution of 'clusterE1E9' and 'clusterE9E21' for misreconstructed and genuine photons.

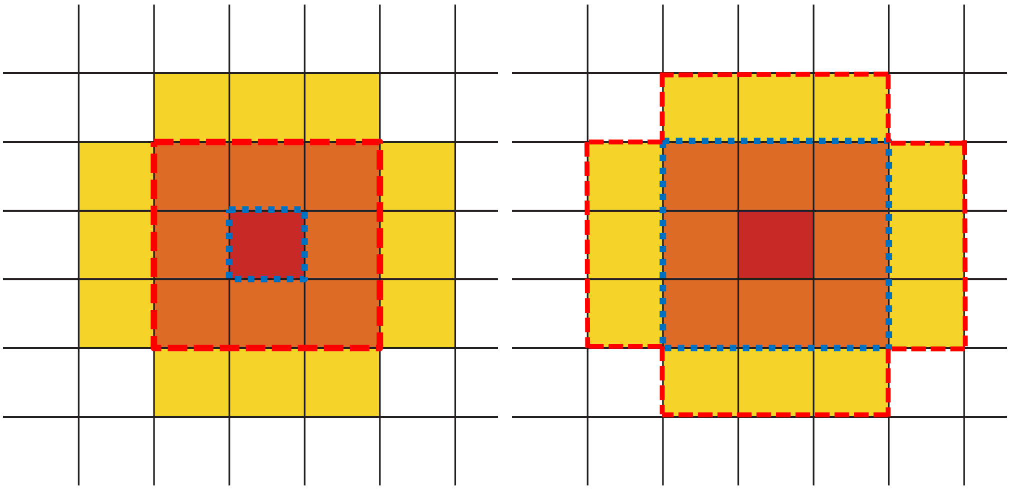


FIGURE 5.10: Diagram of clusterE1E9 (left) and clusterE9E21 (right) where the the blue dotted region encloses the central crystal(s), and the red dotted region encloses the surrounding crystals.

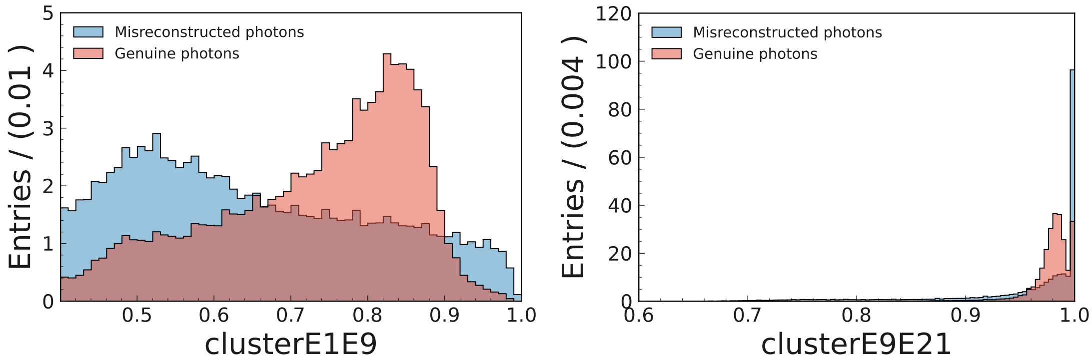


FIGURE 5.11: Comparison between misreconstructed (blue) and genuine (red) photons for clusterE1E9 (left) and clusterE9E21 (right). The ECL cluster variables are clipped at the lower and upper boundaries.

5.4.3 clusterAbsZernikeMoment40, clusterAbsZernikeMoment51 and clusterZernikeMVA

Zernike moments are calculated for each ECLCluster to quantify the circularity of the distribution in a plane. The two Zernike moments of interest are 'clusterAbsZernikeMoment40' and 'clusterAbsZernikeMoment51'. They are calculated per shower in a plane perpendicular to the shower direction using the equation,

$$|Z_{nm}| = \frac{(n+1)}{\pi} \frac{1}{\sum\limits_{i=1}^{n} w_i E_i} \left| \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} w_i E_i \right|$$

where n and m are integer numbers that define the Zernike moment. E_i represents the energy of the i-th crystal in the ECLCluster, while w_i denotes the weight of the i-th crystal after the splitting of the respective ECLConnectedRegion into ECLShowers. R_{nm} represents the related Zernike polynomial, ρ_i represents the radial distance of the i-th crystal in the perpendicular plane, and α_i denotes the polar angle of the i-th crystal in the perpendicular plane. The variable 'clusterAbsZernikeMoment40' and 'clusterAbsZernikeMoment51' corresponds to $|Z_{40}|$ and $|Z_{51}|$, respectively. The Zernike moments for various values of n and m are shown in Figure 5.12. For $|Z_{40}|$, the values can range from 0 to 1.6, while for $|Z_{51}|$, the values can range from 0 to 1.2.

The variable 'clusterZernikeMVA' is the output of a FastBDT training using eleven Zernike moments of the cluster trained to distinguish between showers from photons and K_L^0 particles. The moments incorporated are:

$$|Z_{11}|, |Z_{20}|, |Z_{22}|, |Z_{31}|, |Z_{33}|, |Z_{40}|, |Z_{42}|, |Z_{44}|, |Z_{51}|, |Z_{53}|, |Z_{55}|$$

An MVA is employed because storing all the Zernike moments for each ECL shower would require an excessive amount of disk space.

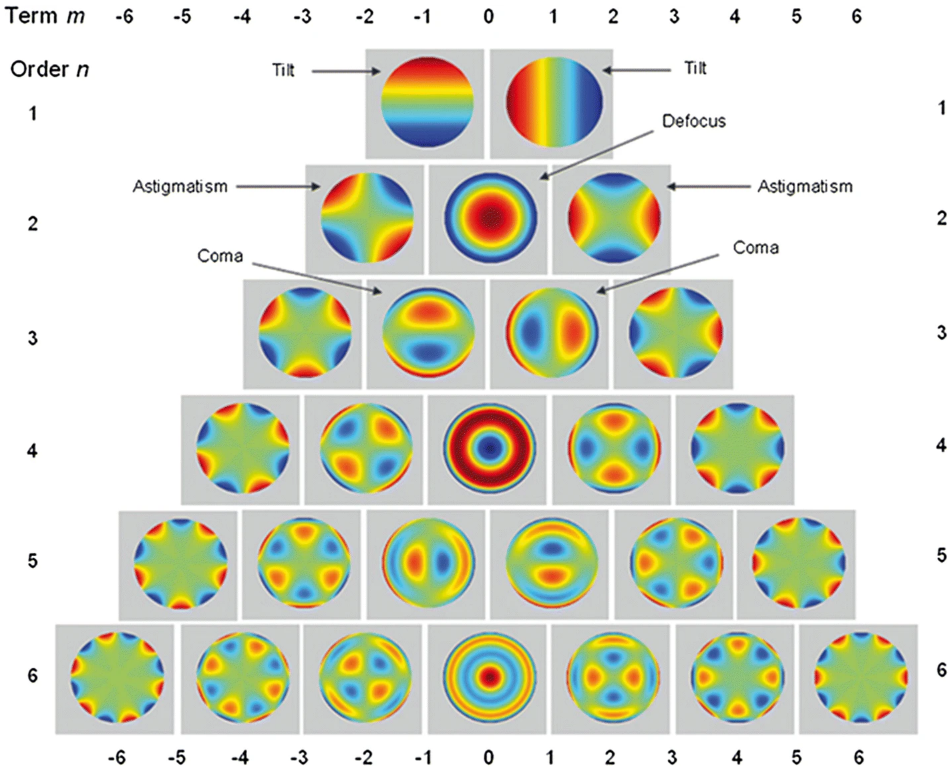


FIGURE 5.12: Zernike polynomials at various orders of n and m [16].

Zernike moments can also be used to distinguish between electromagnetic and hadronic interactions. Typically, hadronic showers exhibit lower values, whereas electromagnetic showers tend to have higher values. Figure 5.13 illustrates the distribution of 'clusterAbsZernikeMoment40', 'clusterAbsZernikeMoment51' and 'clusterZernikeMVA' for misreconstructed and genuine photons.

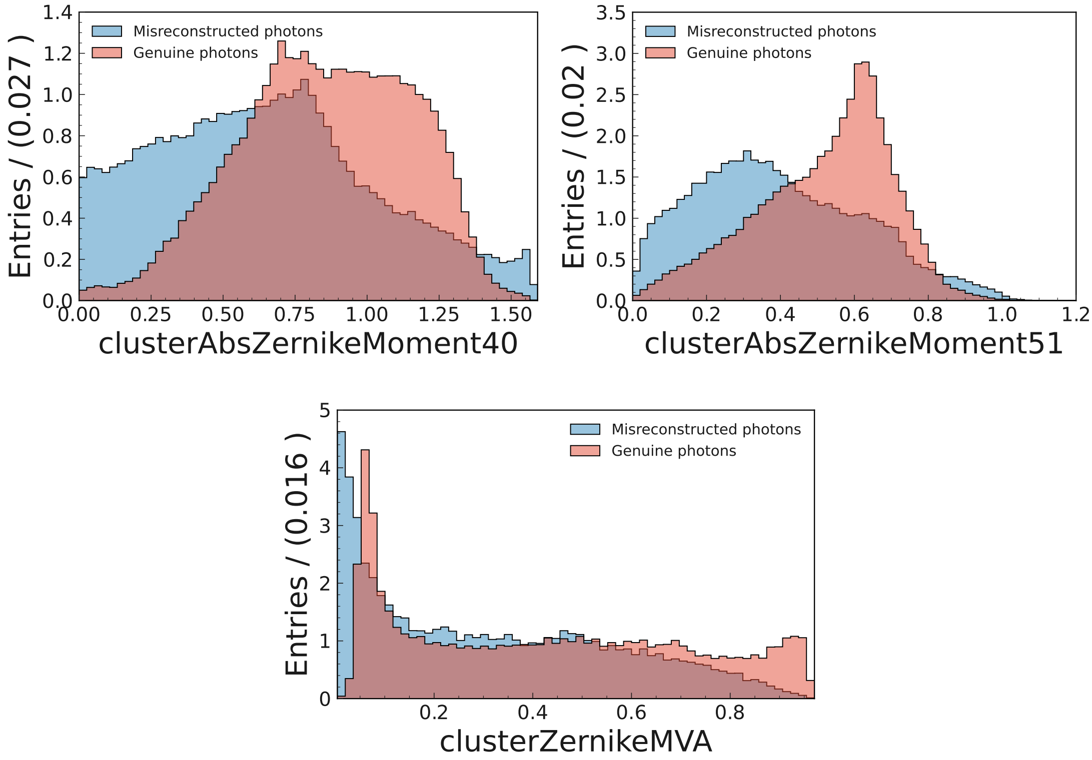


FIGURE 5.13: Comparison between misreconstructed (blue) and genuine (red) photons for clusterAbsZernikeMoment40 (left), clusterAbsZernikeMoment51 (middle) and clusterZernikeMVA (right).

5.4.4 minC2TDist

The variable 'minC2TDist' represents the 'minimum cluster to track distance', which is the shortest distance between an ECL cluster and the nearest track. It is calculated by determining the shortest distance between the extrapolated hits of every track in the event and the position of the ECL shower. If the calculated distance exceeds 250.0 nm, it is capped at that value. If no extrapolated hits are found in the ECL for the event, the value returned is NaN.

This variable is significant as it can be used to identify 'split-off' photons. A 'split-off' photon is a photon radiated by a charged particle (such as a pion or an electron) passing through the detector. This photon is often in close proximity to the original track of the charged particle and can be mistaken for a separate particle. Therefore, a photon candidate that is a split-off from a nearby track is likely to have a small 'minC2TDist' value.

For signal photons, the closest charged track originates solely from the decay of the partner B meson. Consequently, the distribution of minC2TDist' is relatively flat, with the most probable value being a considerable distance away. Meanwhile, ECL clusters that originate from a charged track but were misreconstructed into a photons will tend to peak at lower values. Figure 5.14 illustrates the distribution of 'minC2TDist' for misreconstructed and genuine photons.

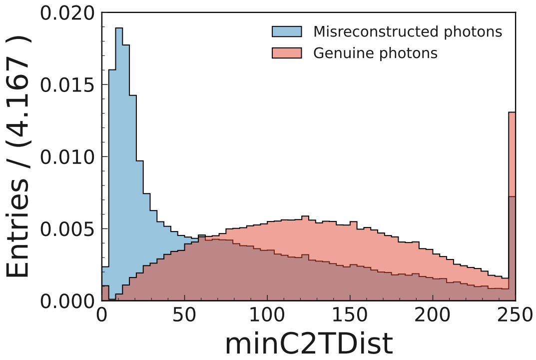


FIGURE 5.14: Comparison between misreconstructed (blue) and genuine (red) photons for minC2TDist.

5.4.5 clusterTiming

The variable 'clusterTiming' represents the time of the ECL cluster relative to the time of the event. It is calculated as clusterTiming = time of the highest energy crystal in the cluster - event t0. The 'time of the highest energy crystal in the cluster' is a time measured by the ECL for the cluster. Only one crystal is used since the accuracy of the time measurement should increase with the energy deposited in the crystal. Consequently, this one crystal should provide the most accurate time measurement among all crystals within the cluster.

The 'event t0' is the time of the event. It takes into account all the different pieces of information to determine the time of the event as a whole. It is measured with respect to the time of the event as determined by the online triggers. It may be provided by the

ECL or a combination of detectors. In the case of the ECL, the 'event t0' is from an $e^+e^- \rightarrow e^+e^-$ event. The large ECL energy deposits from these high-energy electrons set the 'event t0' while any other lower-energy clusters (potentially from beam backgrounds or other sources) may only skew the event t0 value to higher or lower values.

Although this variable is not used directly in the `photonMVA`, it is still a powerful discriminator between signal and background photons, and is exploited to remove very out-of-time photons. For signal, the clusterTiming distribution is centred on approximately zero if the photons (or any ECL clusters) are consistent with coming from the main part of the events. The beam background tends to have a significantly wider distribution. Figure 5.15 illustrates the distribution of 'clusterTiming' for misreconstructed and genuine photons.

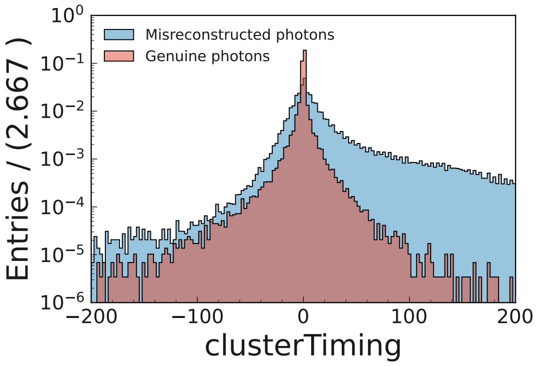


FIGURE 5.15: Comparison between misreconstructed (blue) and genuine (red) photons for clusterTiming. The y-axis in the plot is displayed on a logarithmic scale to enhance the visibility of events far away from zero.

5.4.6 Data-MC agreement

The photon variables that are selected must demonstrate good agreement between data and MC. This is to ensure the consistency of the classifier's performance across both data sets, so that the signal efficiency determined in MC can be used to extract the branching fraction from data.

Only photon variables with a reduced chi-squared value of less than 5 are used. (see Table 5.2). The reduced chi-squared is a measure of the goodness of fit of a model to observed data and is defined as

$$\chi^2_{\text{reduced}} = \frac{1}{N-k} \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2}$$

where N the number of data points, k is the number of fitting parameters, y_i is the observed value of the i$^{\text{th}}$ data point, $\hat{y}_i$ is the predicted value of the i$^{\text{th}}$ data point, and σ_i is the uncertainty associated with the i$^{\text{th}}$ data point. A χ^2_{reduced} value close to 1 indicates a good fit between the model and the data, while values significantly greater than 1 suggest that the simulated distribution of the ECL variable may not accurately replicate the data. The distribution of the photon variables in MC and data is shown in Appendix A.1 while the correlation matrix is shown in Appendix A.2.

ECL Variable	**Reduced χ^2**
clusterHighestE	1.32
clusterZernikeMVA	1.09
minC2TDist	1.95
clusterSecondMoment	3.63
pt	1.05
E	0.93
clusterE1E9	3.64
clusterE9E21	2.79
clusterAbsZernikeMoment40	2.50
clusterAbsZernikeMoment51	1.31

TABLE 5.2: Variables used in `photonMVA` and the reduced χ^2 (24 degrees of freedom) from descending feature importance which is a measure of how much impact a feature has on the model's predictions.

5.5 PhotonMVA

5.5.1 Creating training and testing datasets

Boosted decision trees are prone to overfitting, which occurs when the model excessively learns from statistical fluctuations. To avoid this, the available data are split between a training and testing dataset. During the training phase, the BDT learns patterns and optimises its performance using the training data. The degree to which the BDT can generalise to new, unseen data is assessed by its performance on the testing dataset. This also serves as a check for overfitting, as the performance on the training dataset should not be significantly better than on the testing dataset.

To create samples of genuine and misreconstructed photons, photons originating from signal MC $B^0 \to \pi^0\pi^0$ decays reconstructed with very loose selections applied are used. This is referred to as the `B0:all` particle list. These selections are as follows:

γ: $E > 30$ MeV, $|\texttt{clusterTiming}| < 200$ ns, $\texttt{clusterNHits} > 1.5$,

$0.2967 < \texttt{clusterTheta} < 2.6180$

where E, `|clusterTiming|`, and `clusterNHits` are the variables described in the previous section. The positive direction is given by the electron-beam direction, and the polar angle, `clusterTheta`, is defined with respect to the $+z$ axis. These selections are chosen to mirror the photon selections that will be used in the analysis. This is to remove photons that will not be included in the π^0 reconstruction. These selections are chosen to mirror the photon selections that will be used in the analysis, and remove photons that would not be included in the π^0 reconstruction. Using photons from the $B^0 \to \pi^0\pi^0$ decays ensures that the training is specific to the signal decay mode. This specificity is important because of the unique nature of the final-state particles, which comprise solely of high-momentum photons.

The detector is divided into three regions referred to as the backward endcap ($32.2^\circ < \theta < 128.7^\circ$), barrel ($12.4^\circ < \theta < 31.4^\circ$), and forward endcap ($130.7^\circ < \theta < 155.1^\circ$). The photon variables have been found to be significantly different depending on which region they originated from (see Figure 5.16). As such, a different `photonMVA` is trained for each region. For each of the ECL regions, the classifier is trained and validated using two independent datasets. 80% of the data is used for training, while the remaining 20% is used for testing. Each dataset has an equal number of genuine and misreconstructed photons. The number of misreconstructed photons limits the size of the data set.

The number of genuine/misreconstructed photons in the training and testing datasets for the `photonMVA` in each region is shown in Table 5.3. It was also found that 7.1% of photons from correctly-reconstructed signal events originated from the backward region, 88.4% were from the barrel region, and 4.5% were from the forward region.

Region	**Training Dataset**	**Testing Dataset**
Backward	67,936	16,985
Barrel	316,707	79,177
Forward	52,484	13,122

Table 5.3: Number of genuine/misreconstructed photons in the training and testing datasets for the `photonMVA` in each region.

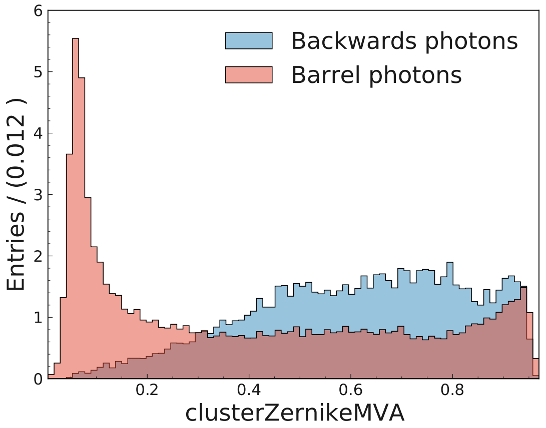


Figure 5.16: Comparison between genuine photons in the backwards (blue) and barrel (red) region for the `clusterZernikeMVA` variable.

5.5.2 Optimised hyperparameters

An improvement in the performance of the `photonMVA` without overfitting was achieved by carefully tuning the default hyperparameters. The number of trees was set to 400 and the number of layers to 6. This corresponds to twice the default values. All other hyperparameters were kept at their default value.

- Depth of the trees=6
- Number of trees=400
- Shrinkage=0.1

To evaluate the performance of the classifier with both default and optimised hyperparameters, the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) curve is used. A ROC curve is a plot that displays the true positive rate against the false positive rate at various classification thresholds. A true positive is defined as an instance where both the model's prediction and the actual class are positive. Conversely, a false positive is an instance where the model incorrectly predicts a positive class when the actual class is negative.

A perfect classifier would have a true positive rate of 1.0, indicating that all signal events are correctly identified, and a false positive rate of 0.0, meaning no signal events are erroneously rejected, across all classification thresholds. A classifier that makes random predictions would have an ROC curve that displays equal true and false positives, resulting in a diagonal line from the bottom left to the top right, as shown in Figure 5.17.

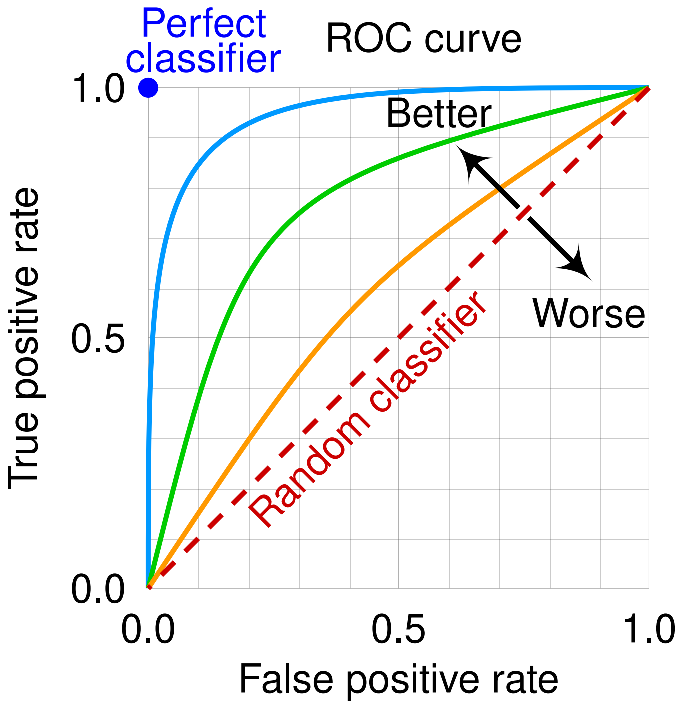


FIGURE 5.17: The Receiver Operating Characteristic (ROC) curve is illustrated with false positives plotted on the x-axis and true positives on the y-axis. The diagonal shows the performance of a random classifier. Three example classifiers are shown in increasing order of performance (orange, green, blue) are shown [17].

In this case, the background rejection is equivalent to the true positive rate, while one minus the signal efficiency represents the false positive rate. Hence, it is more convenient to have the background rejection plotted as a function of signal efficiency. The AUC for both cases is the same. From Table 5.4, it is observed that slightly better performance is achieved with optimised hyperparameters. The performance of the model using optimised parameters in both the training data set and the test data set is presented in Table 5.5. The AUC in the two data sets is identical, indicating that the model was not overfit.

Region	**optimised**	**Default**
Backward	0.922	0.921
Barrel	0.942	0.938
Forward	0.904	0.903

TABLE 5.4: AUC curves in each region for optimised and default FBDT hyperparameters.

Region	**Testing**	**Training**
Backward	0.922	0.922
Barrel	0.942	0.942
Forward	0.904	0.904

TABLE 5.5: AUC curves in each region for optimised FBDT hyperparameters in testing and training data sets.

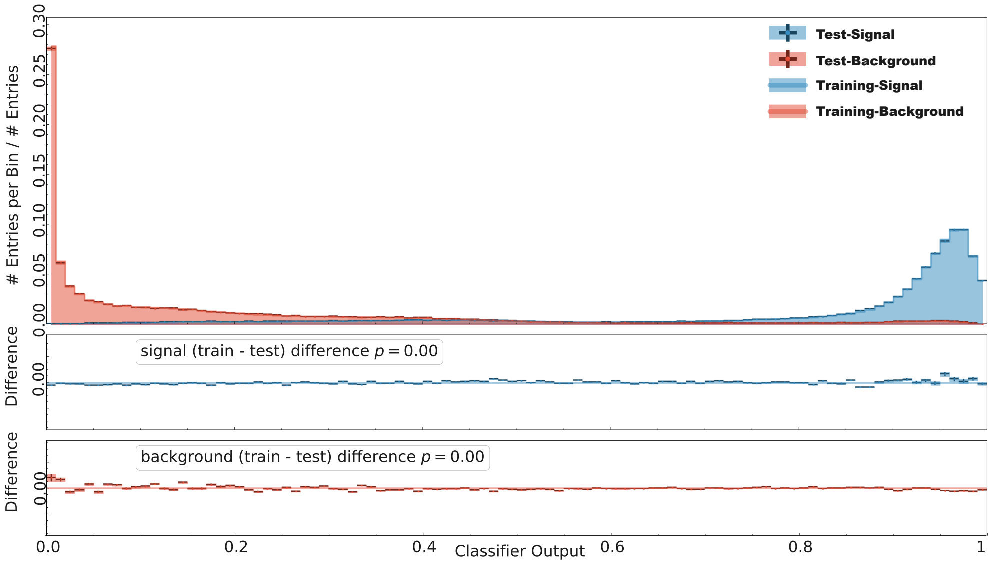


FIGURE 5.18: Comparison of signal and background distribution for the training and testing datasets (top). The difference between the distribution for signal (middle) and background (bottom) are shown below.

To further check for overfitting, the training is applied to the training and testing dataset and the two distributions are compared. If the model is not biased towards the training dataset, both distributions should be almost identical. It was found that on average, the difference between the distributions is almost zero.

5.5.3 PhotonMVA Results

The application of the `photonMVA`to the testing data indicates it is highly effective in suppressing the majority of misreconstructed photons. This is particularly evident in the barrel region, as illustrated in Figure 5.19, which shows the distribution of the `photonMVA` output for both genuine and misreconstructed photons. A clear separation between the two is observable. The results of the training in the backward and forward endcaps is shown in Appendix A.3.

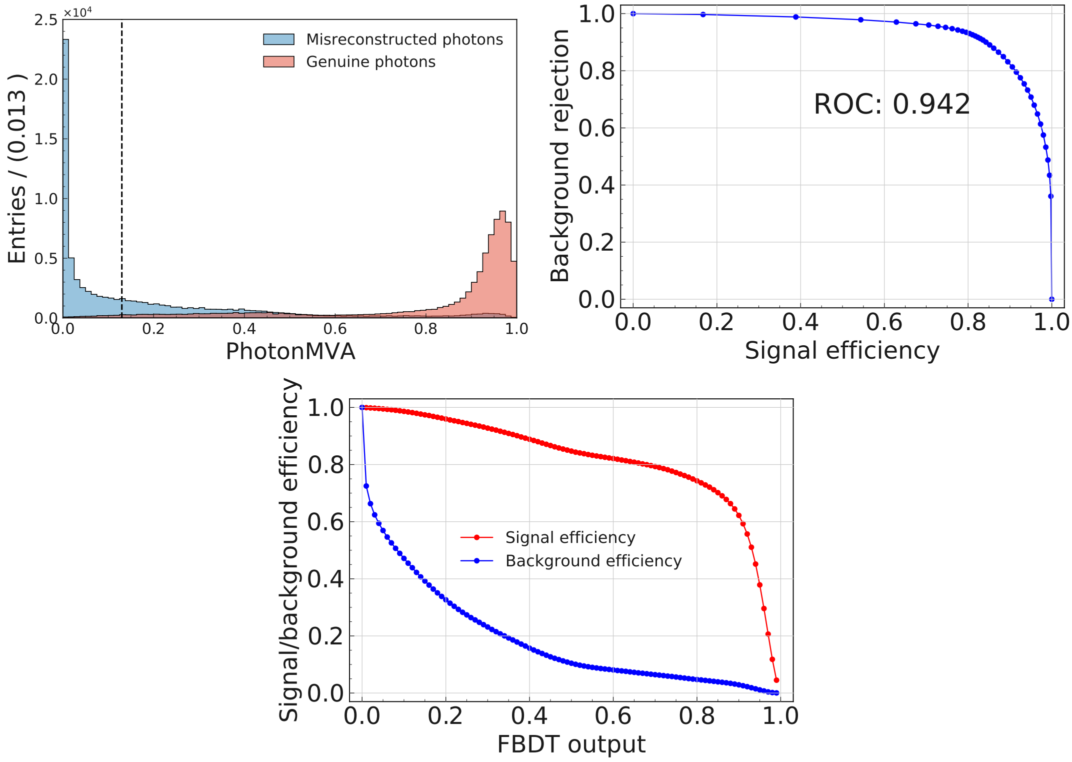


Figure 5.19: Results from the training of photon classifiers in the barrel region. The top left plot shows the distribution of `photonMVA` for both genuine photons (in red) and misreconstructed photons (in blue). The ROC curve is depicted in the top right plot while the bottom plot shows the signal efficiency and background rejection as a function of the `photonMVA` output.

5.5.4 PhotonMVA Validation

The `photonMVA` classifier was trained with signal MC data and validated for real data using the $D^{*+} \to \bar{D}^0(\to K_S^0(\to \pi^+\pi^-)\pi^0)\pi^+$ calibration mode. This specific decay mode is produced from the hadronisation of charm quarks and is expected to provide a substantial sample size in the experimental data due to high levels of continuum background. The impact of the `photonMVA` on both the data and the MC was evaluated by performing a maximum likelihood fit to determine the yield before and after its application. The selection criteria for the calibration mode are described below.

The π^0 reconstruction uses the same final state selections that will be used in the signal decay (see Section 6.1). $\pi^\pm$ candidates were reconstructed from charged particle candidates that were reconstructed within the full CDC polar angle acceptance ($17^\circ < \theta < 150^\circ$), and were located close to the interaction point in the longitudinal ($|dz| < 3.0$ cm) and radial ($|dr| < 0.5$ cm) directions. The reconstructed $\pi^\pm$ candidates were required to have a number of hits in the CDC greater than 20 (`nCDCHits > 20`) to minimise beam background-induced tracks. For the K_s^0 reconstruction, oppositely charged $\pi^{\pm}$ candidates were paired, and it was ensured that they originated from a common space point and had a dipion mass within the range of 0.47–0.53 GeV/c^2. For the reconstruction of D^0, K_s^0 was combined with π^0, and the candidates were required to have masses within the range of 1.82–1.90 GeV/c^2 with momenta in the c.m. frame greater than 2.5 GeV/c. Lastly, D^{*+} candidates were reconstructed by combining the D^0 and π^+ in a kinematic vertex fit. In the kinematic vertex fit, the conservation of energy and momentum and the geometric constraint that all daughter particles originate from a common vertex are used to adjust the measured momenta and vertex position. This adjustment is made in a way that minimises a χ^2 value which represents the difference between the measured and adjusted values, weighted by their uncertainties. This is done to improve the resolution of a decay vertex and the associated particle momenta [74].

The best candidate per event was chosen by selecting the π^0 candidate with the lowest χ^2 value of the mass-constraint diphoton fit. As shown in Figure 5.21, the difference between the mass of the D^+ and D^0, $\Delta M = M_{D^+} - M_{D^0}$, is a powerful discriminator between signal and background. To ensure that D^{*+} candidates were reconstructed with high purity, a requirement of $0.144 < \Delta M < 0.147\,\text{GeV}/c^2$ was imposed. The selections are summarised in Table 5.6 and are depicted in Figure 5.20.

Particle	**Selection**	**Particle retention (%)**
$\pi^{\pm}$	thetaInCDCAcceptance==1	98.7
	nCDCHits > 20	82.7
	$\lvert dz\rvert < 3.0$ cm	94.2
	$\lvert dr\rvert < 0.5$ cm	86.9
	PID > 0.1	90.5
K_s^0	$0.47 < M < 0.53$ GeV/c^2	81.8
D^0	$1.82 < M < 1.90$ GeV/c^2	89.2
D^{*+}	0.144 GeV/$c^2 < \Delta M < 0.147$ GeV/c^2	89.5
	$p_{c.m.} > 2.5$ GeV/c	84.4

Table 5.6: $D^{*+} \to \bar{D}^0(\to K_S^0(\to \pi^+\pi^-)\pi^0)\pi^+$ selections and correctly-reconstructed particle retention. There is significant overlap between selections for each particle, and this list only serves to give a general idea of how much signal is lost during each stage of the reconstruction.

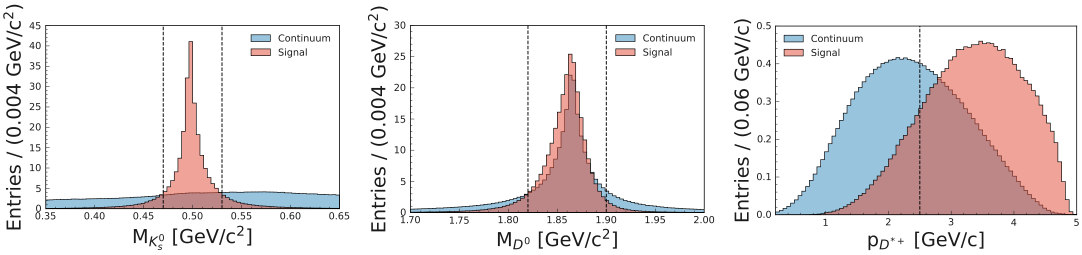


FIGURE 5.20: Comparison between signal (blue) and background (red) for the mass of the K_s^0 (left), mass of the D^0 (middle) and momentum in the center of mass frame of the D^{*+} (right). The black vertical dotted line(s) denotes the applied selection(s).

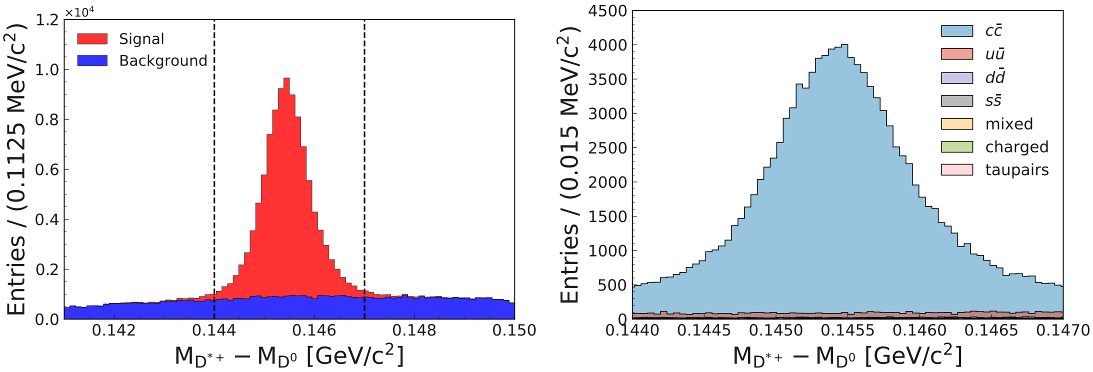


FIGURE 5.21: Stacked histogram of the mass difference between the D^{*+} and D^0 for signal (red) and background (blue) in signal MC (left) and the sample composition of the $D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ produced from hadronisation of charm quarks (right).

5.5.5 Distribution comparison in MC and Data

The `photonMVA` distribution in the MC and 189.9 fb^{-1} of data are compared in Figure 5.22 and found to be in agreement. This agreement is observed for photons originating from both soft and hard π^0. However, the `photonMVA` performs better on hard π^0 as evident by the excess near 0.8 in the barrel region due to soft π^0 as shown in Figure 5.23. This suggests that the `photonMVA` is not restricted to photons from high momentum π^0 mesons in $B^0 \to \pi^0\pi^0$ decays, and is applicable to photons from π^0 mesons of all momentum ranges. It should be noted that there is a data-MC discrepancy at the upper endpoint in the forward region. However, the overall effect is very small since only 4.5% of photons originate from the forward region and only the very highest `photonMVA` outputs would be affected.

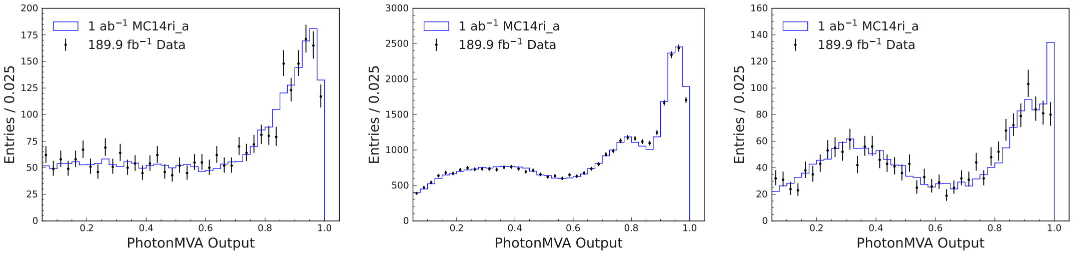


FIGURE 5.22: `PhotonMVA` output for photons originating from the π^0 in the $D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ calibration mode for data (dots) and MC (blocks) for backward (left), barrel (middle) and forward (right) region.

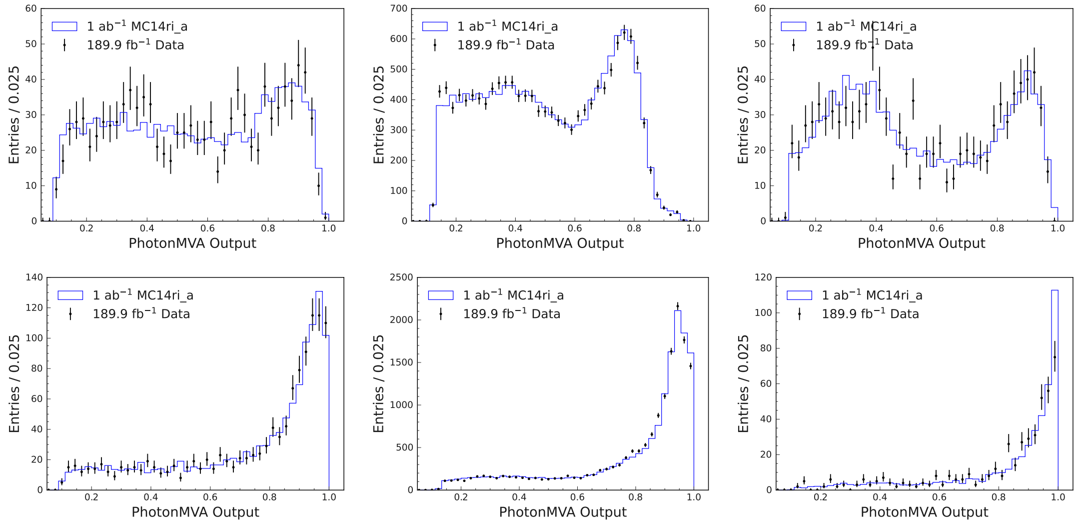


FIGURE 5.23: `PhotonMVA` output for photons originating from soft π^0, defined as $p < 1.5\,\mathrm{GeV}/c$, in the $D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ calibration mode for data (dots) and MC (blocks) for the backward (top left), barrel (top middle) and forward (top right) region. Similarly hard π^0, defined as $p > 1.5\,\mathrm{GeV}/c$, are shown for the backward (bottom left), barrel (bottom middle) and forward (bottom right) region.

5.5.6 Efficiency in Data and MC

The final-state reconstruction is performed with and without implementing a `photonMVA` classification threshold greater than 0.2 in each region. The rationale for this choice will become apparent in Chapter 6. The retention of genuine photons and the rejection of misreconstructed photons resulting from this selection are summarised in Table 5.7.

Region	**`photonMVA` cut**	**Signal γ retention (%)**	**Background γ rejection (%)**
Backward	0.20	95.1	38.0
Barrel	0.20	95.9	32.5
Forward	0.20	95.0	45.9

TABLE 5.7: `photonMVA` cut used for each region with genuine photon retention and misreconstructed photon rejections.

To evaluate the efficiency of the MC-Data, the candidates were restricted to be within the momentum range of the π^0 in $B^0 \to \pi^0\pi^0$ ($p > 1.5\,\mathrm{GeV}/c^2$). The signal yields and backgrounds were then determined using fits to the ΔM distribution. The fit was applied to 1 ab^{-1} of run-dependent MC (MC14ri_a) and 189.9 fb^{-1} of data, as shown in Figure 5.24. The signal was modeled using a double Gaussian, while the background was modeled with a straight line. To validate the fitting procedure, the fitted and truth-matched yields for signal and background were compared, with and without the `photonMVA` (refer to Table 5.8).

The ratio of the yield with and without the `photonMVA` requirement was taken to calculate the efficiency of the requirement. The statistical error of the ratio, $\delta\epsilon$, was calculated using the formula,

$$\delta\epsilon = \frac{1}{n}\left[\sqrt{m(1-\frac{n}{m})} \oplus \sqrt{((\delta n)^2 - n)(1 - \frac{(\delta n)^2 - n}{(\delta m)^2 - m})}\right] \tag{5.5}$$

where m (n) is the yield before (after) the selection. δm (δn) is its statistical error, and $\oplus$ represents a quadratic sum. When the `photonMVA` is applied in MC, the D^{*+} background rejection was found to be $8.9 \pm 0.9\%$ (absolute), and the signal retention was $92.3 \pm 0.2\%$ (absolute). The same PDF parameters from MC are used to fit the experimental data and the background rejection and the signal retention is found to be $13.0 \pm 2.9\%$ and $92.3 \pm 0.4\%$, respectively. This uncertainty is only statistical. The results are summarised in Table 5.9 and show agreement for the signal. Although the background rejection in MC and data does not disagree, the background is a minor component of the fit and would have large systematic uncertainties. It is clear that the overall reduction of events after the application of the `photonMVA` is comparable in MC and data.

	Without photonMVA	**With photonMVA**
Fitted Signal	$6.122 \pm 0.033 \times 10^4$	$5.648 \pm 0.031 \times 10^4$
Truth-match Signal	6.130×10^4	5.677×10^4
Fitted Background	$8.16 \pm 0.23 \times 10^3$	$7.43 \pm 0.22 \times 10^3$
Truth-match Background	8.10×10^3	7.13×10^3

TABLE 5.8: Number of fitted and truth-matched signal and background for MC in reconstructed $D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ decays with and without `photonMVA` applied.

	MC	**Data**
Signal retention	$92.3 \pm 0.2\%$	$92.3 \pm 0.4\%$
Background rejection	$8.9 \pm 0.9\%$	$13.0 \pm 2.9\%$

Table 5.9: Signal retention and background rejection for MC and data in reconstructed $D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ decays due to applied `photonMVA`.

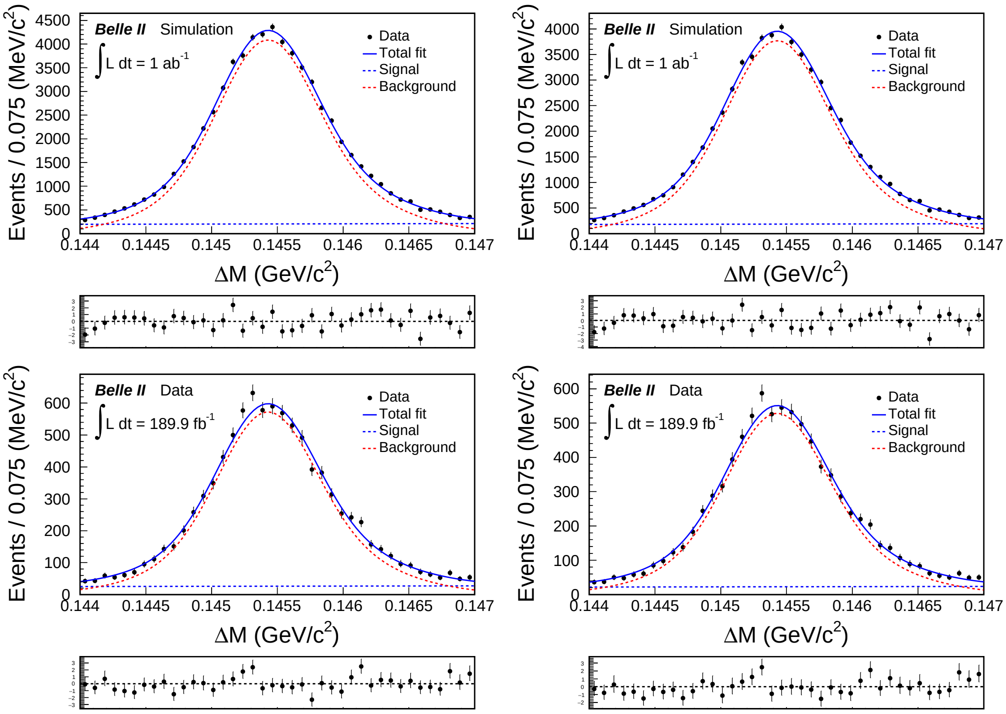


Figure 5.24: Fits for 1 ab^{-1} of MC14ri_a (top) and 189.9 fb^{-1} of data (bottom) for signal (red) and background (blue) without (left) and with (right) the `photonMVA` applied.

Chapter 6

Signal mode reconstruction

This chapter provides a detailed description of the selection and reconstruction of the $B^0 \to \pi^0\pi^0$, and explores the correlation between the M_{bc} and ΔE variables. It also studies the continuum and tau pair events, the two largest sources of background after the initial signal reconstruction.

6.1 Signal mode selection and reconstruction

The process of 'reconstructing' any particle means using the conservation of energy and momentum on its decay products to infer the properties of the original particle. The challenge is that not every pair of photons in the detector comes from a π^0, and not every pair of π^0 mesons comes from a B^0. Therefore, it is essential to apply appropriate selection criteria to filter out particle candidates that are likely not the result of the decay of interest.

The process of reconstructing the signal decay:

$$\begin{aligned} B^0 \to & \pi^0\pi^0 \\ & \quad \pi^0 \to \gamma\gamma, \pi^0 \to \gamma\gamma \end{aligned}$$

begins by creating a list of potential photon candidates, i.e. energy deposits in the ECL not matched to any tracks. Selection criteria, such as the `photonMVA` described previously, are then applied to remove candidates that are unlikely to originate from the signal decay. Then, all possible combinations of two photon candidates are combined to form π^0 candidates. Further selections based on energy, momentum, and invariant mass are applied to reduce the resulting 'combinatorial background'. The combinatorial background here refers to π^0 candidates which are the result of random combinations of any two photons that are unrelated to the signal decay. Finally, all possible combinations of two π^0 candidates are randomly combined into a B^0 candidate, and more selections are applied. The specific selection criteria for the signal mode are described below.

Photon candidates are identified by requiring that the number of crystals in an ECL energy deposition (`clusterNHits`), which can be fractional due to energy splitting with nearby clusters, be greater than 1.5. The cluster timing (`clusterTiming`) is required to be within 200 ns of the estimated event time. This selection range is quite broad to account for the large data-MC discrepancies observed in the modelling of the cluster timing. The most important cut is based on the energy of the photon, as there is an extremely high number of low-energy photon hits in the ECL detector. The energy (`E`) is required to exceed 20.0 MeV in the barrel region, and 22.5 MeV in the forward endcap and backward endcap regions.

The `photonMVA` selection is optimised for all regions using the figure of merit (FOM) defined as $\frac{S}{\sqrt{S+B}}$, where S represents the number of signal events and B represents the number of background events. The $B^0 \to \pi^0\pi^0$ signal is reconstructed using all selections described in Table 6.1, 6.2, and 6.3, except the `photonMVA` requirement, which ranges from 0.1 to 0.9, as shown in the left-hand plot of Figure 6.1. It was observed that the largest FOM lies within the range 0.1 to 0.3. To narrow this range, the same reconstruction was repeated with the `photonMVA` requirement ranging from 0.15 to 0.25, as shown in the right-hand plot of Figure 6.1. A selection of 0.20 was determined to maximise the FOM.

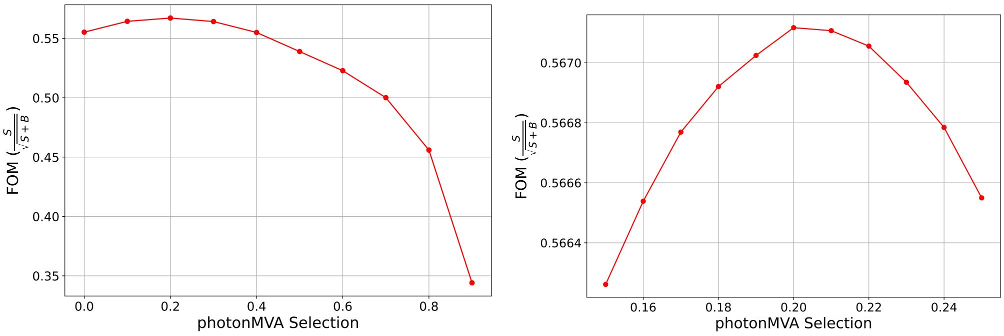


FIGURE 6.1: Figure of merit ($\frac{S}{\sqrt{S+B}}$) as a function of `photonMVA` from 0.1 to 0.9 (left) and 0.15 to 0.25 (right).

Selection	**Signal γ loss (%)**
`photonMVA` $>$ `0.2`	4.0
`E` $>$ `0.03`	1.17
`abs(clusterTiming)` $<$ `200`	0.0330
`clusterNHits` $>$ `1.5`	0.210
`0.2967` $<$ `clusterTheta` $<$ `2.6180`	0.608

TABLE 6.1: Selection for photons and the percentage of correctly reconstructed signal photons lost. There is significant overlap between selections and the loss only provide a general idea of the tightness of each selection.

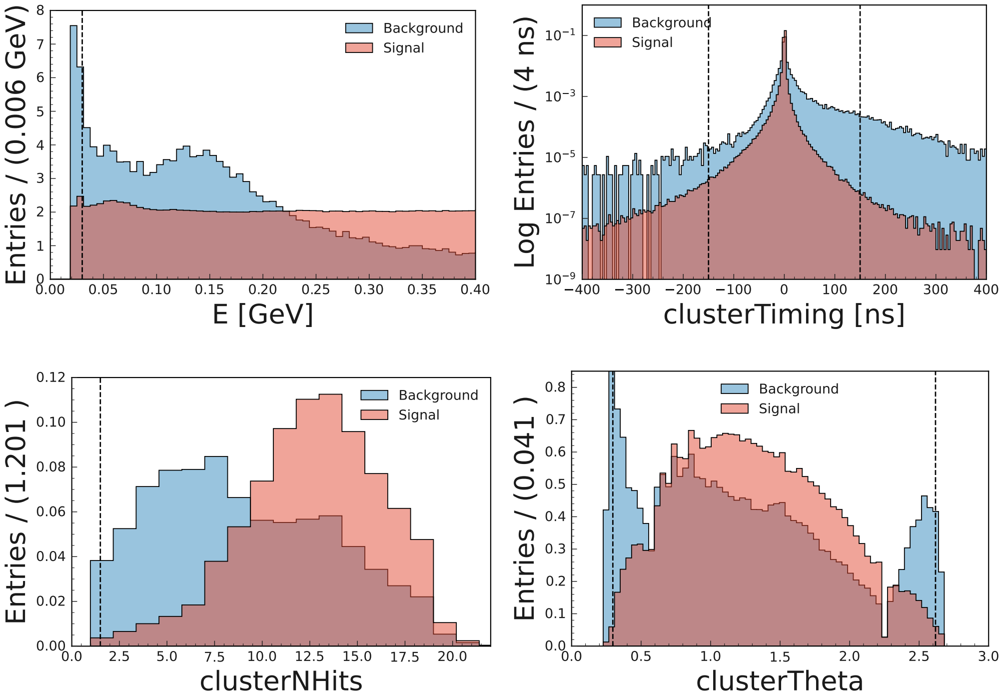


FIGURE 6.2: Comparison between correctly reconstructed (red) and misreconstructed (blue) γ for energy (top left), cluster timing (top right), `clusterNHits` (bottom left) and `clusterTheta` (bottom right). The γ particles are the daughters of π^0 reconstructed with no selection applied.

The selected photons are paired to form π^0 candidates. The π^0 momentum (`p`) in the lab frame is required to be greater than 1.5 GeV/c, and the angle between the momenta of final-state photons (`daughterAngle`) in the lab frame be less than 0.4 radians. The difference in ϕ between final-state photons (`daughterDiffOfPhi`) in the lab frame is also required to be less than 0.4 radians. These requirements suppress the combinatorial background from low-energy photons. The selections for photons and their impact on the percentage of photons that are excluded as a result are summarised in Table 6.1.

The cosine of the helicity angle (`cosHelicityAngleMomentum`), defined as the angle between the higher energy γ direction in the π^0 rest frame and the π^0 direction in the lab frame (see Figure 6.3), is required to be less than 0.99 to reject misreconstructed π^0 mesons, which tend to peak very close to one. The diphoton mass (`M`) is required to be between 0.115 GeV/c^2 and 0.150 GeV/c^2, which corresponds to a range of approximately $+2.0\ \sigma$ and $-2.5\ \sigma$ about the known π^0 mass. The mass requirement is asymmetric as the reconstructed π^0 mass has a slight negative skew due to energy leakage from the ECL calorimeter. The momentum resolution of the π^0 candidates was improved by performing a kinematic fit that constrained their mass to the known value [5]. The selections are shown in Figure 6.4 where the π^0 are the daughters of the B^0 particle candidates reconstructed with no selections applied. The selections for π^0 and their impact on the percentage of π^0 that were excluded as a result are summarised in Table 6.2.

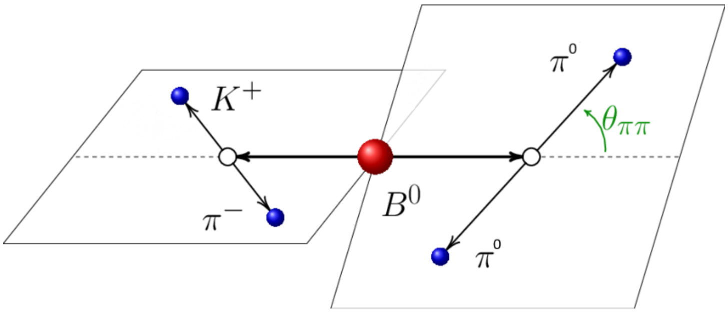


FIGURE 6.3: The helicity angle, $\theta_{\pi\pi}$, is a measure of the angle between the direction of a decay product and the direction of motion of the parent particle in the rest frame of the parent particle.

Selection	**Signal π^0 loss (%)**
`daughterAngle < 0.4`	1.592
`|daughterDiffOfPhi| < 0.4`	1.547
`|cosHelicityAngleMomentum| < 0.99`	0.0004
`p > 1.5`	0.094
`0.115 < M < 0.150`	6.828

TABLE 6.2: Selection for π^0 and the percentage of correctly reconstructed signal π^0 lost. There is significant overlap between selections and the loss only provide a general idea of the tightness of each selection. Here, signal π^0 refers to correctly reconstructed π^0 originating from correctly reconstructed $B^0 \to \pi^0\pi^0$ decays.

Then, all combinations of two π^0 that pass this selection are combined to create signal B^0 candidates. For correctly reconstructed B-meson candidates, ΔE should peak at zero except for resolution. However, observed ΔE distributions peak below zero since energy is lost by electromagnetic interactions in the material before the calorimeter and by energy leakage from the ECL cluster. For instance, when a particle strikes an ECL crystal, the electron shower can leak into the surrounding crystals and may not be recovered. The continuum on the other hand has a flat distribution that decreases

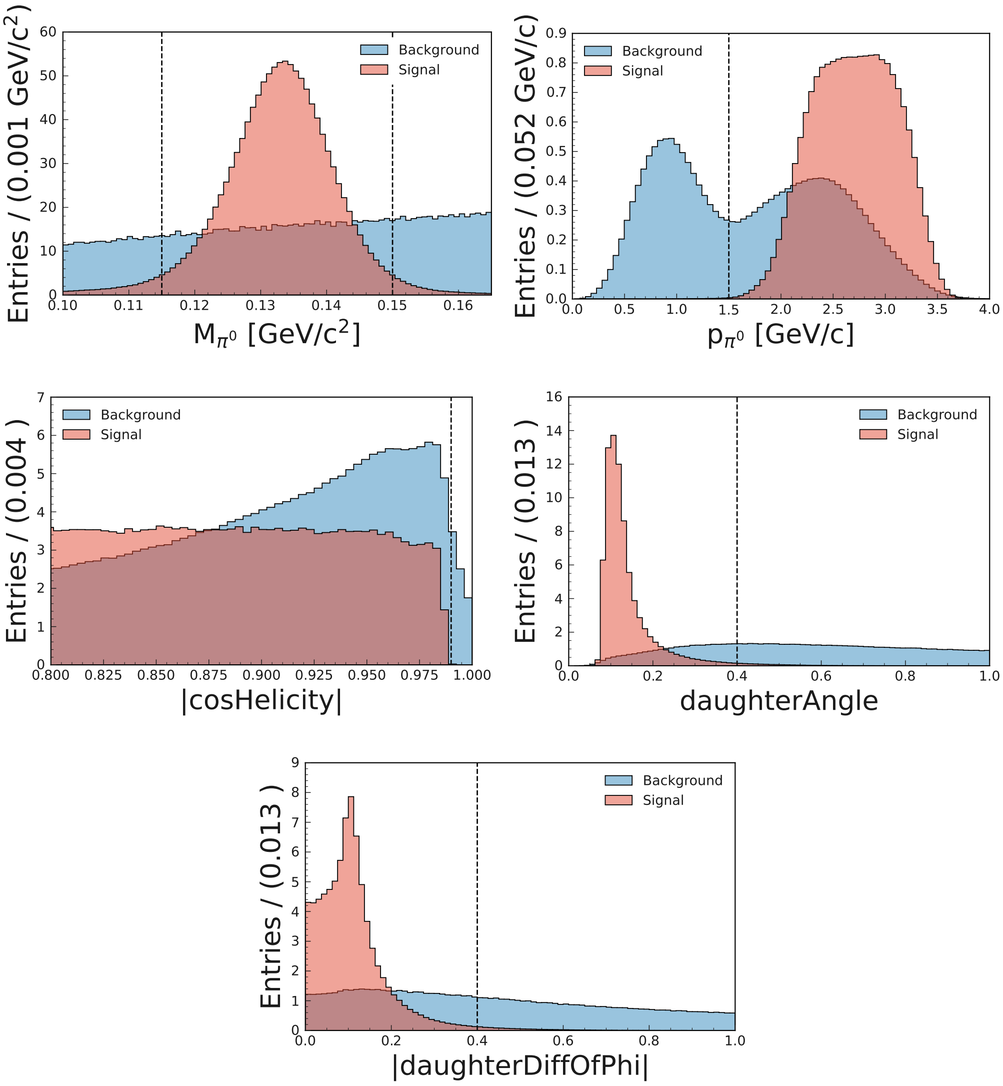


FIGURE 6.4: Comparison between signal (red) and background (blue) π^0 for mass (top left), momentum (top right), `|cosHelicity|` (middle left), `daughterAngle` (middle right) and `|daughterDiffOfPhi|` (bottom). The π^0 are the daughters of $B^0 \to \pi^0\pi^0$ reconstructed with no selection applied. In the momentum plot, the second background peak near 2.5 GeV/c originates from π^0s where one photon is from a genuine hard π^0 and the other is from a tag-side photon.

towards higher values of ΔE. For correctly reconstructed B-meson candidates M_{bc} should peak about the mass of the B meson at 5.28 GeV/c^2 while the continuum has a flat distribution that rapidly decreases to zero at the same point, as shown in Figure 6.5. Candidate B mesons are required to have $5.26 < M_{bc} < 5.29$ GeV/c^2 and $-0.3 < \Delta E < 0.2$ GeV. The selections for B^0 and their impact on the percentage of B^0 that are excluded as a result are summarised in Table 6.3. All selections are summarised in Table 6.4.

Following signal reconstruction, some events can contain multiple B^0 candidates due to

Selection	**Signal B^0 loss (%)**
$5.26 < M_{\mathrm{bc}} < 5.29$ GeV/c^2	0.518
$-0.3 < \Delta E < 0.2$ GeV	3.21

Table 6.3: Selection for B^0 and the percentage of correctly reconstructed signal B^0 lost. Here, signal B^0 refers to correctly reconstructed B^0 originating from correctly reconstructed $B^0 \to \pi^0\pi^0$ decays.

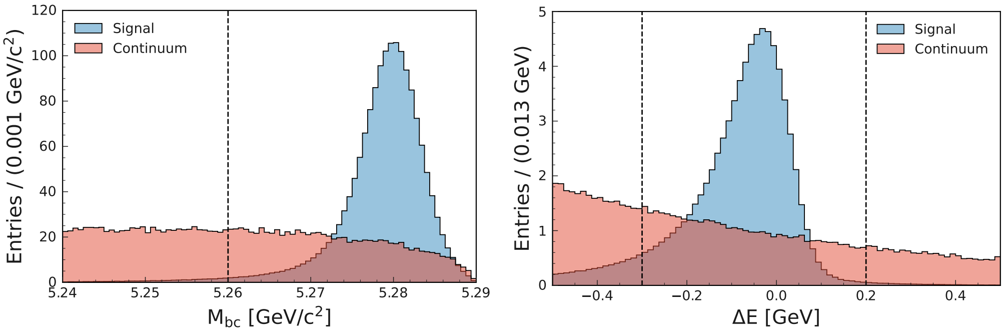


Figure 6.5: Comparison between signal (red) and continuum (blue) π^0 for M_{bc} (left) and ΔE (right)

γ selections	**Signal (10 M)**	**Continuum (1 ab^{-1})**
`photonMVA` $>$ `0.2`	5,922,633	3,187,379
`E` $>$ `0.03 GeV`	5,857,703	2,927,325
`abs(clusterTiming)` $<$ `200 ns`	5,839,798	2,888,862
`clusterNHits` $>$ `1.5`	5,839,401	2,887,834
`0.2967`° $<$ `clusterTheta` $<$ `2.6180`°	5,793,481	2,823,395
π^0 selections		
`daughterAngle` $<$ `0.4`	5,293,971	2,368,319
`|daughterDiffOfPhi|` $<$ `0.4`	5,236,740	2,320,196
`|cosHelicityAngleMomentum|` $<$ `0.99`	5,236,740	2,320,196
`p` $>$ `1.5` GeV/c	5,236,740	2,320,188
`0.115` $<$ `M` $<$ `0.150` GeV/c^2	4,849,344	1,994,135
B^0 selections		
$5.26 < M_{\mathrm{bc}} < 5.29$ GeV/c^2	4,795,532	565,762
$-0.3 < \Delta E < 0.2$ GeV	4,634,033	289,581

Table 6.4: Number of surviving events after each $B^0 \to \pi^0\pi^0$ selections for self-generated signal (10 M) and 1 ab^{-1} of continuum.

the large number of reconstructed π^0 candidates. The $B^0 \to \pi^0\pi^0$ signal, consists of two high momentum π^0 that can only originate from a direct B^0 decay. Hence, the primary source of misreconstruction is from the misreconstructed π^0 due to beam background, or a π^0 from the tag side. Such high energy π^0 particles are quite rare, with only about 0.1-1% of B^0 decays having π^0 as direct daughters. In these rare cases, it is difficult to identify which two π^0 to use to reconstruct the B^0.

Fortunately, the vast majority of events (98.66%) contain just one candidate. For such events, the average multiplicity is 2.03 candidates per event. The candidate with the minimum sum of the absolute deviations of the reconstructed π^0 masses from the known value ($|dM(\pi_1^0)| + |dM(\pi_2^0)|$) is chosen as the best candidate. This requirement is 56% efficient in selecting the correct B^0 candidate.

After reconstruction and best candidate selection, the reconstruction efficiency is 46.4% and the purity is 98.6%. Assuming $\mathcal{B}(B^0 \to \pi^0\pi^0) = 1.59 \times 10^{-6}$, one expects 152 ± 25 signal events in 189.9 fb^{-1}. This number includes self-cross feed, which is expected to be 1.78 ± 0.29, and is currently considered negligible. The total number of events expected for the signal, continuum, $B\overline{B}$, and tau pairs is summarised in Table 6.5.

Luminosity (fb^{-1})	**Signal**	**Continuum**	**$B\overline{B}$**	**Tau pairs**
189.9	152 ± 25	54,992	113	16,139
1000	798 ± 132	289,581	596	84,989

Table 6.5: Number of expected events for signal, continuum, $B\overline{B}$, and tau pairs after the $B^0 \to \pi^0\pi^0$ selections but before continuum suppression for 189.9 fb^{-1} and 1000 fb^{-1}. The uncertainty in the signal is due to the uncertainty of the world-averaged branching fraction.

6.1.1 `photonMVA` sculpting check

Ideally, the shape of the distributions of the fitting variables, M_{bc} and ΔE, should remain unchanged throughout the analysis. However, these variables can be modified or 'sculpted' due to the selection criteria or the analysis method used. This can lead to bias in the distributions, which can mimic a signal where there is none or hide a signal that is actually present.

In this case, sculpting occurs when an overfitted model starts to influence the distribution of a variable that should be uncorrelated with the variable the model is attempting to predict. For example, the model can 'learn' that events with ΔE values close to zero are likely to be signal events. This can lead to a bias towards events with ΔE values close to zero, resulting in peaks in both signal and background distributions. This usually happens if the input variables used to train BDT are highly correlated with the fitting variables. To ensure that the `photonMVA` does not sculpt the distribution of M_{bc}, ΔE, and T_c, each variable was examined before and after the application of the classifier. As shown in Figure 6.6, the distributions are nearly identical, indicating that `photonMVA` is likely not overfitting the fitting variables.

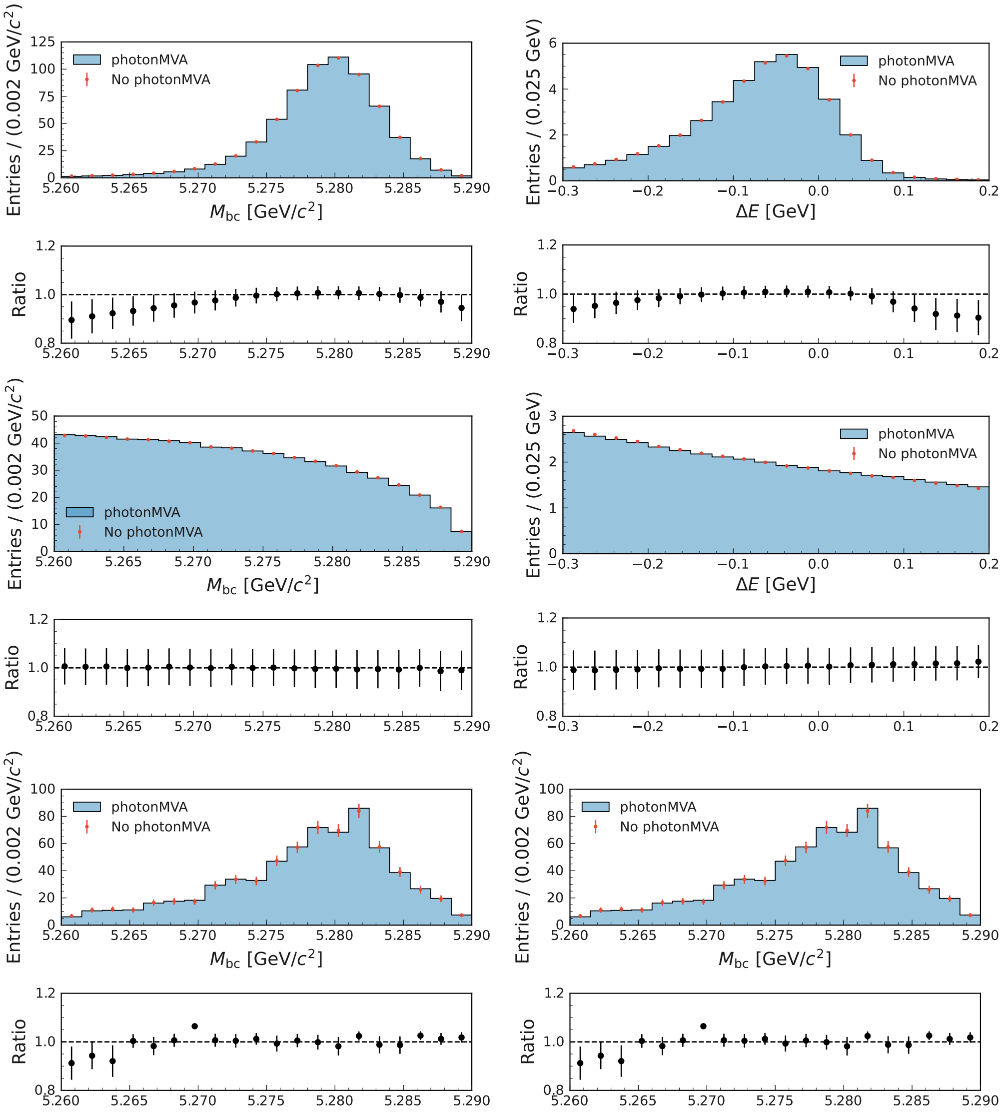


FIGURE 6.6: M_{bc} (left) and ΔE (right) distribution for signal (top), continuum (middle) and $B\overline{B}$ (bottom) with (blocks) and without (dots) the `photonMVA` applied.

6.1.2 Correlation between M_{bc} and ΔE

In principle, ΔE and M_{bc} should be uncorrelated as the reconstructed momentum and energy are independent of each other. However, unlike charged particles, which can be measured with multiple detectors, photons can only be detected and measured using the ECL. Since the energy, and hence the momentum, is measured using only the ECL, there is a slight correlation between ΔE and M_{bc} for correctly reconstructed signal events. This is illustrated in the top left plot of Figure 6.7. The linear correlation is determined to be -8.8%. However, the correlation is actually non-linear, so this value merely serves as an indicator of the degree of correlation.

To investigate the correlation between M_{bc} and ΔE, correctly reconstructed signal events are divided into 4 bins of M_{bc}, with bins defined as [5.26,5.277], [5.277,5.280], [5.280, 5.282] and [5.282,5.29] GeV/c^2, as shown in the top right plot of Figure 6.7.It is evident from the bottom plot of Figure 6.7 that the relationship between ΔE and M_{bc} is non-linear, making it challenging to model analytically. For example, the distribution of ΔE in the tail region of M_{bc} is significantly broader than at the peaks.

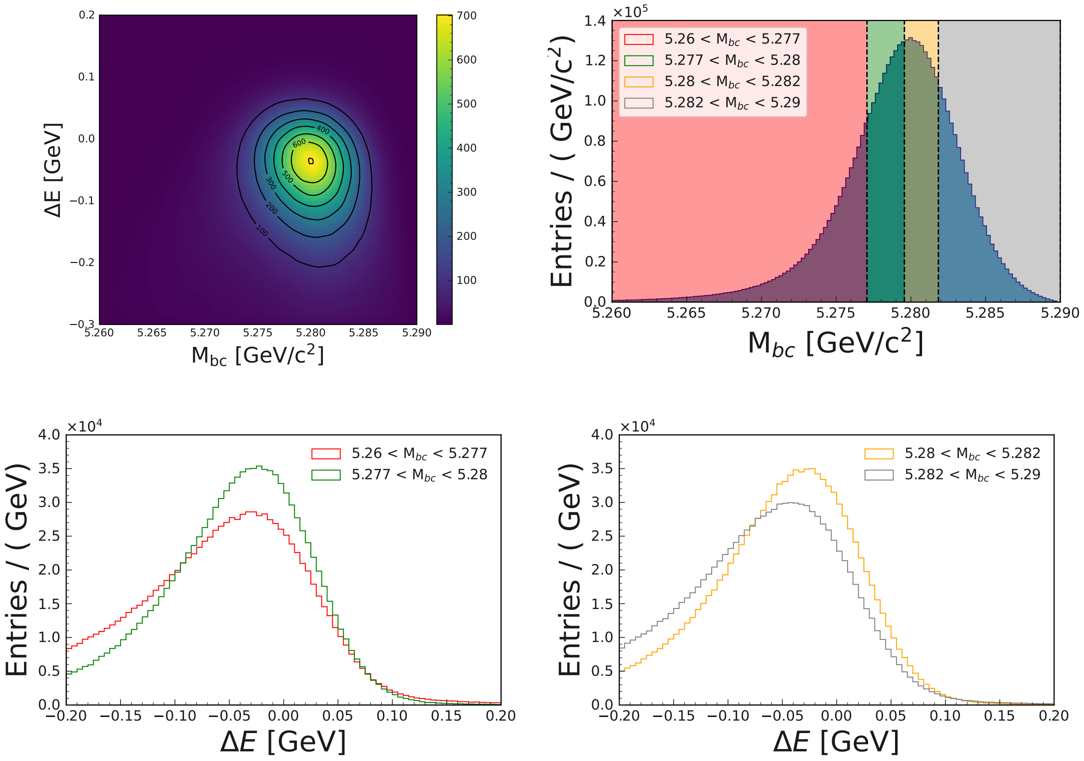


FIGURE 6.7: 2D scatter plot for $M_{\rm bc}$ and ΔE (top left). M_{bc} is divided into four regions with roughly equal number of events (top right), the endpoints are defined as [5.26,5.277] (red), [5.277,5.280] (green), [5.280, 5.282] (yellow) and [5.282,5.29] (gray). ΔE distribution for the red and green region (bottom left), and yellow and gray regions (bottom right).

6.2 Other Background

In this analysis, there are two primary sources of background besides the continuum background: $e^- e^- \to \tau^+ \tau^-$ events, which are known as tau pairs and non-signal $B\overline{B}$ events.

6.2.1 Tau pairs

The cross section of $e^-e^- \to \tau^+\tau^-$ (tau pairs) is 0.919 ± 0.003 nb [75] and is comparable to $e^-e^- \to \Upsilon(4S) \to B\overline{B}$ processes which have a cross section of 1.100 nb. Typically, the contribution of these tau pairs is small when the hadronic skim is applied, as only 15% of the tau pairs are retained due to the requirement on the minimum number of charged tracks (nTracks > 2). However, this analysis does not apply the hadronic skim because it would result in the loss of 23.1% of the signal. This choice leads to a significant number of $e^-e^- \to \tau^+\tau^-$ events being misreconstructed due to π^0 in $\tau^+ \to \pi^+\pi^0\nu_\tau$ decay being misreconstructed as signal. This is a significant source of background as the branching fraction for this decay is quite substantial at 0.2549 ± 0.0009. Approximately 16,000 tau pairs are predicted to be reconstructed in 189.9 fb^{-1} of data. As shown in Table 6.5, this constitutes approximately 30% of the expected continuum events.

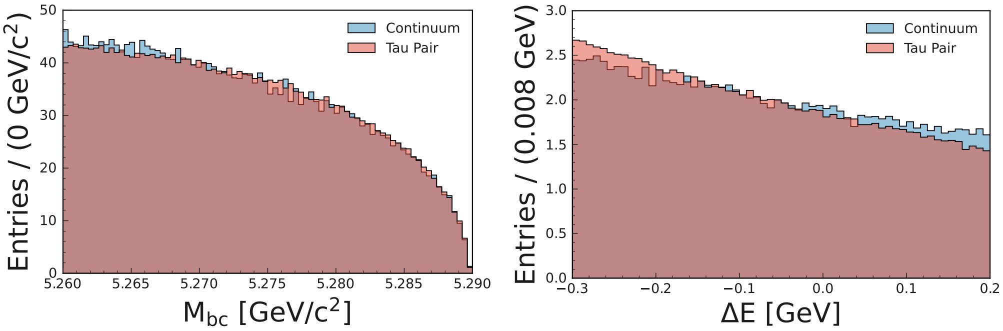


FIGURE 6.8: Comparison between tau pairs (red) and continuum (blue) for the variables M_{bc} (left) and ΔE (right).

6.2.2 $B\bar{B}$ Background

Backgrounds arising from non-signal B decays are rare B decays that proceed via $b \to u, d, s$ transitions, and these are denoted as $B\overline{B}$. The expected number of $B\overline{B}$ decays is estimated by studying two 1 ab^{-1} MC data sets: the $B^0\bar{B}^0$ sample and the B^+B^- sample. Together, these two datasets form what is known as the 'mixed' sample. To investigate the $B\overline{B}$ background, events containing $B^0 \to \pi^0\pi^0$ decays are excluded from the 'mixed' sample. After the signal events are removed, 596 $B\overline{B}$ events are identified in 1 ab^{-1} of MC data. Of these, 539 events (90.4%) of the $B\overline{B}$ decays originate from B^+B^- decays, while the remaining 57 events (9.6%) are due to $B^0\bar{B}^0$ decays.

Further study showed that in the 'charged' sample, 95.9% of decays are due to $B^+ \to \rho^+\pi^0$ decays where the charged pion from the subsequent $\rho^+ \to \pi^+\pi^0$ decay is not reconstructed. In the 'mixed' MC sample, the decays are mainly due to $B^0 \to K_S^0(\to \pi^0\pi^0)\pi^0$ decays where one of the π^0 is unreconstructed. The breakdown of contributing decays for mixed and charged decays is shown in Table 6.6 and Table 6.7, respectively.

Decay	**Count**	**Percentage (%)**
$\boldsymbol{K^0\pi^0}$	18	32.1
2 daughters	10	17.9
$\boldsymbol{\pi^0\pi^0\pi^0}$	9	16.1
$\boldsymbol{\rho^+\rho^-}$	7	12.5
$\boldsymbol{\pi^0\eta}$	2	3.6
Other Decays	10	17.8

Table 6.6: Decays contributing to the mixed $B\overline{B}$ sample along with the number of events and percentage of total. If the decay is not found, the number of daughters is shown.

Decay	**Count**	**Percentage (%)**
$\boldsymbol{\rho^+\pi^0}$	517	95.9
2 daughters	10	1.9
Other Decays	12	2.2

Table 6.7: Decays contributing to the charged $B\overline{B}$ sample along with the number of events and percentage of total. If the decay is not found, the number of daughters is shown.

To simplify the modelling of the 'mixed' background, it was assumed that it follows the same distribution as the dominant decay $B^0 \to K_S^0(\to \pi^0\pi^0)\pi^0$. Similarly, the 'charged' background was modelled using the dominant decay $B^+ \to \rho^+\pi^0$. The $B\overline{B}$ background peaks at similar values of M_{bc} and T_c, but ΔE is shifted to negative values due to the missing energy of the unreconstructed particle, as shown in Figure 6.9. There are an insufficient number of these rare decays in the 'mixed' sample to accurately model the $B\overline{B}$ background. Therefore, for each of the $B^0 \to K_S^0(\to \pi^0\pi^0)\pi^0$ and $B^+ \to \rho^+\pi^0$ decays, 1×10^6 events was locally generated. This a sufficient number of events to accurately estimate the reconstruction efficiency and model the background.

The signal reconstruction and continuum suppression was applied to the 1 ab^{-1} 'mixed' sample and this locally generated $B\overline{B}$ data set. From the 'mixed' sample, 641 $B\overline{B}$ events were reconstructed. This corresponds to approximately 113 $B\overline{B}$ events in 189.9 fb^{-1}. The reconstruction results for the locally generated $B\overline{B}$ sample are presented in Table 11.4. From the table, it can be observed that $B^+ \to \rho^+\pi^0$ and $B^0 \to K_s(\to \pi^0\pi^0)\pi^0$

together contribute a total of 119 ± 14 events, which effectively accounts for the entire $B\overline{B}$ background.

$\boldsymbol{B\bar{B}}$ **decay**	**Branching ratio**	$\mathcal{A}_{CP}$	**Efficiency**	**Estimation**
$B^+ \to \rho^+\pi^0$	$1.09 \pm 0.14 \times 10^{-5}$	0.02 ± 0.11	4.60%	109 ± 14
$B^0 \to K_s(\to \pi^0\pi^0)\pi^0$	$3.04 \pm 0.15 \times 10^{-6}$	0.00 ± 0.13	1.71%	10.7 ± 0.6

TABLE 6.8: Estimation of number of events in 189.9 fb^{-1} and expected $\mathcal{A}_{CP}$ for $B^+ \to \rho^+\pi^0$ and $B^0 \to K_s(\to \pi^0\pi^0)\pi^0$ from efficiency as determined by `MC14ri_a`.

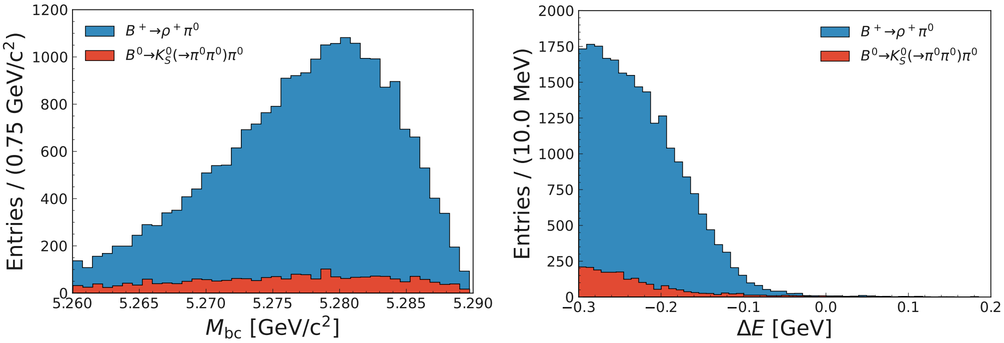


FIGURE 6.9: Stacked histogram for expected $B\overline{B}$ background in 50 ab^{-1} dataset for $M_{\rm bc}$ (left) and ΔE (right) for $B^+ \to \rho^+\pi^0$ (blue) and $B^0 \to K_s(\to \pi^0\pi^0)\pi^0$ (red).

Chapter 7

Continuum Suppression

At energies near the $\Upsilon(4S)$, the cross section for producing $u\bar{u}, d\bar{d}, c\bar{c}$, and $s\bar{s}$ quark-antiquark pairs is three times greater than the $B\overline{B}$ cross section. These quark-antiquark pairs subsequently undergo hadronisation, resulting in the formation of large showers of particles. This 'continuum' background contributes substantially to the uncertainty in $\mathcal{A}_{CP}$ and $\mathcal{B}$ measurements. The $B^0 \to \pi^0\pi^0$ decay, in particular, exhibits a large continuum background due to its small branching ratio.

Fortunately, the majority of continuum events can be suppressed by using the spatial and momentum distribution, or the event shape, in the in the centre-of-mass frame of the particles in the decay. $B\overline{B}$ pairs, being produced in an almost stationary state, emit decay products uniformly in all directions, resulting in a spherically symmetric event shape. In contrast, continuum events are produced from the hadronisation of back-to-back quark pairs with high momentum. These quark pairs are initially moving almost directly away from each other along the beam pipe due to the conservation of momentum. Hence, the hadrons produced in the fragmentation do not deviate much from the flight direction of the quark pairs. This causes the continuum to have a back-to-back 'jet-like' event shape. The spherical and jet-like nature of signal and continuum events, respectively, is shown in Figure 7.1.

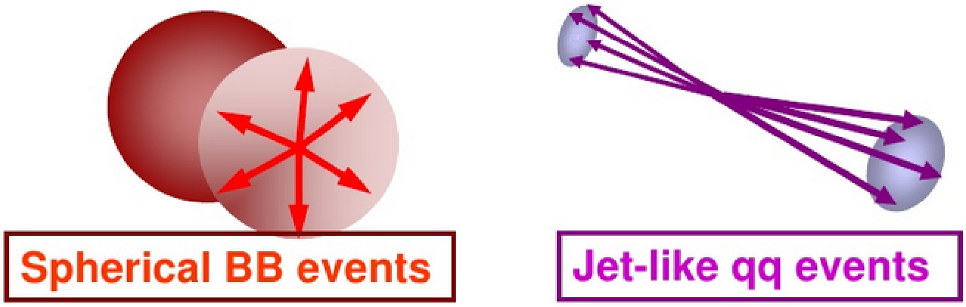


FIGURE 7.1: Diagram of difference in event shape for signal and continuum

7.1 Topological variables

To discriminate against the continuum background, the Fast Boosted Decision Tree (FBDT) classifier is used to combine 24 variables associated with the event topology, which are known to provide statistical discrimination between the B-meson signal and the continuum background. Topological discrimination variables are constructed to quantify the event shape of particle decays in the centre-of-mass frame. These topological variables are also required to have a correlation with ΔE and M_{bc} that is less than 5% to prevent unintentionally sculpting the distributions.

- **Thrust variables:** The thrust variable is used to quantify how 'jet-like' an event is. The thrust axis $\vec{T}$ is the unit vector along which the projection of all the momenta of the final-state particles is maximised. The thrust is defined as

  $$T = \frac{\sum_{i=1}^{N} |\vec{T} \cdot \vec{p_i}|}{\sum_{i=1}^{N} |\vec{p_i}|} \tag{7.1}$$

  where $\vec{p_i}$ is the momentum of the i-th particle for a collection of N particles. Due to the spherical topology of the $B\overline{B}$ events, the signal distribution is expected to be uniform, while the jet-like nature of the continuum events leads to a distribution that peaks at 1.

  Only the thrust of the rest of event (ROE), known as thrustOm, is used. The ROE refers to all particles produced in the event but are not directly involved in signal B candidate reconstruction. The thrust of the signal B candidate is not used as it was observed have a 30% correlation with M_{bc}. The thrust variable, thrustOm, for signal and continuum is shown in Figure 7.2.

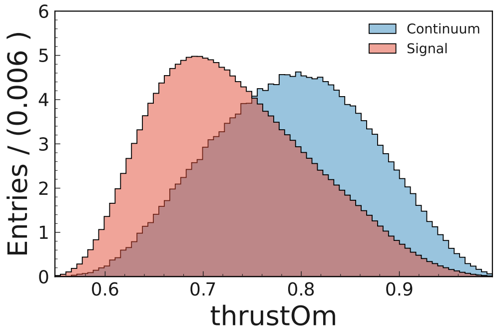


Figure 7.2: Comparison between continuum (blue) and signal (red) events for thrustOm.

- **Kakuno-Super-Fox-Wolfram moments:** The Fox-Wolfram moments [76] are a set of dimensionless numbers that can be used to describe the spherical symmetry of an event. These rotationally invariant moments are derived from the pairwise angles between all tracks in an event.

  $$H_l = \sum_{ij} \frac{|\mathbf{p}_i||\mathbf{p}_j|}{E^2_{\text{event}}} P_l(\cos\theta_{ij}) \tag{7.2}$$

  where p_i and p_j are the momenta of the i^{th} and j^{th} particle in the event, E_{event} is the total energy of the event, θ_{ij} is the angle between the momenta of the i^{th} and j^{th} particle, and P_l is the l^{th} order Legendre polynomial ($l \in [0, 4]$),

  $$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} \left[(x^2 - 1)^l\right] \tag{7.3}$$

  The sum runs over all pairs of particles in the event. Typically, these moments are normalised to H_0, with the second moment, R_2 (see Figure 7.3), defined as:

  $$R_2 = H_2/H_0$$

  While higher moments can offer additional shape discrimination, their power diminishes rapidly. Consequently, they are not utilised in this analysis.

  The discrimination power of the Fox-Wolfram moments diminishes if particles are missing. To account for missing particles, a more sophisticated version of the FW moments, the Kakuno-Super-Fox-Wolfram (KSFW) moments H^g_{xl} was developed [77]. The variable g in the superscript describes the particles considered for the double sum. If one of the sums runs over B candidate daughters and the

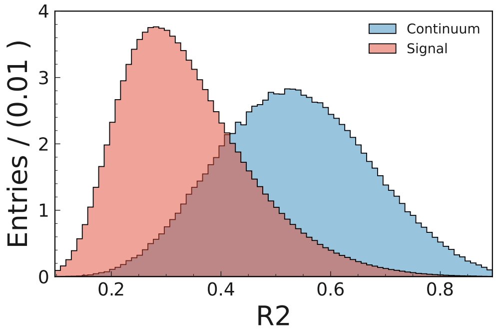


FIGURE 7.3: Comparison between continuum (blue) and signal (red) events for R2.

other sum runs over particles from the rest of the event, the variable g is denoted as 'so'. If both sums consider particles in the ROE, the variable g is denoted as 'oo'. The variable x depends on whether the considered particles are charged 0, neutral 1, or missing 2. It is only used with the superscript 'so'. For example, the KSFW moments H_{02}^{so} is the sum over the B candidate daughters and the ROE particles, where the considered particles are charged, and the second Legendre polynominal is used.

There are a total of 16 possible KSFW moments. However for odd l, one finds that $H_{1l}^{so} = H_{2l}^{so} = 0$. In addition, the KSFW variable, H_0^{oo}, was observed to have a 35.4% correlation with ΔE. Therefore, only the following 13 KSFW moments are used:

$$H_{00}^{so}, H_{02}^{so}, H_{04}^{so}, H_{10}^{so}, H_{12}^{so}, H_{14}^{so}, H_{20}^{so}, H_{22}^{so}, H_{24}^{so}, H_1^{oo}, H_2^{oo}, H_3^{oo}, H_4^{oo}$$

The two most discriminating KSFW moments for signal and continuum are shown in Figure 7.4.

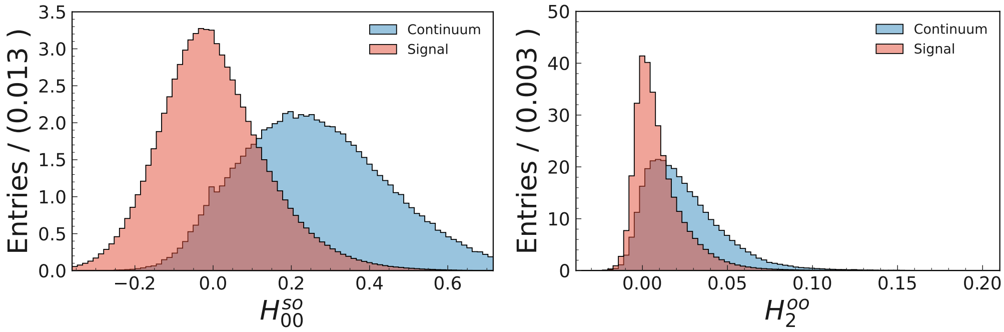


FIGURE 7.4: Comparison between continuum (blue) and signal (red) events for H_{02}^{so} (left) and H_2^{oo} (right).

- **Cleo Cones:** The CLEO cones describe the sum of the absolute value of the momentum of all final-state particles in the event around the thrust axis into concentric cones at angular intervals of 10 degrees. This results in 9 concentric cones, referred to as CleoCone(i), where $i \in \{1, 2, ..., 9\}$. The Cleo Cones are illustrated in the left-hand plot of Figure 7.5. In this analysis, all but the CleoCone(1) were used to train the continuum suppression classifier. It was observed that CleoCone(1) had a 35.4% correlation with ΔE. The most discriminating Cleo Cone is shown in the right-hand plot of Figure 7.5.

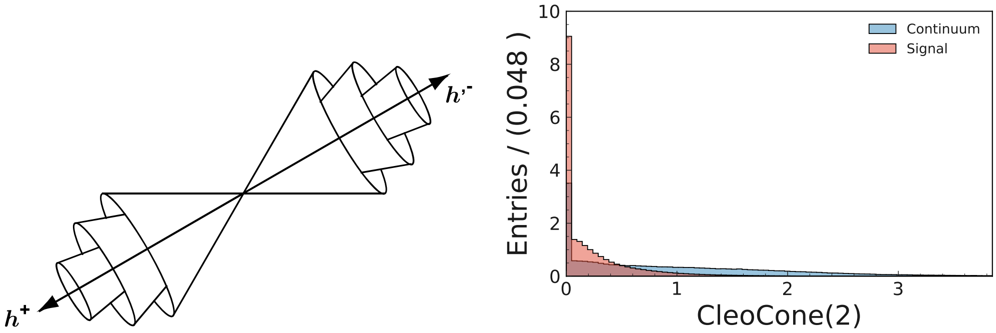


FIGURE 7.5: Illustration of the first three CLEO cones for a $B^0 \to h^+ h^{'-}$ candidate [6] (left) and a comparison between continuum (blue) and signal (red) events for CleoCone(2).

- **cosTBTO:** This variable is the cosine of the angle between the thrust axis of the signal B candidate, TB, and the thrust axis of the ROE, TO. In signal events, the momenta of the final-states particles from the B^0_{sig} and the ROE are uniformly distributed. As a result, the absolute value of the cosine of the angle between them follows a uniform distribution. Conversely, for continuum and tau pair events, the momenta of the particles are strongly collimated in the same direction as the jets. Consequently, the cosine of the angle of between the signal B candidate and ROE is strongly peaked at 1 (see Figure 7.6).

  A similar variable cosTBz, the cosine of the angle between the thrust axis of the signal B candidate and the z-axis, was considered but not included. It was observed that this variable had a -10.8% correlation with ΔE.

Two other variables, the missing mass squared and transverse energy, commonly used to suppress the continuum were considered but not included. The squared missing mass is defined as:

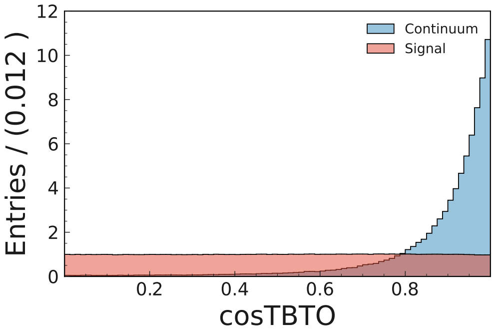


Figure 7.6: comparison between continuum (blue) and signal (red) events for cosTBTO.

$$M_{\text{miss}}^2 = \left(E_{\Upsilon(4S)} - \sum_{n=1}^{N_t} E_n \right)^2 - \left(\sum_{n=1}^{N_t} |p_n| \right)^2 \tag{24}$$

where $E_{\Upsilon(4S)}$ is the energy of the $\Upsilon(4S)$, and E_n and p_n are the energy and momentum of the n-th particle, respectively. It was observed that this variable had a -5.8% correlation with ΔE. The transverse energy

$$E_t = \sum_{n=1}^{N} |(P_t)n|$$

is the scalar sum of the transverse momenta of all particles in the event. It was observed that this variable had a 14.1% correlation with ΔE. The correlations of all considered continuum variables to M_{bc} and ΔE is shown in Table 7.1. Plots of the distribution of signal and continuum for the remaining continuum suppression variables can be found in Appendix B.1.

7.2 Training with simulated data vs sideband

Two methods of obtaining continuum samples to train the continuum suppression are examined; one utilises 1 ab^{-1} of MC, while the other uses a portion of the experimental data known as sideband. It consists of events that satisfy all $B^0 \to \pi^0\pi^0$ selection criteria but are in a signal-depleted region $5.22 < M_{bc} < 5.27\,\text{GeV}/c^2$ and $0.1 < \Delta E < 0.5\,\text{GeV}$. The MC method has the advantage of significantly larger number of events and the benefit of residing in the signal region, while the sideband region, being data-driven, avoids any continuum data-MC discrepancy and also includes tau pairs.

Variables	M_{bc} (%)	ΔE (%)
R2	0.543	3.91
cosTBTO	-0.111	1.15
cosTBz	-1.96	**-10.8**
thrustBm	**30.2**	4.31
thrustOm	0.288	-0.607
E_t	1.25	**14.1**
M^2_{miss}	0.336	**-5.83**
H^{so}_{00}	-0.117	4.38
H^{so}_{02}	0.0283	1.76
H^{so}_{12}	-0.243	0.277
H^{so}_{04}	-0.0785	-0.19
H^{so}_{10}	-0.119	3.2
H^{so}_{14}	-0.363	-0.799
H^{so}_{20}	-0.42	1.43
H^{so}_{22}	-1.21	-1.74
H^{so}_{24}	-0.0197	0.108
H^{oo}_{0}	-0.387	**7.47**
H^{oo}_{1}	0.117	0.313
H^{oo}_{2}	0.501	1.59
H^{oo}_{3}	-0.0452	0.203
H^{oo}_{4}	0.294	0.693
CleoCone(1)	2.5	**35.4**
CleoCone(2)	-1.34	0.443
CleoCone(3)	-0.054	0.905
CleoCone(4)	0.0429	0.983
CleoCone(5)	0.0272	0.73
CleoCone(6)	-0.0662	0.143
CleoCone(7)	-0.0565	-0.525
CleoCone(8)	-0.205	-0.745
CleoCone(9)	-0.00112	-0.775

TABLE 7.1: Variables considered for training the continuum suppression classifier and their correlation to the M_{bc} and ΔE. Variables with correlation above 5% are shown in bold and excluded.

The FBDT described in subsection 5.3.6 is used as the classifier. It is trained to identify statistically significant signal and background features using data split into a training set and a testing set, with 80% of the data used for training and the remaining 20% reserved for testing. Each data set contains an equal number of correctly reconstructed signal and continuum background events. This balanced dataset helps to prevent the classifier from developing a bias towards the class that has more samples.The MC truth variable described in section 5.2, `IsSignal`, is equal to 1 for correctly reconstructed signal events and 0 for the continuum background. It is used to define the training target.

The number of events in the training and testing datasets for the MC and sideband is

	Training Dataset	**Testing Dataset**
MC	463,330	115,834
Sideband	185,196	46,300

Table 7.2: The number of events for the MC/sideband training and testing datasets.

presented in Table 7.2. Two continuum suppression multivariate (CSMVA) classifiers are trained; one for the MC and the other using the sideband region. Then each trained classifier is applied to both sets of testing samples. The results are summarised in Table 7.3. The distribution of the classifier output, the signal/background efficiency, and the ROC curve for both CSMVA classifiers when evaluated on the sideband testing data are shown in Figure 7.7.

	MC Training	**Sideband Training**
Sideband Dataset	0.911	0.948
MC Datasets	0.930	0.923

Table 7.3: The AUC for the MC and sideband training on the MC and sideband testing dataset. The difference in performance due difference in the size of the training dataset is negligible.

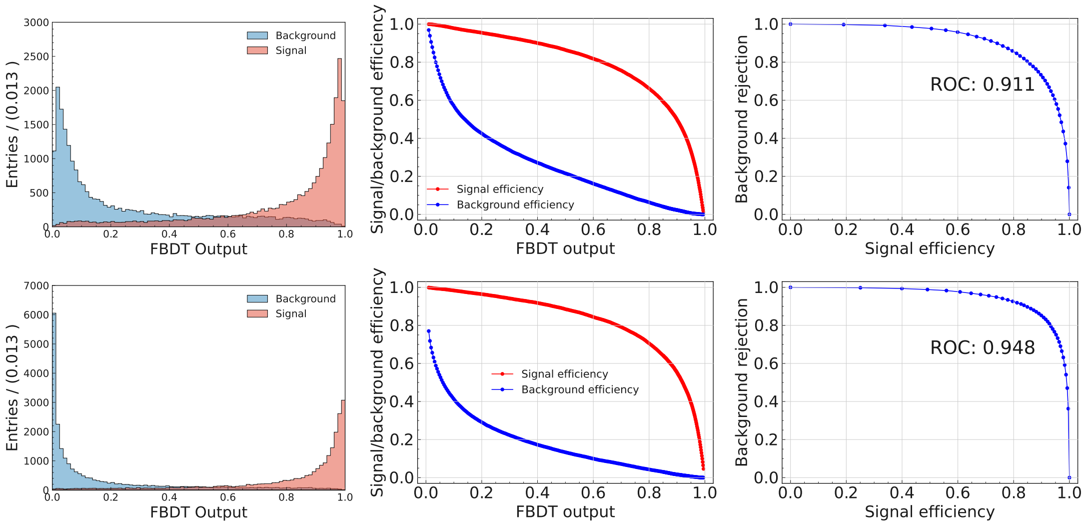


Figure 7.7: Plot of CSMVA output on training test-sample (left), signal/background efficiency (middle) and the ROC curve (right) for MC training (top) and sideband training (bottom) applied to sideband testing dataset.

The sideband training for continuum suppression was used as it demonstrated superior performance for the sideband testing data set and only slightly inferior performance for the MC testing data set compared to MC training. This is most likely due to the discrepancies between data and MC in the CS variables and the inclusion of tau pairs in the training. Furthermore, as more experimental data is collected, the performance of sideband training is expected to improve.

7.2.1 Log transform of the CSMVA output

The output of the CSMVA classifier can be transformed into a Gaussian-like shape, denoted as T_c, according to the following formula:

$$T_c = \log\left[\frac{C - C_{\text{min}}}{C_{\text{max}} - C}\right] \tag{7.4}$$

where C represents the CSMVA classifier output, C_{min} is the chosen CSMVA threshold, and C_{max} is the maximum value of the continuum classifier output and is typically very close to 1. This logarithmic transform is advantageous as it allows the CSMVA classifier information to be easily included in the final fit, enabling further background discrimination.

7.3 Optimising the CSMVA threshold

In a standard Belle II analysis, a figure of merit (FOM) is typically calculated for every possible CSMVA threshold. One commonly used FOM is $S/\sqrt{S+B}$, where S and B represent the number of signal and background events, respectively. The CSMVA requirement that produces the highest FOM is then used to determine the optimal CSMVA selection.

However, this procedure is not as useful in this analysis, as the signal yield and $\mathcal{A}_{CP}$ values are extracted by performing a three-dimensional (M_{bc}, ΔE, T_c) unbinned extended maximum likelihood fit simultaneously to data split into seven bins. The bins with higher values have a larger signal-to-background ratio. The specifics of this split will be discussed in Chapter 8. It is evident that a simple FOM does not account for the additional discriminating power provided by the different bins.

In this analysis, the optimal selection is determined by performing simplified or 'toyMC' studies for different CSMVA thresholds ranging from 0.02 to 0.98 on simulated data. The selection that minimises the $\mathcal{A}_{CP}$ error is chosen, as it is currently the greatest limitation to fully exploiting the $B \to \pi\pi$ isospin relations. ToyMC studies involve the use of simplified models to generate a large number of toy data sets. Each toy data set is then analysed as if it were real data. Each ToyMC contains the signal, continuum, and $B\overline{B}$ components with expected yields (189.9 fb^{-1}) depending on the selection of CSMVA as shown in Figure 7.9. The distribution of these components in M_{bc}, ΔE, and T_c depends on the selection of CSMVA. This is most evident in Figure 7.8 for T_c. The shape of T_c depends on the chosen continuum suppression selection and, as such, new probability density functions (PDF) are determined for each CSMVA selection. The exact fitting procedure and PDF shapes for each component in M_{bc}, ΔE, and T_c are discussed in more detail in Chapter 8.

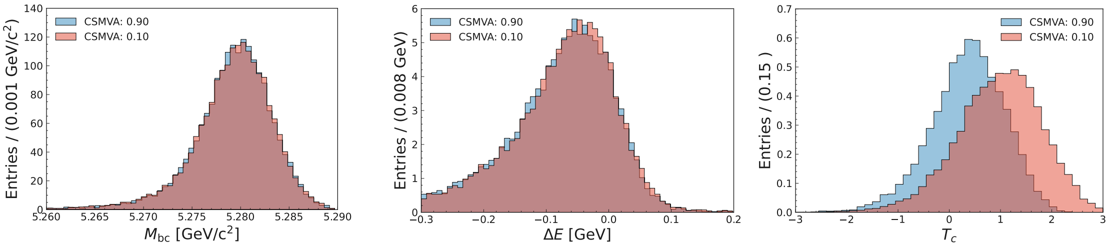


FIGURE 7.8: Distribution of M_{bc} (left), ΔE (middle) and T_c (right) for a continuum suppression selection of 0.10 (red) and 0.90 (blue).

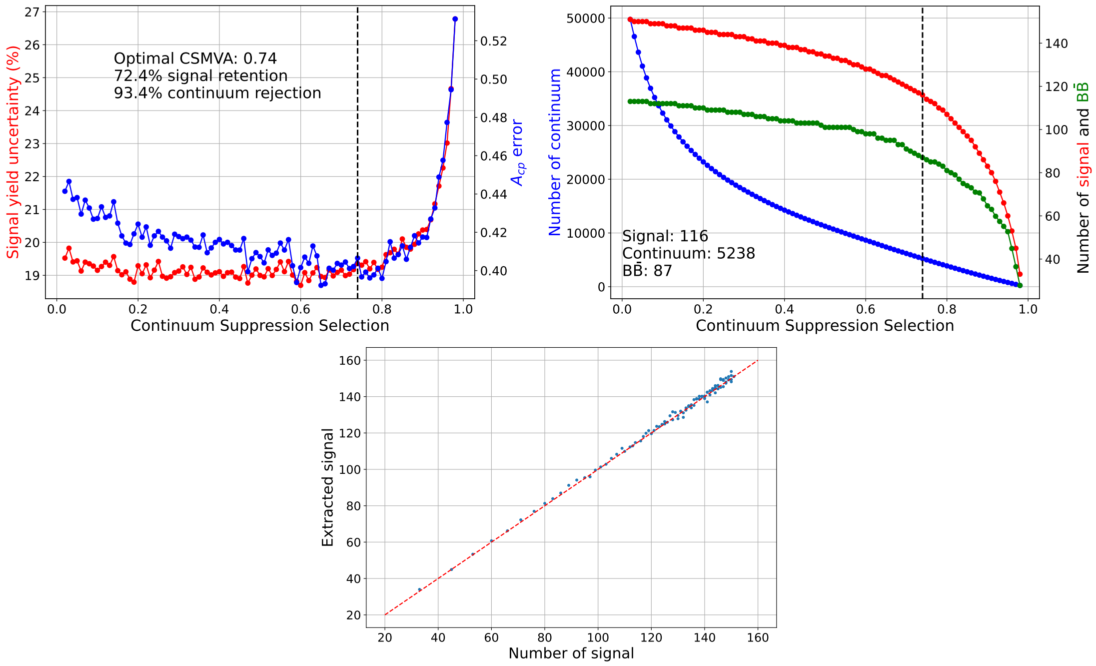


FIGURE 7.9: The mean signal yield percentage uncertainty (red) and $\mathcal{A}_{CP}$ error (blue) from 400 ToyMCs as a function of continuum suppression output (top left). Number of signal (red), continuum (blue) and $B\overline{B}$ (green) as a function of continuum suppression output (top right). The dotted vertical black line at 0.74 is the optimal continuum suppression selection. Number of signal in dataset vs number of extracted signal (bottom), the left-most data point corresponds to a CS selection of 0.98 and the red dotted line has a gradient of 1. All plots use datasets corresponding to 189.9 fb^{-1}.

For each continuum suppression selection, 400 unbinned maximum likelihood fits are performed in 7 bins using a 189.9 fb^{-1} ToyMC dataset. The mean of the signal yield and the $\mathcal{A}_{CP}$ uncertainty is then determined. This is depicted in the top left-hand plot of Figure 7.9. A selection of 0.74 is found to be optimal, rejecting 93.4% of the continuum and retaining 72.4% of the signal for an absolute signal efficiency of 35.5% and purity of 99.0%. This represents the largest value of the continuum suppression that minimises the uncertainty of the signal yield (19.0%) and the error $\mathcal{A}_{CP}$ (0.40).

With this selection, 116 ± 19 signal, 5238 continuum, 87 $B\overline{B}$ and 104 tau pairs events are expected in the 189.9 fb^{-1} dataset, as depicted in the top right plot of Figure 7.9. The extracted yield as a function of the number of signal events, as shown in the bottom plot of Figure 7.9, demonstrates a linear relationship. The level of statistical uncertainty in this study is already comparable with that of the Belle experiment (693 fb^{-1}), which achieved a signal yield and $\mathcal{A}_{CP}$ uncertainty of 19% and 0.36, respectively. A summary of the signal efficiencies at each step of the analysis is shown in Table 7.5.

Luminosity (fb^{-1})	**Signal**	**Continuum**	$\boldsymbol{B\overline{B}}$	**Tau pairs**
189.9	116 ± 19	5238	87	87
1000	607 ± 100	27,585	459	459

Table 7.4: Number of expected events for signal, continuum, tau pairs and $B\overline{B}$ after continuum suppression for 189.9 fb^{-1} and 1000 fb^{-1}. The uncertainty on the signal is due to the uncertainty of the PDG branching fraction.

Selection	**Signal efficiency (%)**
`No selections`	78.1
`Skimming`	58.0
`Signal selections`	46.4
`Continuum Suppression`	35.5

Table 7.5: Summary of signal efficiency for major steps in the analysis.

7.4 Validation of continuum suppression

The validation of the CSMVA sideband training is performed using 17.95 fb^{-1} of off-resonance data, which contains only continuum events. The same $B^0 \to \pi^0\pi^0$ selection, as previously discussed, is applied. This results in the number of candidates before and after the CSMVA selection being 736 and 69, respectively. The background rejection is found to be $90.6 \pm 3.3\%$, which is in line with the expectation from MC (93.5%).

It is also observed that the CSMVA classifier is independent of the signal features and does not create artificial peaks in the continuum background, as illustrated in Figure 7.10. However, the slope of the continuum background in ΔE is slightly less steep after applying the classifier. This should not bias the fitted yields, as the continuum PDF parameters are determined from the data sideband.

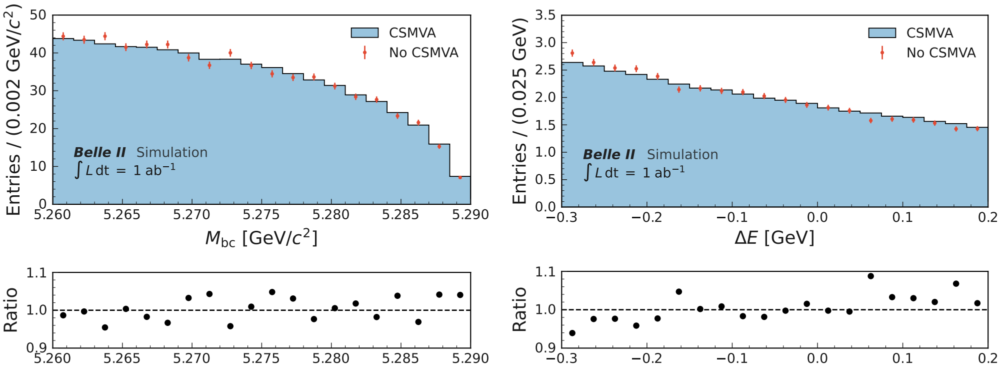


FIGURE 7.10: Plot of M_{bc} (left) and ΔE (right) for simulated continuum background before (blue) and after (red) a continnum suppression (CS) selection of 0.74 is applied.

7.4.1 Effect of photonMVA on yields after CSMVA

To assess the impact of the `photonMVA` on signal, continuum, and $B\overline{B}$ yield, the analysis selections and continuum suppression were performed without the `photonMVA` > 0.2 criteria applied. The results, summarised in Table 7.6, show that the inclusion of the `photonMVA` increases the purity of the signal from 98.0% to 98.6%, although with a 4.0% decrease in yield. Meanwhile, the number of continuum and BB events is reduced by 8.4% and 7.5%, respectively.

	Signal (10 M)	**Continuum (1 ab^{-1})**	**$B\overline{B}$ (1 ab^{-1})**
without photonMVA	3,701,337	28,323	496
with photonMVA	3,553,844	25,954	459

TABLE 7.6: Number of signal, continuum and $B\overline{B}$ after the continumm suppression with and without the `photonMVA` applied. The signal comes from 10,000,000 events while the continuum and $B\overline{B}$ comes 1 ab^{-1} of MC data.

Chapter 8

Extracting the branching fraction and $\mathcal{A}_{CP}$

To determine the number of signal, continuum, and $B\overline{B}$ events and extract $\mathcal{A}_{CP}$, probability density functions (PDFs) that characterise the M_{bc}, ΔE and T_c distributions of each component are constructed. This is necessary because it is not possible to remove all the background while maintaining a signal sample large enough for a statistically significant measurement. These PDFs are mathematical models that describe the probability distribution of each component. The process of 'fitting' these PDFs to the data essentially means determining the parameters of the underlying process that are most likely to generate the observed data. The yield of each component, along with $\mathcal{A}_{CP}$, constitutes some of these parameters. Importantly, it must be emphasised that the validity of this procedure relies on the correctness of the model. That is, the parameters of the fit reflect nature only if the model accurately captures the underlying physical processes.

The data analysis conducted in this work used ROOT [78], a comprehensive data processing framework developed by the World-Wide High Energy Physics community, predominantly written in C++. For the fitting procedures, RooFit, a specialised package within the ROOT environment is used. RooFit is specifically designed for data modelling and fitting, providing flexible and powerful tools to define PDFs that can be fitted to data. These fits are used to extract the total number of signal events and the time-integrated *CP* asymmetry of the $B^0 \to \pi^0\pi^0$ decay.

8.1 Measurement of CP violation

A $B^0\overline{B}^0$ pair, produced in the decay of a $\Upsilon(4S)$, exists as a coherent quantum state. Einstein-Bose statistics demands that the total wave function that describes a state of identical bosons must be symmetric. Given that the B meson possesses a spin of 0 while the $\Upsilon(4S)$ has a spin of 1, it follows that the orbital angular momentum (L) between the pair of B mesons is equal to 1. Therefore, the angular part of the wavefunction, $(-1)^{L=1} = -1$, is antisymmetic and the entangled state is antisymmetric,

$$|B_1^0(t_1), B_2^0(t_2)\rangle = \frac{1}{\sqrt{2}}\left(|B_1^0(t_1)\rangle\bar{B}_2^0(t_2)\rangle - \bar{B}_1^0(t_1)\rangle|B_2^0(t_2)\rangle\right) \tag{8.1}$$

where t_1 and t_2 represent the times at which the decays of B_1^0 and B_2^0 occur, respectively. The probability of having identical neutral B mesons depends on the difference between these two times, $\Delta t = t_1 - t_2$. Assuming that the decay rate difference for the B meson system is negligible ($\Delta\Gamma = 0$), Equation 8.1 can be rewritten using Equation 2.56 and Equation 2.57 as follows:

$$|B_1^0(t_1), B_2^0(t_2)\rangle = \frac{1}{\sqrt{2}}e^{-\Gamma_2(t_1+t_2)}\left[\cos\left(\frac{\Delta m\Delta t}{2}\right)(|B^0\rangle|B^0\rangle - |B^0\rangle|B^0\rangle)\right. \tag{8.2}$$

$$\left.+i\sin\left(\frac{\Delta m\Delta t}{2}\right)\frac{p}{q}|B^0\rangle|B^0\rangle - \frac{q}{p}|B^0\rangle|B^0\rangle\right] \tag{8.3}$$

Until one of the mesons decays at $\Delta t = 0$, the system contains exactly one B^0 and one $\bar{B}^0$. However, it is impossible to determine which particle is the B^0 and which is the $\bar{B}^0$ until one of them decays. Let us consider the scenario where one of the mesons, denoted as B^0, decays at time $t_1 = t$ into a flavour-specific state, represented as f. At this point, the wave function of the system collapses, and the flavour of the other meson, B^0, must be the opposite of B^0 at the moment of decay. This occurs instantaneously regardless of the distance between the two B mesons. This phenomenon is an example of the Einstein-Podolsky-Rosen (EPR) paradox and allows the initial flavour of the B^0 to be determined. The Belle II Flavour Tagger, which will be discussed in Section 8.2, is a multivariate machine learning algorithm used to identify the flavour of the B^0. After the decay of the B, the B is free to oscillate between its antiparticle and itself until it decays to a CP-eigenstate, f_{CP}, at time $t_2 = t_{CP}$. Consequently, it is possible to have events with B^0B^0 or $\bar{B}^0\bar{B}^0$ mesons in the final state. The time-dependent decay rates can be written as

$$
\begin{aligned}
\Gamma(f_{CP}, f) &= \left|\langle f_{CP}, f|\mathcal{H}|B^0_{CP}(t_{CP}), B^0(t)\rangle\right|^2 \\
&= \frac{1}{4}e^{-\Gamma(t_{CP}+t)}\left|A_{CP}\right|^2\left|A\right|^2\left|\frac{p}{q}\right|^2 \cdot [1-\cos(\Delta m\Delta t) \\
&\quad +2\mathrm{Im}(\lambda_{CP})\sin(\Delta m\Delta t) + (1+\cos(\Delta m\Delta t))\left|\lambda_{CP}\right|^2\Big]
\end{aligned}
\tag{8.4}
$$

$$
\begin{aligned}
\Gamma(f_{CP}, \bar{f}) &= \left|\langle f_{CP}, \bar{f}|\mathcal{H}|B^0_{CP}(t_{CP}), B^0(t)\rangle\right|^2 \\
&= \frac{1}{4}e^{-\Gamma(t_{CP}+t)}\left|A_{CP}\right|^2\left|A\right|^2 \cdot [1+\cos(\Delta m\Delta t) \\
&\quad -2\mathrm{Im}(\lambda_{CP})\sin(\Delta m\Delta t) + (1-\cos(\Delta m\Delta t))\left|\lambda_{CP}\right|^2\Big]
\end{aligned}
\tag{8.5}
$$

where the decay amplitude $A = \langle f|\mathcal{H}|B^0\rangle = \langle \bar{f}_{\text{tag}}|\mathcal{H}|\bar{B}^0\rangle$ and $A_{CP} = \langle f_{CP}|\mathcal{H}|B^0\rangle = \langle \bar{f}_{CP}|\mathcal{H}|\bar{B}^0\rangle$.

After normalising Equation 8.4 and Equation 8.5 to unity over the entire region $-\infty < \Delta t < \infty$, the time-dependent probability of a flavour tag $q = +1(-1)$ corresponding to $B^0 = B^0(\overline{B}^0)$ can be expressed as follows.

$$
\mathcal{P}(\Delta t, q) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}}\left[1 + q(\mathcal{A}_{CP}\cos(\Delta m\Delta t) + \mathcal{S}_{CP}\sin(\Delta m\Delta t))\right] \tag{8.6}
$$

where τ_{B^0} is the B^0 lifetime. The violation parameters CP, $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$, can be obtained by fitting the Δt distribution for $q = +1$ and $q = -1$. However, for $B^0 \to \pi^0\pi^0$ decays, the absence of charged tracks makes it impossible to reconstruct the vertex location of the B^0 with the picosecond precision required for a time-dependent CP analysis. This time requirement is set by the lifetime of the B meson. Nevertheless, the direct CP-violation parameter, $\mathcal{A}_{CP}$, can still be determined by performing a time-integrated CP-violation analysis. By integrating Equation 8.6 over the range $-\infty < \Delta t < \infty$, the time-integrated probability is given by

$$
\mathcal{P}(q) = \frac{1}{2}\left[1 + q \cdot \mathcal{A}_{CP} \cdot \frac{1}{(\tau_{B^0}\Delta m)^2 + 1}\right] = \frac{1}{2}\left[1 + q \cdot \mathcal{A}_{CP} \cdot (1 - 2 \cdot \chi_d)\right] \tag{8.7}
$$

where χ_d is the time-integrated $B^0 - \overline{B}^0$ mixing probability parameter from Equation 2.60.

However, it is important to note that Equation 8.7 is based on the assumption of perfect tagging. In reality, the Belle II Flavour Tagger may not always assign the correct flavour tag, or it may fail to assign a flavour altogether. The fraction of signal events that are incorrectly tagged, known as the wrong-tag fraction, is denoted by w. Meanwhile, the fraction of events to which a flavour tag can be assigned, referred to as the tagging efficiency, is represented by ϵ. Furthermore, there may be a slight difference for ϵ and w between B^0 and $\overline{B}^0$ due to charge asymmetries in detection and reconstruction. To incorporate the impacts of tagging efficiency and the wrong-tag fraction for B^0 and $\overline{B}^0$, ϵ and w are defined as

$$\epsilon = \frac{\epsilon_{B^0} + \epsilon_{\bar{B}^0}}{2}, \quad \frac{w_{B^0} + w_{\bar{B}^0}}{2} \tag{8.8}$$

and the difference is defined as

$$\Delta\epsilon = \epsilon_{B^0} - \epsilon_{\bar{B}^0}, \quad \Delta w = w_{B^0} - w_{\bar{B}^0} \tag{8.9}$$

where the subscript is the true flavour of the B meson. For example, w_{B^0} is the fraction of true B^0 mesons that were wrongly classified as $\bar{B}^0$. The difference in the wrong-tag fraction between positive and negative tags is given by Δw, while the difference in the tagging efficiency between the positive and negative tags is given by $\Delta\epsilon$. The observed probability, $\mathcal{P}(q)$, can be expressed in terms of the efficiencies and wrong tag fractions of B^0 and $\overline{B}^0$.

$$\begin{aligned}
\mathcal{P}_{\text{obs}}(q=+1) &= \varepsilon_{B^0}(1 - w_{B^0}) \cdot P(q=+1) + \varepsilon_{B^0} w_{B^0} \cdot \mathcal{P}(q=-1) \\
\mathcal{P}_{\text{obs}}(q=-1) &= \varepsilon_{B^0}(1 - w_{B^0}) \cdot P(q=-1) + \varepsilon_{B^0} w_{B^0} \cdot \mathcal{P}(q=+1)
\end{aligned}$$

which can be written in terms of w, $\mu = \frac{\Delta\varepsilon}{2\varepsilon}$, and Δw as:

$$\begin{aligned}
P(q)_{\text{obs}} = \frac{1}{2}\big[&1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) \\
&- [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot (1 - 2\chi_d)\mathcal{A}_{CP}\big]
\end{aligned} \tag{8.10}$$

Putting this all together, the PDF for the signal component is

$$\begin{aligned}
\mathcal{P}^s(M_{\text{bc}}, \Delta E, T_c, q) &= \mathcal{P}^s_{\text{obs}}(q) \cdot \mathcal{P}^s(M_{\text{bc}}, \Delta E, T_c) \\
&= [1 - q\Delta w + q\mu(1 - 2w) \\
&+ [q(1 - 2w) + \mu(1 - q\Delta w)](1 - 2\chi_d)\mathcal{A}_{CP}] \\
&P^s(M_{\text{bc}}, \Delta E, T_c)
\end{aligned} \tag{8.11}$$

Therefore, once the mistagging parameters, w, Δw, and μ are known, a fit to the data divided into positive and negative q values can be used to determine the value of $\mathcal{A}_{CP}$ that best represents the data.

8.2 Flavour Tagging

Measurements of CP violation require the full reconstruction of a B^0_{sig} and the determination of the flavour of the accompanying B^0_{tag} meson at the time of its decay, a task known as flavour tagging. The Belle II Flavour Tagger [79] is an essential tool for this purpose and uses a category-based multivariate machine learning algorithm to determine the quark flavour of B mesons. This is possible because many decay modes provide flavour signatures through flavour-specific final states. Flavour signatures are characteristics of B decay products that are correlated with the flavour of the neutral B meson. For example, in semileptonic decays such as $\bar{B}^0 \to D^{*+}l^-\bar{\nu}_l$, a negatively charged lepton unambiguously tags a $\bar{B}^0$. The different flavour signatures are sorted into thirteen tagging categories. Some categories focus on primary leptons from B decays, while others rely on secondary leptons produced from the decay of charmed mesons and baryons. Additionally, some categories consider the highest momentum charged particle as a flavour tag. The thirteen categories for the flavour tagger and their corresponding target particles are as follows:

- **Electron, Muon, and Kinetic lepton**: exploit the signatures provided by primary leptons from B decays occurring via transitions $b \to cl^-\nu_l$, or $b \to ul^-\nu$, where c corresponds to an electron, muon or both depending on the category. For instance, a positively charged lepton tags a $\bar{B}^0$ meson, and a negatively charged one tags a B^- meson.

- **Intermediate Electron, Intermediate Muon, and Intermediate Kinetic Lepton**: exploit flavour signatures from secondary leptons produced through the decay of charmed mesons and baryons that occurs via transitions $b \to c \to s(d)l^+\nu_l$.

- **Kaon**: exploits the signature from kaons originating from decays of charmed mesons and baryons produced via $b \to c \to s$ transitions.

- **Slow Pion**: exploits the flavour signatures of secondary pions from decays $B^0 \to D^+(\to D^0\pi^+)X^-$.

- **Kaon-Pion**: exploits the flavour signatures of decays containing both a kaon and a slow pion.

- **Fast Hadron**: exploits the flavour signatures of kaons and pions from the W boson in $b \to cW^-$ or $b \to uW^-$ decays, and from one-prong decays of primary tau leptons from $b \to \tau^-(\to h^-\nu_\tau)\nu_\tau X$ transitions, where h^- stands for a π^- or a K^-.

- **Maximum p^***: this category is based on selecting the charged particle with the highest momentum in the $\Upsilon(4S)$ frame and using its charge as a flavour tag.

- **Fast-Slow-Correlated (FSC)**: exploits the flavour signatures of both slow pions and high-momentum primary particles.

- **Lambda**: exploits the additional flavour signatures provided by the Λ baryons from the $b \to c \to s$ transitions. A Λ baryon indicates a B^0, and a $\bar{\Lambda}$ indicates a $\bar{B}^0$.

The different flavour signatures enhance the performance of the flavour tagger and make it more robust to misidentification. More information can be found in Ref. [80].

The category-based flavour tagger employs a two-level procedure, consisting of an event level for each category followed by a combiner level, as shown in Figure 8.1. Initially, at the event level, the flavour tagger creates a list of particles for each type of reconstructed track, including electrons, muons, kaons, pions, and protons/lambda. A range of flavour-tagging input variables are then computed for each track, using the candidate particle list and information from the ECL and KLM.

The input for the category-based flavour tagger encompasses kinematic and vertex information, track-hit data, and charged-particle identification (PID) information. The algorithm configuration used in this work comprises 186 input variables for all categories collectively. For each category, FBDT classifiers employ these input variables, such as missing momentum and particle identification information, to assign a classifier output, labelled 'RightCategory', to each particle candidate. The 'RightCategory' ranges from 0 to 1, corresponding to the probability of the target providing the correct tagging.

In each category, only the information from the particle candidate with the highest 'RightCategory' is used and multiplied by the flavour charge to create 'qp'. Consequently, there are 13 'qp' values corresponding to flavour information from each of the categories.

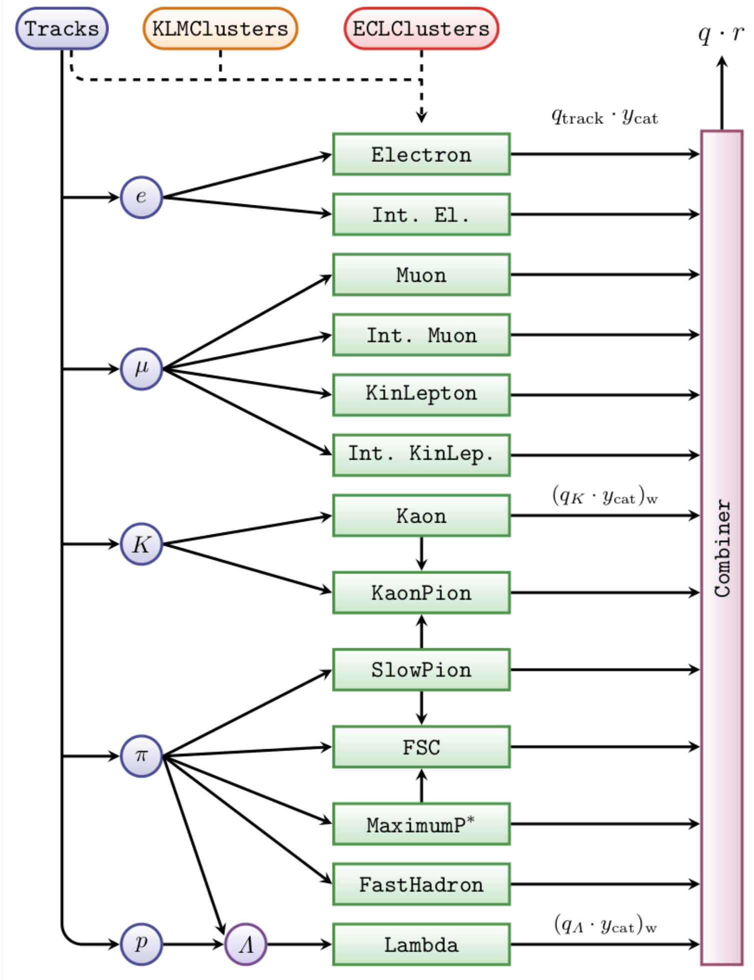


FIGURE 8.1: Overview of the category-based flavor tagger. The tracks on the tag side are used to create e, μ, K, π, and p candidate lists. Each category considers the list of candidates belonging to its own targets. The different categories are represented by green boxes, and the combiner by a magenta box.

At the combiner level, all the ‘qp’ values of each categories are combined in another FBDT to produce a single number $q \cdot r$ that represents the likelihood of the B^0_{tag} possessing a specific b-flavour. The value $q = +1$ tags a B^0, while $q = -1$ tags a $\bar{B}^0$. The value of r represents the algorithm’s confidence in the assigned q value. It is defined as $r = 1 - 2w$, where w is the fraction of incorrectly tagged events and ranges from zero, indicating no flavour distinction between B^0 and $\bar{B}^0$, to one for a definitive flavour assignment. For example, a sample of events where $r = 0$ would have an equal number of correctly and incorrectly tagged events. Conversely, a sample of events where $r = 1$ would have no incorrectly tagged event.

8.2.1 Fitting in bins of $q \cdot r$

The continuum background peaks at zero and then gradually decreases near $q \cdot r = \pm 1$, as shown in Figure 8.2. This is because the continuum background contains no B mesons, and hence the flavour tagger algorithm should ideally produce a $q \cdot r$ output close to zero. In contrast, the distribution of $q \cdot r$ for signal events rapidly increases near $q \cdot r = \pm 1$. Hence, it is evident that when the flavour tagger is confident in its assignment of flavour of the tag-side, the continuum is greatly reduced. The fraction of events in each r bin for signal and background is presented in Table 8.1 and Table 8.2, respectively.

It may seem logical to impose a binary selection on the $q \cdot r$ distribution. However, this would dramatically decrease the precision of any signal measurement, given that a large fraction of the signal is close to $q \cdot r = 0$. To incorporate the discriminating power of the $q \cdot r$ information, the data are divided into 7 bins. The bins are spaced so that each bin has an approximately equal number of candidates, these edges of the bin r are: 0.0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.875, and 1.0, which is shown as different coloured bands in Figure 8.2. The bins with higher r values will possess a larger signal-to-background ratio, thus contributing more significantly to the precision of the final measurement. Therefore, the signal yield and $\mathcal{A}_{CP}$ values are determined by performing a three-dimensional (M_{bc}, ΔE, T_c) unbinned extended maximum likelihood fit simultaneously to events in the seven intervals of r. Consequently, $q \cdot r$ could be regarded as the 'fourth' dimension of the fit.

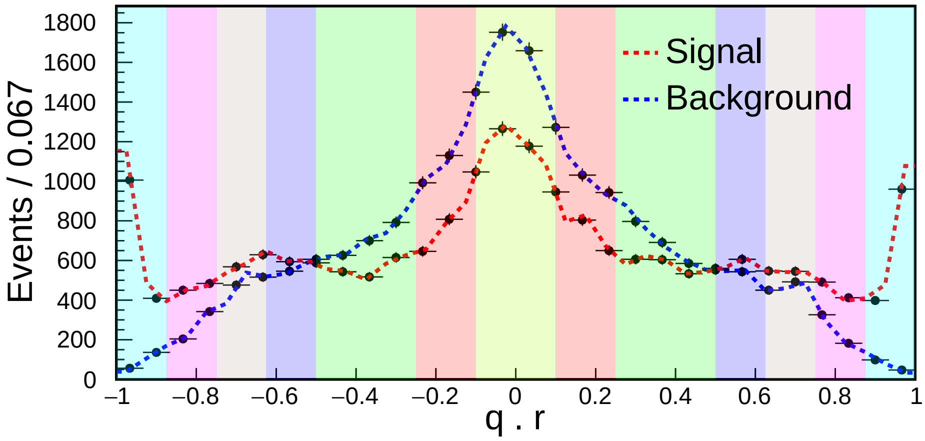


FIGURE 8.2: The generated $q \cdot r$ for 20000 events with $A_{CP} = 0.0$ (dots) overlaid with a RooHistPDF using 40 bins and 2nd order extrapolation for signal (red) and continuum background (blue). The colored bands corresponding to the 7 bins of $q \cdot r$.

The fit is performed in 7 bins of r to use the flavour-tagging information to help distinguish between signal and continuum events.

The unbinned extended maximum likelihood fit uses a PDF made up of signal and several background components. Fitting refers to the process of adjusting the parameters of a model to best fit the observed data. The meaning of the other terms is as follows:

- **Maximum likelihood:** Maximum likelihood estimation is a statistical method used to estimate the parameters of a probability distribution. It uses the concept of a likelihood, which quantifies the probability of observing a specific set of data given certain parameters. For several statistically independent points that follow the same PDF, the total likelihood is the product of the likelihood at each individual point.

  $$\mathcal{L}(\vec{\theta}) = \prod_{i=1}^{N} f\left(x_i; \vec{\theta}\right). \tag{8.12}$$

  represents the total likelihood function $\mathcal{L}(\vec{\theta})$, which quantifies the probability of observing a specific set of N data points $\vec{x} = (x_1, ..., x_N)$ given a set of k parameters $\vec{\theta} = \theta_1, ..., \theta_k)$. The data points are assumed to be distributed according to the PDF $f(x; \vec{\theta})$. The product runs over all observed events x_i.

  The goal of maximum likelihood estimation is to find the set of parameters that maximise the total likelihood. In practise, it is more computationally efficient to take the logarithm of the total likelihood and multiply it by -2. This approach, known as minimising the negative log-likelihood, yields the same extremum and simplifies the calculation.

- **Extended:** The term ‘extended’ refers to the inclusion of the total number of events as an additional parameter in the likelihood function. The total likelihood is calculated as the product of the Poisson probability of observing the total number of events and the product of the likelihoods for each individual data point.

  $$\mathcal{L}(\vec{\theta}) = \frac{e^{-\mu}}{N!} \prod_{i=1}^{N} \mu f(x_i; \vec{\theta}) \tag{8.13}$$

  where N is the observed number of events and μ is the expected number of events.

- **Unbinned**: This means that the data is not divided into bins or intervals. Instead, each individual data point is used in the analysis and the total likelihood is evaluated at each data point. This also avoids bias or loss of information due to the choice of bin size and boundaries.

- **Fit**: The fitting procedure employs MINUIT2 [81], using MIGRAD [82] as the minimiser and MINOS for asymmetric uncertainty calculations. Following each fit, both the statuses of MIGRAD and MINOS, and the completeness and accuracy of the covariance matrix are checked. If any of these checks fails, the parameters that were allowed to vary or 'float' during the fit are reset to their initial values. These parameters are then randomised according to a Gaussian distribution centred around their initial value, with a standard deviation equivalent to their uncertainty. The fit is re-tried up to a maximum of 10 times. If a satisfactory fit is not achieved within these attempts, the fitting process is considered to have failed.

Therefore, the total likelihood function for all components across all r bins is given by

$$\mathcal{L} = \frac{e^{-\sum_j N^j}}{\prod_k N_k!} \times \prod_k \left[\prod_{i=1}^{N_k} \left(\sum_j f_k^j N^j P_k^j (M_{\text{bc}}^i, \Delta E^i, T_c^i, q^i) \right) \right], \tag{8.14}$$

where i is the number of candidates, j is the sample component in terms of signal (s), continuum (c), and $B\bar{B}$, and k indicates the r interval. Here, N^j denotes the yield for component j, N_k denotes the number of candidates in the kth bin, f_k^j is the fraction of candidates in the kth bin for the jth component, and $P_k^j(M_{\text{bc}}^i, \Delta E^i, T_c^i, q^i)$ is the probability density function (PDF) to have the ith event of the jth component in the kth bin. The values of f_k^{c} implicitly include a factor of one-half due to the division of the data into positive and negative q values for each r intervals.

The PDF for the signal, continuum, and $B\overline{B}$ component is given by Equation 8.11, but divided into 7 bins of r.

$$\begin{aligned} P_k^c(M_{\text{bc}}, \Delta E, T_c, q) =& [1 - q\Delta w_k + q\mu_k(1 - 2w_k) \\ & + [q(1 - 2w_k) + \mu_k(1 - q\Delta w_k)](1 - 2\chi_d)\mathcal{A}_{CP}] \\ & P^s(M_{\text{bc}}, \Delta E, T_c), \end{aligned} \tag{8.15}$$

where w_k is the fraction of signal events incorrectly tagged (wrong-tag), Δw_k is the difference in the wrong-tag fraction between positive and negative tags, and $\mu_k = \Delta\epsilon_k / 2\epsilon_k$ is the asymmetry of the tagging efficiency. Here, ϵ_k is the tagging efficiency and $\Delta\epsilon_k$ is the difference in the tagging efficiency between positive and negative tags. The free parameters of the fit are the yields N^s and N^c, as well as the CP-violation parameter, $\mathcal{A}_{CP}$. All remaining parameters are fixed. PDFs of the various components do not share any common parameters.

8.2.2 Mistagging Parameters

The flavour PDFs $P_j(q)_{obs}$ (see Equation 8.10) have the same form for the signal and background sources with independent parameters, $\varepsilon_k, w_k, \Delta w_k, \mu_k$ for the signal, continuum, and $B\overline{B}$ components. As previously mentioned, j represents the sample component in terms of signal (s), continuum (c), and $B\bar{B}$, while k indicates the r interval. This leads to a total of $4 \times 3 \times 7 = 84$ parameters associated with the Belle II Flavour Tagger's mistagging that must be estimated to accurately extract $\mathcal{A}_{CP}$. Furthermore, to determine the individual yield of the signal, continuum, and $B\overline{B}$, the fraction of events in each r bin, f_k^j for each component must also be determined. This adds additional $3 \times 7 = 21$ parameters.

The fraction of signal events in each r interval (f_k^s), along with w_k, Δw_k, and $\Delta\epsilon_k$, are fixed to values obtained from a fit to $B^0 \to D^{(*)-}h^+$ decays, where h^+ stands for a π^+ or K^+, following Ref. [80]. The mistagging parameters used in this work are summarised in Table 8.1.

Bin	$\boldsymbol{\epsilon}$	$\boldsymbol{w}$	$\boldsymbol{\Delta w}$	$\boldsymbol{\mu}$
1	19.0 ± 0.3	49.47 ± 0.78	3.0 ± 1.6	-1.0 ± 1.7
2	17.1 ± 0.3	42.41 ± 0.75	4.3 ± 1.5	-0.4 ± 1.7
3	21.3 ± 0.3	31.41 ± 0.65	0.8 ± 1.3	3.4 ± 1.5
4	11.3 ± 0.3	18.33 ± 0.82	4.2 ± 1.6	1.5 ± 2.0
5	10.7 ± 0.3	16.48 ± 0.8	3.4 ± 1.6	3.4 ± 2.1
6	8.2 ± 0.2	9.32 ± 0.82	2.8 ± 1.6	-2.0 ± 2.3
7	12.4 ± 0.2	1.74 ± 0.47	2.4 ± 0.93	-1.8 ± 1.8

TABLE 8.1: Mistagging parameters and uncertainty. ϵ is the bin fraction, w is the wrong tag fraction, Δw is the difference in wrong tag fraction between positive and negative b-flavor tags, and μ is the difference in tagging efficiency between positive and negative b-flavor tags. All numbers are in percentages.

The fraction of $B\overline{B}$ events in each r interval ($f_k^{B\overline{B}}$), along with w_k, Δw_k, and $\Delta\epsilon_k$ is obtained using the generated $B\overline{B}$-only MC described in subsection 6.2.2. This is important because the $B\overline{B}$ background originates from both neutral and charged B mesons. The mistagging parameters determined in Ref. [80] have different values depending on whether the tag-side B meson is a B^0 or a B^+. A weighted average, which depends on the number of charged or neutral B mesons, would be required. In any case, the CP asymmetry of the $B\overline{B}$ background is expected to be zero and should not affect the extracted $\mathcal{A}_{CP}$ of the signal decay.

The fraction of events in each r bin for the continuum background (f_k^{c}) is obtained by using experimental sideband data. This approach is taken to account for the difference in the $q \cdot r$ distribution between continuum and tau pairs and any potential data-MC discrepancy, as shown in Figure 8.11 and Figure 8.3, respectively. The minor difference in the $q \cdot r$ bin fraction between the MC and experimental sideband data, as shown in Table 8.2, suggests that the contribution of the tau pair and the data-MC discrepancy is also small. The experimental sideband $q \cdot r$ distributions of the continuum are validated by comparing the $q \cdot r$ bin fractions with the off-resonance data. The results, which are shown in Table 8.3, demonstrate good agreement. All mistagging parameters are set to zero since there are no B mesons in continuum events.

Bin	**Continuum MC bin fraction (%)**	**Sideband Data bin fraction (%)**
1	32.2	32.1
2	25.6	26.3
3	22.3	23.3
4	9.5	8.7
5	7.1	6.1
6	2.6	2.5
7	0.62	0.83

TABLE 8.2: $q \cdot r$ bin fraction for continuum in MC (left) and in experimental sideband data (right).

$\boldsymbol{\lvert q \cdot r \rvert}$ **bin edges**	**MC14a** (%)	**Data** (%)	**Number of events**
0.0-0.1	24.4	23.5 ± 1.8	173
0.1-0.25	23.2	25.0 ± 1.8	184
0.25-0.5	25.7	27.6 ± 1.9	203
0.5-0.625	10.9	10.6 ± 1.2	78
0.625-0.75	9.4	8.0 ± 1.0	59
0.75-0.875	4.6	4.5 ± 0.8	33
0.875-1.0	1.5	0.8 ± 0.3	6

TABLE 8.3: The $q \cdot r$ bin fractions for `MC14ri_a` (148,307 events) and off-resonance data (736 events) where the uncertainty is statistical.

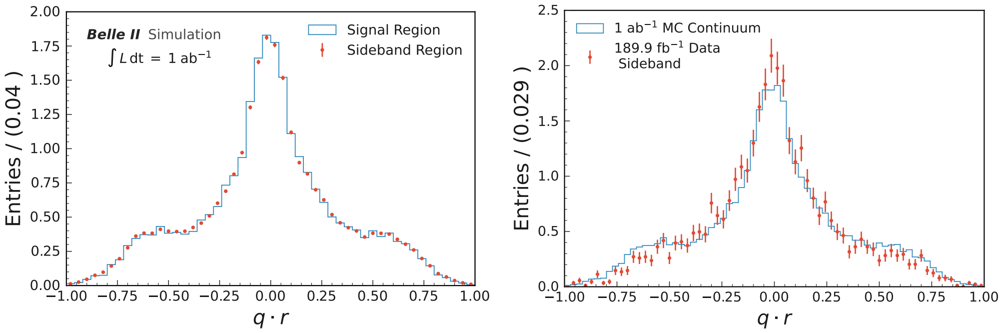


FIGURE 8.3: $q \cdot r$ distribution of continuum in the signal (blocks) and sideband (dots) region in 1 ab^{-1} of MC (left), and 1 ab^{-1} of MC (blocks) in the signal region and 189.9 fb^{-1} of experimental data (dots) in the sideband region (right).

8.3 Fitting with Probability Density Functions

The full PDF is the sum of the individual M_{bc}, ΔE, and T_c PDFs for the signal, continuum, and $B\overline{B}$ components across each of the seven r bins. These PDFs are designed to account for any correlation between the fitting variables, as well as any dependence on r. Moreover, they are constructed to accommodate any discrepancies between MC and actual data. The PDF shapes used to model the various components are discussed below and summarised in Table 8.4.

Component	$\boldsymbol{M_{\mathrm{bc}}}$	$\boldsymbol{\Delta E}$	$\boldsymbol{T_c}$
Signal	2D KDE	2D KDE	Bifurcated Guassian + Gaussian
Continuum	8 ARGUS	1st order Chebychev	Bifurcated Gaussian + Gaussian
$B\bar{B}$	2 ARGUS	KDE	Bifurcated Gaussian + Gaussian

TABLE 8.4: The PDF shapes are used to fit the signal, continuum and $B\overline{B}$ components of the signal mode.

8.3.1 Modelling signal

In the previous Belle [28] study, complicated analytical functions were used to account for the M_{bc} and ΔE correlations. In this work, the correlation between M_{bc} and ΔE is taken into account with a two-dimensional kernel density estimation (KDE). KDEs models the distribution of an arbitrary input data set as a superposition of density functions at each data point, called kernels. It is a non-parametric way to estimate the PDF of a random variable. Given independent and identically distributed samples $x_1, ..., x_n$ drawn from a distribution with unknown density f, the shape of f is estimated as

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(x - x_i\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{8.16}$$

where K is the kernel and h is a parameter known as the bandwidth, which defines the 'smoothness' of the PDF. A kernel with subscript h is called the scaled kernel and is defined as $K_h(x) = 1/hK(x/h)$. The kernel typically is a symmetric non-negative function that integrates to one. The kernel assigns weights to x_i, depending on their distance from x. As shown in Figure 8.4, the idea is to compute the contribution of each point in the data set to the estimated density at any given point in the feature space. A lower bandwidth means that only points very close to the current position are given any significant weight, leading to a high sensitivity to the input data. Conversely, a higher bandwidth results in a shallow kernel where distant points can contribute, leading to a density estimate that is very smooth.

In this work, the Gaussian distribution is used as the kernel function, with the width of the Gaussian serving as the bandwidth. The bandwidth is modified based on the local event density, an approach known as adaptive kernel estimation. For instance, in high-density areas where the statistical uncertainty is low, a small bandwidth is chosen to preserve detailed shape information. Conversely, in low-density areas where uncertainty is high, a large bandwidth is used to smooth out statistical fluctuations.

In addition, the data points are mirrored about the observable boundaries to account for edge effects occurring when the kernel function extends beyond the boundaries of the data. This can lead to biased and inaccurate density estimates near the boundaries of the data set. Creating a mirror image of the data at the boundaries essentially extends the data set, allowing the kernel function to operate fully even at the edges. The mirrored data points ensure that the kernel function has sufficient data to work with at the boundaries, improve the modelling behaviour at the edges for distributions that are not close to zero at the edge. These boundary conditions are necessary to properly account for the normalisation of the estimated density function, $\hat{f}_h(x)$.

The two-dimensional M_{bc}-ΔE KDE created by ROOT cannot be stored and exported to be used in a fit. This is a technical limitation of the software package. To address this, the following method is used to store the data:

1. 2D KDEs are generated with the `RooNDKeysPdf` function

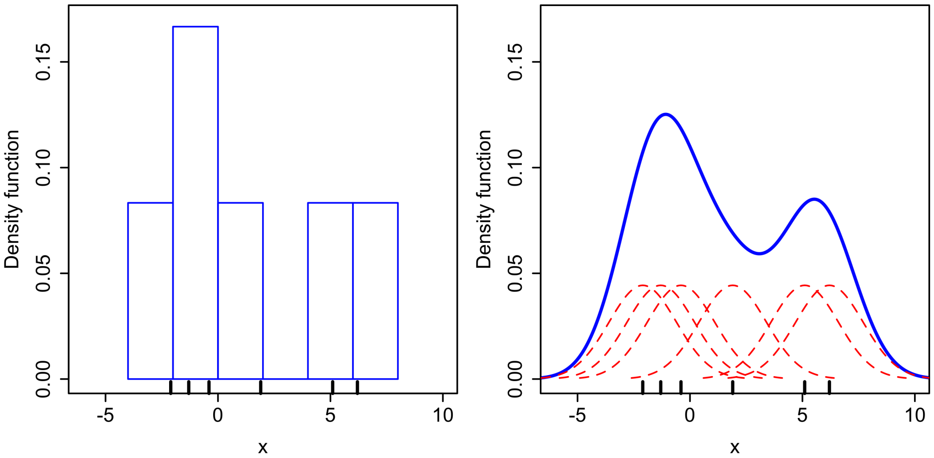


FIGURE 8.4: Comparison of the histogram (left) and kernel density estimate (right) created using the same six data points, represented as lines on the horizontal axis. The kernels corresponding to each data point are the red dashed curves, and the kernel density estimate is the blue curves. Plots taken from [18].

2. The KDE are converted to `TH2` histograms using `createHistogram` method with 320×320 binning to preserve as much detail as possible
3. The `TH2` histograms are converted to a `RooDataHist`
4. The `RooDataHist` are converted to a `RooHistPdf` so that is can be used by `RooFit`.
5. The `RooHistPdf` are saved to a `RooWorkspace` in a ROOT file

The T_c PDF is constructed from the sum of a bifurcated Gaussian and a Gaussian. This combination was found to be the most effective in modelling the distribution. The M_{bc}-ΔE KDE and T_c PDF parameters were determined using $B^0 \to \pi^0\pi^0$ decays reconstructed in signal-only MC. The fits to the simulated data are shown in Figure 8.5.

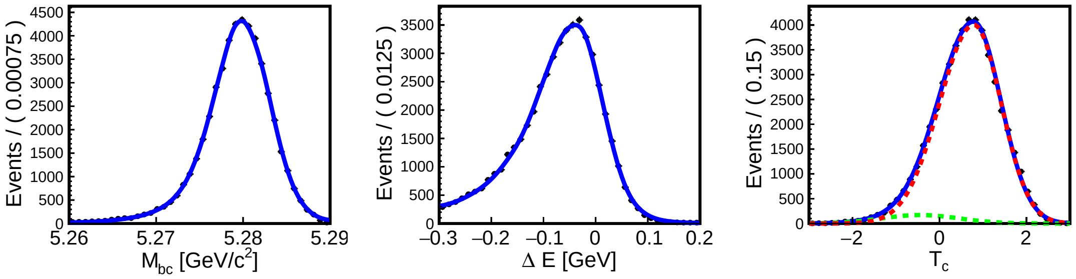


FIGURE 8.5: Signal mode PDFs for M_{bc}, ΔE and T_c fitted to signal-only MC. T_c is composed of a bifurcated Gaussian (red) and a Gaussian (green).

8.3.2 Modelling continuum

In the control mode, a discrepancy is observed between MC and data near the M_{bc} endpoint for the continuum. The M_{bc} endpoint in data was significantly smaller than the estimated 5.29 GeV/c^2 expected from MC. To eliminate the possibility of sculpting due to signal-based continuum suppression, the sideband regions of both MC and data were compared before continuum suppression was applied to them. As shown in the left-hand plot in Figure 8.6, the Data-MC discrepancy was still present.

An investigation was then carried out to determine whether this discrepancy could be attributed to one of the selections. This is done by reconstructing the control mode using only standard particle selections. Despite this, the discrepancy between Data and MC persists. Moreover, it is found that the M_{bc} distribution of the optimised selection and the selections using standard particles are identical, as shown in the right-hand plot of Figure 8.6.

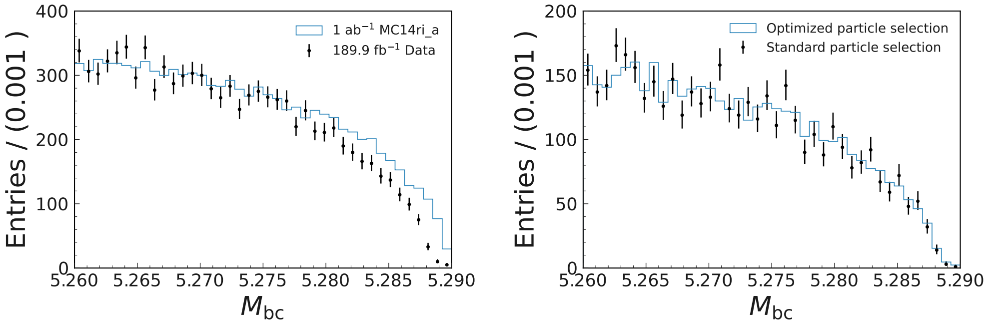


FIGURE 8.6: Data (dots) and MC (blocks) in the sideband region (left) before the continuum suppression is applied. The control mode reconstruction using only standard particle list (dots) and optimised selection previously described (blocks) in the sideband region (right).

Finally, an investigation was conducted to determine whether the discrepancy depends on the run period. Plots of M_{bc} are shown for data collected during 2020 and 2021 in Figure 8.7. The data for 2020 was found to align closely with MC. However, the remaining experiments, corresponding to data from 2021, clearly exhibit a discrepancy between data and MC. This discrepancy is found to be the result of the beam energy being slightly lower than anticipated, which causes the M_{bc} endpoint to be below 5.29 GeV/c^2.

To model the MC-Data discrepancy in M_{bc} the continuum component M_{bc} is modelled by 8 ARGUS functions with evenly spaced endpoints from 5.287 to 5.290 where the

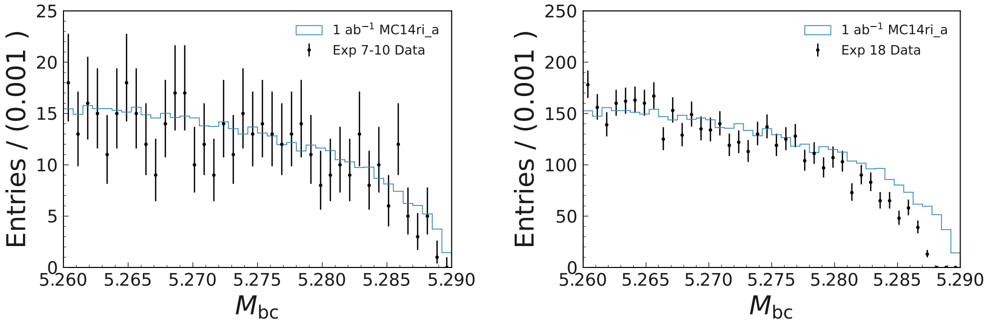


FIGURE 8.7: Comparison of run-independent MC (blocks) and data (dots) for experiments 7-10 which was collected in 2020 (left) and experiment 18 which was collected in 2021 (right) in the sideband region. MC is normalized to the data.

fraction of each PDF equals the fraction of events in the corresponding centre-of-mass energy, $\mathrm{E_{CMS}}$ bin. The fraction of events in each bin can be found in Figure 8.8.

$$\mathrm{PDF_{continuum}} = \sum_{i=0}^{7} h_i \cdot f(M_{bc}, \chi, E_i) \tag{8.17}$$

where i denotes the $\mathrm{E_{CMS}}$ bin, h_i is the fraction of events belonging to the $\mathrm{E_{CMS}}$ bin, f is an ARGUS function, χ is the ARGUS shape parameter and E_i is the ARGUS endpoint.

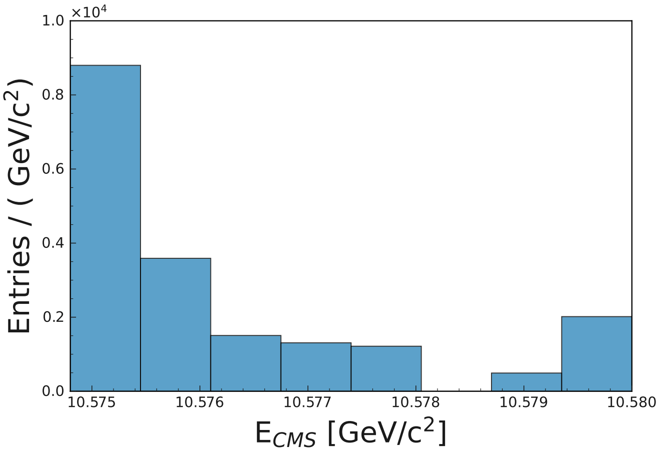


FIGURE 8.8: Histogram of the centre-of-mass energy E_{CMS} of the experimental sideband with 8 bins. The fraction of events belonging in each bin equals the fraction of each of the 8 ARGUS functions that compromise the continuum M_{bc} PDF.

The ΔE PDF is a first-order Chebychev function, while the T_c PDF is constructed from the sum of a bifurcated Gaussian and a Gaussian.

According to MC, the continuum in the sideband and signal region is expected to be identical, as shown in the left plots of Figure 8.9. Therefore, the continuum background parameters are extracted from the experimental data sideband region defined as

$5.22 < M_{\rm bc} < 5.27\,{\rm GeV}/c^2$ and $0 < \Delta E < 0.5\,{\rm GeV}$. The range is limited since getting too far away from the signal region may introduce continuum phenomenology that is significantly different from the one that is relevant in the signal. The $B\bar{B}$ background has negligible contribution as within both ranges, only 0.07% survive, and $B\bar{B}$ is expected to be only 1.1% of the total background. The fits to the sideband data are shown in Figure 8.10.

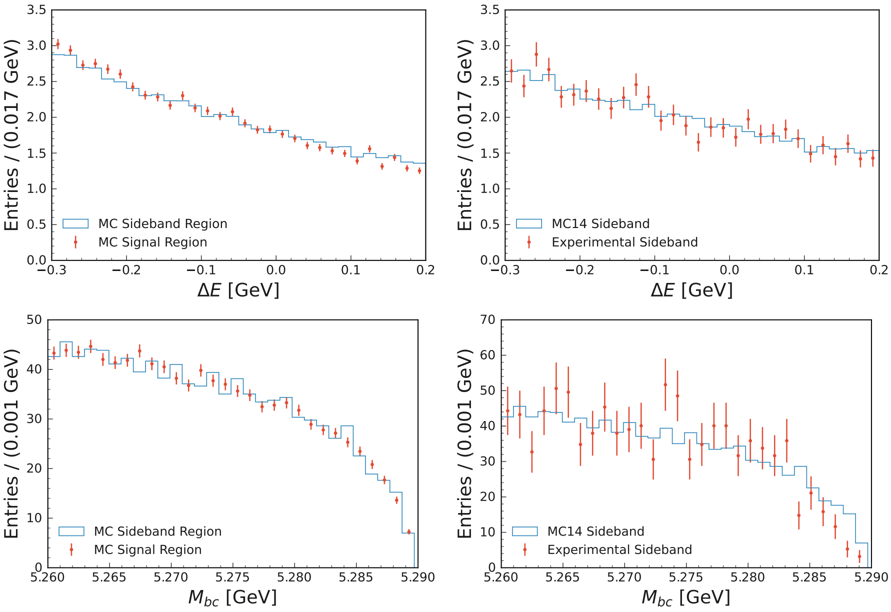


FIGURE 8.9: ΔE (top) and M_{bc} (bottom) for sideband (blocks) and signal (dots) region in MC (left), and for MC (blocks) and Data (dots) in the sideband region (right).

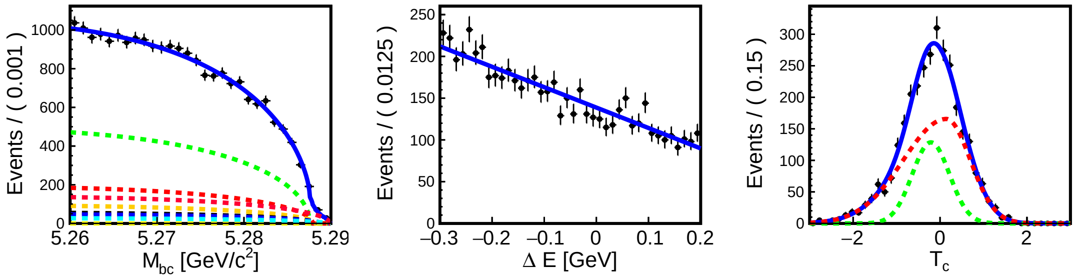


FIGURE 8.10: Distribution of ΔE (left), $M_{\rm bc}$ (middle) and T_c (right) for sideband region ($0 < \Delta E < 0.5$ GeV or $5.22 < M_{\rm bc} < 5.27$ GeV/c^2) used to extract sideband parameters.

8.3.3 Modelling tau pairs

The jet-like event kinematics of tau pairs is very similar to those of continuum. Consequently, the majority of tau pair events are eliminated by the continuum suppression.

However, the number of tau pairs is not accurately modelled, as indicated by the discrepancy between the data and MC in the ratio between tau pairs and the continuum in the sideband distribution of R2 (refer to Figure 8.12). Therefore, it is challenging to accurately determine the number of tau pairs using MC, making it difficult to correctly fit both the continuum and the tau pair. Nonetheless, after the application of continuum suppression, most tau pairs are removed, and their total number amounts to only 3% of the continuum. Therefore, given their nearly identical M_{bc}, ΔE, and T_c distributions (as seen in Figure 8.11), it is reasonable to consider the tau pair component as part of the continuum background. This simplification is factored into the systematic uncertainty of the continuum parameterisation.

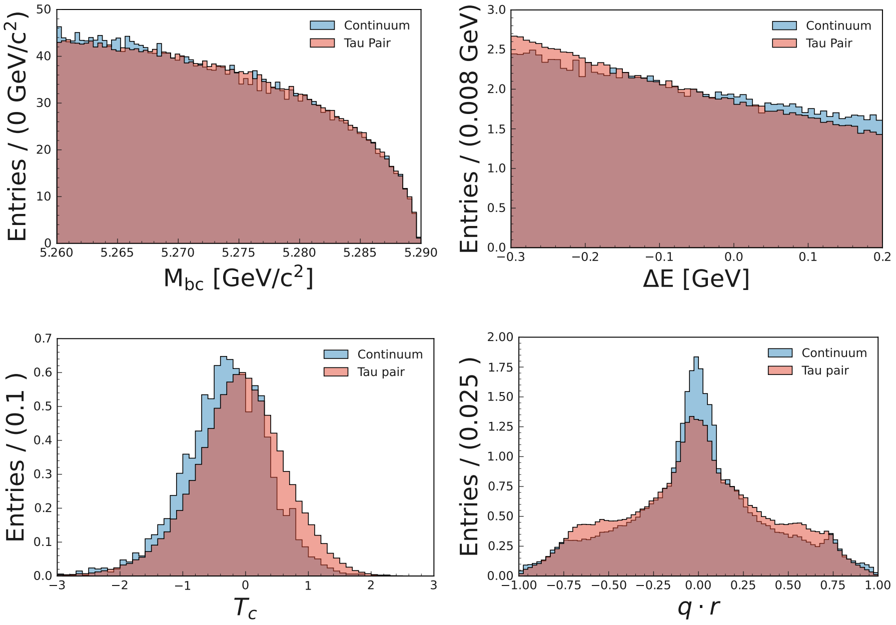


FIGURE 8.11: Comparison between tau pairs (red) and continuum (blue) for M_{bc} (top left), ΔE (top right), T_c (bottom left) and $q \cdot r$ (bottom right).

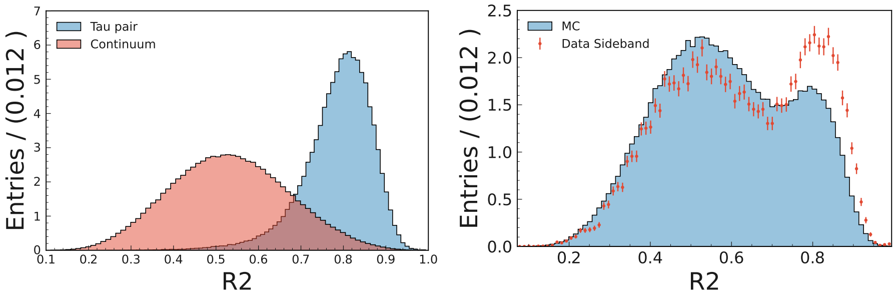


FIGURE 8.12: Comparison between tau pairs (red) and continuum (blue) for R2 (left) and R2 distribution for MC (blocks) and Data (dots) in the sideband region (right).

8.3.4 Modelling $B\overline{B}$

For the $B\bar{B}$ component, a KDE is used to model ΔE, while two ARGUS functions are employed to model $M_{\rm bc}$. This approach is taken because the $M_{\rm bc}$ endpoints are not accurately represented by a KDE. It is important to account for this, as the signal and $B\overline{B}$ both peak around 5.28 GeV/c^2 and their yields are comparable. The T_c PDF is the sum of a Gaussian and a bifurcated Gaussian. The PDF parameters of the $B\overline{B}$ are determined using MC that contains decays of $B^+ \to \rho^+\pi^0$ and $B^0 \to K_s(\to \pi^0\pi^0)\pi^0$ decays in proportions that reflect expected values. Fits to the $B\overline{B}$ only data are shown in Figure 8.13.

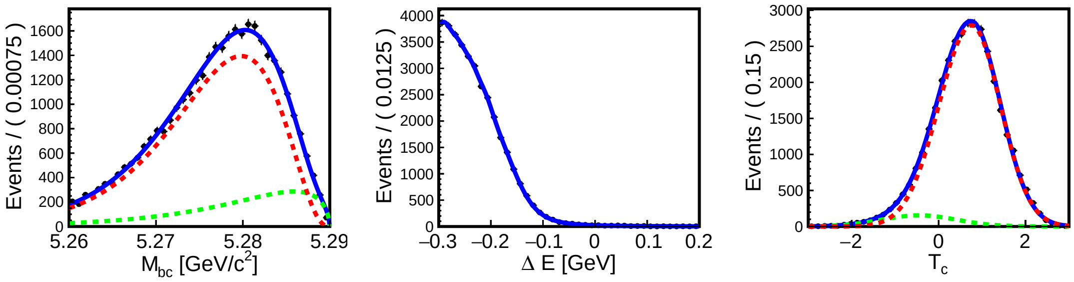


FIGURE 8.13: PDFs for $M_{\rm bc}$, ΔE and T_c for $B\bar{B}$ (bottom). T_c is composed of a bifurcated Gaussian (red) and a Gaussian (green).

8.3.5 Dependence of fitting variables to r bins

Using MC, the distribution of M_{bc}, ΔE, and T_c for the signal, continuum, and $B\overline{B}$ components is examined to determine their correlation with r. The plots for the bins defined as $0.0 < |r| < 0.1$, $0.5 < |r| < 0.626$, and $0.875 < |r| < 1.0$ are presented in Figure 8.15. The M_{bc} and ΔE distributions of the signal and $B\overline{B}$ are found to be independent of the r bins. However, for the continuum, it is observed that the slope of ΔE depends on the r bins, with low r bins having the smallest slope. Unfortunately, this correlation cannot be properly modelled using the 1 ab^{-1} data set. As illustrated in Figure 8.14, there is an insufficient number of continuum events in the higher bins of r for an accurate fit. As a result, the continuum ΔE parameters are extracted from the sideband and used for all r bins. This assumption of identical continuum ΔE distributions in each r bin will be taken into account in the systematic uncertainty.

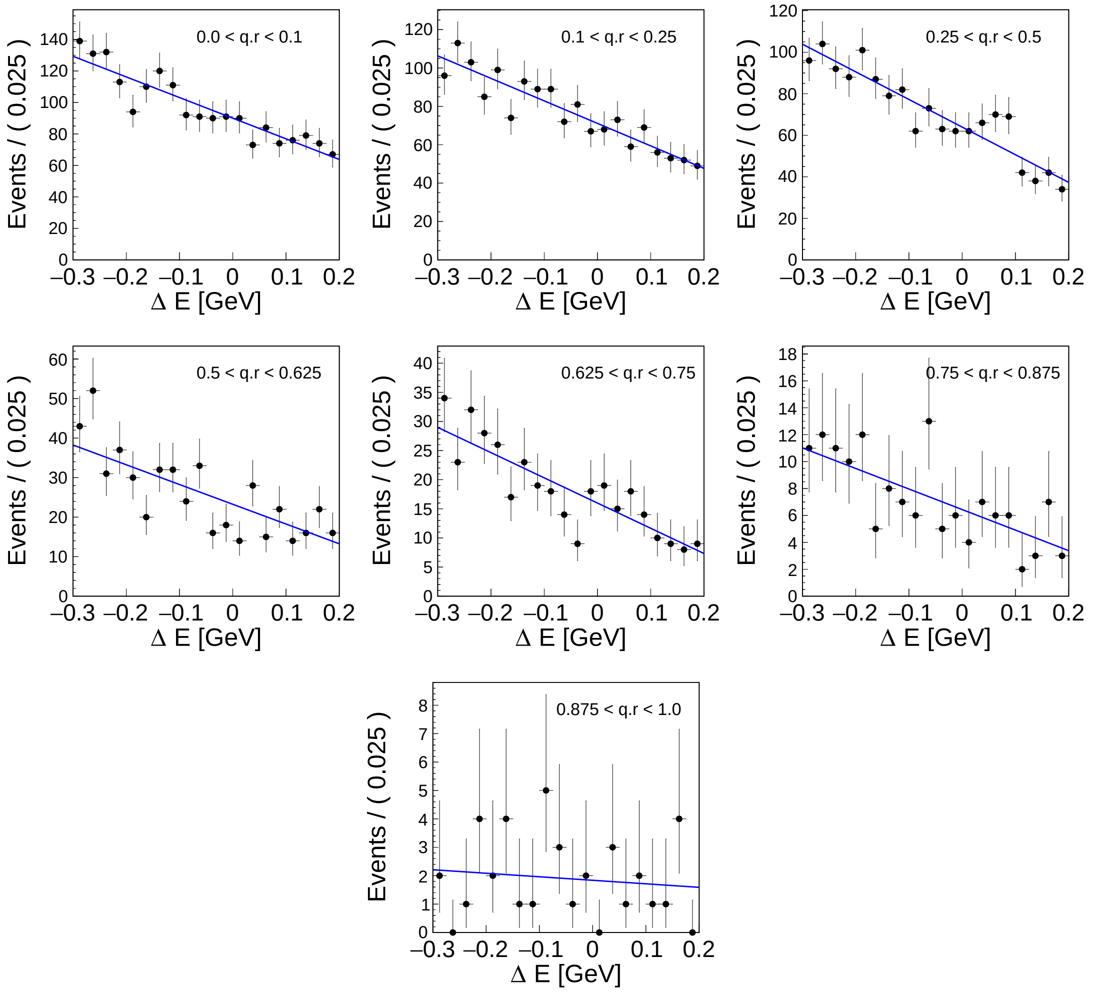


FIGURE 8.14: Fits to ΔE for the experimental sideband in each r bins. There are insufficient statistics at higher r bins for a reliable fit.

In addition, r and T_c are observed to be correlated for the signal and $B\overline{B}$ components. As such, different PDF parameters for each r bin are extracted from MC and used in all fits. The T_c PDF used in each r bin is shown in Appendix C.1). The correlation between T_c and r is expected, since T_c depends on the continuum suppression output and well-tagged B^0 candidates are more likely to have continuum suppression outputs close to 1. A summary of whether the PDF parameters used in the PDF fits are identical or vary for each $q \cdot r$ bin is shown in Table 8.5.

Component	$\boldsymbol{M_{\mathrm{bc}}}$	$\boldsymbol{\Delta E}$	$\boldsymbol{T_c}$
Signal	Identical	Identical	Different
Continuum	Identical	Identical	Identical
$B\bar{B}$	Identical	Identical	Different

TABLE 8.5: If the M_{bc}, ΔE and T_c PDF shapes are identical or different in each $q \cdot r$ bin for the signal, continuum and $B\overline{B}$.

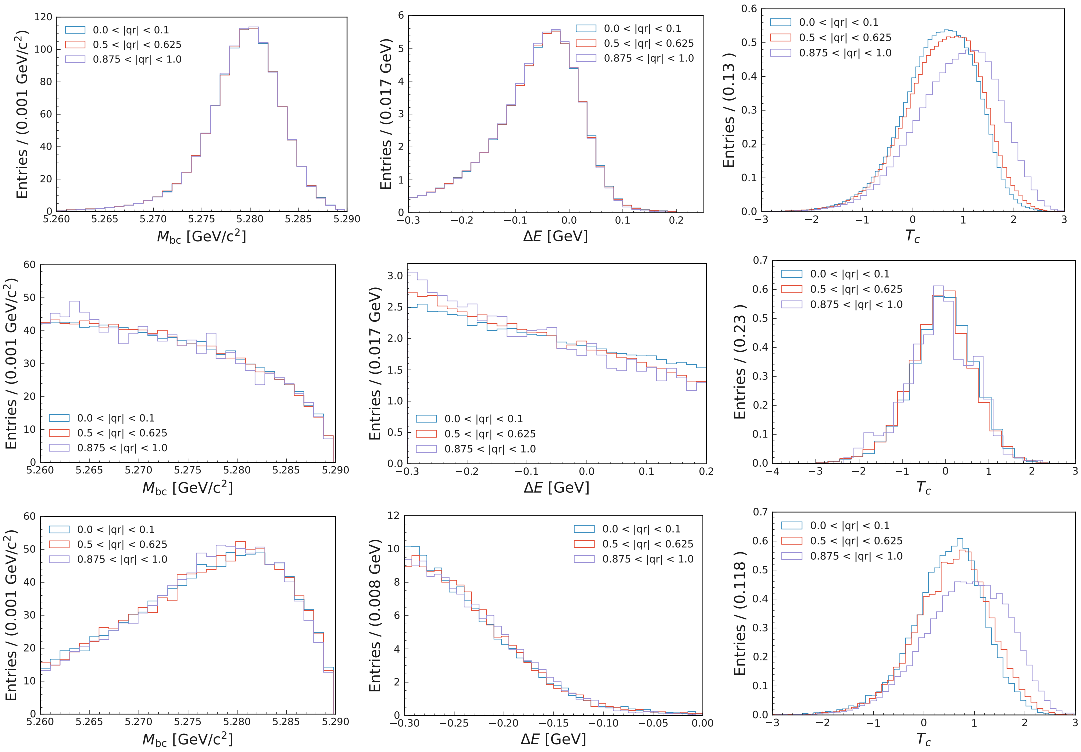


FIGURE 8.15: Normalised distributions for M_{bc} (left), ΔE (middle) and T_c (right) for signal (top), continuum (middle) and $B\bar{B}$ (bottom) for $q \cdot r$ bins defined as $0.0 < |r| < 0.1$, $0.5 < |r| < 0.626$ and $0.875 < |r| < 1.0$.

8.3.6 Asymmetry of $q \cdot r$ in sideband

The large continuum background is expected to be around 50 times larger than the signal, and hence it is important to check for any continuum flavour tagger output asymmetry. Ideally, the $q \cdot r$ distribution of the continuum background should be symmetrical. However, instrumental asymmetries in the detection of particles and antiparticles can lead to a skewed $q \cdot r$ distribution, which in turn biases the extracted $\mathcal{A}_{CP}$ of the signal. The asymmetry of $q \cdot r$ of the continuum, $\mathcal{A}_{\text{cont}}$, is calculated using the following formula:

$$\mathcal{A}_{\text{cont}} = \frac{N(\overline{B}^0) - N(B^0)}{N(B^0) + N(\overline{B}^0)} \tag{8.18}$$

where $N(B^0)$ and $N(\overline{B}^0)$ are the number of continuum events with $q \cdot r > 0$ and $q \cdot r < 0$, respectively. The $q \cdot r$ distribution of the continuum background in the sideband and in the signal regions is shown in Figure 8.3. Both distributions were found to agree. Therefore, it is assumed that the $\mathcal{A}_{CP}$ in the sideband and signal region of the experimental data are identical, as shown in Table 8.6. To account for this instrumental asymmetry, an $\mathcal{A}_{CP}$ term in the continuum PDF is included. This $\mathcal{A}_{CP}$ term is equal to the value extracted from the sideband, $\mathcal{A}_{\text{cont}}$,

$$P_i^c(M_{\text{bc}}, \Delta E, T_c, q) = [1 + q \cdot \mathcal{A}_{CP}]P^c(M_{\text{bc}}, \Delta E, T_c) \tag{8.19}$$

	$\boldsymbol{\mathcal{A}}_{\textbf{cont}}$
MC Signal Region	-0.024 ± 0.006
MC Sideband Region	-0.025 ± 0.003
Data Sideband Region	-0.033 ± 0.002

Table 8.6: The extracted $\mathcal{A}_{\text{cont}}$ values for the sideband and signal region in 1 ab^{-1} of MC, as well as for the sideband region in 189.9 fb^{-1} of data. The error is statistical only.

8.4 ToyMC Fits

ToyMC data sets serve as valuable tools for testing and understanding the fitting procedure before applying them to real, and often more complex, data. They provide a controlled environment in which the fitting procedure can be validated by generating data that follow some model, then applying a PDF to these data, and checking if the results match the known properties. They can also be used to estimate the expected statistical uncertainty. Finally, ToyMC data sets can be used to understand the systematic uncertainties in an analysis. This is done by varying the input parameters of the ToyMC data and studying the effect of these variations on the final result.

8.4.1 Creating ToyMC data sets

To construct the ToyMC data set for the validation of the fitting procedure, a source of signal, continuum, and $B\overline{B}$ events is needed. For the signal, events can be sampled without replacement from self-generated MC. This approach should not introduce bias, as a few hundred events are selected from a large dataset of 3.5×10^6 signal events. Furthermore, sampling directly from the generated signal MC means that the data set will inherit the underlying $\mathcal{A}_{CP}$, making it simple to create many ToyMC data sets with varying $\mathcal{A}_{CP}$ values. However, the generation of continuum and $B\overline{B}$ events requires a different approach. These events must be generated from PDFs, since directly sampling 189.9 fb^{-1} of data from the 1000 fb^{-1} `MC14ri_a` dataset could potentially introduce bias into the fitting process.

The continuum and $B\overline{B}$ samples are generated in the same way. To generate these samples with the correct proportions in each r bin, a PDF is created for the $q \cdot r$ distribution. This is achieved by first converting the $q \cdot r$ data into a histogram with 40 bins. Then, the histogram is converted to a continuous PDF using a technique known as second-order interpolation. This is used to estimate the values between the discrete bins of the histogram. It means fitting a quadratic function to each set of three adjacent bins and using this function to estimate the values between the bins. The result is shown in Figure 8.2, and is able to reproduce the $q \cdot r$ distribution of background events. For each generated event, a $q \cdot r$ value is sampled from the $q \cdot r$ PDF and assigned to the generated data point. This ensures that the ToyMC data set has the correct proportions in each bin.

For the $B\overline{B}$ component, this procedure is more complicated, since the T_c parameters for $B\overline{B}$ vary as a function of r. Therefore, the challenge lies in generating events with the correct T_c distribution in the appropriate proportion of $q \cdot r$ bins. To accomplish this for a single event, the $q \cdot r$ point is initially sampled and the corresponding PDF parameters for the sampled $q \cdot r$ value are used to generate a $B\overline{B}$ event. This procedure is repeated for each generated event to ensure that the ToyMC data set has the correct PDF in each bin with the correct proportion.

Assuming the world-average $\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0)$ is accurate, $\mathcal{B}(B^0 \to \pi^0\pi^0) = 1.59 \times 10^{-6}$, and a signal efficiency of 35.5%, approximately 116 ± 19 signal with $\mathcal{A}_{CP} = 0.33$ is expected in a 189.9 ab^{-1} data set. For the background, 5238 continuum and 87 $B\overline{B}$ events are expected.

8.4.2 Fitting to ToyMC data sets

Using the procedure described above, data sets were created corresponding to 189.9 ab^{-1} of data with different input signal yields and $\mathcal{A}_{CP}$ values. The signal yield and $\mathcal{A}_{CP}$ were determined by simultaneous fit in 7 bins of $q \cdot r$ in three dimensions: M_{bc}, ΔE, and T_c. The only parameters allowed to float are the signal and continuum yields along with the $\mathcal{A}_{CP}$. The $B\overline{B}$ yield, on the other hand, was fixed to the expected yield from $B^+ \to \rho^+\pi^0$ and $B^0 \to K_S^0(\to \pi^0\pi^0)\pi^0$ MC. This decision to fix the $B\overline{B}$ yield improves the precision of the fit and is a reasonable approach because the $B\overline{B}$ is largely made up of two decay modes that account for almost all the background $B\overline{B}$.

An example fit, with a fitted value of 122 ± 24 signal, 5234 ± 76 continuum and $\mathcal{A}_{CP} = 0.31 \pm 0.42$ is shown in Figure 8.16. These plots are signal-enhanced, meaning that when one variable is displayed, the selections on other variables are applied. The signal-enhanced region is defined as $5.275 < M_{bc} < 5.285\,\mathrm{GeV}/c^2$, $-0.1 < \Delta E < 0.05\,\mathrm{GeV}$, and $0 < T_c < 3$. For example, if the displayed variable is M_{bc}, then only events that satisfy $-0.1 < \Delta E < 0.05\,\mathrm{GeV}$ and $0 < T_c < 3$ will be shown in the plot. Without these signal-enhanced plots, the signal would be visible above the background.

It should also be noted that to produce these signal-enhanced plots, the M_{bc}-ΔE signal KDE must be recreated as it cannot be imported for plotting purposes. However, in scenarios where no plot is required and only the result is of interest, the KDE can be imported, which can save several hours of processing time.

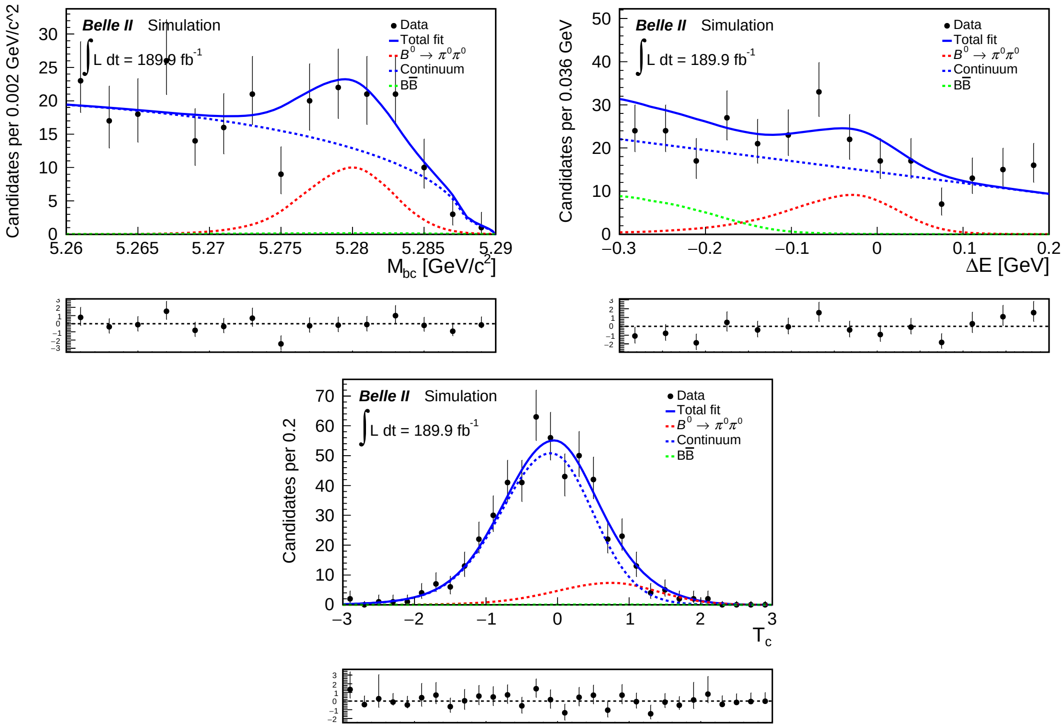


FIGURE 8.16: Fits to ΔE (left), M_{bc} (middle) and T_c (right) for ToyMC corresponding to 189.9 fb^{-1} in the signal-enhanced region.

8.4.3 ToyMC Results

ToyMC fits are repeated 1000 times with yields randomised by a Poisson distribution around the expected number. The properties of the fitter are studied by calculating the pull value of each fit and examining the distribution. The pull distribution is the error-weighted distribution of estimators around the true value defined as:

$$\text{Pull}(\theta_i) = \frac{\hat{\theta}_i - \theta_i}{\hat{\sigma}_{\hat{\theta}_i}}$$

where θ_i is the fit parameter (the signal yield or A_{CP}), $\hat{\theta}_i$ is the estimate of the parameter, and $\hat{\sigma}_{\hat{\theta}_i}$ is the estimate of its uncertainty. An unbiased fitter is one where the mean of the pull distribution is consistent with 0 and the width is consistent with unity. Figure 8.17 shows the distribution of the signal yield and $\mathcal{A}_{CP}$, along with their corresponding pulls, obtained from ToyMC fits. The input signal yield value was set to 115, and the $\mathcal{A}_{CP}$ value was set to 0.0, which are the signal yield and $\mathcal{A}_{CP}$ expected in the full 189.9 fb^{-1} data set. A Gaussian function is fitted to the pull to estimate the means and width of the pull distributions. The mean and width of the signal yield pulls are 0.007 ± 0.046 and 1.002 ± 0.034, respectively, indicating that there is no systematic bias in the fitter. Similarly, the mean and width of the pull of the $\mathcal{A}_{CP}$ values are 0.000 ± 0.044 and 0.971 ± 0.033, respectively.

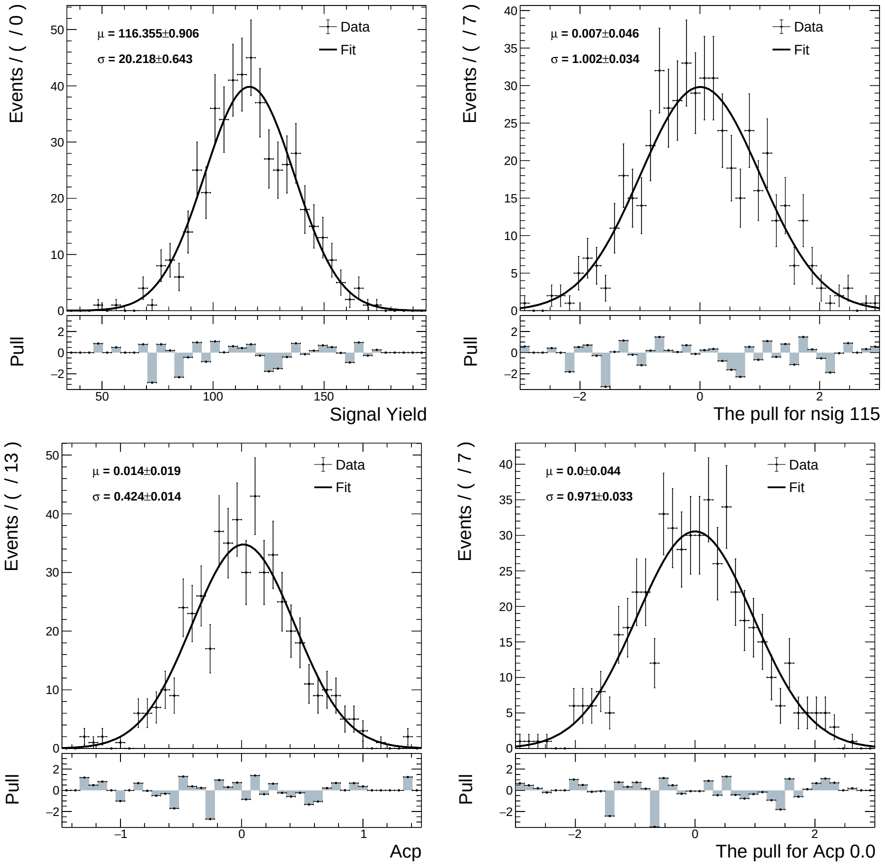


FIGURE 8.17: A Gaussian is applied to 1000 fitted yields (top left) and $\mathcal{A}_{CP}$ (bottom left) along with their corresponding pull distribution to the right for a 189.9 fb^{-1} dataset (Yield=115, $\mathcal{A}_{CP} = 0.0$).

8.4.4 Linearity Check

To verify that the extracted signal yield is unbiased, the number of signal events used in the ToyMC fits ranges from 80 to 160, in increments of 5. This range represents two standard deviations from the expected signal yield. Similarly, to verify that the extracted $\mathcal{A}_{CP}$ is unbiased, the signal events are drawn from datasets generated with A_{cp} ranging from -1 to $+1$, in increments of 0.2. In Figure 8.18, the fitted yield and $\mathcal{A}_{CP}$ are plotted against their respective input values, with an accompanying line of best fit. The best-fit line for the signal yield has a gradient of 0.995 ± 0.006 and an intercept of 1.392 ± 0.767, while the best-fit line for $\mathcal{A}_{CP}$ has a gradient of 1.001 ± 0.013 and an intercept of -0.021 ± 0.008. Both lines are almost consistent with a straight line with a gradient of 1 and an intercept of 0. This indicates that there are only small biases in the fitter for both the signal yield and $\mathcal{A}_{CP}$. These biases, which originate from imperfect signal modelling, are taken into account in the systematic uncertainties.

8.4.5 Significance Estimation

The expected significance of a 189.9 fb^{-1} dataset is estimated through ToyMC studies. Wilk's theorem is a fundamental result in statistical hypothesis testing and can be used to define a significance. It states that as the number of events approaches ∞, the distribution of the test statistic $-2\log(\Lambda)$ asymptotically approaches the χ^2 distribution. Here, Λ is the likelihood ratio, $NLL_{\text{null}}/NLL_{\text{free}}$, where NLL_{free} is the negative log-likelihood of the full fit and NLL_{null} is the negative log-likelihood of the fit where the signal yield is set to zero while the background yields are free to float. The χ^2 distribution has degrees of freedom equal to the difference in the number of parameters in the model with all parameters set free and the number of parameters in the model with some parameters fixed. This result is useful because it allows a p-values to be derived. This p-value is the probability of observing the test statistic obtained in data assuming that the null hypothesis is true.

In the case of $B^0 \to \pi^0\pi^0$ the test statistic is first converted to the probability for a given χ^2 with two degrees of freedom (p-value). The degrees of freedom equal two instead of one because setting the signal yield to zero also means that $\mathcal{A}_{CP}$ must equal zero. The p-value is then converted to a significance level using the quantile function

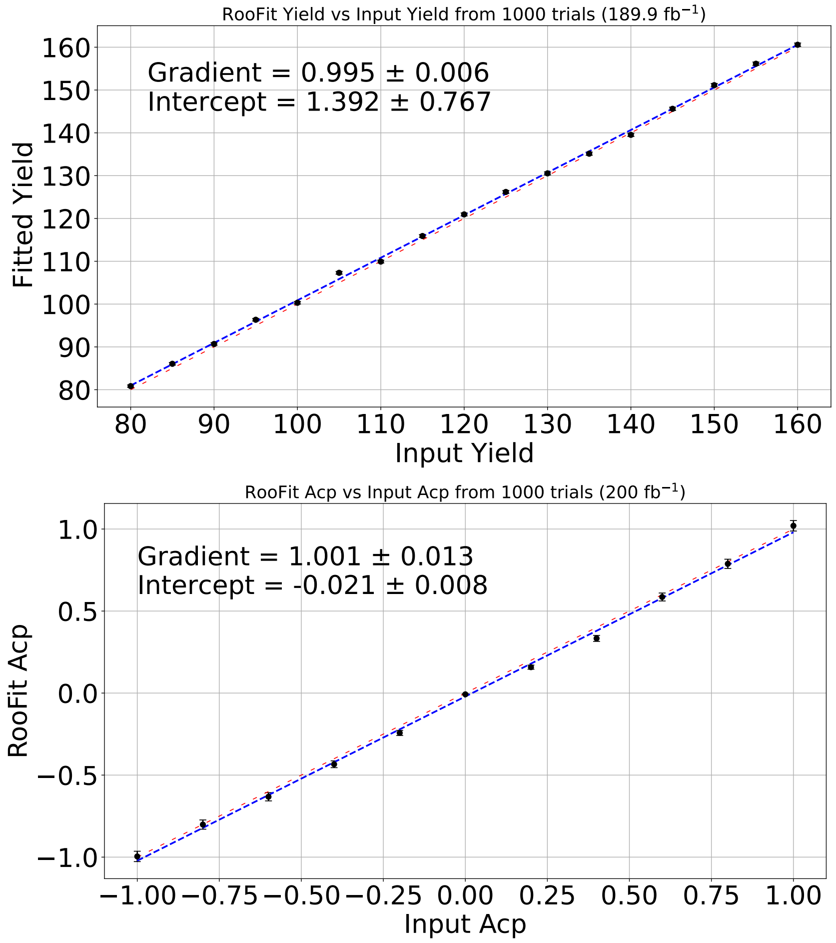


Figure 8.18: Linearity plot for $B^0 \to \pi^0\pi^0$ signal yield (left) and A_{CP} (right) for a 189.9 fb^{-1} dataset. The dotted red line represents a perfect fitter while the solid blue line is the best fit.

of the standard normal distribution. The quantile function is also known as the inverse cumulative distribution function as it takes a probability value as input and returns the corresponding value from the distribution.

A total of 1000 fits are performed on data sets with the number of signal events ranging from 80 to 160, in increments of 5. The average significance as a function of input yield is shown in Figure 8.19. Even when the signal yield is two standard deviations lower, a minimum significance of 6 is projected with a 189.9 fb^{-1} data set. For comparison, the Belle experiment achieved a statistical significance of 3.4 with a signal yield of $25.6^{+9.3}_{-8.4}$ using a 140 fb^{-1} data set [35].

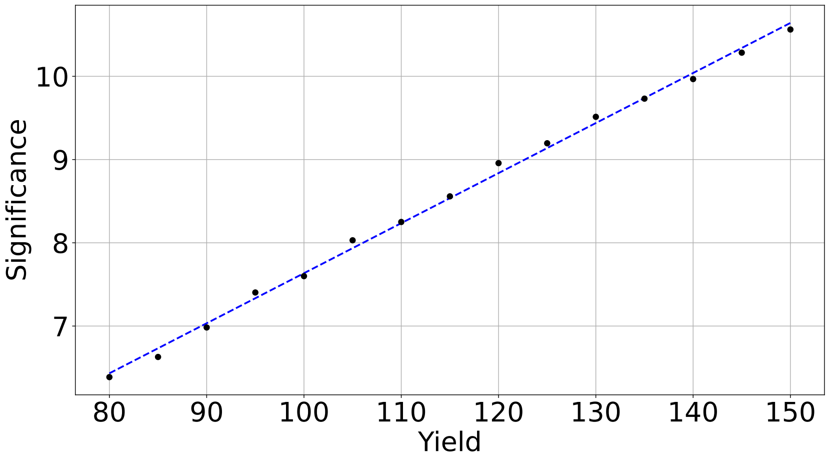


FIGURE 8.19: Average significance as a function of the number of $B^0 \to \pi^0\pi^0$ signal events for 1000 fits in 189.9 fb^{-1} dataset.

8.4.6 Sensitivity Plots

The branching fraction and $\mathcal{A}_{CP}$ of the $B^0 \to \pi^0\pi^0$ mode from 200 fb^{-1} to 50 ab^{-1} is estimated with 100 fits and compared to the world average. As shown in Figure 8.20, the statistical uncertainty on the branching fraction and $\mathcal{A}_{CP}$ is expected to surpass Belle with a 240 fb^{-1} data set. This is about one third of Belle's 693 fb^{-1} dataset. The world average branching fraction and $\mathcal{A}_{CP}$ is expected to be surpassed at 300 fb^{-1} and 500 fb^{-1} respectively. The difference between these estimates arises because Babar reported a $\mathcal{A}_{CP}$ measurement with very low statistical and systematic uncertainties relative to Belle, while the uncertainties of the branching fraction are comparable.

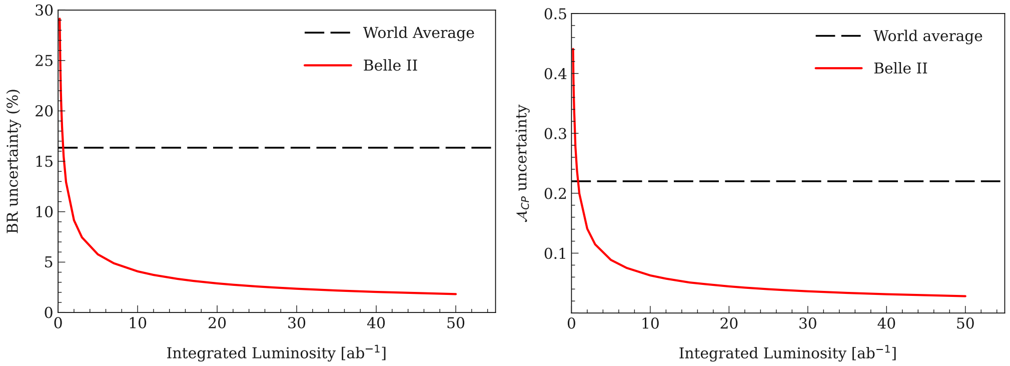


FIGURE 8.20: Sensitivity plots for the branching fraction (left) and $\mathcal{A}_{CP}$ (right). The uncertainties include both statistical and systematic. The systematic uncertainties is conservatively estimated to scale with the inverse square root of the luminosity.

8.4.7 Comparison with no $q \cdot r$ bin fit

Finally, to verify that the use of 7 bins of $q \cdot r$ provides superior performance, a total of 1000 ToyMC fits are performed without any $q \cdot r$ binning. Data sets of 189.9 fb^{-1} were used, consisting of 116 signal events with the $\mathcal{A}_{CP}$ value set to -0.4. As shown in Table 8.7, fitting without incorporating the $q \cdot r$ bins has a significantly higher statistical uncertainty, particularly in the uncertainty on $\mathcal{A}_{CP}$. It is important to note that this comparison only accounts for the statistical uncertainty. A more comprehensive study would include systematic uncertainties, where it is anticipated that the single-bin fit would exhibit fewer such uncertainties. Nevertheless, it is evident that a simultaneous fit in bins of $q \cdot r$ provides a more precise measurement.

	7 bins	**No bin**
Signal yield uncertainty (%)	18.1	19.7
$\mathcal{A}_{CP}$ uncertainity	0.40	0.69

TABLE 8.7: Comparison of the signal yield and $\mathcal{A}_{CP}$ uncertainty for 1000 ToyMC corresponding to 189.9 fb^{-1}. The uncertainty is statistical only.

Chapter 9

Fit validation with the control mode

To validate the $B^0 \to \pi^0\pi^0$ analysis, the analysis procedure is repeated on the $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ control mode. This mode was chosen because it contains two π^0 in the final state and has properties similar to the signal mode. Given that $\mathcal{B}(B^0 \to D^0\pi^0)$ equals $2.6 \pm 0.24 \times 10^{-4}$ and $\mathcal{B}(D^0 \to K^-\pi^+\pi^0)$ equals $13.9 \pm 0.5\%$, the overall branching fraction is calculated to be $3.64 \pm 0.33 \times 10^{-5}$. As a result, the number of events in the control mode is expected to be 10 to 20 times larger than that of the signal, with the $\mathcal{A}_{CP}$ expected to be 0. Before applying the $B^0 \to \pi^0\pi^0$ analysis to the experimental data, a similar procedure is carried out on the control mode. If the results of the control mode match previously established measurements, it gives confidence that the B meson reconstruction, continuum suppression, fitting, etc. used to measure the signal mode are also accurate. Otherwise, the procedure can be adjusted to correct for any issues, such as unexpected data-MC discrepancies.

9.1 Control mode selections

The selections for all photons and π^0 are identical to the signal mode, except that the $p_{\pi^0} > 1.5\,\mathrm{GeV}/c$ requirement is lifted. This is because the π^0 from the D^0 has a significantly lower momentum compared to the π^0 from the B^0, as shown in Figure 9.1. The charged kaons and pions employ 'standard' selections that are recommended for use due to their high data-MC agreement. For both particles, the transverse impact parameter, $|dr|$, is required to be less than 0.5 cm. This parameter measures the signed distance to the point of closest approach (POCA) to the beam interaction point in the $r - \phi$ plane. The longitudinal impact parameter, $|dz|$, is required to be less than 2 cm. This parameter measures the z-coordinate of the POCA to the beam interaction point. The polar angle θ of the kaon is required to fall within the range of 17° to 150°. This angle is measured with respect to the beam axis, with 0° being along the beam and 90° being perpendicular to it. Finally, the number of CDC hits associated with the track is required to be greater than 20. These hits are points where the particle has interacted with the detector, and a higher number of hits generally indicates a more reliable track reconstruction.

The particle identification (PID) system combines information from the TOP, ARICH, SVD, CDC, KLM, and ECL subdetectors. The track information from each subdetector is used to determine the likelihood of each charged particle hypothesis: electron, muon, pion, kaon, proton, and deuteron. The difference in log-likelihood between two particle hypotheses can be used to construct a combined likelihood ratio, $\mathcal{L}(\alpha/\beta)$, according to

$$\mathcal{L}(\alpha : \beta) = \frac{\prod_i \mathcal{L}(\alpha)}{\prod_i \mathcal{L}(\alpha) + \prod_i \mathcal{L}(\beta)} \tag{9.1}$$

where α and β represent two different particle types, and i extends over subdetectors. A $\mathcal{L}(\alpha/\beta)$ greater than 0.5 means that a track is more likely to originate from a particle of type α than from a particle of type β. For π^+ candidates, the likelihood ratio $\mathcal{L}(K : \pi)$ is required to be less than 0.4. For K^- candidates, the likelihood ratio $\mathcal{L}(K : \pi)$ is required to be greater than 0.6.

The K^+, π^-, and π^0 candidates are combined to form D^0 candidates. The D^0 mass (`M`) is required to be between 1.84 GeV/c^2 and 1.88 GeV/c^2, which corresponds to a range of approximately one sigma about the known D^0 mass. The momentum resolution of the D^0 candidates was improved by performing a kinematic fit that constrained their mass to the known value (`massKFit`). These D^0 meson candidates are then combined with the remaining π^0 candidates to form B^0 meson candidates. The selections for the charged particle, D^0, and B^0 are summarised in Table 9.1.

After applying these selections, the candidate multiplicity equals 1.05. To select the optimal candidate, we choose the one with the minimum deviation of the reconstructed

Particle	Selection
π^+	Transverse impact parameter $\lvert dr \rvert < 0.5$ cm
	Longitudinal impact parameter $\lvert dz \rvert <$ 2 cm
	Binary PID between kaons and pions < 0.4
	Polar angle θ within the range $17^\circ < \theta < 150^\circ$
	Number of CDC hits associated to the track > 20
K^-	Transverse impact parameter $\lvert dr \rvert < 0.5$ cm
	Longitudinal impact parameter $\lvert dz \rvert <$ 2 cm
	Binary PID between kaons and pions > 0.6
	Polar angle θ within the range $17^\circ < \theta < 150^\circ$
	Number of CDC hits associated to the track > 20
D^0	$1.84 < \mathrm{M} < 1.88$ GeV/c^2
	`massKFit` $\chi^2 > 0$
B^0	$5.26 < M_{\mathrm{bc}} < 5.29$ GeV/c^2
	$-0.2 < \Delta E < 0.2$ GeV

Table 9.1: Control mode selections for π^+, $^-$, D^0, and B^0 particles.

invariant mass of π^0 from the known value for the π^0 originating from the B^0, denoted as $|dM(\pi^0)|$. This π^0 is selected due to its kinematic similarity to the neutral pions from the signal mode. This selection method proves to be 56.6% efficient in choosing the correct B^0.

The signal efficiency is 12.9% and the number of truth-matched candidates over the total number of candidates, or purity, is 88.0%. A dataset of 189.9 fb^{-1} is expected to yield 653 ± 44 control events, 14,093 continuum events, and 2026 $B\bar{B}$ events. The continuum and $B\overline{B}$ background are estimated using the 1 ab^{-1} `MC14ri_a` data while the control mode is estimated using the control-only MC sample. As a result of the relatively large number of charged particles in the final state, the tau pairs can be easily removed, making their contribution negligible.

The majority of the background is due to rare $B\bar{B}$ decays which are skewed to negative values of ΔE but peaks in M_{bc} and T_c in the same region as the signal. The other background, known as self-cross feed (SxF) comes from $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ events being misreconstructed. This occurs when a particle from the tag side is wrongly included in the control mode reconstruction, i.e. $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ candidates from control-only MC that are not correctly reconstructed.

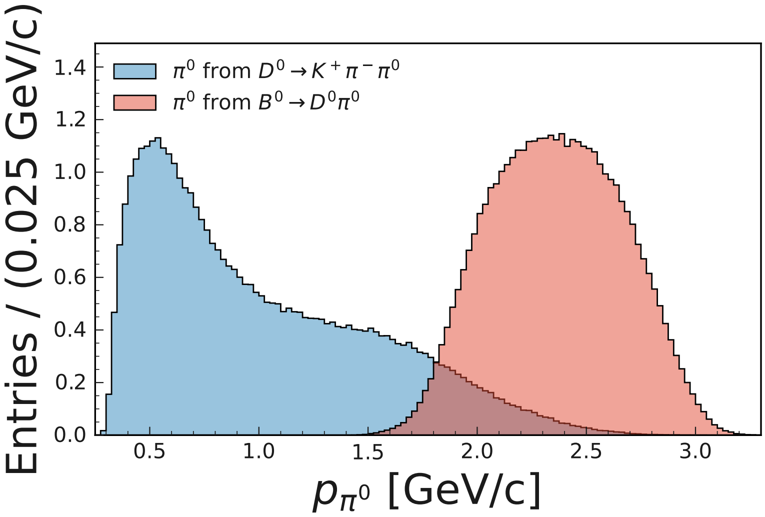


FIGURE 9.1: Momentum for π^0 from $D^0 \to K^+\pi^-\pi^0$ (blue) and $B^0 \to D^0\pi^0$ (red) for correctly reconstructed $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ candidates.

9.2 Continuum Suppression

To validate the signal-mode continuum suppression and estimate its systematic uncertainty, signal-mode continuum suppression training is applied to the control mode. The results are shown in Figure 9.2 and an AUC of 0.917 is achieved. For comparison, when the continuum suppression is retrained, an improved AUC curve of 0.929 is obtained. This discrepancy arises primarily because the continuum selected by the signal criteria is predominantly composed of $u\bar{u}$ and $d\bar{d}$ events, while the continuum selected by the control criteria is mainly composed of $u\bar{u}$ and $c\bar{c}$ events. This difference in the composition of the continuum background between the signal and the control is illustrated in Table 9.3 and Figure 9.3. However, this is not an issue since the focus is on the agreement between data and MC, and not on the performance of classifier.

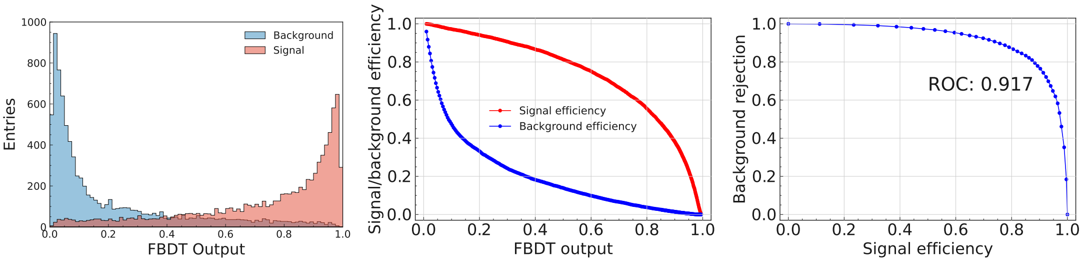


FIGURE 9.2: CSMVA output on training test-sample (left), signal/background efficiency (middle) and the ROC curve (right) for signal-mode CS training applied to the control mode.

The continuum suppression retains 62.5% of control mode events while rejecting 91.9% of the continuum background. The total efficiency of the control mode is 9.0% and the truth-matched fraction is 89.0%. Therefore, in a 189.9 fb^{-1} data set, approximately 616 ± 42 control mode events are expected. For the background, MC predicted a total of 3966 continuum and 2197 $B\overline{B}$ events.

Selection	**Control efficiency (%)**
`Skimming`	58.0
`Control selections`	12.9
`Continuum Suppression`	9.0

Table 9.2: Summary of control efficiency for major steps in the analysis.

Background	**Signal**	**Control**
$u\bar{u}$	180,586	34,546
$d\bar{d}$	50,252	8288
$s\bar{s}$	13,734	6202
$c\bar{c}$	22,388	42,265

Table 9.3: Breakdown of the continuum background for the signal and control mode.

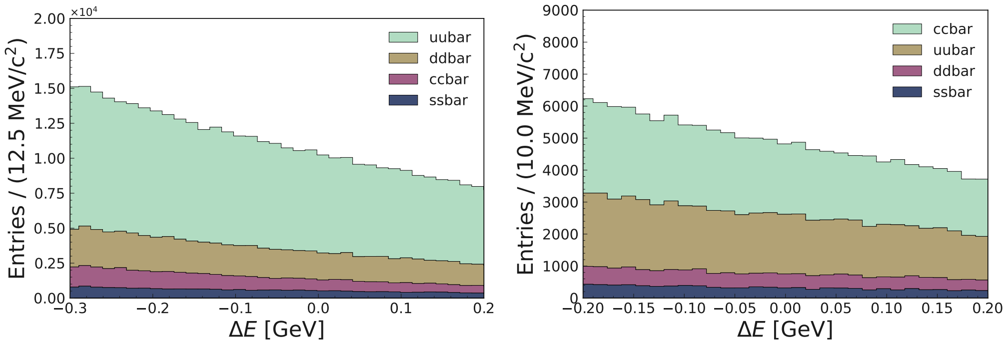


Figure 9.3: Continuum background composition for the signal (left) and control mode (right).

9.3 ToyMC Fits

The Probability Density Functions (PDFs) used to model M_{bc}, ΔE and T_c for the control, continuum, $B\overline{B}$ and self-cross feed components are similar to the ones used in the signal mode. A correlation of 10.2% between M_{bc} and ΔE is observed in the control mode, which is accounted for using a 2D Kernel Density Estimation (KDE). The continuum component M_{bc} is modelled by a double ARGUS function, while a first order Chebychev function is used for ΔE. The parameters for the continuum PDF are derived from the experimental sideband region, as depicted in Figure 9.5.

For the $B\bar{B}$ component, ΔE is modelled with a KDE, while $M_{\rm bc}$ is modelled with two ARGUS functions. This approach is adopted due to the inadequate modelling of the $M_{\rm bc}$ endpoints by a KDE. It is crucial to account for this, given that both the signal and $B\overline{B}$ peak around 5.28 GeV/c^2 and the yields of $B\overline{B}$ and signal are comparable.

The T_c PDF for all components is the sum of a Gaussian and a bifurcated Gaussian. This is summarised in Table 9.4. An unbinned maximum likelihood fit was performed on the control, continuum, and generic MC datasets to extract the PDF parameter shapes. The fit results are shown in Figure 9.4.

Component	$\boldsymbol{M_{\rm bc}}$	$\boldsymbol{\Delta E}$	$\boldsymbol{T_c}$
Signal	2D KDE	2D KDE	Bi-Gaussian + Gaussian
Continuum	8 ARGUS	1st order Chebychev	Bi-Gaussian + Gaussian
$B\bar{B}$	Double ARGUS	KDE	Bi-Gaussian + Gaussian
Crossfeed	ARGUS + Gaussian	KDE	Bi-Gaussian + Gaussian

TABLE 9.4: The PDF shapes used to fit the signal, continuum, $B\overline{B}$ and self-cross feed components of the control mode.

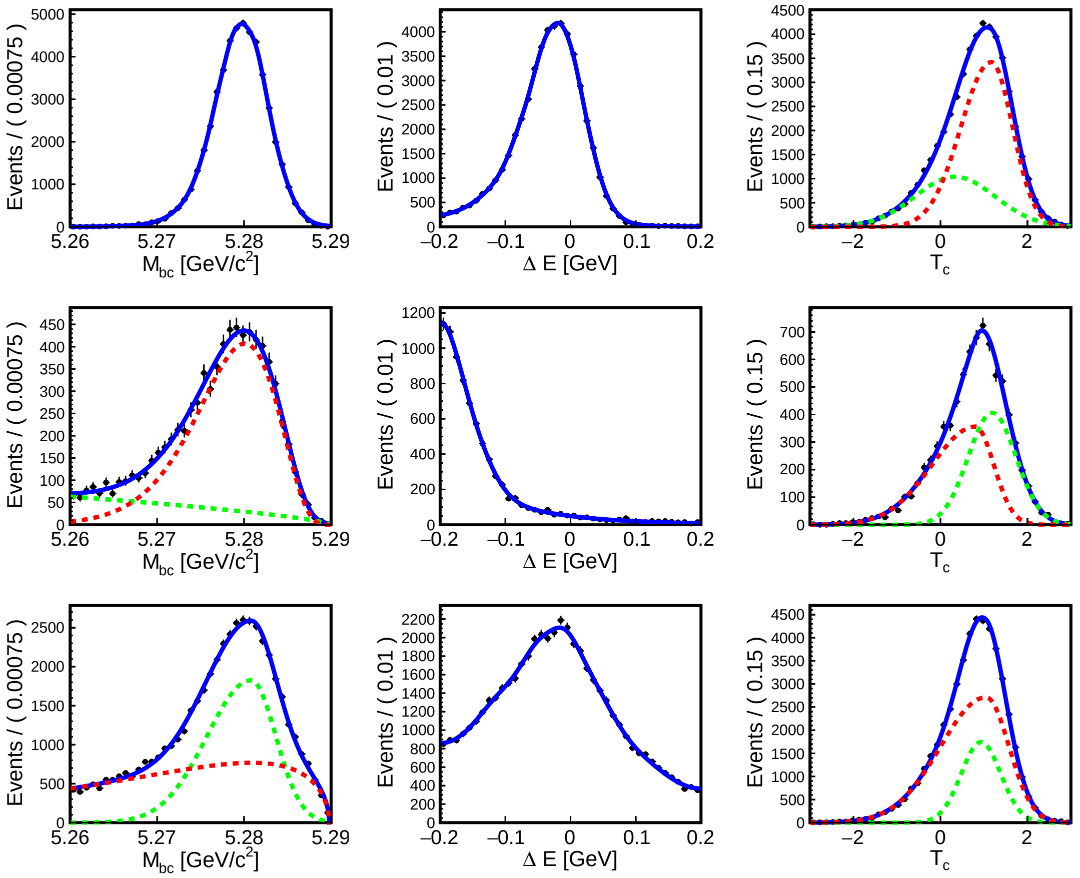


FIGURE 9.4: Control mode PDFs for $M_{\rm bc}$, ΔE and T_c for signal (top), $B\bar{B}$ (middle) and self-cross feed (bottom) fitted to MC.

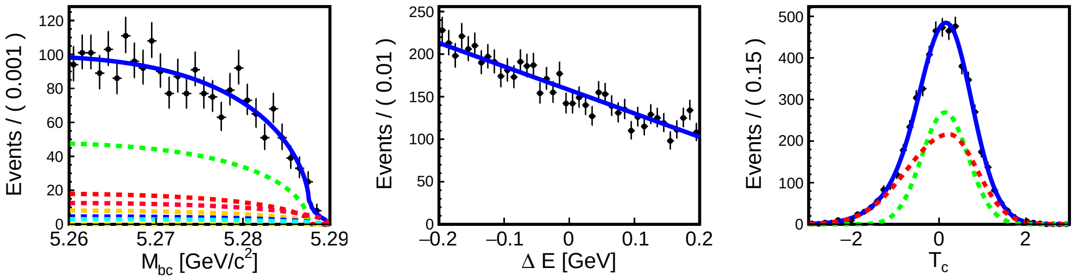


FIGURE 9.5: Control mode PDFs for $M_{\rm bc}$, ΔE and T_c for continuum, fitted to experimental sideband data.

Given that the self-cross feed component peaks in the same regions as the control mode, MC is used to constrain the self-cross feed yield to be 12.3% of the signal yield. The procedure to create ToyMC data sets for the control mode is identical to that for the signal mode. The only difference is the addition of the self-cross feed, which is sampled without replacement from the control-only MC. In addition, the $B\overline{B}$ yield is allowed to float since unlike the signal mode, the $B\overline{B}$ background is not dominated by a single decay mode and therefore cannot be accurately estimated.

The control yield and $\mathcal{A}_{CP}$ values are determined by performing a three-dimensional (M_{bc}, ΔE, T_c) unbinned extended maximum likelihood fit simultaneously to events in the seven intervals of r. All PDF parameters are fixed with only the control, continuum, and $B\overline{B}$ yields and the $\mathcal{A}_{CP}$ allowed to float. This fitting procedure is applied to a ToyMC data set corresponding to 189.9 fb^{-1}, where the expected yield comprises 616 signal events with $A_{CP} = 0.0$ and 3966 continuum events. An example of a fit, which yields 649 ± 38 signal events with $\mathcal{A}_{CP} = 0.10 \pm 0.16$ and 3965 ± 73 continuum events, is shown in Figure 9.6.

9.3.1 Linearity Check

The fit is repeated 1000 times for an input yield ranging from 400 to 600, in increments of 20. To test the linearity of the procedure, a best-fit line is applied to the extracted yield as a function of the input yield, as shown in Figure 9.7. The best-fit line for the control yield has a gradient of 1.011 ± 0.003 and an intercept of 0.03 ± 1.99, while the best-fit line for $\mathcal{A}_{CP}$ has a gradient of 1.033 ± 0.001 and an intercept of -0.008 ± 0.007. Both lines are almost consistent with a straight line with a gradient of 1 and an intercept

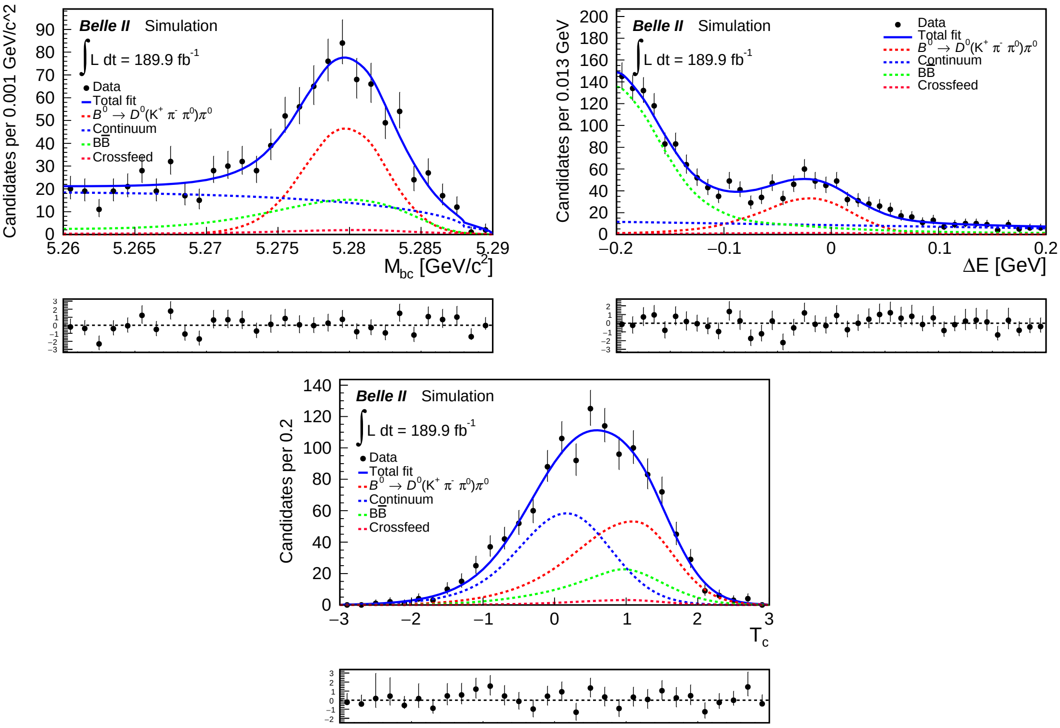


FIGURE 9.6: Signal enhanced distribution of ΔE (left), $M_{\rm bc}$ (middle) and T_c (right) for $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ control mode in 189.9 fb^{-1} ToyMC. The projection of a simultaneous unbinned maximum likelihood fit is overlaid.

of 0. Similarly to the signal mode, there are only small biases for both the control yield and $\mathcal{A}_{CP}$ due to the imperfect modelling of the various components.

9.4 Corrections to MC and data

9.4.1 Photon Energy Bias

In the experimental data, a small but consistent negative shift in ΔE is observed for decays involving neutral pions when compared to MC. To address this, a 'Photon Energy Bias Correction' is applied exclusively to the data. As a check of this correction, a data-driven method is employed using the $e^+e^- \to D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ mode. The measured π^0 momentum and the π^0 momentum predicted using energy-momentum conservation are compared to estimate the effectiveness of the corrections.

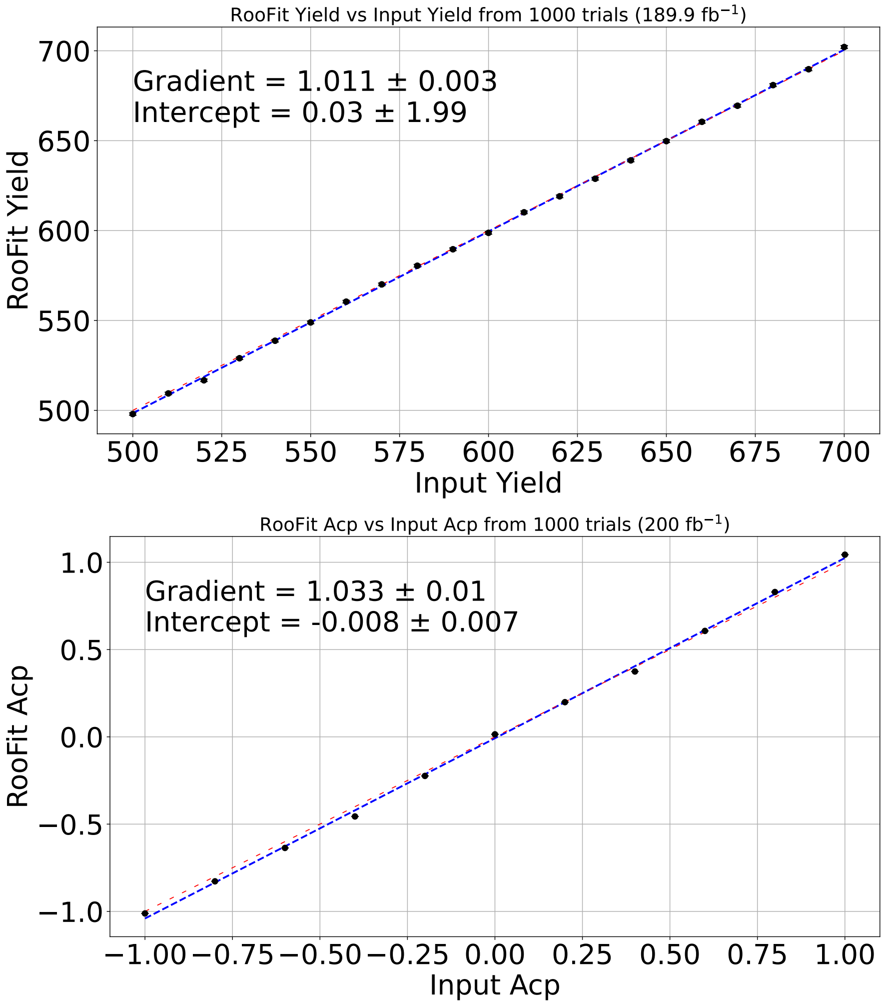


FIGURE 9.7: Linearity plot for $B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$ control yield (left) and A_{CP} (right) for a 189.9 fb^{-1} dataset. The dotted red line represents a perfect fitter while the solid blue line is the best fit.

To reconstruct the final-state particles, all selections are identical to those described in Section 5.5. The π^0 momentum can be predicted from energy-momentum conservation:

$$\begin{aligned}
(M_{D^0})^2 &= (E_{K_S^0} + E_{\pi^0})^2 - (\vec{p_{K_S^0}} + \vec{p_{\pi^0}})^2 \\
&= E^2_{K_S^0} + 2E_{K_S^0}E_{\pi^0} + E^2_{\pi^0} - p^2_{K_S^0} - p^2_{\pi^0} - 2p_{K_S^0}p_{\pi^0}\cos\theta \\
&= M^2_{K_S^0} + M^2_{\pi^0} + 2E_{K_S^0}E_{\pi^0} - 2p_{K_S^0}p_{\pi^0}\cos\theta
\end{aligned}$$

Using $E_{\pi^0} = \sqrt{M^2_{\pi^0} + p^2_{\pi^0}}$, and solving exactly for p_{π^0}:

$$p_{\pi^0} = \frac{CM \pm \sqrt{4C^2K^2m^2 - 4K^4m^2 + K^2M^2}}{2(C^2 - K^2)} \tag{9.2}$$

where $M = M^2_{K_S^0} + M^2_{\pi^0} - M^2_{D^0}$, $C = p_{K_S^0}\cos\theta$, m is the mass of π^0 and $K = E_{K_S^0}$.

On average, the predicted π^0 momentum is closer to the true momentum as shown in Figure 9.9. This is because the momentum/energy resolution of $\pi^\pm$ is excellent at Belle II and is more precise than the neutral measurements. Furthermore, in the experimental data, the π^0 momentum is less likely to be affected by ECL-related miscalibration, since the momentum of charged particles is determined primarily by the CDC. It should be noted that the difference between the predicted π^0 momentum and the true π^0 momentum is skewed in the positive direction. This is due to the charged $\pi^\pm$ losing momentum as they travel through the detector. Therefore, the predicted π^0 momentum is more likely to be overestimated.

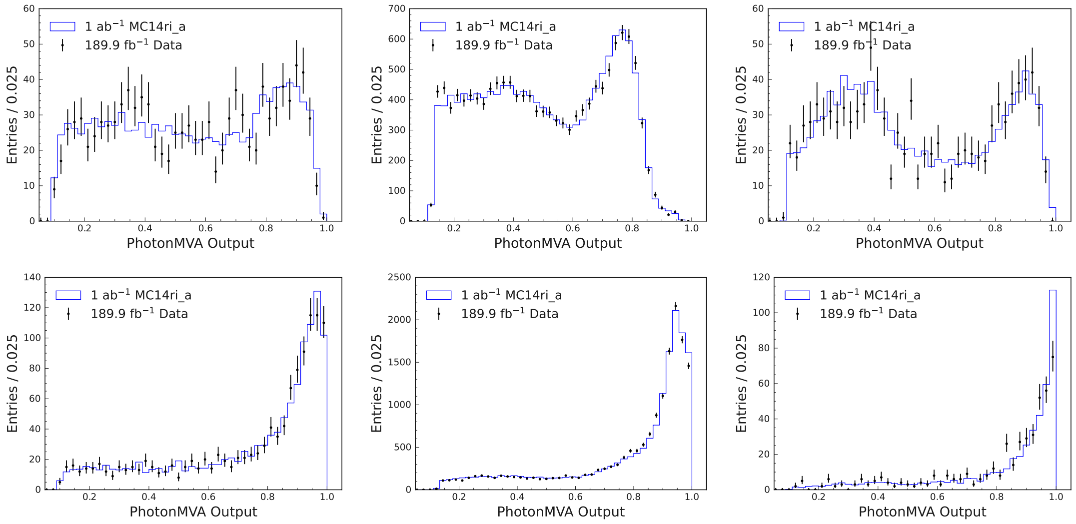


FIGURE 9.8: `PhotonMVA` output for photons originating from soft π^0, defined as $p < 1.5\,\mathrm{GeV}/c$, in the $D^{*+} \to D^0(\to K_s(\to \pi^+\pi^-)\pi^0)\pi^+$ calibration mode for data (dots) and MC (blocks) for the backward (top left), barrel (top middle) and forward (top right) region. Similarly, hard π^0, defined as $p > 1.5\,\mathrm{GeV}/c$, are shown for the backward (bottom left), barrel (bottom middle) and forward (bottom right) regions.

9.4.1.1 Calculating the correction factor

The predicted π^0 momentum as a function of the measured π^0 momentum is shown in Figure 9.10, using 1 ab^{-1} of `MC14ri_a` and 189.9 fb^{-1} of data. A line of best-fit is applied, with the intercept set to zero, and the gradient is extracted. In the momentum range of the signal π^0 (1.7 to 3.8 GeV/c), the gradient of MC and data are 1.0116 and 1.0122, respectively. This difference represents the data-MC discrepancy. To match the MC gradient to the data one, a correction factor of $1.0113/1.0122 = 0.9994$ is applied to the π^0 momentum, but only in MC. If the ‘Photon Energy Bias Corrections’ is not

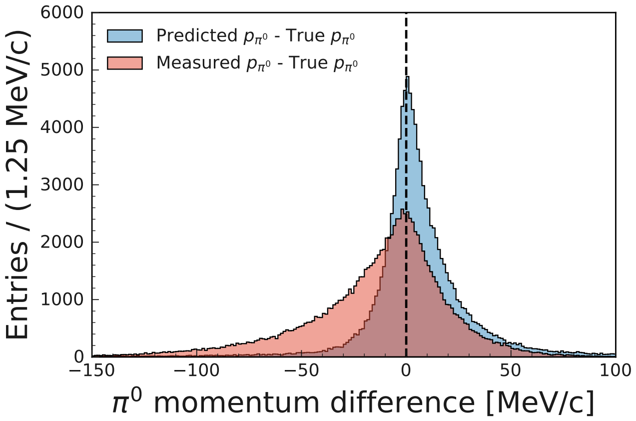


FIGURE 9.9: Difference between the predicted π^0 momentum and the true π^0 momentum (blue) vs the difference between the measured π^0 momentum and the true π^0 momentum (red).

applied, the Data-MC discrepancy is much larger since the gradient in Data is 1.0140 and hence the correction factor is $1.0113/1.0140 = 0.9976$.

To correct for the remaining Data-MC discrepancy so that the MC gradient equals the Data one, in MC only the correction factor of 0.9994 is applied to the π^0 momentum and ΔE is recalculated:

$$\Delta E = (E_{\pi_1^0}^{\text{modified}} + E_{\pi_2^0}^{\text{modified}} + E_{\text{others}}) - E_{\text{CMS}}/2 \tag{9.3}$$

where $E_{\pi^0}^{\text{modified}} = \sqrt{(p_{\pi^0} \times C)^2 + m_{\pi^0}^2}$ is the modified π^0 energy, C is the 'correction factor', E_{others} is all other particles, and E_{CMS} is the energy in the center of mass frame. This shifts the simulated ΔE by approximately -1 MeV, as shown in the left plot of Figure 9.11). This is the ΔE distribution that is expected to be observed in the experimental data and is used instead of the ΔE previously described. This modification to the ΔE shape has a minimal impact on the overall analysis procedure, as the 'Photon Energy Bias Corrections' effectively addresses any potential issues.

9.4.1.2 Control mode correction

The $B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$ control mode contains two π^0 with different momentum ranges. The momentum range of the π^0 that is the daughter of the B^0 is between 1.5 and 3 GeV/c and corresponds to a correction factor of 0.9997. The momentum range of the π^0 that is the granddaughter of the B^0 is less than 2.5 GeV/c and corresponds to a correction factor of 0.9994. The nominal and modified ΔE distribution is shown

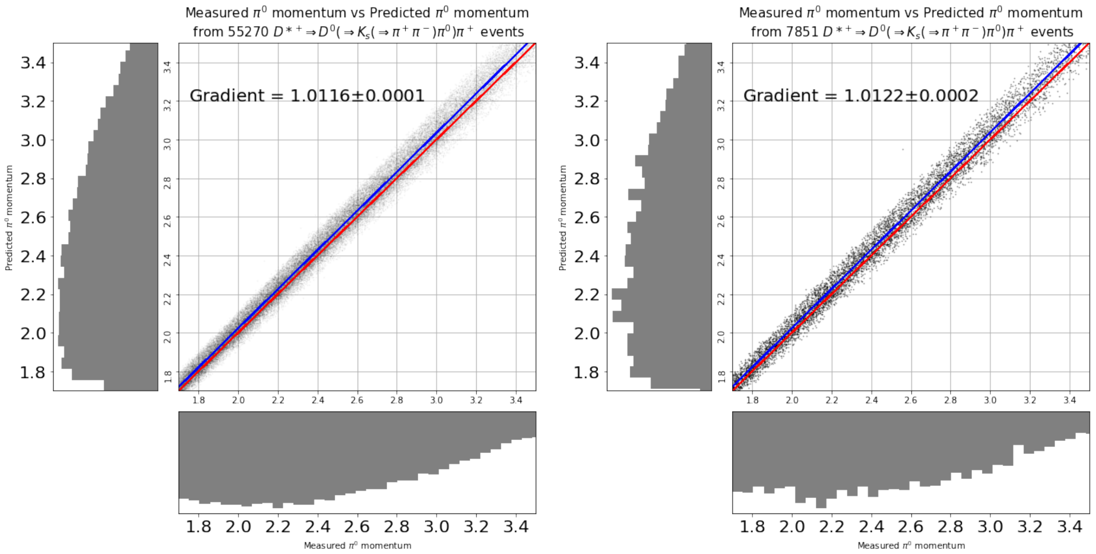


FIGURE 9.10: Predicted vs measured π^0 momentum in MC (left) and data (right) in the signal π^0 momentum range. The red dashed line has a gradient of one, while the blue dashed line is the best fit.

on the right graph of Figure 9.11. This is the ΔE distribution of the control mode that is expected to be observed in the experimental data, and is used instead of the ΔE previously described. A crystal ball is fitted to the new and nominal ΔE distribution and the shift in the control mode is determined to be 0.6 MeV. The results are summarised in Table 9.5.

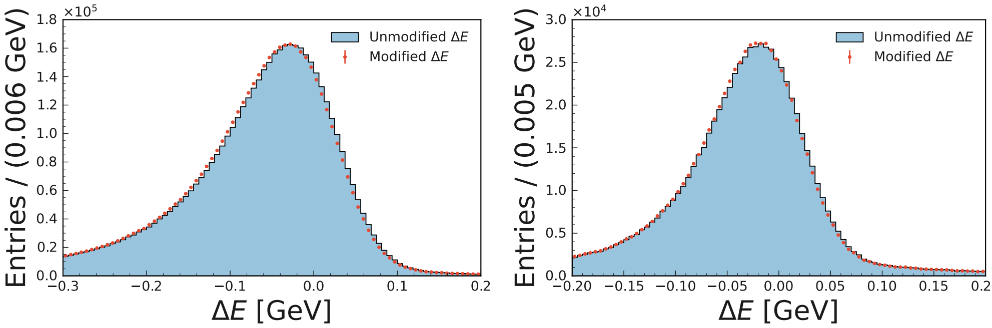


FIGURE 9.11: Unmodified (blue) and modified (red) ΔE for the signal (left) and control (right) mode.

	Correction Factor(s)	**ΔE Shift**
Signal	0.9994	-1 MeV
Control	0.9994 & 0.9997	-0.6 MeV

TABLE 9.5: Summary of π^0 correction factor and the corresponding ΔE shift.

9.4.2 Width of ΔE

The width of ΔE depends highly on the precise reconstruction of photons in the ECL. Various factors, such as the miscalibration of the crystals, can cause a discrepancy in width between MC and data. To determine the ΔE width that best fits the data, a 'fudge factor' is typically applied to the width of the PDF that models ΔE. However, since ΔE and M_{bc} are modelled with a 2D KDE, there are no parameters that can be changed to freely modify the width. The MC data set must be altered to create a new KDE with different widths. To do this, each value of ΔE is altered according to

$$E_{\text{new}} = E_{\text{peak}} + sF(E_{\text{measured}} - E_{\text{peak}}) \tag{9.4}$$

where E_{new} is the new ΔE value, E_{peak} is the position of the ΔE peak, determined with a Crystal Ball Fit, sF is a scaling factor on the ΔE width (which will be determined), and E_{measured} is the measured ΔE value. The values of ΔE closer to the mean will shift less toward the mean, while values farther away will shift more, resulting in a narrower peak. Then a new KDE is generated from this modified MC data set, as shown in Figure 9.12. Furthermore, since the modification is merely a scaling factor, the correlation remains unaffected.

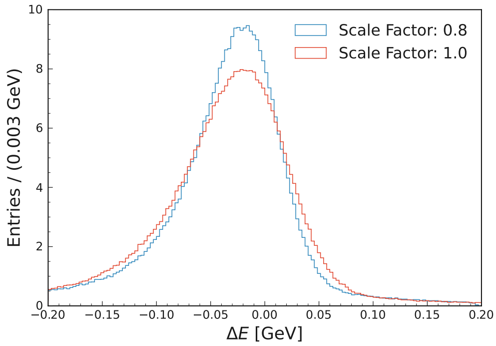


FIGURE 9.12: ΔE with (blue) and without (red) scale factor applied.

To determine the optimal scaling factor, control mode KDEs are created with scaling factors varying from 0.7 to 1.3, in increments of 0.005. These KDEs are used to fit the experimental control mode data, and the negative log-likelihood (NLL) is determined. As shown in the left-hand plot of Figure 9.13, a scaling factor of 0.950 is the minimum NLL and corresponds to the scaling factor that provides the best fit.

To validate that this scaling factor can be consistently determined, 200 ToyMC datasets are created with the signal generated from KDEs from with scaling factor 0.90, 0.85, 0.80 and 0.75, and the process of determining the scaling factor is repeated for each dataset. As shown in the right-hand plot of Figure 9.13, on average, the optimal scaling factor can be consistently determined. The scaling factor of 0.950 will be applied to the width of the ΔE PDF in fits to real data for both the signal and control modes.

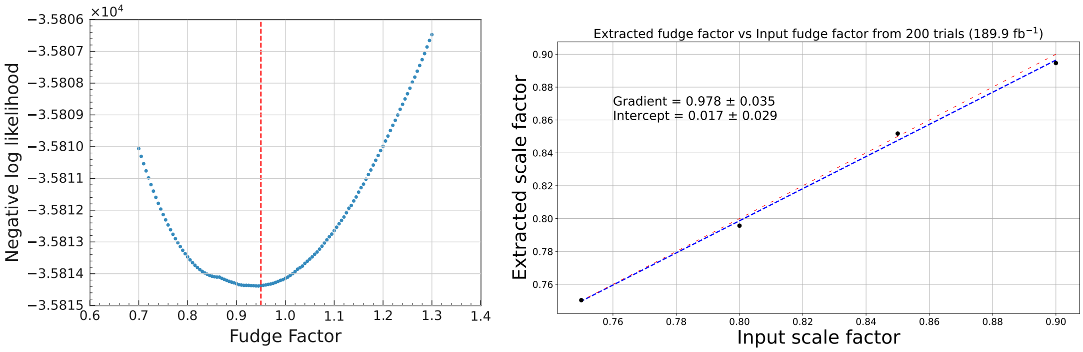


FIGURE 9.13: Negative log-likelihood for different scaling factors on experimental control mode data (top left). Distribution of optimal scaling factors for an input scaling factor of 0.20 (top right). Linearity plot of scaling factor (bottom).

Chapter 10

Fitting to experimental data

10.1 Control mode result

The control mode is fitted with the 'Photon Energy Bias' and the scaling factor, described in Subsection 9.4, applied. The fitted yield is 582 ± 32, while the expected yield is 616 ± 42. The branching fraction is calculated using

$$\mathcal{B}(B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0) = \frac{N_s(1 + f^{+-}/f^{00})}{2\ \varepsilon\ N_{B\overline{B}}}, \tag{10.1}$$

where N is the signal yield obtained from the fits, ϵ is the signal efficiency (see Table 9.2) and $N_{B\overline{B}}$ is the number of produced $B\overline{B}$ pairs, which is provided by dedicated studies and found to be $(198.0 \pm 3.0) \times 10^6$. The term f^{+-}/f^{00} is the ratio of the branching fractions for the decay of $\Upsilon(4S)$ to B^+B^- and $B^0\bar{B}^0$. The ratio f^{+-}/f^{00} is determined to be $1.065 \pm 0.012 \pm 0.019 \pm 0.047$ [83], where the first and second uncertainties are statistical and systematic, respectively, and the third uncertainty is due to the assumption of isospin symmetry in $B \to J/\psi(\to \ell\ell)K$, where $\ell = e$ or μ. All parameters except the signal, continuum, and $B\overline{B}$ yields, and the signal $\mathcal{A}_{CP}$ are fixed.

The branching fraction is determined to be

$$\mathcal{B}(B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0) = (3.66 \pm 0.21) \times 10^{-5}$$

and the direct CP asymmetry to be

$$\mathcal{A}_{CP}(B^0 \to \bar{D}^0\pi^0) = 0.01 \pm 0.16$$

. The uncertainties for the control mode measurements are statistical only. The most precise measurements of $\mathcal{B}(B^0 \to \bar{D}^0\pi^0) = (2.70 \pm 0.12) \times 10^{-4}$ come from Belle [84], while the world average value of $\mathcal{B}(\bar{D}^0 \to K^+\pi^-\pi^0) = (14.4 \pm 0.5)\%$ is taken from the Particle Data Group [5]. Therefore, the expected $\mathcal{B}$ measurement is $\mathcal{B}(B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0)_{\text{expected}} = (3.89 \pm 0.22) \times 10^{-5}$. Meanwhile, the SM prediction for the $\mathcal{A}_{CP}$ value is zero. The fitted branching and $\mathcal{A}_{CP}$ data agree with the expected values within the uncertainty. The results are summarised in Table 10.1. Figure 10.1 shows the signal-enhanced projections of the fits to data. The signal-enhanced region is defined as $5.275 < M_{bc} < 5.285\,\text{GeV}/c^2$, $-0.10 < \Delta E < 0.05\,\text{GeV}$, and $0 < T_c < 3$; for each plot, the selection on the plotted variable is not applied. On average, these signal-enhanced regions contain 47% of signal decays but only 11% of background.

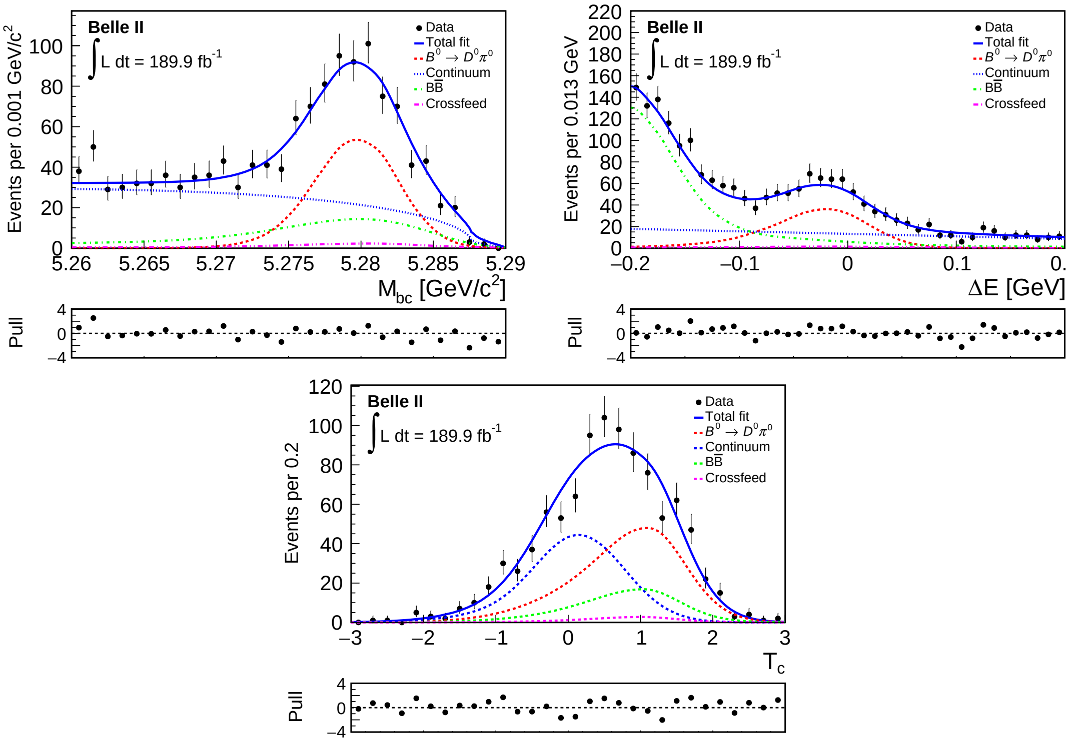


FIGURE 10.1: Distributions of M_{bc} (left), ΔE (middle), and T_c (right) for the $B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0$ candidates, for all seven r bins combined. The result of the fit to the data is shown as a solid blue curve. The fit components are shown as a red dashed curve (signal), blue dotted curve (continuum background), green dash-dotted curve ($B\bar{B}$ background), and magenta solid-dotted curve (crossfeed). The plots are signal-enhanced, which correspond to candidates with $5.275 < M_{bc} < 5.285\,\text{GeV}/c^2$, $-0.10 < \Delta E < 0.05\,\text{GeV}$, and $0 < T_c < 3$. When the respective variable is displayed, the selections on that variable are not applied. The difference between observed and fit value divided by the uncertainty from the fit (pulls) is shown below each distribution.

Decay Mode	**Expected BF**	**Fitted BF**	**Expected $\mathcal{A}_{CP}$**	**Fitted $\mathcal{A}_{CP}$**
$B^0 \to D^0\pi^0$	$3.75 \pm 0.24 \times 10^{-5}$	$3.66 \pm 0.21 \times 10^{-5}$	0	0.01 ± 0.16

TABLE 10.1: The expected and fitted Branching Fraction (BF) and $\mathcal{A}_{CP}$ for the $B^0 \to D^0(\to K^-\pi^+\pi^0)\pi^0$ control mode.

10.2 Signal mode results

With all corrections applied to MC and data, and the validation of the signal procedure using the control mode, the signal mode can now be 'unblinded'. This 'unblinding' means that the signal mode analysis has effectively addressed or taken into account any potential issues identified during the validation with the control mode. As such, a refined signal mode analysis can now be applied to the full 189.9 fb^{-1} data set.

With a signal efficiency of 35.5%, the expected number of signal events is 116 ± 16. The fit procedure, as described in Chapter 8, is applied to 3177 selected $B^0 \to \pi^0\pi^0$ events. Only the signal yield, $\mathcal{A}_{CP}$ and the continuum yield are allowed to float, while the $B\overline{B}$ yield is fixed to 87. A signal yield of 93 ± 18 is obtained. The branching fraction is then calculated using

$$\mathcal{B}(B^0 \to \pi^0\pi^0) = \frac{N_s(1 + f^{+-}/f^{00})}{2\ \varepsilon\ N_{B\bar{B}}\ \mathcal{B}(\pi^0 \to \gamma\gamma)^2}, \tag{10.2}$$

where N_s is the signal yield, ε is the signal reconstruction and selection efficiency, $N_{B\bar{B}}$ is the number of $B\bar{B}$ pairs produced, $\mathcal{B}(\pi^0 \to \gamma\gamma)$ is the $\pi^0 \to \gamma\gamma$ branching fraction, and $f^{+-}/f^{00} = 1.065 \pm 0.012 \pm 0.019 \pm 0.047$ is the ratio of the branching fractions for the decay of $\Upsilon(4S)$ to B^+B^- and $B^0\bar{B}^0$. Inserting the values $N_s = 93\pm18$, $\varepsilon = (35.5\pm4.7)\%$, $N_{B\bar{B}} = (198.0 \pm 3.0) \times 10^6$, and $\mathcal{B}(\pi^0 \to \gamma\gamma) = (98.823 \pm 0.034)\%$ [5], the branching fraction is calculated to be [1]

$$\mathcal{B}(B^0 \to \pi^0\pi^0) = (1.38 \pm 0.27 \pm 0.22) \times 10^{-6}$$

where the first and second uncertainties are statistical and systematic, respectively. The systematic uncertainties will be discussed in Chapter 11. The uncertainty in ε is due to the systematic uncertainty associated with π^0 reconstruction and continuum classifier efficiency. The measured $\mathcal{B}$ agrees with the previous average value [5] value, $1.59 \pm 0.26 \times 10^{-6}$, within their uncertainties. These results are more closely aligned with those of Belle than with those of BaBar, as can be seen in Table 10.2.

The direct CP violation parameter is measured to be

$$\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0) = +0.14 \pm 0.46 \pm 0.07$$

where the first and second uncertainties are statistical and systematic, respectively. Similarly, the measured $\mathcal{A}_{CP}$ agrees with the previous average value, 0.33±0.22, within their uncertainties. Like the branching fraction measurement, the results are more closely aligned with those of Belle than with those of BaBar. The Belle, the BaBar, the previous average, and the new average when this result is included are shown in Figure 10.9.

	$\mathcal{B}(B^0 \to \pi^0\pi^0)$	$\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0)$
Belle	$1.31 \pm 0.19 \pm 0.19$	$0.14 \pm 0.36 \pm 0.10$
BaBar	$1.83 \pm 0.21 \pm 0.13$	$0.43 \pm 0.26 \pm 0.05$
PDG value	1.59 ± 0.26	0.33 ± 0.22
Belle II value	$1.36 \pm 0.26 \pm 0.19$	$0.14 \pm 0.46 \pm 0.07$

Table 10.2: The branching fraction ($\times 10^{-6}$) and A_{CP} of $B^0 \to \pi^0\pi^0$ as measured at Belle and BaBar, their averaged value [5] and our result.

The distributions for ΔE, M_{bc}, and T_c, overlaid with fit projections, are displayed in Figure 10.2 for the 189.9 fb^{-1} dataset. Figure 10.4 shows similar plots, but only includes the top 3 r bins. Figure 10.3 presents the signal-enhanced plots, split between positive and negative q tags. The signal-enhanced region is defined by $5.275 < M_{bc} < 5.285\,\mathrm{GeV}/c^2$, $-0.1 < \Delta E < 0.05\,\mathrm{GeV}$, and $0 < T_c < 3$, depending on the variable plotted. These cuts retain 50% of the signal while eliminating 96% of the continuum background.

10.2.1 Significance

The signal significance is calculated using the test statistic $2(\mathcal{L}_m - \mathcal{L}_0)$, where $\mathcal{L}_m$ denotes the log-likelihood for the observed yield and $\mathcal{L}_0$ is computed by setting the signal yield to zero. Ignoring the systematic uncertainty, the test-statistic is determined to be 56.3 with two degrees of freedom. Assuming that Wilks' theorem is valid, a significance of 7.1 standard deviations is calculated. The high significance is attributed to the fact that the fitting is also performed in bins of $q \cdot r$, which effectively discriminates between signal and background. To illustrate this, in Figure 10.5 only events in the three highest

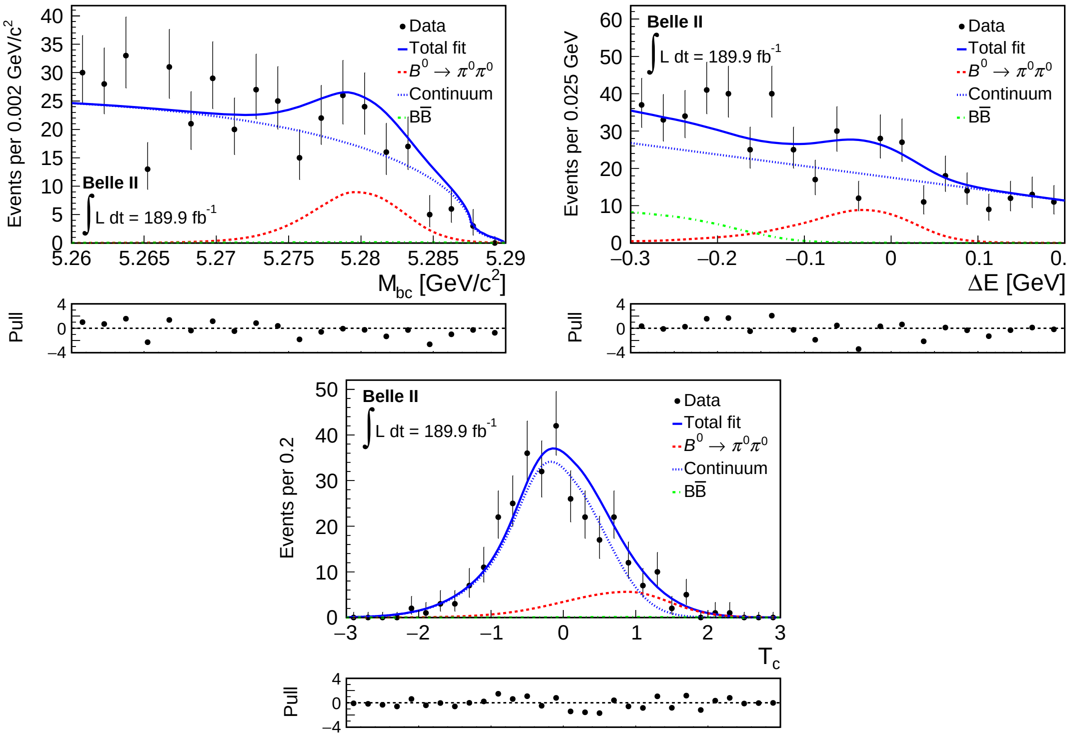


FIGURE 10.2: Distributions of M_{bc} (left), ΔE (middle), and T_c (right) for the $B^0 \to \pi^0\pi^0$ candidates reconstructed in 189.9 fb^{-1} of Belle II data. The result of a fit to the sample is shown as a solid blue line. The fit components are shown as blue dashed line (signal), red dashed line (continuum background), and green dashed line ($B\overline{B}$ background). The plots are shown as signal-enhanced projections. The normalised residuals are shown below each distribution.

r bins are shown along with a looser definition for the signal-enhanced region. The loose signal-enhanced region is defined as $5.275 < M_{bc} < 5.285$, $-0.2 < \Delta E < 0.1$, and $-0.5 < T_c < 3$. It is clear that within the highest r bins, the number of continuum events is greatly reduced.

To account for the systematic uncertainty, both the statistical and systematic components are taken into consideration.

1. **Statistical:** The negative log-likelihood (NLL) of the unbinned maximum likelihood fit as a function of the signal yield is calculated. This is achieved by refitting the data with the signal yield fixed from 0 to 180, as shown in Figure 10.7.

2. **Systematic:** A Gaussian distribution is used, the width of which is determined by the additive systematic uncertainties. Additive systematic uncertainties refer to uncertainties that are added or subtracted directly from the measured value. In contrast, multiplicative systematic uncertainties are proportional to the measured value. The additive systematic uncertainties are summarised in Table 10.3.

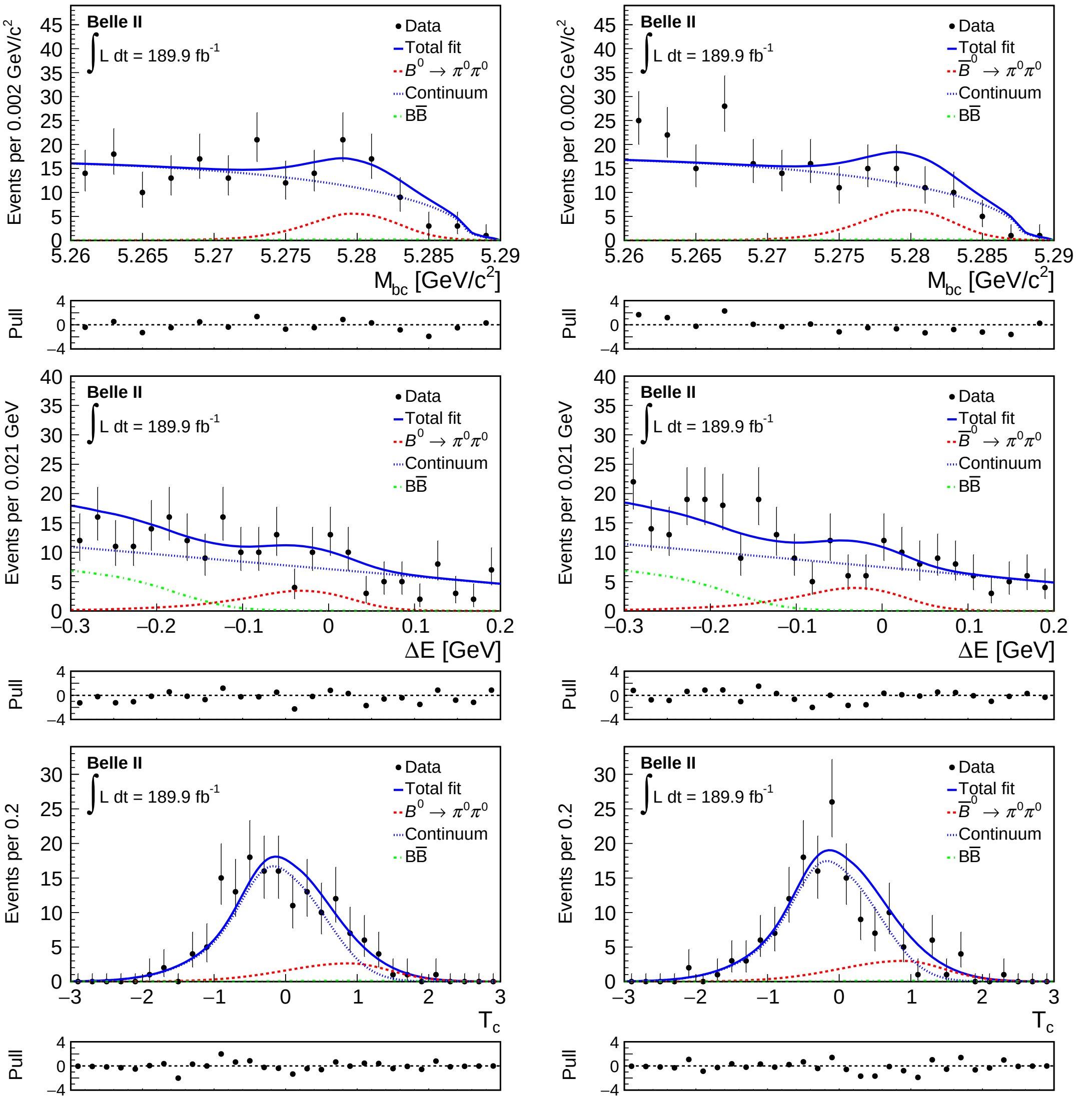


FIGURE 10.3: Distributions of M_{bc} (left), ΔE (middle), and T_c (right) for the $B^0 \to \pi^0\pi^0$ candidates reconstructed in 189.9 fb^{-1} of Belle II data for events with positive (top) and negative (bottom) q tags. The result of a fit to the sample is shown as a solid blue line. The fit components are shown as blue dashed line (signal), red dashed line (continuum background), and green dashed line ($B\overline{B}$ background). The plots are shown as signal-enhanced projections. The normalised residuals are shown below each distribution.

The width is calculated as the product of the best-fit yield (93) and the additive systematic uncertainty (8.6%), resulting in a value of 8.

The statistical component is then converted into a likelihood and convolved with the systematic component. This results in a significantly smaller negative log-likelihood when the yield is zero, a change illustrated in Figure 10.6. The test statistic $2(\log \mathcal{L}_m - \log \mathcal{L}_0) = 32.0$ and a total significance of 5.2 standard deviations is calculated.

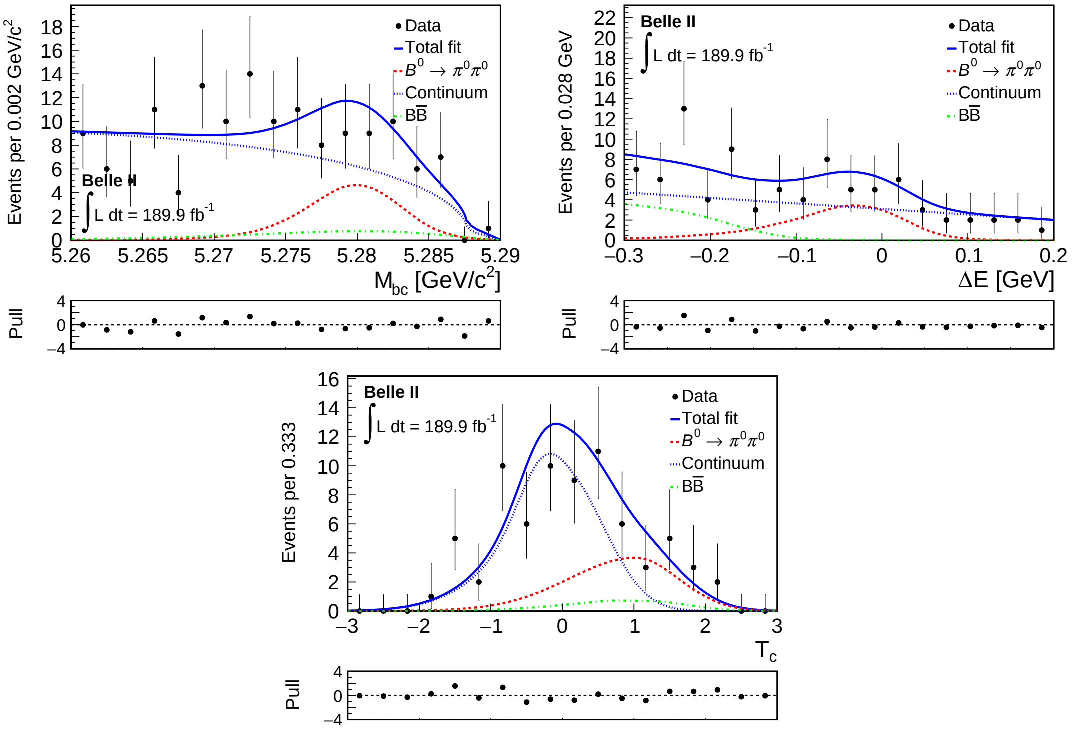


FIGURE 10.4: Distributions of M_{bc} (left), ΔE (middle), and T_c (right) for the $B^0 \to \pi^0\pi^0$ candidates reconstructed in 189.9 fb^{-1} of Belle II data for events in the highest 3 bins of r. The result of a fit to the sample is shown as a solid blue line. The fit components are shown as blue dashed line (signal), red dashed line (continuum background), and green dashed line ($B\overline{B}$ background). The normalised residuals are shown below each distribution.

Source	$\mathcal{B}$ **(%)**
Continuum parametrization	7.4
Fixed $B\overline{B}$ background yield	2.3
Signal $q \cdot r$ bin fractions	2.2
Knowledge of the photon-energy scale	2.0
Assumption of independence of ΔE from $q \cdot r$	1.8
Choice of signal model	1.3
Branching fraction fit bias	1.0
Best candidate selection	0.2
Total	8.6

TABLE 10.3: Additive branching fraction systematic uncertainties for the signal mode. The total is calculated by adding all systematic uncertainties in quadrature.

For an additional verification of the estimated significance, the NLL is plotted as a function of the signal yield to examine its symmetry. If the curve is symmetric, it indicates that the statistical uncertainties have been correctly estimated. As shown in Figure 10.7, the curve appears to be approximately symmetric around the best-fit value for the signal yield, both for one and two sigma.

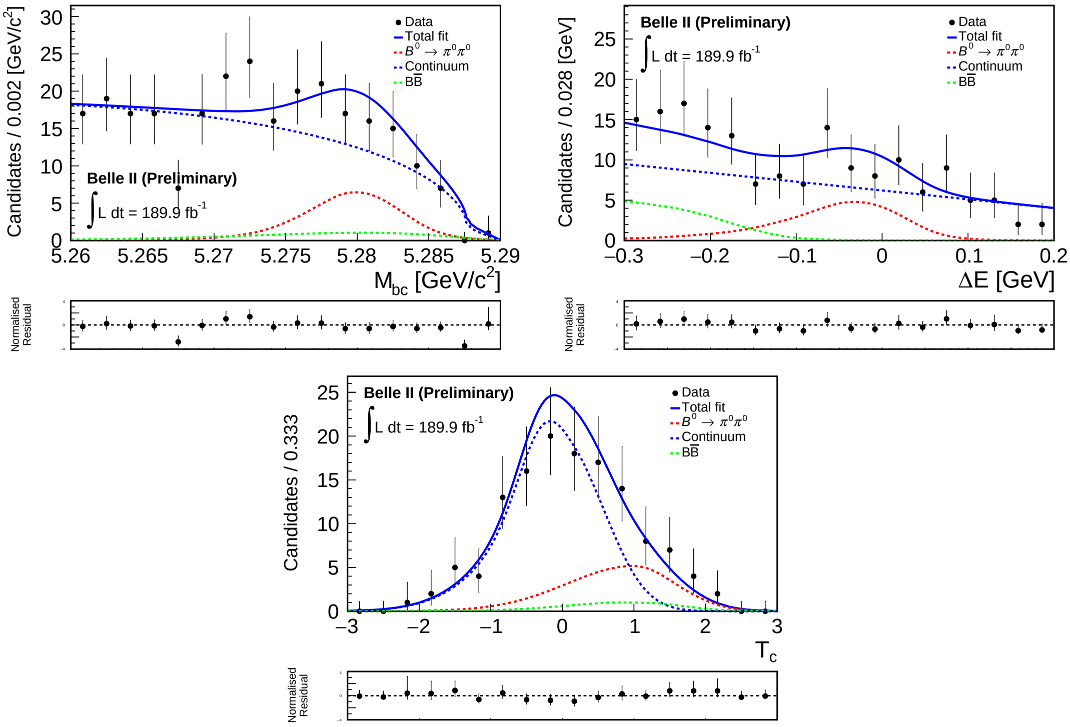


FIGURE 10.5: Distributions of M_{bc} (left), ΔE (middle), and T_c (right) for the $B^0 \to \pi^0\pi^0$ candidates reconstructed in 189.9 fb^{-1} of Belle II data for events in the best 3 $q \cdot r$ bins. This demonstrates that within the highest r bins, the number of continuum events is greatly reduced. The result of a fit to the sample is shown as a solid blue line. The fit components are shown as blue dashed line (signal), red dashed line (continuum background), and green dashed line ($B\overline{B}$ background). The normalised residuals are shown below each distribution.

Furthermore, 100,000 ToyMC experiments under the null hypothesis (i.e. assuming that there is no signal, only background) are performed. The test-statistic $-2(\mathcal{L}_m - \mathcal{L}_0)$ for these simulations is calculated and shown in Figure 10.8. In data, the test-statistic is 56.3, while the maximum value reached in the 100,000 background-only ToyMC experiments is 10.5. When the observed value of the test statistic is compared with the distribution obtained from the ToyMC experiments, it can be seen that the probability of observing a test statistic as extreme as the observed one under the null hypothesis is extremely unlikely. This suggests that the true significance is indeed very high.

10.2.2 Determination of ϕ_2

The measured $\mathcal{B}$ and $\mathcal{A}_{CP}$ for $B^0 \to \pi^0\pi^0$ in this analysis are averaged with the values obtained by the Belle and BaBar experiments. The averaging procedure utilises the same method employed by the Particle Data Group [5]:

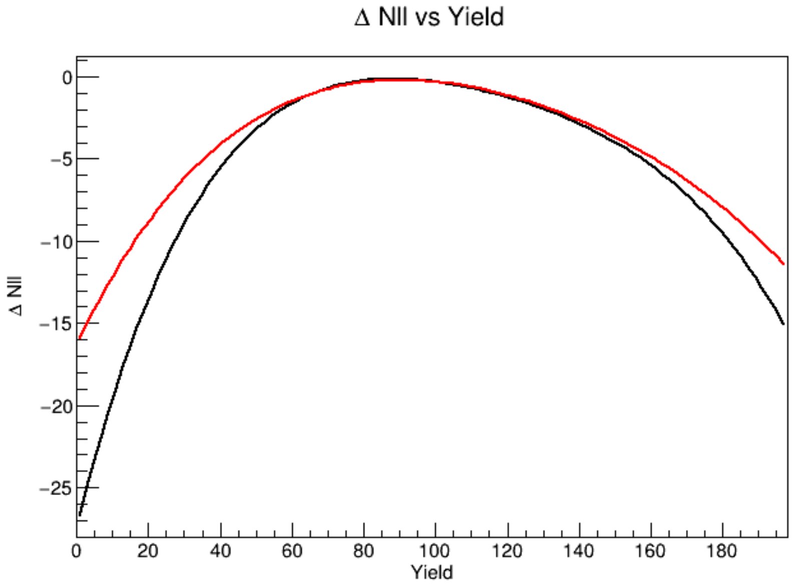


FIGURE 10.6: The negative log-likelihood (NLL) as a function of the signal yield without systematic uncertainties included (black), and with systematic uncertainties included (red). The NLL is rescaled so that the minimum NLL equals zero.

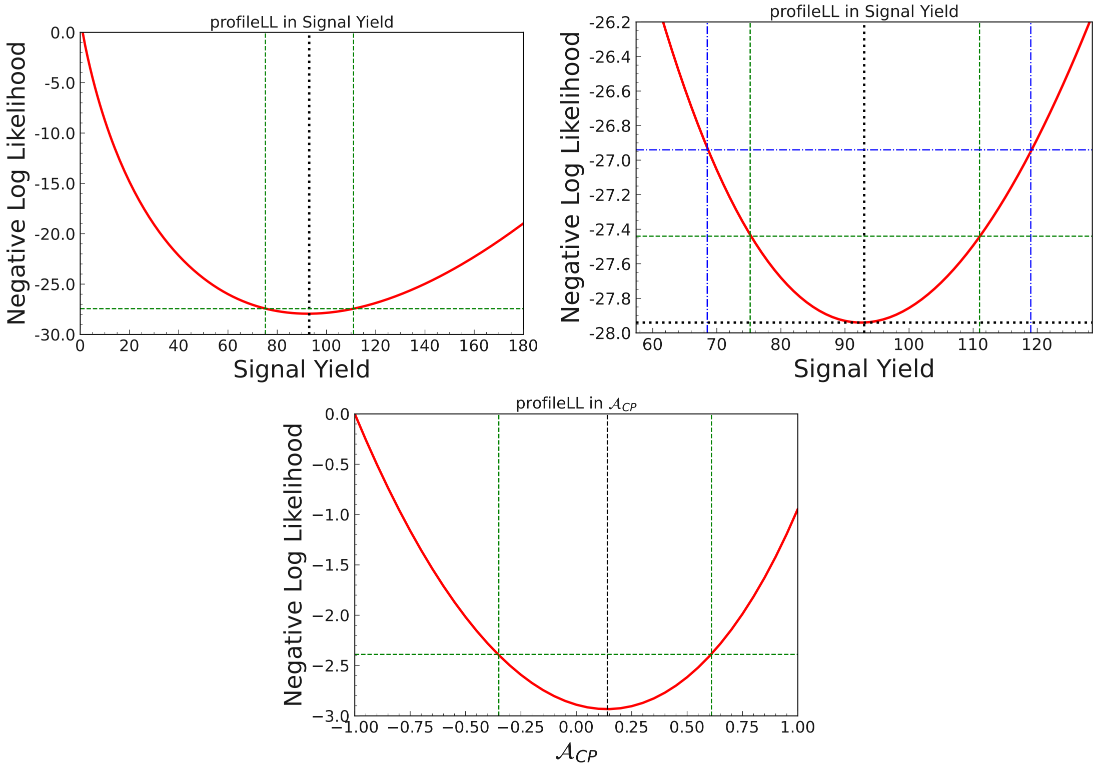


FIGURE 10.7: The negative log-likelihood (NLL) as a function of the signal yield (left) and $\mathcal{A}_{CP}$ (bottom). The NLL is rescaled so that the maximum NLL equals zero. The NLL vs signal yield plot zoomed in around the minimum is also shown (right). The green and blue dotted lines corresponds to one and two sigma respectively.

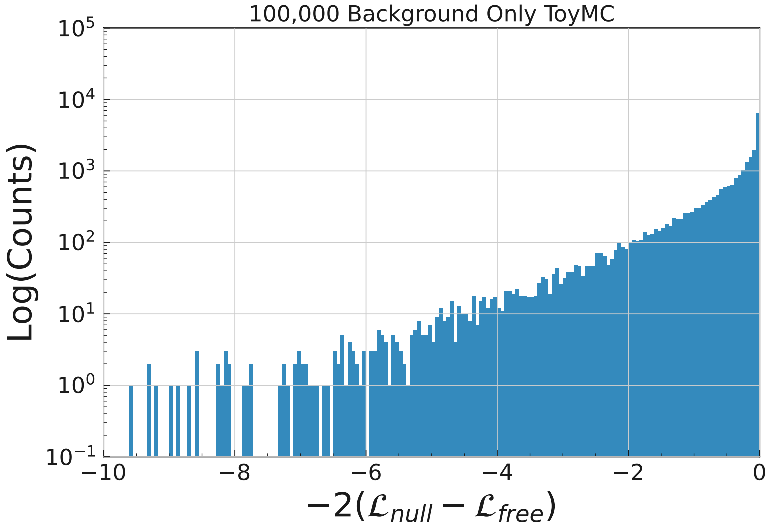


FIGURE 10.8: Histogram of $-2(\mathcal{L}_m - \mathcal{L}_0)$ for 100,000 background only ToyMC.

1. Calculate a weighted average and error as

$$x \pm \delta x = \frac{\sum_i w_i x_i}{\sum_i w_i} \pm \left(\sum_i w_i\right)^{-1/2}, \tag{10.3}$$

   where $w_i = 1/(\delta x_i)^2$. Here x_i and δx_i are the value and error reported by the ith experiment, and the sums run over the N experiments. Each result is weighted by the inverse of its squared uncertainty, which gives more weight to more precise measurements.

2. Calculate $\chi^2/(N-1) = \sum w_i(x - x_i)^2/(N-1)$. $N-1$ is the expectation value of χ^2 if the measurements are from a Gaussian distribution.

3. If the value $\chi^2/(N-1)$ is less than or equal to 1, the result is accepted. If it is much greater than 1, the value is typically rejected. But if the value is only slightly greater than 1, the error is increased by a scaling factor $S = \left[\chi^2/(N-1)\right]^{1/2}$.

This procedure corrects for measurements with underestimated uncertainties. For the signal yield, the χ^2 value is 1.17 and hence the naive weighted error is scaled by 1.08. The χ^2 value of the $\mathcal{A}_{CP}$ is 0.26, and hence no scaling factor is applied. The updated world-average result is

$$\mathcal{B}(B^0 \to \pi^0\pi^0)_{\text{average}} = (1.54 \pm 0.17) \times 10^{-6}$$
$$\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0)_{\text{average}} = 0.30 \pm 0.19$$

This updated average, along with the world-averaged measurements of $\mathcal{B}$ and time-dependent CP violation for $B^0 \to \pi^+\pi^-$ and $B^+ \to \pi^+\pi^0$ is used to constrain ϕ_2 using the isospin analysis described in subsection 2.6.3. The result is shown in Fig. 10.10. The world-averaged values exclude $13.5^\circ < \phi_2 < 77^\circ$ at the 68% confidence level and exclude $18.5^\circ < \phi_2 < 72^\circ$ at the 95% confidence level. These new results exclude $13^\circ < \phi_2 < 77.5^\circ$ at the 68% confidence level and exclude $15.0^\circ < \phi_2 < 75.5^\circ$ at the 95% confidence level. This represents a relative increase in precision of 2.0%. The main limiting factor to improving precision remains the uncertainty of $\mathcal{A}_{CP}$.

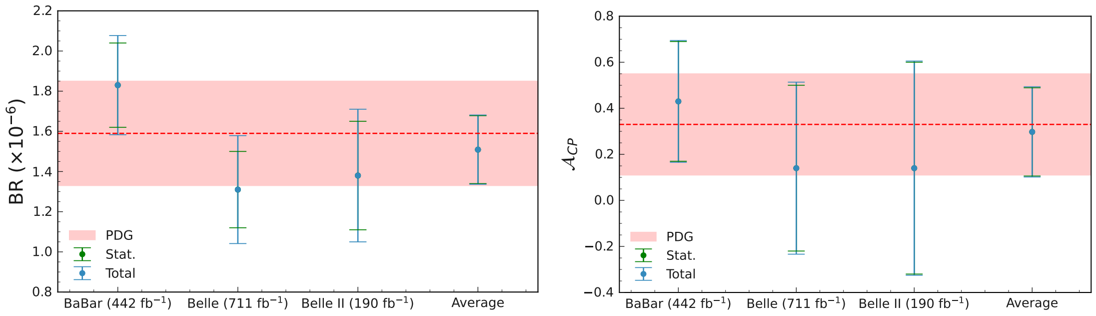


FIGURE 10.9: The branching fraction (left) and A_{CP} (right) of the $B^0 \to \pi^0\pi^0$ decay as measured at Belle and BaBar, their averaged value (red band), the new result, and the average of all three experiments.

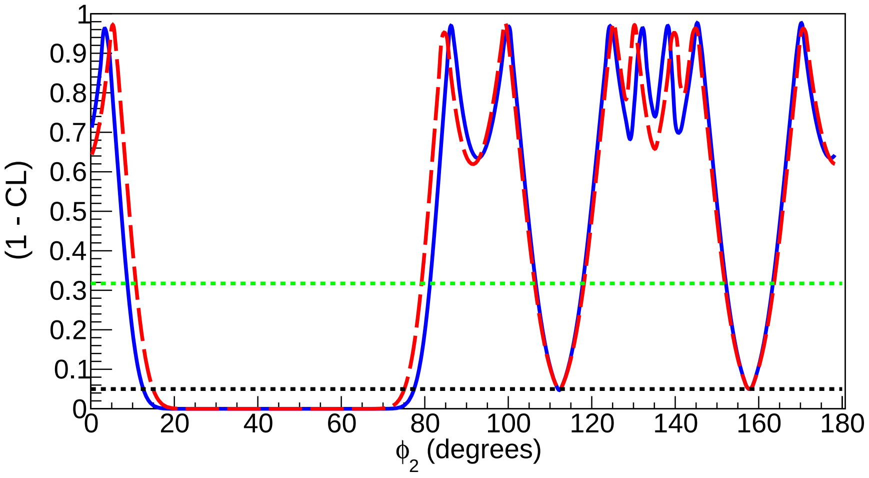


FIGURE 10.10: Scan of the confidence level for ϕ_2. The dashed red curve shows the previous constraint from world-averaged values [5], the solid blue curve includes our new results. The updated results for $B^0 \to \pi^0\pi^0$ exclude $13^\circ < \phi_2 < 77.5^\circ$ at the 68% confidence level (green dashed line) and $15.0^\circ < \phi_2 < 75.5^\circ$ at 95% confidence level (black dashed line).

Chapter 11

Systematic uncertainties

Systematic uncertainties are errors that are not related to the size of the data sample but arise from the measurement procedure, experimental setup, and theoretical models used in the analysis. Unlike statistical uncertainties, systematic uncertainties often cannot be reduced by collecting more data. Instead, they must be estimated and taken into account in the analysis.

In this analysis, which uses a 189.9 fb^{-1} dataset, the dominant uncertainty was statistical. However, as the luminosity of SuperKEKB continues to increase and more experimental data are collected, the systematic uncertainties will soon surpass the statistical ones. Therefore, it is important at this stage to carefully study these uncertainties and take as many as possible into account.

The largest sources of systematic errors arise from; differences between MC simulations and actual data, uncertainties in fixed parameters, the choice of particular analysis techniques, and the operational performance of the detector. Some of these uncertainties are common to many Belle II analyses and have been extensively studied, with their values provided for use. Other systematic uncertainties are specific to this analysis and need to be examined in more detail. Table 11.1 summarises the systematic uncertainty, with more information provided in the following subsections.

Source	$\boldsymbol{\mathcal{B}}$ **(%)**	$\boldsymbol{\mathcal{A}_{CP}}$
π^0 reconstruction efficiency	11.6	n/a
Continuum parametrisation	7.4	0.02
Continuum classifier efficiency	6.5	n/a
$1 + f^{+-}/f^{00}$	2.5	n/a
Fixed $B\overline{B}$ background yield	2.3	0.01
Fixed signal r bin fractions	2.2	0.01
Knowledge of the photon-energy scale	2.0	n/a
Assumption of independence of ΔE from r	1.8	< 0.01
Number of $B\overline{B}$ meson pairs	1.5	n/a
Choice of $(M_{bc}, \Delta E)$ signal model	1.3	0.02
Fixed continuum r bin fraction	1.1	< 0.01
Branching fraction fit bias	1.0	n/a
Best candidate selection	0.2	< 0.01
Mistagging parameters	n/a	0.05
Potential non-zero $B\overline{B}$ background $\mathcal{A}_{CP}$	n/a	0.03
$\mathcal{A}_{CP}$ fit bias	n/a	0.02
Continuum $q \cdot r$ asymmetry	n/a	0.01
Total	16.2	0.07

Table 11.1: Summary of systematic uncertainties. Sources are assumed to be independent to be independent, and the total is calculated by adding all systematic uncertainties in quadrature.

11.0.1 π^0 reconstruction efficiency

The systematic uncertainty associated with possible data simulation discrepancies is studied using the $D^{*+} \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^+$ and $D^{*+} \to \bar{D}^0(\to K^+\pi^-)\pi^+$ decays. In the reconstruction of both decays, the selection of charged particles is identical to those used in the control mode, and π^0 uses the same selection as in the signal mode. K^+, π^-, and π^0 candidates were combined to form D^0 meson candidates. The mass of $D^0 \to K^+\pi^-\pi^0$ candidates (M) is required to be between 1.75 GeV/c^2 and 2.00 GeV/c^2. Meanwhile, the mass of $D^0 \to K^+\pi^-$ candidates is required to be between 1.80 GeV/c^2 and 1.92 GeV/c^2

These D^0 candidates were then combined with a charged pion to form D^{*-} candidates. The mass difference between the D^{*+} and D^0 mesons is required to have $0.1448 < M[K^+\pi^-\pi^0\pi^-] - M[K^+\pi^-\pi^0] < 0.1461\,\text{GeV}/c^2$ and $0.1447 < M[K^+\pi^-\pi^0] - M[K^+\pi^-] < 0.1467\,\text{GeV}/c^2$. The momentum of the D^{*-} candidates in the c.m. frame was required to be larger than 2.5 GeV/c for $K^+\pi^-\pi^0$ candidates and larger than 2.17 GeV/c for $K^+\pi^-$ candidates.

These modes are chosen because they contain high-momentum π^0 mesons that have different π^0 reconstruction efficiencies compared to low-momentum π^0 mesons. The π^0 reconstruction efficiency was determined using known branching fractions and the observed yields of the $D^{*-} \to D^0(\to K^+\pi^-\pi^0)\pi^+$ and $D^{*-} \to D^0(\to K^+\pi^-)\pi^-$ decays. The signal yields of the two control channels were compared to determine the π^0 reconstruction efficiency. It was assumed that their signal yields are expressed as

$$N(K^-\pi^+\pi^0) = N_{D^0}B(D^0 \to K^-\pi^+\pi^0)B(\pi^0 \to \gamma\gamma)\epsilon_{K\pi}\epsilon_{D^*}\epsilon_{D^0}\epsilon_{\pi^0}, \tag{11.1}$$

and

$$N(K^-\pi^+) = N_{D^0}B(D^0 \to K^-\pi^+)\epsilon_{K\pi}\epsilon_{D^*}\epsilon_{D^0}, \tag{11.2}$$

where $N(K^-\pi^+\pi^0)$ and $N(K^-\pi^+)$ are the signal yields of the $D^+ \to D^0(\to K^-\pi^+\pi^0)\pi^+$ and $D^+ \to D^0(\to K^-\pi^+)\pi^+$ decays respectively; N_{D^0} is the number of D^0 produced; $\epsilon_{K\pi}$ is the reconstruction efficiency of the charged kaon-pion pair; ϵ_D and ϵ_{D^0} are the efficiency of the D and D^0 candidate selections, respectively; and ϵ_{π^0} is the reconstruction efficiency of the neutral pion. It was assumed that the efficiency to reconstruct the $K^+\pi^-$ system from the D meson was the same across both channels. The ratio of the two yields, combined with the known branching fractions, allows for the determination of ϵ_{π^0}:

$$\epsilon^{\pi^0} = \frac{N(K^-\pi^+\pi^0)}{N(K^-\pi^+)} \cdot \frac{\mathcal{B}(\bar{D}^0 \to K^+\pi^-)}{\mathcal{B}(\bar{D}^0 \to K^+\pi^-\pi^0) \cdot \mathcal{B}(\pi^0 \to \gamma\gamma)} \tag{11.3}$$

The signal yields are determined from fits to the D^0 mass distributions of each mode. The uncertainty of the π^0 reconstruction efficiency, which also includes the uncertainty associated with the `photonMVA`, is determined to be 3.4% for each π^0.

The π^0 reconstruction efficiency as a function of momentum is also measured using $\tau^- \to 3\pi\pi^0\nu$ and $\tau^- \to 3\pi\nu$ decays. A difference of 4.7% in efficiency is observed between the measurement based on D decays and the measurement based on τ leptons. This difference increases the systematic uncertainty for a total of $\sqrt{3.4^2 + 4.7^2} = 5.8\%$ per pion. The total systematic uncertainty associated with the π^0 reconstruction efficiency is then 11.6%, as there are two pions and their errors are fully correlated. This systematic study is performed centrally and made available for all Belle II analyses.

11.0.2 Continuum parametrisation

The systematic uncertainty of the continuum modelling involves estimating the systematic uncertainty associated with each of the eight data-driven continuum PDF parameters. The contribution of each parameter is determined by refitting on MC with the parameter used in the continuum PDF individually fluctuated by either their positive or negative errors. Subsequently, all other parameters were shifted in accordance with their correlation with the fluctuating parameter. The shift in all other parameters was calculated using the formula

$$y_{\text{shifted}} = \frac{\Delta x}{x} \times \sigma_{xy} \times y_{\text{old}}$$

where x is the parameter that is shifted by its uncertainty, y represents the remaining parameters requiring a shift, and σ_{xy} denotes the correlation between x and y. This procedure results in 16 alternative sets of parameters and their corresponding uncertainties, with each of the eight continuum parameters producing two sets, one for positive fluctuations and one for negative fluctuations.

A total of 1000 ToyMC studies were performed for the set of parameters related to the positive and negative fluctuation for each of the eight continuum parameters. Only the largest deviation from the nominal value is used, instead of both positive and negative fluctuations. This is true for all systematic uncertainties that involve fluctuating a parameter by its uncertainty and repeating the fit. The difference between the alternative and the nominal yield was averaged and is shown in Table 11.2.

The systematic uncertainty is the sum in quadrature of the change in signal yield and $\mathcal{A}_{CP}$ for each parameter. The yield and $\mathcal{A}_{CP}$ uncertainties shown in Table 11.2 were added in quadrature, leading to an uncertainty in the branching fraction and $\mathcal{A}_{CP}$ of 7.4% and 0.02, respectively.

11.0.3 Continuum classifier efficiency

The systematic uncertainty of continuum suppression arises from the possible data-MC discrepancy after applying the continuum classifier selections. This discrepancy is expected to be small, since the continuum suppression is trained using experimental sideband data. The systematic uncertainty is investigated by fitting two separate $B^0 \rightarrow$

Parameter	**Yield Uncertainty (%)**	**$\mathcal{A}_{CP}$ Uncertainty**
Tc Fraction	4.38	0.013
Tc Mean Gaussian 1	5.12	0.017
Tc Width Gaussian	2.50	0.008
Tc Mean Gaussian 2	1.38	0.004
Tc Left Width Bi-Gaussian	0.90	0.002
Tc Right Width Bi-Gaussian	0.93	0.002
$M_{\mathbf{bc}}$ ARGUS shape	0.36	0.001
ΔE Chebyshev	0.31	0.001
Total	7.4	0.02

TABLE 11.2: The average difference in yield and $\mathcal{A}_{CP}$ between the nominal sideband parameters and parameters that have been fluctuated according to their uncertainty, with all other parameters are also altered based on their correlation to the fluctuated parameter. The values are added in quadrature and taken as the systematic uncertainty.

$D^0(\to K^-\pi^+\pi^0)\pi^0$ control mode samples; one that passes the CSMVA selection (`CSMVA > 0.74`) and one that fails (`CSMVA < 0.74`) with the same classifier that was used in the $B^0 \to \pi^0\pi^0$ analysis. The fits are shown in Figure 11.1.

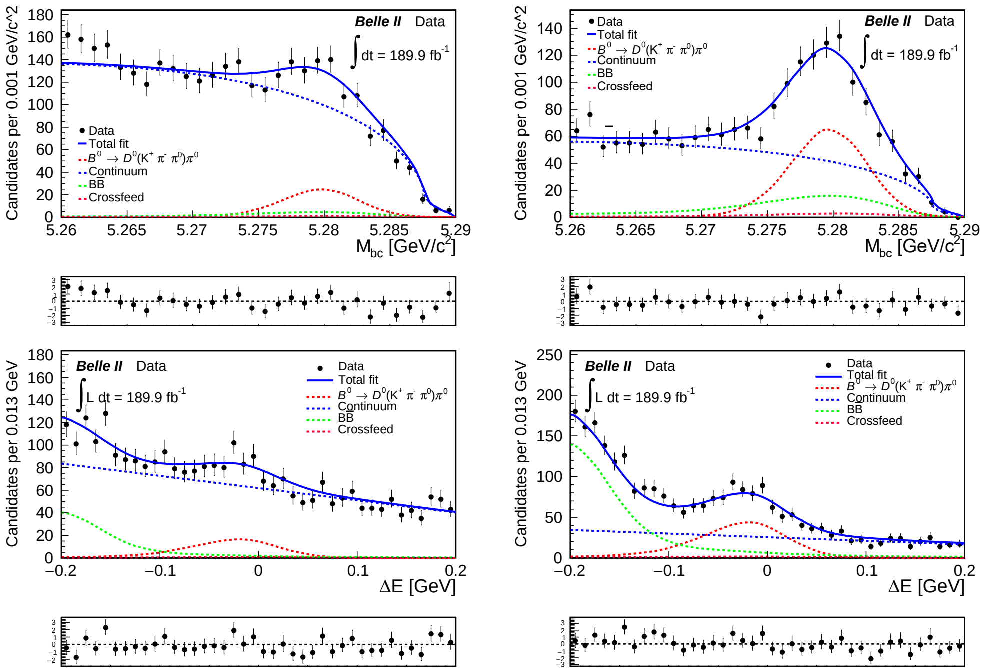


FIGURE 11.1: Control mode fits for M_{bc} (top) and ΔE (bottom) for a dataset that failed (left) and passed (right) the 0.74 continuum suppression selection.

Following this, the efficiency in data and MC, $\epsilon_{\text{MC/Data}}$, is determined. This efficiency is defined as the ratio of signal events that pass the selection to the total signal events, as expressed in the following equation:

$$\epsilon_{\text{MC/Data}} = \frac{\text{Signal events that pass the selection}}{\text{Total signal events (pass and fail the selection)}} \tag{11.4}$$

Due to the observed discrepancy in M_{bc} between data and run-independent MC, as detailed in discussed subsection 8.3.2, run-dependent MC is used. The efficiency in data, ϵ_{Data}, was found to be 0.6928 ± 0.0386, while the efficiency in MC was determined to be $\epsilon_{\mathrm{MC}} = 0.7201 \pm 0.0207$. These uncertainties are solely statistical. The results are summarised in Table 11.3.

Efficiency	$\boldsymbol{B^0 \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^0}$
ϵ_{Data}	0.6928 ± 0.0386
ϵ_{MC}	0.7201 ± 0.0207

TABLE 11.3: Signal efficiency for data and run MC in a 189.9 and 1000 fb^{-1} datasets respectively.

The ratio between the efficiency in data and MC, denoted as $\frac{\epsilon_{\mathrm{Data}}}{\epsilon_{\mathrm{MC}}}$, is determined to be 1.04 ± 0.065. Given the consistency of the efficiency ratio in data and MC, the efficiency ratio is treated as 1 and the statistical uncertainty of the efficiency ratio is assigned as the systematic uncertainty. Consequently, the systematic uncertainty of the continuum classifier for the branching fraction is 6.5%.

11.0.4 Ratio between charged and neutral B mesons $(1 + f^{+-}/f^{00})$

The ratio between the number of charged and neutral B mesons produced in $\Upsilon(4S)$ decays, denoted as f^{+-}/f^{00}, is expected to be very close to one due to the charged and neutral B mesons having very similar masses and the conservation of isospin. The $\Upsilon(4S)$ has an isospin of zero, and hence its decay products must also have a total isospin of zero. This can be achieved in two ways: either by producing a B^+ and a B^-, or by producing a B^0 and a $\bar{B}^0$. Therefore, under the assumption of perfect isospin conservation and equal neutral and charged B meson masses, the $\Upsilon(4S)$ should decays into charged B mesons exactly as often as it decays into neutral B mesons.

However, a recent measurement by the Belle experiment [83] using $B^+ \to J/\psi(l^+l^-)K^+$ and $B^0 \to J/\psi(l^+l^-)K^0$ samples, where $J/\psi(l^+l^-)$ stands for $J/\psi \to l^+l^-$ (l represents e or μ), with 711×10^6 fb^{-1} of data, has observed a slight discrepancy in this assumption. These decay modes are used because the possible contribution of isospin symmetry breaking from rescattering in $B \to J/\psi(ll)K$ is expected to be small in the SM [85], of the order of λ^3.

The yields of $B^+ \to J/\psi(l^+l^-)K^+$ and $B^0 \to J/\psi(l^+l^-)K^0$ are given as

$$N^+_{\text{sig}} = 2N_{B\bar{B}} f^{+-} \epsilon^+ B[B^+ \to J/\psi(l^+l^-)K^+],$$
$$N^0_{\text{sig}} = 2N_{B\bar{B}} f^{00} \epsilon^0 B[B^0 \to J/\psi(l^+l^-)K^0],$$

where N^+_{sig}, N^0_{sig}, ϵ^+, and ϵ^0 are the signal yields and reconstruction efficiencies of charged and neutral B mesons, respectively; $N_{B\bar{B}}$ is the number of $B\bar{B}$ pairs. Assuming isospin invariance in $B \to J/\psi(l^+l^-)K$, this leads to

$$f^{+-}/f^{00} = \frac{N^+_{\text{sig}} \epsilon^0 \tau^0}{N^0_{\text{sig}} \epsilon^+ \tau^+}$$

The Belle analysis determined f^{+-}/f^{00} to be $1.065 \pm 0.012 \pm 0.019 \pm 0.047$, where the first and second uncertainties are statistical and systematic, respectively, and the third uncertainty is due to the assumption of isospin symmetry in $B \to J/\psi(\to \ell\ell)K$, where $\ell = e$ or μ. This uncertainty arises from taking $\lambda^3 = 1.1\%$ as the uncertainty due to the isospin symmetry assumption in the decay amplitude, which leads to a 4.4% uncertainty. All uncertainties are added in quadrature, leading to a f^{+-}/f^{00} systematic uncertainty of 0.052. Therefore, the systematic uncertainty in the branching fraction due to $1 + f^{+-}/f^{00}$ is 2.5%.

11.0.5 Fixed $B\bar{B}$ background yield

The PDF shape parameters of the $B\bar{B}$ background is modelled using the expected yield of the two dominant contributions, the $B^+ \to \rho^+\pi^0$ and $B^0 \to K_s(\to \pi^0\pi^0)\pi^0$ decay modes. In addition, the number of $B\bar{B}$ background in the fits is fixed to the expectation from MC. The knowledge of the true number of $B\overline{B}$ events is limited by the uncertainties on the measured branching fraction of both decays. The estimated uncertainty of both decays, combined according to their relative contribution, is 12.8%. To account for the uncertainty in the branching fraction, 1000 ToyMCs are generated. Two fits are performed, one in which the $B\overline{B}$ yield is fixed to the MC expectation (87) and one in which the yield of the $B\bar{B}$ yield is fixed to values one standard deviation higher (98) and lower (76). Only the largest deviation from the nominal value is used, instead of both positive and negative fluctuations. The mean difference between the alternative and nominal yields is taken as the systematic uncertainty. The same is done for the

$\mathcal{A}_{CP}$ value. The systematic uncertainty related to the fixing the $B\bar{B}$ background yield for the branching fraction and $\mathcal{A}_{CP}$ is 2.3% and 0.01, respectively.

$B\bar{B}$ decay	Branching ratio	$\mathcal{A}_{CP}$	Efficiency	189.9 fb^{-1} estimation
$B^+ \to \rho^+\pi^0$	$1.09 \pm 0.14 \times 10^{-5}$	0.02 ± 0.11	4.60%	109 ± 14
$B^0 \to K_s(\to \pi^0\pi^0)\pi^0$	$3.04 \pm 0.15 \times 10^{-6}$	0.00 ± 0.13	1.71%	10.7 ± 0.6

TABLE 11.4: Estimation of number of events in 189.9 fb^{-1} and expected $\mathcal{A}_{CP}$ for $B^+ \to \rho^+\pi^0$ and $B^0 \to K_s(\to \pi^0\pi^0)\pi^0$ from efficiency as determined by `MC14ri_a`.

11.0.6 Fixed signal r bin fractions

The fraction of signal events in each r bin, ϵ_i ($i = 1, ..., 7$), is provided by the Belle II Flavour Tagger group [80]. Given that the final fit is performed simultaneously in 7 bins of r, accurate values of ϵ_i are crucial for extracting the branching fraction and $\mathcal{A}_{CP}$ of the signal mode. These values must be fixed for a stable fit.

The value of the r bin fractions is determined by reconstructing B mesons that decay into hadronic final states with branching fractions of 10^{-5} or greater, i.e. $B^0 \to D^{*-}h^+$ and $B^0 \to D^-h^+$ where $h = \pi, \rho$. The value and uncertainty of the r bin fractions is evaluated by fitting the ΔE distribution of these hadronic final states and are shown in Table 8.1. These uncertainties account for many different sources of errors, such as ΔE PDF parametrisation, possible flavour mixing of the background, and any bias in the fit model as determined in MC.

The systematic uncertainty associated with the fixed signal fractions for r bins is determined by refitting MC with the signal fractions fluctuated by their one-standard-deviation uncertainties. Since the sum of the r bin fractions must sum to one, the value of the final bin, $0.875 < q \cdot r < 1.0$, is defined as $1 - \sum \epsilon_i$. A total of 1000 ToyMC studies were performed for each of the seven r fraction values. The difference between the alternative and the nominal yield was averaged and is shown in Table 11.5. The systematic uncertainty is the sum in quadrature of the change in signal yield for each bin. The systematic uncertainty related to fixing the signal r bin fractions for the branching fraction is 2.2%.

Bin	**Yield Uncertainty(%)**
1	0.85
2	0.56
3	0.46
4	0.32
5	0.20
6	0.04
7	0.01
Total	2.2

Table 11.5: Contribution of each continuum r bin fraction parameters to the $\mathcal{B}$ systematic uncertainty.

11.0.7 Knowledge of the photon-energy scale

To account for any data-MC discrepancy in the ΔE width, the scaling factor that best fit the control mode was determined and applied to the signal mode. To evaluate the systematic uncertainty associated with the ΔE width corrections, a total of 1000 ToyMC datasets were generated. Each of these data sets was subjected to two separate ToyMC studies: one with the optimal scaling factor, as described in Section 9.4.2, applied, and the another without it. The difference between the nominal and alternative signal yield was then calculated. The standard deviation of the differences from the 1000 ToyMC experiments was used to estimate the systematic uncertainty. This led to a systematic uncertainty in the branching fraction of 1.8%.

To estimate the systematic uncertainty associated with the ΔE peak, a 1000 ToyMC studies was performed with and without the ΔE shift. The average was taken as the systematic uncertainty, resulting in a systematic uncertainty for the branching fraction of 0.46%. Hence, the overall systematic uncertainty related to the ΔE peak shift and width scaling factor for the branching fraction was 1.9%.

In addition, a 'Photon Energy Bias Correction' was provided and applied exclusively to experimental data. To account for any uncertainty, two additional corrections were prepared: one with the errors shifted upward, and the other with them shifted downward. This is shown in Figure 11.2. After unblinding, the data was refitted using both corrections and the difference between the nominal and alternative signal yield was determined. The systematic uncertainty related to the 'Photon Energy Bias Correction' for the branching fraction was 2.0%.

To avoid double counting the systematic uncertainty due to the ‘Photon Energy Bias Correction’, and the ΔE peak shift and width scaling factor, the larger of the two was taken as the systematic uncertainty. Therefore, the systematic uncertainty related to the knowledge of the photon-energy scale is 2.0%.

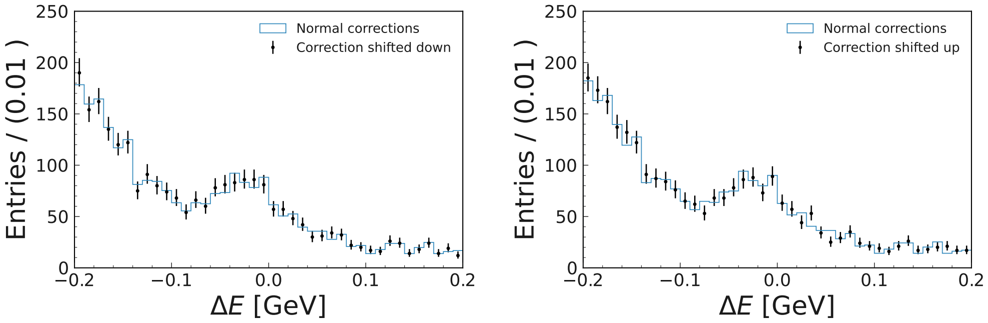


FIGURE 11.2: The distribution of ΔE of the control mode with ‘Photon Energy Bias Correction’ shifted down (left) and up (right).

11.0.8 Assumption of independence of ΔE for r

As shown in Figure 8.14, there are insufficient events in the sideband region for the higher $q \cdot r$ bins for a stable fit. As a result, the continuum ΔE PDF parameter was extracted from the entire sideband and applied across all $q \cdot r$ bins. To account for this, the ΔE slope for each $q \cdot r$ bin was estimated using MC. The deviation from the nominal slope and the fraction of continuum events in each r bin are shown in Table 11.6 and Figure 8.14.

To evaluate the systematic uncertainty, a total of 1000 ToyMC datasets were generated. Each of these datasets was subjected to two separate ToyMC studies: one using identical continuum ΔE PDF parameters and another using the varying ones. The difference between the alternative and the nominal yield was averaged. The same process was repeated for the $\mathcal{A}_{CP}$ values. The systematic uncertainty related to assuming that the continuum ΔE PDF parameters are indepedent of the r bins for the branching fraction and $\mathcal{A}_{CP}$ is 1.8% and 0.003, respectively. This systematic uncertainty is relatively small, because the largest deviation occurs in the higher bins of $q \cdot r$ where only a small fraction of the continuum is present.

$\|q\cdot r\|$ bin edges	Deviation	Bin Fraction
0.0-0.1	1.0205	0.323
0.1-0.25	1.0183	0.257
0.25-0.5	1.0141	0.223
0.5-0.625	1.009	0.094
0.625-0.75	0.9897	0.071
0.75-0.875	0.9310	0.026
0.875-1.0	0.8438	0.006

TABLE 11.6: The deviation from the ΔE continuum slope used in the data and fraction of events for each $q \cdot r$ bin.

11.0.9 Number of $B\overline{B}$ meson pairs

The calculation of the branching fraction uses the number of $B^0\bar{B}^0$ pairs which is computed using the integrated luminosity, the $e^+e^- \rightarrow \Upsilon(4S)$ cross-section, and the known value of the $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ cross-section. The current method is to use the run-dependent MC to count the number of $B\overline{B}$, which is assigned a systematic uncertainty of 1.5%. The uncertainty in $\mathcal{B}(\Upsilon(4S) \rightarrow B^0\bar{B}^0)$ is taken into account in the calculation of the branching fraction.

$$N_{BB} = \frac{N^{had}_{on-res} - R_{lumi} \cdot N^{had}_{off-res} \cdot k}{\epsilon_{BB}} \tag{11.5}$$

where N^{had}_{on-res} is the number of selected hadronic events in the on-resonance data, $R_{lumi} \cdot N^{had}_{off-res} \cdot k$ is the estimated number of non-BB events in the on-resonance data, and ϵ_{BB} is the efficiency of the hadronic event selection criteria for BB events. R_{lumi} is the ratio between on-resonance and off-resonance luminosity $L_{on-res}/L_{off-res}$, $N^{had}_{off-res}$ is the number of selected hadronic events in off-resonance data, and k takes into account the variation in non-BB efficiencies and cross sections with beam energy.

Using Equation 11.5, the total number of $B\overline{B}$ pairs is found to be:

$$N_{B\overline{B}} = (198.0 \pm 3.0) \times 10^6 \tag{11.6}$$

The relative uncertainty on N_{BB} is 1.5% and is dominated by systematic uncertainties due to: data/MC disagreement effects entering ϵ_{BB} and k determination (0.5%), limited knowledge of luminosities of on-resonance and off-resonance samples defining R_{lumi} (0.9%), and the limited number of off-resonance data samples available to compute k (1.14%).

11.0.10 Choice of (M_{bc}, ΔE) signal model

The correlation between M_{bc} and ΔE is taken into account using a two-dimensional kernel density estimation (KDE). To estimate the systematic uncertainty associated with the choice of the KDE model, an alternative model using analytical functions is employed. The M_{bc} and ΔE distributions are fitted with uncorrelated Crystal Ball functions, as shown in Figure 11.3. A total of 1000 ToyMC datasets were generated. Each of these datasets was subjected to two separate ToyMC studies: one using the 2D KDE model and another using the analytical signal PDF model. The difference between the alternative and the nominal yields was averaged. The same process was repeated for the $\mathcal{A}_{CP}$ values. The averaged differences between the values of the signal yield and $\mathcal{A}_{CP}$ using the KDE and analytical signal PDF models are treated as the systematic uncertainty. The systematic uncertainty related to the choice of the (M_{bc}, ΔE) signal model for the branching fraction and $\mathcal{A}_{CP}$ is 1.3% and 0.02, respectively.

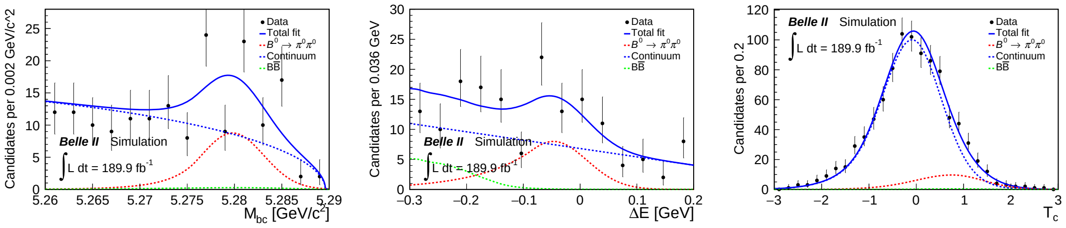


FIGURE 11.3: M_{bc} (left), ΔE (middle) and T_c (right) for $B^0 \to \pi^0\pi^0$ mode in 189.9 fb^{-1} ToyMC. The M_{bc} and ΔE are modelled using analytical Crystal Ball functions.

11.0.11 Branching ratio and $\mathcal{A}_{CP}$ fit bias

The linearity test in Figure 8.18 shows a small bias for the $\mathcal{A}_{CP}$ and signal yield. The best-fit line for the signal yield has a gradient of 0.995 ± 0.006 and an intercept of 1.392 ± 0.767, while the best-fit line for $\mathcal{A}_{CP}$ has a gradient of 1.001 ± 0.013 and an intercept of -0.021 ± 0.008. As these biases are of the same order as their uncertainty, they are not corrected for and instead taken as a systematic uncertainty. The systematic uncertainty related to this bias for the branching fraction and $\mathcal{A}_{CP}$ is 1.0% and 0.02, respectively.

11.0.12 Best Candidate Selection

In this analysis, if an event had multiple candidates, the one with the minimum deviation of the two π^0's reconstructed invariant masses from the world average was selected. To evaluate the systematic uncertainty related to this choice, 1000 ToyMC datasets were generated: half in which the best B^0 candidate was selected and another in which the B^0 candidates were instead selected randomly. To ensure that the generated samples are identical except for the choice of best candidate, 1000 numbers are randomly generated and used as the seed for the event generation. Then, ToyMC studies were performed on both samples, and the differences between the alternative and the nominal yields in each were averaged. The same process was repeated for the $\mathcal{A}_{CP}$ values. The systematic uncertainty related to the best candidate selection for the branching fraction and $\mathcal{A}_{CP}$ was 0.2% and 0.002, respectively.

11.0.13 Mistagging Parameters

In the analysis of experimental data, the mistagging, asymmetry, and the discrepancy between mistagging for B^0 and $\bar{B}^0$ terms, along with their uncertainties for each bin, are determined from the data. Each parameter is individually fluctuated by its uncertainty, as shown in Table 8.1, and the fit is repeated for 1000 ToyMC studies. It is observed that most of the parameters exhibit no correlations or only minor correlations. The difference between the alternative and nominal $\mathcal{A}_{CP}$ is calculated and added in quadrature, which is then considered as the systematic uncertainty.

These uncertainties also account for potential biases in the FlavorTagger due to the signal side. The contribution of each mistagging parameter is documented in Table 11.7. As expected, the highest bins of $q \cdot r$, where the majority of the signal is located, contribute the most. The average $\mathcal{A}_{CP}$ uncertainty for each of the 21 parameters is 0.041. The systematic uncertainty for the branching fraction and $\mathcal{A}_{CP}$ is 0.3% and 0.05, respectively.

11.0.14 $B\overline{B}$ Background $\mathcal{A}_{CP}$

The two dominant $B\overline{B}$ backgrounds, $B^+ \to \rho^+\pi^0$ and $\to K_S^0(\to \pi^0\pi^0)\pi^0$ are assumed to have $\mathcal{A}_{CP} = 0.0$. However, the uncertainties in $\mathcal{A}_{CP}$ are relatively large, as shown in

Bin	w	Δw	μ
1	0.017	0.0021	0.00071
2	0.0091	0.0024	0.0013
3	0.0071	0.0045	0.003
4	0.0086	0.0079	0.0066
5	0.012	0.023	0.019
6	0.0096	0.014	0.013
7	0.0073	0.012	0.014

TABLE 11.7: Contribution of each mistagging parameter to the $\mathcal{A}_{CP}$ systematic uncertainty.

Table 11.4, and the potential $\mathcal{A}_{CP}$ in the $B\overline{B}$ background can affect the extracted value of $\mathcal{A}_{CP}$ $(B^0 \to \pi^0\pi^0)$.

To estimate the systematic error, $B^+ \to \rho^+\pi^0$ is generated with $\mathcal{A}_{CP} = -0.09, 0.13$ and $\to K^0_S(\to \pi^0\pi^0)\pi^0$ with $\mathcal{A}_{CP} = -0.13, 0.13$. These values correspond to one standard deviation from their known values. Using 1000 ToyMC, the $\mathcal{A}_{CP}$ for the two $B\overline{B}$ backgrounds is varied by one standard deviation from their known values. The data are then refitted with all possible combinations of $\mathcal{A}_{CP}$ values for the background processes. The mean of the $\mathcal{A}_{CP}$ for each of the different possible $\mathcal{A}_{CP}$ is determined, as shown in Table 11.8. The largest difference between the averaged $\mathcal{A}_{CP}$ values between zero $B\overline{B}$ $\mathcal{A}_{CP}$ and all other possible $B\overline{B}$ $\mathcal{A}_{CP}$ is 0.021. This value is used to conservatively estimate the systematic uncertainty for the $\mathcal{A}_{CP}$.

$\mathcal{A}_{CP}$ $(B^+ \to \rho^+\pi^0)$	$\mathcal{A}_{CP}$ $(\to K^0_S\pi^0)$	**Average Extracted** $\mathcal{A}_{CP}$
0	0	-0.339
-0.09	-0.13	-0.351
0.13	0.13	-0.318
-0.09	0.13	-0.354
0.13	-0.13	-0.327

TABLE 11.8: The possible $\mathcal{A}_{CP}$ depending on the $\mathcal{A}_{CP}$ in the $B\overline{B}$ background.

11.0.15 Continuum $q \cdot r$ asymmetry

As described in subsection 8.3.6, a CP asymmetry of -0.033 ± 0.002 is extracted from the continuum data sideband and incorporated into the fit. To estimate the systematic error, the signal data are re-refitted with the continuum $\mathcal{A}_{CP}$ fluctuating by one standard deviation. The alternative $\mathcal{A}_{CP}$ value is 0.15, the difference between this value and the nominal value of 0.14 is considered as the systematic uncertainty. The systematic uncertainty related to the continuum $q \cdot r$ asymmetry for $\mathcal{A}_{CP}$ is 0.01.

Chapter 12

Conclusion

This thesis presents a study of the branching fraction and direct CP asymmetry of the $B^0 \to \pi^0\pi^0$ decay mode. This analysis utilises an electron-positron collision sample containing 198×10^6 $B\overline{B}$ pairs, collected at Belle II.

Simulated samples are employed to optimise event selection criteria, compare observed data distributions with expectations, study background sources, and determine the probability density functions used to fit the data. A machine learning algorithm, trained using information from the electromagnetic calorimeter, is used to statistically distinguish misreconstructed photons from genuine photons. The continuum is suppressed using a similar machine learning algorithm that uses information associated with the event topology. Another machine learning algorithm uses information from the B meson accompanying the signal B meson to assign a probability r for a signal B to be a B^0 or a $\bar{B}^0$.

A fit to the sample composition is used to identify the signal through the kinematic variable, M_{bc} and ΔE, and the log-transform of the continuum suppression T_c . Discrepancies between the simulated and experimental data are corrected, and $\mathcal{B}$ and $\mathcal{A}_{CP}$ are determined by performing a three-dimensional (M_{bc}, ΔE, T_c) unbinned extended maximum likelihood fit simultaneously to events in the seven intervals of r.

The branching fraction and direct CP asymmetry for the $B^0 \to \pi^0\pi^0$ decay is measured to be

$$\mathcal{B}(B^0 \to \pi^0\pi^0) = (1.38 \pm 0.27 \pm 0.22) \times 10^{-6} \tag{12.1}$$

$$\mathcal{A}_{CP}(B^0 \to \pi^0\pi^0) = 0.14 \pm 0.46 \pm 0.07 \tag{12.2}$$

where the first uncertainty is statistical and the second uncertainty is systematic. These measurements agree with previous measurements. The branching fraction uncertainty is similar in size to those reported by the Babar and Belle collaboration, despite using a sample 2.4 and 4.0 times smaller, respectively. These improvements are due to a 60% higher signal efficiency with approximately 40% less background. The higher efficiency and lower background result from improved photon timing, BDT-based photon selection, and data-driven continuum suppression.

These results are averaged with previous measurements of $\mathcal{B}$ and $\mathcal{A}_{CP}$ for $B^0 \to \pi^0\pi^0$ and use an isospin analysis together with known measurements of $\mathcal{B}$ and time-dependent *CP* violation for $B^0 \to \pi^+\pi^-$ and $\mathcal{B}$ and $\mathcal{A}_{CP}$ for $B^+ \to \pi^+\pi^0$ to constrain ϕ_2. The updated results for $B^0 \to \pi^0\pi^0$ exclude $8.5^\circ < \phi_2 < 82.0^\circ$ at the 68% confidence level and $12.5^\circ < \phi_2 < 78.0^\circ$ at 95% confidence level.

Although the current data set is relatively small, a comprehensive analysis of the systematic uncertainties has been performed. It shows that these systematic uncertainties are already on par, or even smaller than those of previous studies. Furthermore, as the sample size increases, both statistical and systematic uncertainties, which are primarily data driven, are expected to decrease. This reduction will pave the way for future high-precision measurements of $B^0 \to \pi^0\pi^0$, thereby allowing for further constraints on the parameter space of new physics.

In particular, the uncertainty of the $B^0 \to \pi^0\pi^0$ inputs, $\mathcal{B}$ and $\mathcal{A}_{CP}$ are the greatest limitation to exploiting the isospin relationships which would allow the least known angle of the unitarity triangle ϕ_2 to be measured. Improved measurements of ϕ_2 will not only test the unitarity of the CKM matrix but also provide robust constraints on potential flavour-structure extensions of the standard model. The results of this thesis demonstrate the improved precision of Belle II and its potential to impose strong constraints on ϕ_2

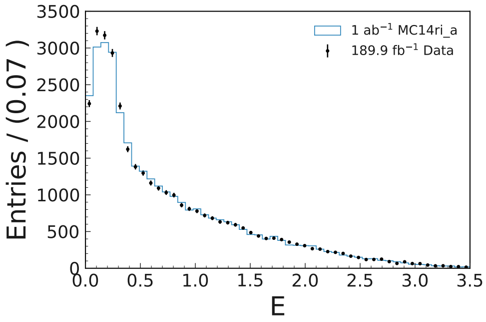
Appendix A

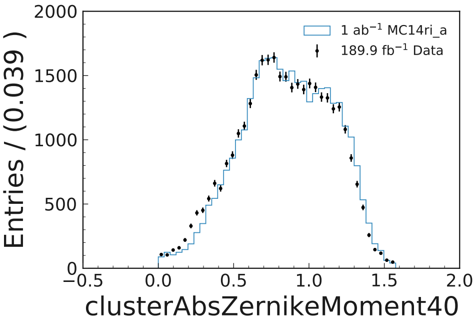
PhotonMVA

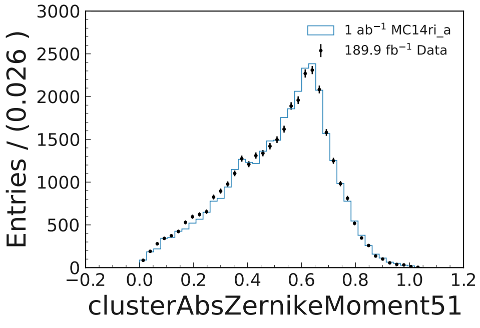
Plots related to the `PhotonMVA` described in Chapter 5 are shown here.

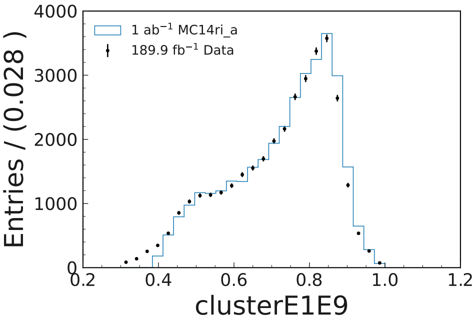
A.1 Data-MC comparison of photon variable

The photon variables that are selected as inputs for the `photonMVA` must demonstrate good agreement between data and MC. This is to ensure the consistency of the classifier's performance across both data sets. The normalised distributions for 189.9 fb^{-1} of data and 1 ab^{-1} of run-indepedent MC are shown below.









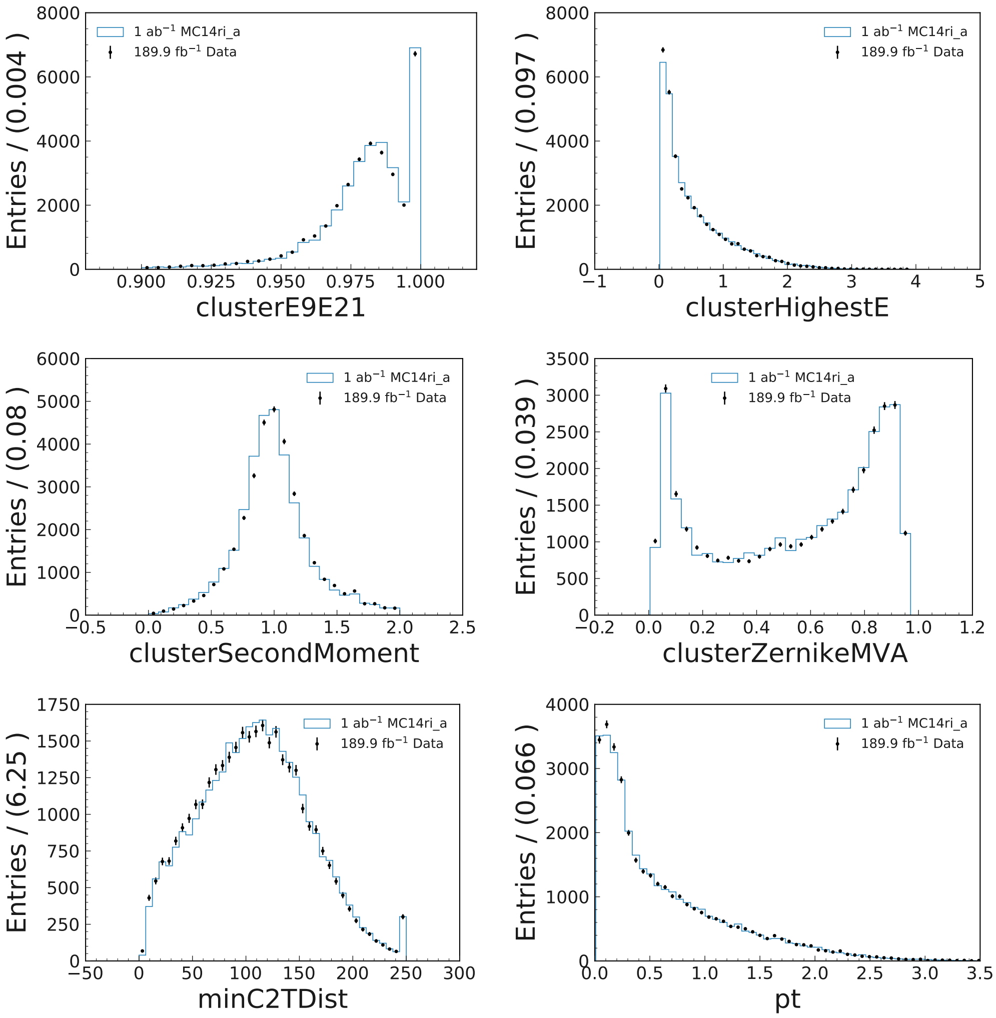
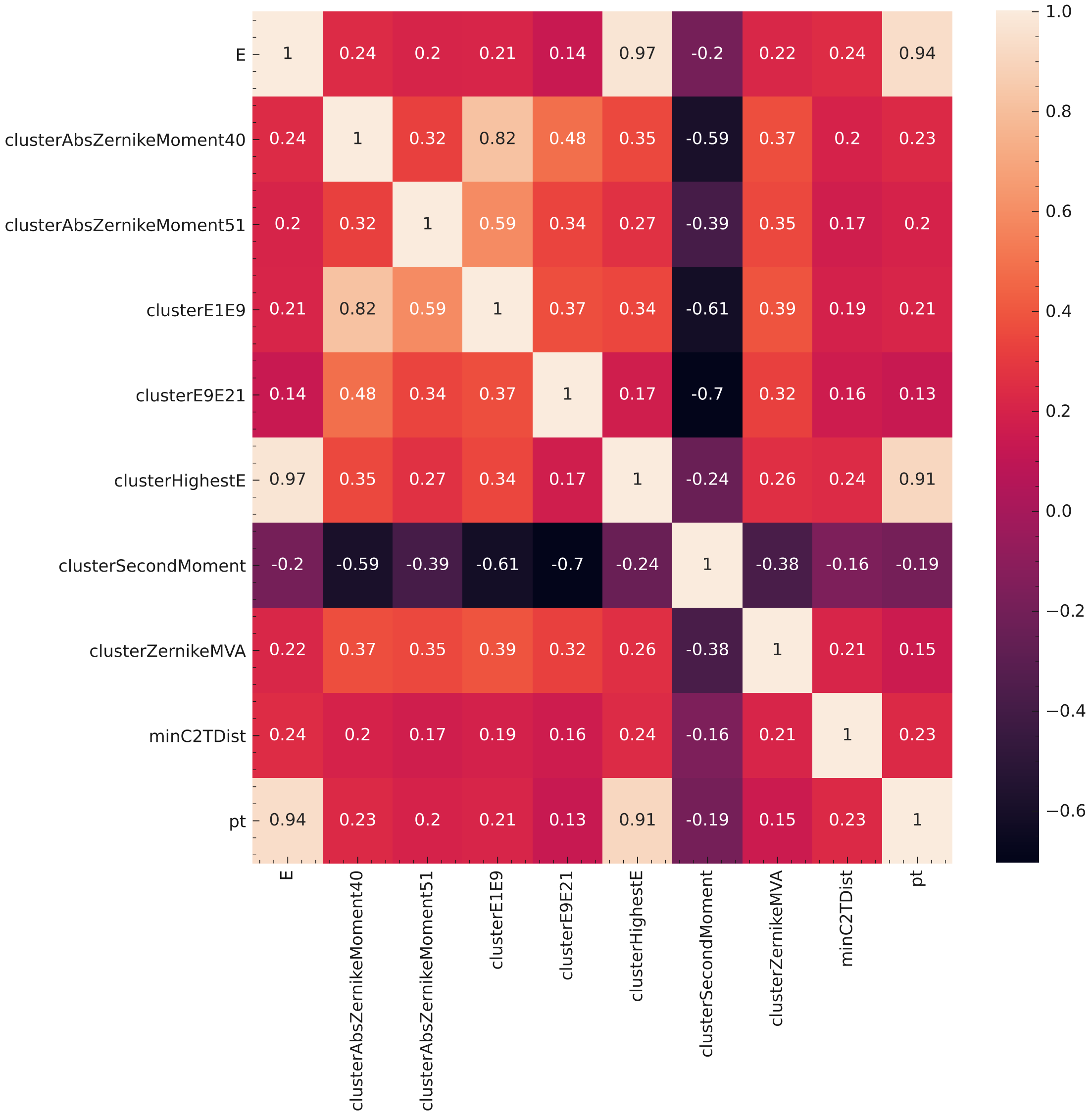


Figure A.1: Data vs MC ECL variables distribution comparison

A.2 ECL Variable Correlation

The correlation of the photon variables that are selected as inputs for the `photonMVA` is shown below. The correlations between many of these variables are relatively large because they characterise similar quantities, such as how the photon energy is distributed between ECL crystals.



	E	clusterAbsZernikeMoment40	clusterAbsZernikeMoment51	clusterE1E9	clusterE9E21	clusterHighestE	clusterSecondMoment	clusterZernikeMVA	minC2TDist	pt
E	1	0.24	0.2	0.21	0.14	0.97	-0.2	0.22	0.24	0.94
clusterAbsZernikeMoment40	0.24	1	0.32	0.82	0.48	0.35	-0.59	0.37	0.2	0.23
clusterAbsZernikeMoment51	0.2	0.32	1	0.59	0.34	0.27	-0.39	0.35	0.17	0.2
clusterE1E9	0.21	0.82	0.59	1	0.37	0.34	-0.61	0.39	0.19	0.21
clusterE9E21	0.14	0.48	0.34	0.37	1	0.17	-0.7	0.32	0.16	0.13
clusterHighestE	0.97	0.35	0.27	0.34	0.17	1	-0.24	0.26	0.24	0.91
clusterSecondMoment	-0.2	-0.59	-0.39	-0.61	-0.7	-0.24	1	-0.38	-0.16	-0.19
clusterZernikeMVA	0.22	0.37	0.35	0.39	0.32	0.26	-0.38	1	0.21	0.15
minC2TDist	0.24	0.2	0.17	0.19	0.16	0.24	-0.16	0.21	1	0.23
pt	0.94	0.23	0.2	0.21	0.13	0.91	-0.19	0.15	0.23	1

FIGURE A.2: Correlation of the `photonMVA` ECL variables from photons that originate from correctly-reconstructed $B^0 \to \pi^0\pi^0$ decays.

A.3 Remaining PhotonMVA Results

For each region of the ECL, a different classifier is trained. The results of the training in the backward and forward endcaps are shown below.

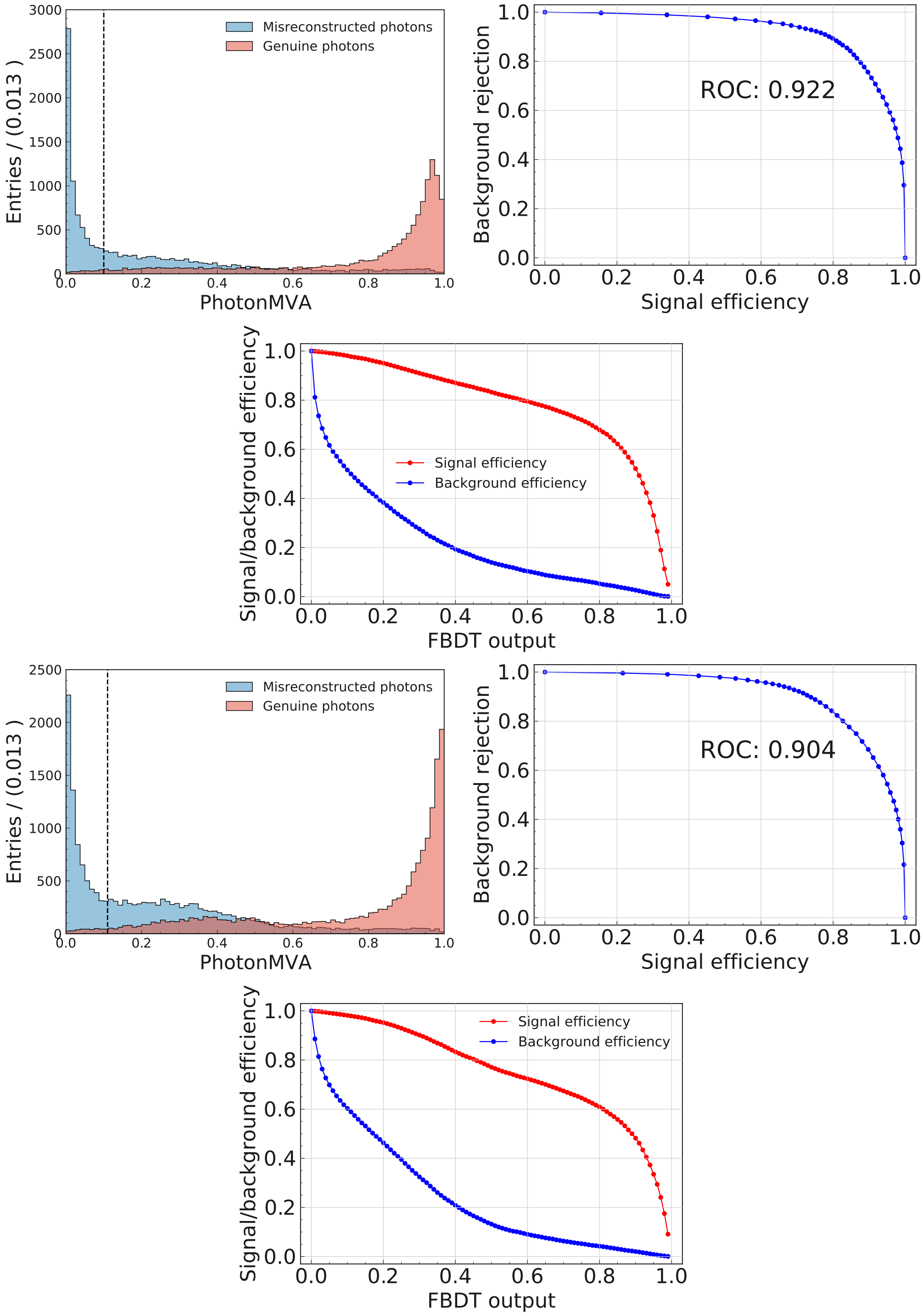


Figure A.3: FBDT output on testing data (left) for signal (red) and background (green) where the black line is the `photonMVA` selection, Receiver Operating Characteristic Curve (middle), and signal-retention and background-rejection (right) for the `photonMVA` in the backward region (top) and forward region (bottom).

Appendix B

Continuum Suppression

B.1 Continuum Suppression Training Variables

All variables considered for the training continuum suppression classifier, but not shown previously, are displayed below.

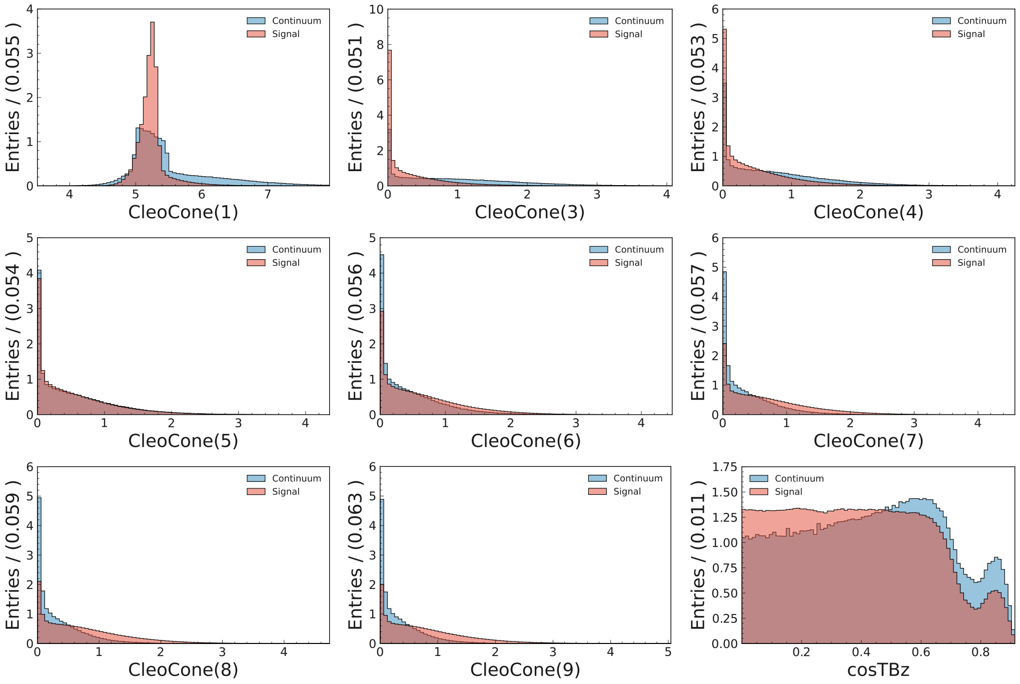


Figure B.1: Variables used for FastBDT Continuum Suppression training for signal (blue) and background (red)

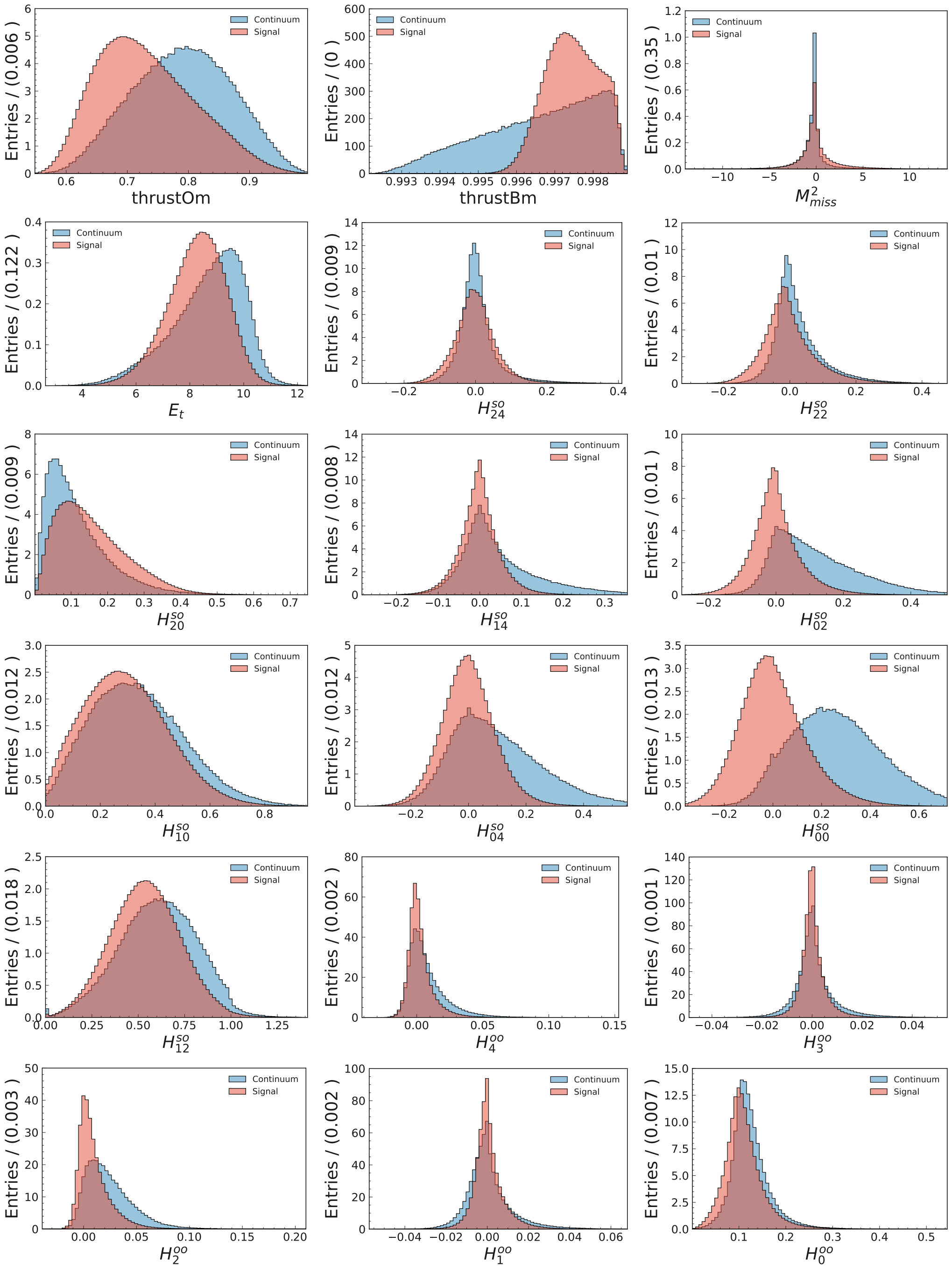


FIGURE B.2: Variables used for FastBDT Continuum Suppression training for signal (blue) and background (red)

Appendix C

Probability Density Function

The r and T_c are observed to be correlated for the signal and $B\overline{B}$ components. The different Probability Density Functions fitted for each r bin from MC are shown below.

C.1 Signal and $B\overline{B}$ fits to T_c for each $q \cdot r$ bin

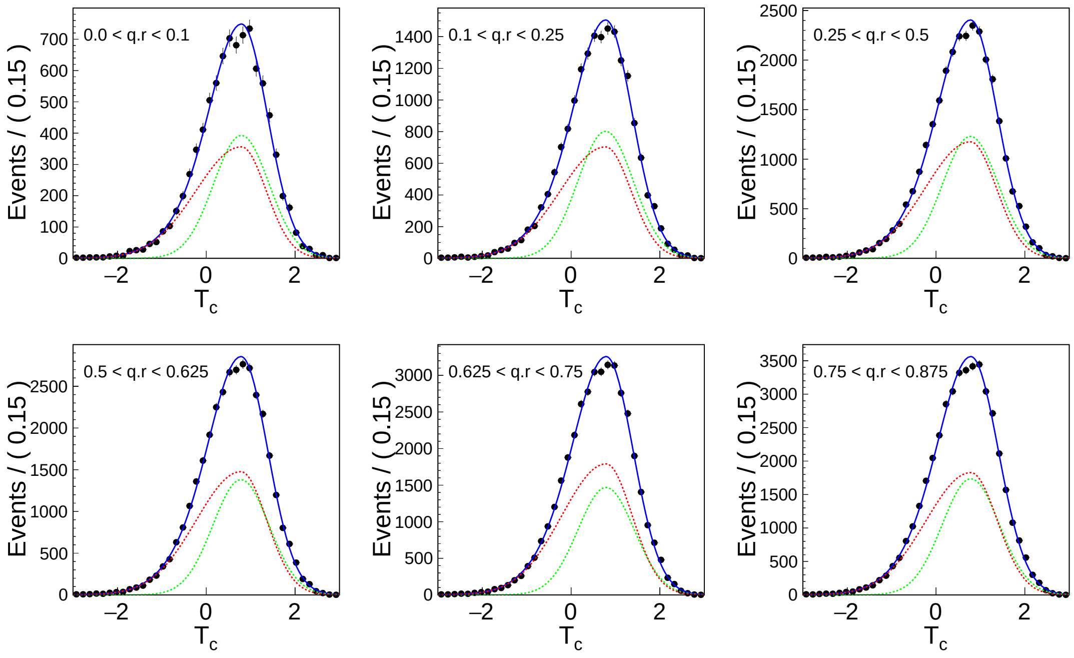


FIGURE C.1: Fits to T_c for MC in each $q \cdot r$ bins for signal.

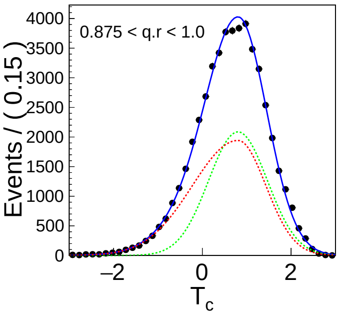


Figure C.1: Fits to T_c for MC in each $q \cdot r$ bins for signal continued.

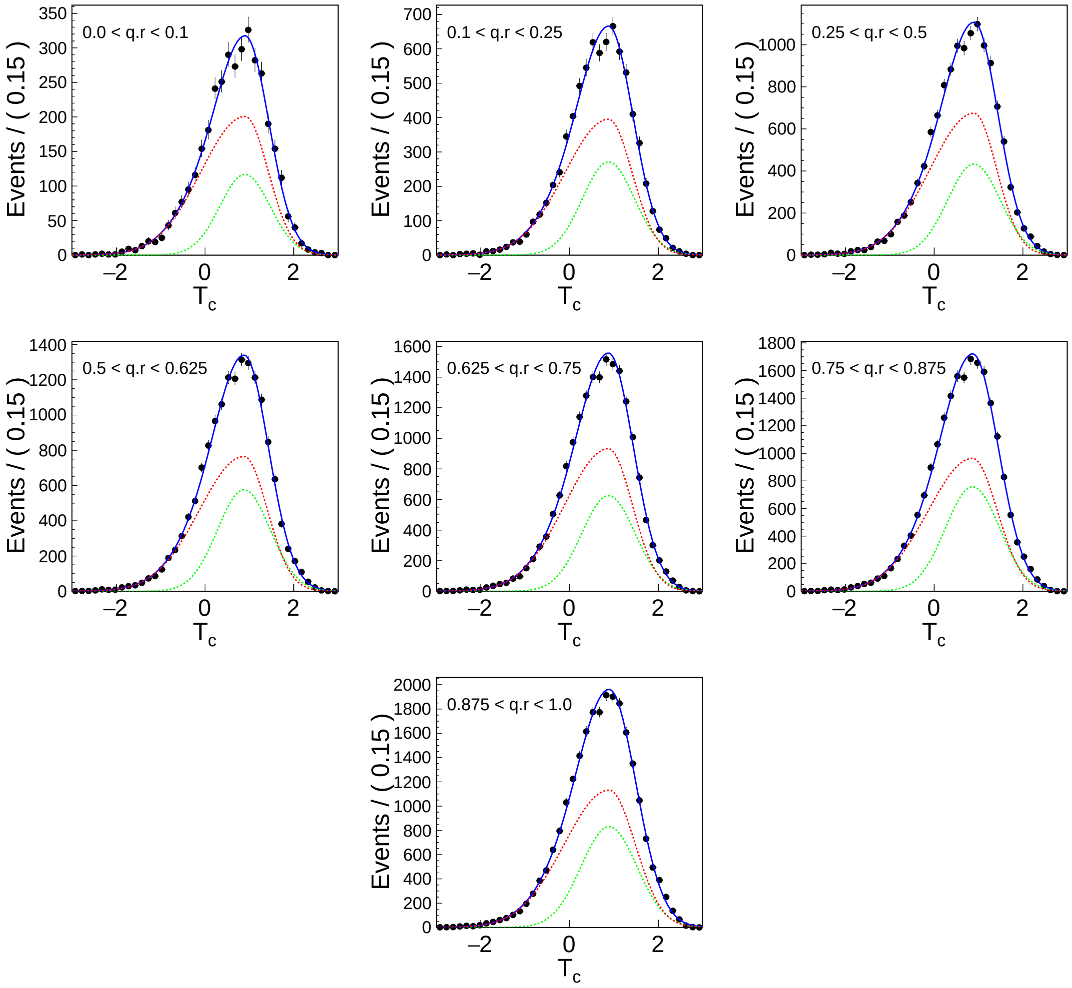


Figure C.2: Fits to T_c for MC in each $q \cdot r$ bins for $B\overline{B}$.

Bibliography

[1] F. Abudinén et al. Measurement of the branching fraction and CP asymmetry of $B^0 \to \pi^0\pi^0$ decays using $198 \times 10^6 B\bar{B}$ pairs in Belle II data. *Phys. Rev. D*, 107(11):112009, 2023. doi:10.1103/PhysRevD.107.112009.

[2] Standard Model of Elementary Particles. https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg.

[3] KEK. Integrated luminosity of b factories. URL: http://belle.kek.jp/bdocs/lumi_belle.png.

[4] ATLAS Collaboration. ATLAS Physics Briefing: ATLAS measures Higgs boson's total width. General Photo, 2022. URL: https://cds.cern.ch/record/2841059.

[5] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2022(8), 08 2022. 083C01. doi:10.1093/ptep/ptac097.

[6] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J.*, C74:3026, 2014. doi:10.1140/epjc/s10052-014-3026-9.

[7] Marina Artuso, Elisabetta Barberio, and Sheldon Stone. B Meson Decays. *PMC Phys. A*, 3:3, 2009. doi:10.1186/1754-0410-3-3.

[8] Michael Gronau and David London. Isospin analysis of CP asymmetries in B decays. *Phys. Rev. Lett.*, 65:3381–3384, 1990. doi:10.1103/PhysRevLett.65.3381.

[9] J. Charles et al. *Eur. Phys. J. C*, 41:1–131, 2005.

[10] S. I. Eidel'man, A. V. Nefediev, P. N. Pakhlov, and V. I. Zhukova. Super-factory of bottomed hadrons Belle II. *Usp. Fiz. Nauk*, 191(5):492–521, 2021. arXiv:2012.05147, doi:10.3367/UFNe.2020.10.038847.

[11] Donald E. Groom, Nikolai V. Mokhov, and Sergei I. Striganov. Muon stopping power and range tables 10-MeV to 100-TeV. *Atom. Data Nucl. Data Tabl.*, 78:183–356, 2001. `doi:10.1006/adnd.2001.0861`.

[12] V.M. Aulchenko, V.N. Zhilich, V.V. Zhulanov, and et al. Structure and algorithm of electronics of a multichannel crystal calorimeter for a high-rate trigger. *Optoelectron. Instrument. Proc.*, 51:31–38, 2015. `doi:10.3103/S8756699015010057`.

[13] T. Keck et al. The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, 3(1):6, 2019. `doi:10.1007/s41781-019-0021-8`.

[14] Abdelgader Alamrouni, Fidan Aslanova, Sagiru Mati, Afaf Jibril, Sani Abba, Hamza Maccido, and Abdullahi Usman. Multi-Regional Modeling of Cumulative COVID-19 Cases Integrated with Environmental Forest Knowledge Estimation: A Deep Learning Ensemble Approach. *International Journal of Environmental Research and Public Health*, 19:738, 01 2022. `doi:10.3390/ijerph19020738`.

[15] Hai Tao, Sinan Salih, Atheer Oudah, Sani Abba, Ameen Ameen, Salih Awadh, Omer A. Alawi, Reham Mostafa, Surendran Udayar Pillai, and Zaher Yaseen. Development of new computational machine learning models for longitudinal dispersion coefficient determination: case study of natural streams, United States. *Environmental Science and Pollution Research*, 29, 05 2022. `doi:10.1007/s11356-022-18554-y`.

[16] P. Kaur, H.S. Pannu, and A.K. Malhi. Plant disease recognition using fractional-order zernike moments and svm classifier. *Neural Computing and Applications*, 31:8749–8768, 2019. `doi:10.1007/s00521-018-3939-6`.

[17] Wikimedia Commons. Roc curve, 2018. URL: `https://commons.wikimedia.org/wiki/File:Roc_curve.svg`.

[18] Wikimedia Commons. Comparison of 1d histogram and kde, 2010. URL: `https://commons.wikimedia.org/wiki/File:Comparison_of_1D_histogram_and_KDE.png`.

[19] M. B. Gavela, P. Hernandez, J. Orloff, O. Pene, and C. Quimbay. Standard model CP violation and baryon asymmetry. Part 2: Finite temperature. *Nucl. Phys.*

B, 430:382–426, 1994. arXiv:hep-ph/9406289, doi:10.1016/0550-3213(94)00410-2.

[20] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. doi:10.1070/PU1991v034n05ABEH002497.

[21] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2π Decay of the K_2^0 Meson. *Phys. Rev. Lett.*, 13:138–140, 1964. doi:10.1103/PhysRevLett.13.138.

[22] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973. doi:10.1143/PTP.49.652.

[23] Ashton B. Carter and A. I. Sanda. CP Violation in B Meson Decays. *Phys. Rev. D*, 23:1567, 1981. doi:10.1103/PhysRevD.23.1567.

[24] Kazuo Abe et al. Observation of large CP violation in the neutral B meson system. *Phys. Rev. Lett.*, 87:091802, 2001. doi:10.1103/PhysRevLett.87.091802.

[25] Bernard Aubert et al. Measurement of CP violating asymmetries in B^0 decays to CP eigenstates. *Phys. Rev. Lett.*, 86:2515–2522, 2001. doi:10.1103/PhysRevLett.86.2515.

[26] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi:10.1093/ptep/ptac097.

[27] J. P. Lees et al. Measurement of CP Asymmetries and Branching Fractions in Charmless Two-Body B-Meson Decays to Pions and Kaons. *Phys. Rev. D*, 87(5):052009, 2013. doi:10.1103/PhysRevD.87.052009.

[28] T. Julius et al. Measurement of the branching fraction and CP asymmetry in $B^0 \to \pi^0\pi^0$ decays, and an improved constraint on ϕ_2. *Phys. Rev. D*, 96(3):032007, 2017. doi:10.1103/PhysRevD.96.032007.

[29] G. Breit and E. Wigner. Capture of Slow Neutrons. *Phys. Rev.*, 49:519–531, 1936. doi:10.1103/PhysRev.49.519.

[30] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963. `doi:10.1103/PhysRevLett.10.531`.

[31] Lincoln Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51:1945, 1983. `doi:10.1103/PhysRevLett.51.1945`.

[32] Y. Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016. *Eur. Phys. J. C*, 77(12):895, 2017. `doi:10.1140/epjc/s10052-017-5058-4`.

[33] J. Charles, O. Deschamps, S. Descotes-Genon, and V. Niess. Isospin analysis of charmless B-meson decays. *Eur. Phys. J. C*, 77(8):574, 2017. `doi:10.1140/epjc/s10052-017-5126-9`.

[34] Bernard Aubert et al. Observation of the decay $B^0 \to \pi^0\pi^0$. *Phys. Rev. Lett.*, 91:241801, 2003. `doi:10.1103/PhysRevLett.91.241801`.

[35] S. H. Lee et al. Evidence for $B^0 \to \pi^0\pi^0$. *Phys. Rev. Lett.*, 91:261801, 2003. `doi:10.1103/PhysRevLett.91.261801`.

[36] Martin Beneke and Matthias Neubert. QCD factorization for $B \to PP$ and $B \to PV$ decays. *Nucl. Phys. B*, 675:333–415, 2003. `doi:10.1016/j.nuclphysb.2003.09.026`.

[37] Cai-Dian Lu, Kazumasa Ukai, and Mao-Zhi Yang. Branching ratio and CP violation of $B \to \pi\pi$ decays in perturbative QCD approach. *Phys. Rev. D*, 63:074009, 2001. `doi:10.1103/PhysRevD.63.074009`.

[38] Ling-Lie Chau, Hai-Yang Cheng, W. K. Sze, Herng Yao, and Benjamin Tseng. Charmless nonleptonic rare decays of B mesons. *Phys. Rev. D*, 43:2176–2192, 1991. [Erratum: Phys.Rev.D 58, 019902 (1998)]. `doi:10.1103/PhysRevD.43.2176`.

[39] Michael Gronau, Oscar F. Hernandez, David London, and Jonathan L. Rosner. Broken SU(3) symmetry in two-body B decays. *Phys. Rev. D*, 52:6356–6373, 1995. `doi:10.1103/PhysRevD.52.6356`.

[40] M. Beneke, G. Buchalla, M. Neubert, and Christopher T. Sachrajda. QCD factorization for $B \to \pi\pi$ decays: Strong phases and CP violation in the heavy quark limit. *Phys. Rev. Lett.*, 83:1914–1917, 1999. `doi:10.1103/PhysRevLett.83.1914`.

[41] M. Beneke and S. Jager. Spectator scattering at NLO in non-leptonic b decays: Tree amplitudes. *Nucl. Phys. B*, 751:160–185, 2006. doi:10.1016/j.nuclphysb.2006.06.010.

[42] Volker Pilipp. Hard spectator interactions in $B \to \pi\pi$ at order α_s^2. *Nucl. Phys. B*, 794:154–188, 2008. doi:10.1016/j.nuclphysb.2007.10.028.

[43] Ya-Lan Zhang, Xue-Yan Liu, Ying-Ying Fan, Shan Cheng, and Zhen-Jun Xiao. $B \to \pi\pi$ decays and effects of the next-to-leading order contributions. *Phys. Rev. D*, 90(1):014029, 2014. doi:10.1103/PhysRevD.90.014029.

[44] Yeo-Yie Charng and Hsiang-nan Li. Weak phases from the $B \to \pi\pi$, $K\pi$ decays. *Phys. Rev. D*, 71:014036, 2005. doi:10.1103/PhysRevD.71.014036.

[45] Satoshi Mishima and Tadashi Yoshikawa. Large electroweak penguin contribution in $B \to K\pi$ and $\pi\pi$ decay modes. *Phys. Rev. D*, 70:094024, 2004. doi:10.1103/PhysRevD.70.094024.

[46] Yeo-Yie Charng and Hsiang-nan Li. Weak phases from topological amplitude parametrization. *Phys. Lett. B*, 594:185–195, 2004. doi:10.1016/j.physletb.2004.04.083.

[47] Cong-Feng Qiao, Rui-Lin Zhu, Xing-Gang Wu, and Stanley J. Brodsky. A possible solution to the $B \to \pi\pi$ puzzle using the principle of maximum conformality. *Phys. Lett. B*, 748:422–427, 2015. doi:10.1016/j.physletb.2015.07.044.

[48] Hsiang-nan Li and Satoshi Mishima. Possible resolution of the $B \to \pi\pi$, πK puzzles. *Phys. Rev. D*, 83:034023, 2011. doi:10.1103/PhysRevD.83.034023.

[49] Yun-Feng Li and Xian-Qiao Yu. Revisiting the $B^0 \to \pi^0\pi^0$ decays in the perturbative QCD approach. *Phys. Rev. D*, 95(3):034023, 2017. doi:10.1103/PhysRevD.95.034023.

[50] Hai-Yang Cheng, Cheng-Wei Chiang, and An-Li Kuo. *Phys. Rev. D*, 91(1):014011, 2015. doi:10.1103/PhysRevD.91.014011.

[51] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J. C*, 74:3026, 2014. doi:10.1140/epjc/s10052-014-3026-9.

[52] M. Bona et al. SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report. 5 2007. arXiv:0709.0451.

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

[54] I. Adachi, T. E. Browder, P. Križan, S. Tanaka, and Y. Ushiroda. Detectors for extreme luminosity: Belle II. *Nucl. Instrum. Meth. A*, 907:46–59, 2018. doi:10.1016/j.nima.2018.03.068.

[55] V. Aulchenko et al. Electromagnetic calorimeter for Belle II. *J. Phys. Conf. Ser.*, 587(1):012045, 2015. doi:10.1088/1742-6596/587/1/012045.

[56] Valerio Bertacchi et al. Track finding at Belle II. *Comput. Phys. Commun.*, 259:107610, 2021. doi:10.1016/j.cpc.2020.107610.

[57] E. Kovalenko. On-line luminosity measurements at Belle II. *JINST*, 15(06):C06067, 2020. doi:10.1088/1748-0221/15/06/C06067.

[58] Satoru Yamada, Ryosuke Itoh, Katsuro Nakamura, Mikihiko Nakao, Soh Y. Suzuki, Tomoyuki Konno, Takeo Higuchi, Zhen'an Liu, and Jingzhou Zhao. Data Acquisition System for the Belle II Experiment. *IEEE Trans. Nucl. Sci.*, 62(3):1175–1180, 2015. doi:10.1109/TNS.2015.2424717.

[59] Yoshihito Iwasaki, ByungGu Cheon, Eunil Won, Xin Gao, Luca Macchiarulo, Kurtis Nishimura, and Gary Varner. Level 1 trigger system for the Belle II experiment. *IEEE Trans. Nucl. Sci.*, 58:1807–1815, 2011. doi:10.1109/TNS.2011.2119329.

[60] Dehui Sun, Zhen'an Liu, Jingzhou Zhao, and Hao Xu. Belle2Link: A Global Data Readout and Transmission for Belle II Experiment at KEK. *Phys. Procedia*, 37:1933–1939, 2012. doi:10.1016/j.phpro.2012.01.036.

[61] Thomas Geßler, Wolfgang Kühn, Jens Sören Lange, Zhen'An Liu, David Münchow, Björn Spruck, and Jingzhou Zhao. The ONSEN Data Reduction System for the Belle II Pixel Detector. *IEEE Trans. Nucl. Sci.*, 62(3):1149–1154, 2015. doi:10.1109/TNS.2015.2414713.

[62] Florian Bernlochner, Bruno Deschamps, Jochen Dingfelder, Carlos Marinas, and Christian Wessel. Online Data Reduction for the Belle II Experiment using DATCON. *EPJ Web Conf.*, 150:00014, 2017. doi:10.1051/epjconf/201715000014.

[63] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001. `doi:10.1016/S0168-9002(01)00089-4`.

[64] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015. `doi:10.1016/j.cpc.2015.01.024`.

[65] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003. `doi:10.1016/S0168-9002(03)01368-8`.

[66] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II Core Software. *Comput. Softw. Big Sci.*, 3(1):1, 2019. `doi:10.1007/s41781-018-0017-9`.

[67] Piotr Golonka and Zbigniew Was. PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays. *Eur. Phys. J.*, C45:97–107, 2006. `doi:10.1140/epjc/s2005-02396-4`.

[68] S. Jadach, B. F. L. Ward, and Z. Was. The Precision Monte Carlo event generator K K for two fermion final states in e+ e- collisions. *Comput. Phys. Commun.*, 130:260–325, 2000. `doi:10.1016/S0010-4655(00)00048-5`.

[69] S. Jadach, Z. Was, R. Decker, and Johann H. Kuhn. The tau decay library TAUOLA: Version 2.4. *Comput. Phys. Commun.*, 76:361–380, 1993. `doi:10.1016/0010-4655(93)90061-G`.

[70] Sato, Yo, Cunliffe, Sam, Meier, Frank, and Zupanc, Anze. Monte carlo matching in the belle ii software. *EPJ Web Conf.*, 251:03021, 2021. `doi:10.1051/epjconf/202125103021`.

[71] Thomas Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, 1(1):2, 2017. `doi:10.1007/s41781-017-0002-8`.

[72] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2nd edition, 2009.

[73] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001. `doi:10.1214/aos/1013203451`.

[74] J. F. Krohn et al. Global decay chain vertex fitting at Belle II. *Nucl. Instrum. Meth. A*, 976:164269, 2020. `arXiv:1901.11198`, `doi:10.1016/j.nima.2020.164269`.

[75] S. Cunliffk Belle II. Physics Cross Section. 2020. URL: `https://confluence.desy.de/display/BI/Physics+Cross+Sections+Table`.

[76] Geoffrey C. Fox and Stephen Wolfram. Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes. *Phys. Rev. Lett.*, 41:1581, 1978. `doi:10.1103/PhysRevLett.41.1581`.

[77] Kazuo Abe et al. Measurement of branching fractions for $B \to \pi\pi, K\pi$ and KK decays. *Phys. Rev. Lett.*, 87:101801, 2001. `doi:10.1103/PhysRevLett.87.101801`.

[78] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997. `doi:10.1016/S0168-9002(97)00048-X`.

[79] Fernando Abudinén. The Belle II flavor tagger. *EPJ Web Conf.*, 214:06032, 2019. `doi:10.1051/epjconf/201921406032`.

[80] F. Abudinén et al. B-flavor tagging at Belle II. *Eur. Phys. J. C*, 82(4):283, 2022. `doi:10.1140/epjc/s10052-022-10180-9`.

[81] M. Hatlo, F. James, P. Mato, L. Moneta, M. Winkler, and A. Zsenei. Developments of mathematical software libraries for the LHC experiments. *IEEE Trans. Nucl. Sci.*, 52:2818–2822, 2005. `doi:10.1109/TNS.2005.860152`.

[82] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975. `doi:10.1016/0010-4655(75)90039-9`.

[83] Seema Choudhury 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*, 107(3):L031102, 2023. `doi:10.1103/PhysRevD.107.L031102`.

[84] T. Bloomfield et al. Measurement of the branching fraction and CP asymmetry for $B \to \bar{D}^0\pi$ decays. *Phys. Rev. D*, 105(7):072007, 2022. `doi:10.1103/PhysRevD.105.072007`.

[85] Robert Fleischer and Thomas Mannel. General analysis of new physics in $B \to J/\psi K$. *Phys. Lett. B*, 506:311–322, 2001. `doi:10.1016/S0370-2693(01)00346-X`.