Search for the lepton flavour violating decay $B^0 \to \ell^\pm \tau^\mp$ at Belle II

PRIYANKA CHEEMA

THE UNIVERSITY OF SYDNEY

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

School of Physics
Faculty of Science
The University of Sydney
Australia

8 May 2026

Abstract

Despite the successes of the Standard Model, its limitations provide strong indications that new physics beyond the Standard Model must exist. Today, numerous avenues are explored to search for any evidence of new physics, one of which is rare particle decays. Such decays are promising as their suppressed decay rates allow for any enhancements from new physics contributions to be more readily seen. One such rare decay of interest is $B^0 \to \ell^\pm \tau^\mp$ ($\ell = e, \mu$) which is a lepton flavour violating decay making it forbidden within the conventional Standard Model framework. However, within an extended Standard Model that includes neutrino oscillations, which there is strong evidence for, the decay can occur but at a rate far beyond any current or future experimental sensitivity. Therefore, observation of $B^0 \to \ell^\pm \tau^\mp$ would be strong and clear evidence of new physics beyond the Standard Model. This study presents the first search for $B^0 \to \ell^\pm \tau^\mp$ using 365 fb^{-1} of data collected by the Belle II experiment from 2019-2022. The Belle II experiment is situated around the interaction point of the SuperKEKB collider which is an asymmetric e^+e^- collider designed to mass produce B mesons. The primary aim of the study presented in this thesis is to develop an analysis technique for future hadronic tag-based $B^0 \to \ell^\pm \tau^\mp$ searches at Belle II that yields an improved sensitivity over previous analysis strategies employed by other experiments. This is achieved by partitioning the τ decay into leptonic and hadronic modes, and designing separate signal extraction methods that are optimised for each as opposed to employing a global strategy for all τ decays. Focusing only on the leptonic τ decays for now, upper limits on the $B^0 \to \ell^\pm \tau^\mp$ branching fraction are derived using a hybrid cut-and-count approach based on the Rolke-Lopez interval method. The results from just the leptonic τ decays have a sensitivity that is competitive with the current world-leading upper limits using only a fraction of the data.

Statement of Student Contribution

The analysis presented in this thesis was conducted within the Belle II Analysis Software Framework called `basf2` [1] which is developed and maintained by the Belle II collaboration, of which I am a member. All the Monte Carlo and data samples used for the work presented here were generated, processed and validated by multiple members of both the Belle II Data Production and Software groups. All the necessary corrections applied to data and Monte Carlo samples within this study were derived from individual studies conducted within the Belle II Performance group. Such studies were made by various members from multiple institutions, and in instances where public documentation of such studies exist, they are cited in this thesis. Large parts of this analysis were performed using computing resources from either the Belle II central computing system or grid computing system. These systems pool computing resources from nearly all Belle II institutions, and are developed and maintained by members of the Belle II Distributed Computing group in addition to local computing experts at each contributing institution. The following list includes all other specific contributions in more detail:

- Most of the plots presented here were made according to the Belle II plotting template. A python package that implements this template was privately developed by Dr. William Sutcliffe, Dr. Henrik Junkerkalefeld, Dr. Markus Prim and Dr. Maximilian Welsch [2].
- The lepton identification corrections applied for the data-MC agreement study presented in Section 6.4, in addition to the coverage plots shown in Appendix E, were applied using a privately developed python package by Dr. William Sutcliffe, Dr. Moritz Bauer and Dr. Michael Eliachevitch [3].
- The C++ module used to generate the EvtGen-skimmed Monte Carlo samples introduced in Section 6.5.2 was written by me but based on the pre-existing code for the `GeneratorPreselection` module in `basf2` [1].
- The E_{extra} cleanup classifier study presented in Chapter 5 was performed in collaboration with Dr. Racha Cheaib, Dr. Savino Longo and Kylie Amirie. My contributions to the work involved improving and extending an existing study designed by R. Cheaib and K. Amirie, and working with R. Cheaib and S. Longo to test the classifiers and solve issues related to their performance and optimisation. I was responsible for including the classifiers in `basf2` and maintaining them by providing regular updates to the classifier weights.

Aside from the contributions mentioned above, I wrote all other python scripts and Jupyter notebooks that were used to perform the work presented here, including specifically the following: event reconstruction and application of analysis cuts and vetoes; background estimations; likelihood fits and toy ensemble tests; optimisation studies; calculation of upper limits.

I certify that this thesis contains work carried out by myself except where otherwise acknowledged.

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Bruce and Chia-Ling, for their incredible support, immense patience, and invaluable guidance. Under your supervision, I have learned not only the intricacies of particle physics but also what it truly means to be a dedicated researcher. Although I have spent only a few years in this field, it is already clear to me how rare it is to have supervisors like you. To Bruce, thank you for giving me the opportunity to pursue a PhD with you. I can only hope that future students will recognise how fortunate they are to benefit from your mentorship. I would also like to extend my heartfelt thanks to my close friends, both from my university days and those I met along the way. Whether through catch-ups at the pub, board game nights, adventures around Japan, or cosy dinners, your companionship has kept me grounded and helped me stay sane throughout this journey. Above all, I am deeply grateful to my mother for always standing by my side and for working tirelessly to provide me with every opportunity under the Sun throughout my life. I definitely could not have made it to this moment without your constant support and love. To my brother, thank you for your continuous support and for sharing your pearls of PhD wisdom. Having now walked this path myself, I am even prouder of your own achievements. I would also like to thank my Pumpkin for your unwavering love, care, and encouragement. Your ability to lift my spirits with bike rides, walks up to the castle, and peaceful Sundays has been a priceless source of comfort. Lastly, I would like to acknowledge my late father, who always took pride in my academic accomplishments. You were there to see me start my academic journey seven years ago, and I wish you could have been here to see me reach this moment.

The research reported in this thesis was supported by the award of a Research Training Program scholarship to the PhD Candidate.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The theory of the Standard Model represents the culmination of several decades work from theoretical and experimental particle physicists who endeavoured to understand the foundations of matter. The theory describes the fundamental particles that constitute all visible matter, and their interactions with the fundamental forces. It is one of the most successful scientific theories to date, and since its completion in 1978, the Standard Model theory has withstood countless experimental tests and even predicted the existence of new particles long before they were discovered in particle colliders [4][5][6]. The most famous example of this was the discovery of the Higgs boson by the ATLAS and CMS experiments at the Large Hadron Collider in 2012. [7][8]. Though the Standard Model provides an expression for the Higgs boson mass in terms of the coupling parameter from the Standard Model scalar potential, the coupling parameter is free and therefore the Higgs boson mass is not predicted by theory. However, by fitting and consistency arguments, the range 116 - 127 GeV/c^2 can be obtained, which is consistent with the value measured by the ATLAS and CMS experiments [7][8]. Nevertheless, it is well-known that the Standard Model theory is incomplete. Its inability to account for dark matter or gravity, the source of neutrino masses, or the matter-antimatter asymmetry in the universe, to name just a few examples, provides strong indications that there is new physics beyond the Standard Model [9][10]. Numerous theories that extend the Standard Model by introducing new particles or interactions have been proposed, with many experiments today focused on searching for these new particles as well as providing measurements that test Standard Model predictions to detect where any new physics contributions might exist. Although no concrete evidence for new physics has been found thus far, certain precision measurements exhibit tension with Standard Model predictions. The most notable of these measurements include the anomalous muon magnetic moment [11][12] and lepton flavour universality tests [13].

A fertile area to search for new physics are rare decays as they are suppressed within the Standard Model and thus exhibit branching fractions of $\mathcal{O}(10^{-5})$ or smaller. This allows for any enhancements in the branching fraction as a result of new physics contributions to be more readily observed. The focus of this thesis is on $B^0 \to \ell^\pm \tau^\mp$ ($\ell = e, \mu$) which is an example of a charged lepton flavour violating decay, a class of decays that are yet to be observed. Albeit forbidden in the absence of neutrino masses, charged lepton flavour violating decays are allowed as a result of neutrino oscillations [14]. However, due to the neutrino mass scale, such decays are rare and expected to have incredibly small branching fractions, well below any current or future experimental sensitivity. Therefore, observation of any charged lepton flavour violating decay would only result from an enhanced branching fraction due to new physics contributions, and would thus be concrete evidence of physics beyond the Standard Model. Previous searches have been performed for $B^0 \to \ell^\pm\tau^\mp$ at a variety of particle colliders, with the current world-best values $\mathcal{B}(B^0 \to e^\pm\tau^\mp) < 1.6 \times 10^{-5}$ and $\mathcal{B}(B^0 \to \mu^\pm\tau^\mp) < 1.2 \times 10^{-5}$ coming from searches by the LHCb and Belle experiments respectively [15][16].

The search presented in this thesis is the first $B^0 \to \ell^\pm\tau^\mp$ search conducted at the Belle II experiment. An upgraded successor to the Belle detector, the Belle II detector is located at the SuperKEKB e^+e^- collider in Japan. It is purpose-built as a B-factory and features an extensive and diverse physics program, centred on precision measurements for flavour physics. Since commencing collisions in 2018, the Belle II detector has recorded an integrated luminosity of $\int \mathcal{L}dt$ = 504 fb^{-1}. An integrated luminosity between 10 – 50 ab^{-1} is anticipated to greatly improve the branching fraction upper limits for $B^0 \to \ell^\pm\tau^\mp$, with the potential to probe as low as $\mathcal{O}(10^{-7})$. Therefore, given the amount of data currently recorded, the purpose of the search presented in this thesis is to explore new avenues for improved sensitivity in $B^0 \to \ell^\pm\tau^\mp$ searches, and set the analysis foundations for future Belle II searches that will have access to significantly more data. A challenge for future $B^0 \to \ell^\pm\tau^\mp$ searches, and for other Belle II searches involving neutrinos, is the increase in background as SuperKEKB reaches higher instantaneous luminosities. This background directly degrades the power of a key analysis variable for these searches called E_{extra}. As this variable is used in the analysis presented in this thesis, a secondary goal was to develop E_{extra} multivariate classifiers that can be used to improve the robustness of this variable against higher backgrounds for future searches.

The layout of this thesis is as follows. Chapter 2 introduces the theoretical framework of the Standard Model, and briefly covers the theoretical concepts that pertain to flavour violation in the quark and lepton sectors. The latter half is focused on the $B^0 \to \ell^\pm\tau^\mp$ decay, its role in searching for new physics, and an overview of the methodologies and results of past experimental searches. Chapter 3 describes the setup of the SuperKEKB collider and the Belle II detector, and outlines the various particle reconstruction methods. Chapter 4 provides discussions on the techniques used for the E_{extra} classifier study and $B^0 \to \ell^\pm\tau^\mp$ analysis. Chapter 5 discusses the work performed to develop clean-up classifiers for the E_{extra} quantity. Chapter 6 presents the entirety of the $B^0 \to \ell^\pm\tau^\mp$ analysis, and concludes with the results on the branching fraction upper limit. Lastly, Chapter 7 is the conclusion, and includes an overview of points of further work for this $B^0 \to \ell^\pm\tau^\mp$ analysis.

All values and measurements presented in this thesis use natural units whereby $\hbar = c = 1$. Unless otherwise specified, charge-conjugates are implied.

Chapter 2

Theoretical Overview

The role of this chapter is to briefly introduce the theoretical foundations of the Standard Model, with a focus on the concepts from flavour physics that pertain to the work presented in this thesis. This is done in Section 2.1 and Section 2.2. Following this is a focused discussion on the $B^0 \to \ell^{\pm}\tau^{\mp}$ decay, in both the context of the Standard Model as well as beyond the Standard Model theories, in Section 2.3. Additionally, the section covers the current status of experimental searches for $B^0 \to \ell^{\pm}\tau^{\mp}$, and provides insight into the various analysis strategies used, with particular emphasis given to the experimental methodologies that are also employed in the search presented here.

2.1 Flavour in the Standard Model

In particle physics, the term flavour refers to the different types of quarks and leptons that exist in the Standard Model. The aim of flavour physics is to study processes that change the flavour of quarks and leptons, and as will be shown in Section 2.1.1, flavour physics played an instrumental role in shaping the Standard Model framework. To this day its importance remains, with flavour physics studies holding an integral place in the current particle physics landscape, by testing Standard Model expectations through precision measurements, and probing sectors that are promising areas of new physics.

2.1.1 Brief History of Flavour

Throughout the 1950s and early 1960s, the initial theoretical building blocks of the Standard Model were taking shape, and by the end of 1962, four fundamental particles, namely two generations of leptons, had already been discovered: the electron and electron neutrino, and the muon and muon neutrino [4]. In this sense, unbeknownst to particle physicists at the time, flavour was already part of the Standard Model. However, the idea of flavour was only formally introduced with the advent of the quark model, which was independently built by Murray Gell-Mann and George Zweig in 1964 [4][17]. This model was primarily made to explain the observed patterns and symmetries when organising the "zoo" of newly discovered hadrons using certain quantum numbers according to the Eightfold Way developed by Gell-Mann and Yuval Ne'eman in 1961 [18] [19]. The quark model described three new particles — the *up*, *down* and *strange* quarks — with the intention that all observed hadrons were in fact bound states of various combinations of these quarks, and not fundamental in nature as originally thought. Initially, the quark model was met with skepticism, and it was questioned if quarks are real particles, or whether the model was just a mathematical abstraction to help explain an array of unanswered questions. But in 1968, confirmation of the existence of *up* and *down* quarks came from deep inelastic scattering experiments conducted at the Standard Linear Accelerator Centre (SLAC) which showed point-like objects inside of protons as predicted by the quark model [20][21]. By extension, the existence of the *strange* quark was also implied, and in 1969, Gell-Mann won the Nobel prize for this model.

The fourth quark — the *charm* quark — was first theorised in 1970 by Sheldon Glashow, John Iliopoulos and Luciano Maiani as part of the Glashow-Iliopoulos-Maiani (GIM) mechanism theory [22]. Their proposed mechanism introduced a fourth quark to describe the observed suppression of decays involving flavour changes of the *strange* quark mediated by neutral currents, and helped explain why decays such as $K_L^0 \to \mu^+\mu^-$ were unobserved at the time despite its expected branching fraction [4]. The *charm* quark "fascinated and pleased" Glashow, Iliopoulos and Maiani as it created symmetry in the number of quarks and leptons known at the time, hence its name [4]. Four years after the GIM mechanism was proposed, the existence of the *charm* quark was experimentally verified by two independent experiments at SLAC and Brookhaven National Laboratory (BNL), both of whom discovered the first bound *charm* quark system, labelled as the J/ψ meson [23][24]. Following this, the GIM mechanism was accepted as part of the Standard Model framework. The J/ψ discovery, announced in November of 1974, sparked the November Revolution, a period in particle physics which yielded new particle discoveries and the exploration of additional quarks which would eventually lead to a completion of the Standard Model [4].

By the time of the November Revolution, two additional quarks were postulated to exist — the *bottom* and *top* quarks. This was a result of the theory proposed by Makoto Kobayashi and Toshihide Maskawa [25] in order to explain CP violation in the kaon system which was observed by James Cronin, Val Fitch, and colleagues in 1964 [26]. The Kobayashi-Maskawa (KM) theory involved extending an existing 2x2 matrix devised by Nicola Cabibbo that described the strength of transitions between *up*, *down* and *strange* quarks [27]. The extension of the Cabbibo matrix into a 3x3 matrix allowed for a parametrisation involving a complex phase which Kobayashi and Maskawa realised can explain the observed CP violation. Their theory grew in popularity after the discovery of the tau lepton (and its presumed neutrino) in 1975 by experimentalists at SLAC [28]. The discovery broke the symmetry in the number of quarks and leptons, a symmetry that could be restored should two additional quarks be discovered. In 1977, four years after Kobayashi and Maskawa published their theory, proton-nucleus collisions at Fermilab observed an Υ meson resonance which is a bound *bottom* quark state [4]. The discovery of the *bottom* quark led to an urgency to find the *top* quark, and a race between experiments at CERN and Fermilab began. However, due to its large mass, it took 18 years for the *top* quark to be discovered, with Fermilab winning the race and announcing the discovery in 1995 [29]. With this discovery, the flavour sector of the Standard Model was made complete.

Experiments studying flavour physics largely shifted focus to B mesons following the discovery of the *bottom* quark, in part to test the KM theory prediction that a certain amount of CP violation should exist in the B meson system. In the 1990s, two new detectors were built, the Belle detector located at the KEKB collider in Japan, and the BaBar detector located at the PEP-II collider at SLAC. Both of these experiments were specifically designed to allow for the measurement of CP violation in the B meson system [14]. In 2001, both the Belle and BaBar collaborations published results that not only confirmed the existence of CP violation in the B meson system, but that the amount of CP violation measured was within the constraints of the KM theory [30][31]. Kobayashi and Maskawa received the Nobel prize in 2008 in recognition of the instrumental role played by the KM theory in helping to shape the modern day Standard Model. Today, the legacy of past flavour physics experiments continues, with current experiments such as Belle II, BES-III and LHCb providing measurements from the *bottom* and *charm* quark sectors with unprecedented precision, and performing searches for any hint of physics beyond the Standard Model.

2.1.2 The Standard Model

The Standard Model (SM) theory is an example of a quantum field theory, with the SM Lagrangian density consisting of terms comprised of quantum fields which are operator-valued mathematical objects defined at every point in space-time. Particles in the SM are described using their own quantum field, with physical particles conceptualised as excitations of their respective field. Each field (or particle) is described using mass and energy terms in the SM Lagrangian, with all possible field interactions described using products of various fields. Figure 2.1 shows all the fundamental particles described by the SM Lagrangian. The fundamental forces are mediated by vector (spin-1) bosons which are otherwise known as gauge bosons as they stem from the SM Lagrangian being invariant under gauge (local phase) transformations. The two massless gauge bosons, the gluon (g) and photon (γ), mediate the strong and electromagnetic force respectively. As a result, the range of these forces is infinite, though in practice the strong force does not act over such a large range due to its confining nature. This arises from quantum chromodynamics (QCD) which is the theory describing the strong force [32]. The massive gauge bosons in the SM are the $W^{\pm}$ and Z^0 bosons which together mediate the weak force. In the SM theory, the electromagnetic and weak forces are unified and described by electroweak theory [33][34][35]. The remaining boson in the SM is the scalar (spin-0) Higgs boson which is responsible for giving mass to the massive fundamental particles. This is achieved through the Higgs mechanism [36][37] in the case of the $W^{\pm}$ and Z^0 bosons, while for the fermions, mass is achieved through the Yukawa couplings [32]. The fundamental constituents of all visible matter are fermions (spin-$\frac{1}{2}$ particles). All fermions interact with the weak force, though only certain subsets interact with the strong and electromagnetic forces. The fermions that interact with the strong force are called quarks (q) and there exists six quark flavours: up (u), down (d), charm (c), strange (s), top (t), bottom (b). All quarks have fractional electric charge, with the up, charm and top quarks having a charge of $+\frac{2}{3}$ while the down, strange and bottom quarks have $-\frac{1}{3}$ charge. As quarks are electrically charged, they also interact with the electromagnetic force. Quarks carry a strong charge called colour of which there are three, *red*, *blue* and *green*, which enables their coupling to the strong force. As no free quarks have ever been observed, it is implied, though never proven by QCD theory, that all free particles are required to be colour neutral. Quarks form bound systems which are collectively called hadrons, of which there are two types: mesons ($q\bar{q}$ systems) and baryons (qqq or $\bar{q}\bar{q}\bar{q}$ systems). Exotic kinds of mesons and baryons also exist in the form of tetraquarks ($qq\bar{q}\bar{q}$ systems) and pentaquarks ($qqqq\bar{q}$ systems) respectively. Leptons refer to the fermions that do not interact with the strong force. The leptons consist of six flavours, namely the electron (e), muon (μ) and tau (τ) along with their neutrino counterparts, the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). The charged leptons have an electric charge of -1 and therefore interact with the electromagnetic force, while the neutrinos are electrically neutral and only couple to the weak force, hence why they are incredibly difficult to detect. For example, the inclusive cross-section for neutrinos at energies of 1 GeV is $\approx 10^{-38}$ cm^2 [38]. For this reason, neutrinos produced in particle colliders go undetected and are often labelled as *missing energy*. Neutrinos are massless in the canonical SM theory, however experiments have since demonstrated this to be false, as discussed in Section 2.2. For all fermions, their antimatter counterparts (denoted $\bar{q}$ for antiquarks and $\bar{\nu}$ for antineutrinos) are identical aside from their quantum numbers (such as electric charge) which are oppositely signed. As depicted in Figure 2.1, the fermions are organised into three generations, with the first generation containing the lightest and most stable fermions, and the primary constituents of all visible matter in the universe. Conversely, the second and third generations are increasingly heavier and therefore less stable. Most notably, the top quark is the heaviest ($M = 172.69$ GeV) and has the shortest mean lifetime (0.5×10^{-24} s), making it the only quark that is unable to hadronise and form a bound state before it decays [39].

In the canonical SM theory, where neutrinos are massless, flavour violation only occurs in the quark sector and is mediated by the electroweak force, in contrast to the strong force which conserves flavour. Following spontaneous symmetry breaking through the Higgs mechanism [36][37], the $W^{\pm}$ bosons only couple to the fermions with negative chirality (left-handed fermions) which form the following doublets of the $SU(2)$ symmetry group

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L \begin{pmatrix} c \\ s' \end{pmatrix}_L \begin{pmatrix} t \\ b' \end{pmatrix}_L \begin{pmatrix} e \\ \nu_e \end{pmatrix}_L \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L$$

Though positive chirality (right-handed) charged fermions exist, they form singlets of the $SU(2)$ group and do not appear in the relevant interaction terms in the SM Lagrangian, while massless right-handed neutrinos do not exist in the SM theory [32]. In the doublets above, it should be noted that the *down-type* quark for each doublet contains their weak eigenstates (d', s', b') which is different to their mass

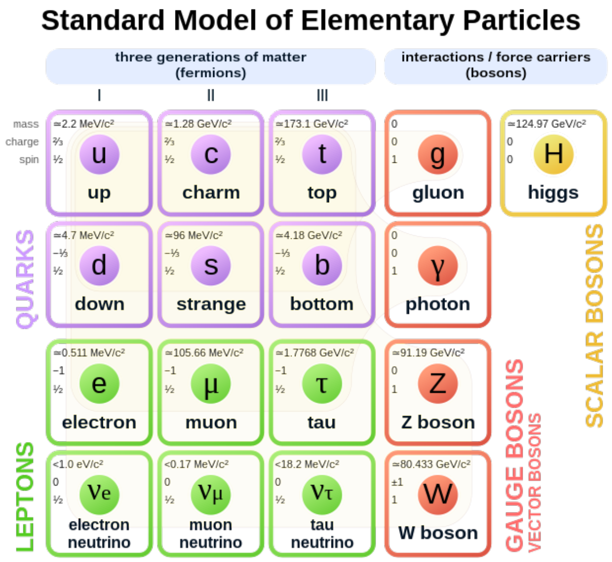


FIGURE 2.1. The fundamental particles in the Standard Model. The mass, charge and spin of each particle is provided. Note that the masses of the quarks are only approximate as none have ever been experimentally observed in isolation. [40]

eigenstates, that is their physical states (d, s, b). The different bases are related through the CKM quark-mixing matrix (V_{CKM}) given in Equation 2.1 where each element of the matrix $V_{ij} \in \mathbb{C}$ relates to the amplitude of the flavour transition between quarks i and j (the probability of the transition would be dependent on $|V_{ij}|^2$). The SM theory offers no prediction of what each individual CKM element should be so they must be experimentally measured, with the current experimental values provided in Equation 2.2. It should be noted that elements for transitions among *up-type* and *down-type* quarks (for example $c \to u$ transitions) do not exist in the CKM matrix.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.1}$$

$$|V_{CKM}| = \begin{pmatrix} 0.97435 & 0.22500 & 0.00369 \\ 0.22486 & 0.97349 & 0.04182 \\ 0.00857 & 0.04110 & 0.999118 \end{pmatrix} \tag{2.2}$$

From KM theory, the CKM matrix can be parameterised using three mixing angles, and a complex phase that is responsible for CP violating processes in the SM [25]. The SM requires the CKM matrix to be unitary, from which the constraints $\sum_i V_{ij}V_{ik}^* = \delta_{jk}$ and $\sum_j V_{ij}V_{jk}^* = \delta_{ik}$ can be derived. Here,

i, j, k are flavours of quarks, and δ_{jk}, δ_{ik} are the Kronecker-Delta functions. From these summations, six null combinations are obtained, each defining a unitarity triangle in the complex plane. The most commonly referred to equation is given in Equation 2.3, which corresponds to the unitarity triangle shown in Figure 2.2.

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{2.3}$$

The SM predicts the CKM matrix to be unitarity, with current experimental results agreeing with the prediction within uncertainties [41]. Precise measurements of each CKM matrix element remains a key aspect of flavour physics experiments, as it allows for the amount of CP violation predicted by the SM to be verified.

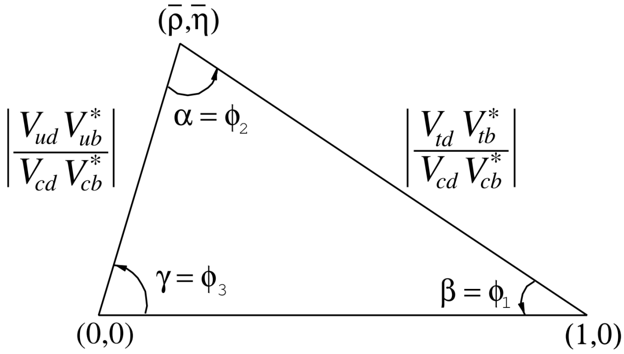


FIGURE 2.2. The unitarity triangle in the complex $(\bar{\rho}, \bar{\eta})$ plane drawn according to Equation 2.3. The parameters $\bar{\rho}$ and $\bar{\eta}$ arise from a parametrisation of the CKM matrix which can be found in detail in [41].

Mixing in the quark sector allows for additional tests of the SM based on lepton flavour universality. This idea stems from the couplings between the weak force and each of the left-handed lepton doublets of $SU(2)$. The SM predicts equal couplings among all the doublets, and this is commonly tested by measuring ratios such as $R(D)$ and $R(D^*)$ which are both defined according to Equation 2.4 where $\ell = e, \mu$. Currently, measurements of these two ratios yield a combined tension around 3.2σ with the SM [42].

$$R(D^{(*)}) = \frac{\mathcal{B}(\overline{B} \to D^{(*)}\tau^-\bar{\nu}_\tau)}{\mathcal{B}(\overline{B} \to D^{(*)}\ell^-\bar{\nu}_\ell)} \tag{2.4}$$

On the other hand, other anomalies involving different ratios such as $R(K)$, which is defined using Equation 2.4 but with K mesons instead of $D^{(*)}$ mesons, have since been resolved as a result of improved analysis strategies [43]. Despite this, lepton flavour universality remains an area of immense interest in the community.

2.2 Lepton Flavour Violation

As mentioned in Section 2.1.2, the SM allows for quark flavour violation, with the flavour mixing of quarks described by the CKM matrix. Conversely, in the lepton sector, the SM conserves individual lepton numbers L_e, L_μ and L_τ, thereby forbidding all lepton flavour violation (LFV) processes. This conservation follows from the assumption that neutrinos are massless. However, this is now shown to be false, with a series of experimental results observing neutrino mixing which demonstrates a non-zero mass difference between different pairs of neutrinos. It proves that neutral lepton flavour violating (NLFV) processes are in fact allowed. Such knowledge implies the existence of charged lepton flavour violating processes (CLFV), however CLFV has yet to be observed.

Neutrino mixing was first hinted at by the initial results of the Homestake experiment in 1968 which showed an unexpected deficit in solar neutrinos [44]. This became known as the solar neutrino problem. For the next 30 years, further results from the Homestake experiment in addition to the results of other experiments built to investigate this problem continued to show a solar neutrino deficit. It was not until 1998 that real-time solar neutrino measurements from the Super Kamiokande experiment provided the first evidence of neutrino mixing by measuring a deficit in the expected number of muon neutrinos from the atmosphere [45]. The deficit was consistent with theoretical predictions for mixing between the muon and tau neutrino flavours. Shortly after, in 2001, the SNO experiment verified the findings of Super Kamiokande, and provided evidence of mixing involving the electron neutrino flavour as well, thereby solving the solar neutrino problem [46]. In 1962, before any such experimental evidence, neutrino mixing was theorised to exist by Bruno Pontecorvo [47]. His idea was later revived by Ziro Maki, Masami Nakagawa, and Shoichi Sakata who made a mixing matrix for neutrinos analogous to the CKM matrix called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [48]. The matrix relates the mass eigenstates (ν_1, ν_2, ν_3) of the neutrino with its flavour eigenstates $(\nu_e, \nu_\mu, \nu_\tau)$ as shown in Equation 2.5.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{2.5}$$

Each component $U_{\alpha i}$ in the PMNS matrix represents the amplitude of each mass eigenstate $i = 1, 2, 3$ that is used to define a given flavour eigenstate $\alpha = e, \mu, \tau$. The PMNS matrix can be used to derive the probability of a neutrino oscillating from one flavour to another, with the equation for two-neutrino oscillation given in Equation 2.6 [32].

$$P_{\alpha\rightarrow\beta,\alpha\neq\beta} = \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2 L}{4E}\right) \tag{2.6}$$

Here, α and β each denote one of the neutrino flavours, while θ is the mixing angle between the two neutrino flavours, L is the oscillation distance and E is the neutrino energy. Present in the equation is the Δm^2 term which is the mass difference between the α and β flavour states. Therefore, by measuring neutrino oscillations between each pair of flavours, the mass differences between each generation can be deduced.

Neutrino oscillations violate lepton flavour as individual lepton numbers are not conserved. An example demonstrating this is the $\nu_\mu \to \nu_\tau$ oscillation where $L_\mu = 1$ and $L_\tau = 0$ prior to the oscillation, but after $L_\mu = 0$ and $L_\tau = 1$. There are two different types of neutrinos that would account for neutrino masses in the SM: Dirac neutrinos or Majorana neutrinos. Although the exact theory behind each is not pertinent to discussions here, it is relevant to note that processes involving Dirac neutrinos conserve the *total* lepton number $L = L_e + L_\mu + L_\tau$, while Majorana neutrinos do not, and such neutrinos would therefore yield different NLFV processes than their Dirac counterparts. Though it is not yet known if neutrinos are of the Dirac or Majorana type, there is hope that studying processes such as neutrino-less double β decay can help elucidate the answer.

NLFV in the form of neutrino mixing implies that flavour violation in the charged lepton sector must also be present. This is due to the electroweak coupling between neutral and charged leptons. However, due to the incredibly small masses of neutrinos, the resulting Yukawa couplings are $\mathcal{O}(10^{-12})$ which leads to SM branching fractions for CLFV processes that are beyond any current or future experimental sensitivity (an example would be $\mu^- \to e^-\gamma$ which has a predicted branching fraction of 10^{-54} in the SM). Therefore, any observation of CLFV is clear and unambiguous evidence of new physics contributions.

2.3 The $B^0 \to \ell^\pm\tau^\mp$ Decay

The decay $B^0 \to \ell^\pm\tau^\mp$ ($\ell = e, \mu$) is an example of a CLFV decay as it does not conserve the individual lepton numbers L_ℓ ($\ell = e, \mu$) and L_τ. It is worth noting that despite this, the total lepton number is conserved, and so the decay can proceed with both Dirac and Majorana neutrino types. In principle, $B^0 \to \ell^\pm\tau^\mp$ can occur through neutrino mixing as depicted in the Feynman diagram in Figure 2.3. However, as discussed in Section 2.2, the expected branching fraction is suppressed as it is dominated by the scale of the neutrino-mass differences Δm^2 (see the Equation 2.6). Contributions to the branching fraction from physics beyond the SM allow the decay rate to reach potentially measurable rates at B-factories such as SuperKEKB, hence the interest in this mode. The situation is improved with the presence of a tau lepton in the final state as opposed to a lighter lepton (an example of this would be $B^0 \to e^+\mu^-$) as it naturally leads to higher branching fractions. This is because the decay rates for such processes are proportional to m_ℓ^2 ($\ell = e, \mu, \tau$), even in the context of new physics (see Equation 2.7 and 2.8). Therefore, $B^0 \to \ell^\pm\tau^\mp$ presents as a useful opportunity to search for new physics, with its sensitivity to new physics discussed in Section 2.3.1. Searches for $B^0 \to \ell^\pm\tau^\mp$ have been ongoing at various particle collider experiments since 1994. With advancements in collider technologies, and increased data, the upper limit on the branching fraction for $B^0 \to \ell^\pm\tau^\mp$ has been made increasingly stringent with each experiment. An overview of the past experimental searches is given in Section 2.3.2 and is needed to provide context for the Belle II search presented in this thesis.

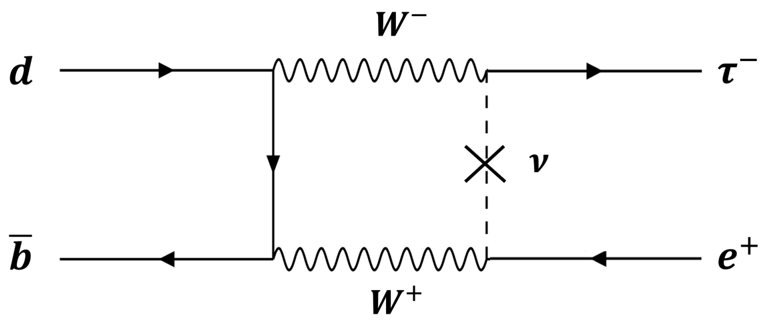


FIGURE 2.3. Feynman diagram for a $B^0 \to e^+\tau^-$ decay proceeding through a neutrino oscillation. The cross indicates the oscillation of neutrino flavour. The internal fermion line on the left of the box represents any of the up-type quarks.

2.3.1 Sensitivity For New Physics

Some beyond the SM theories that lead to enhanced branching fractions for $B^0 \to \ell^\pm\tau^\mp$ include multi-Higgs models [49], generic supersymmetric models [50], Z' models [6] and models involving leptoquarks [51] to name a few [6]. As not all models have been studied in regards to $B^0 \to \ell^\pm\tau^\mp$, what follows is a discussion of two studies that specifically discuss the role of leptoquarks and supersymmetry in the $B^0 \to \ell^\pm\tau^\mp$ decay.

In a 2018 paper by Alexander Smirnov, enhancements to $B^0 \to \ell^\pm\tau^\mp$ by models involving vector leptoquarks were studied [52]. Leptoquarks are hypothetical bosons that couple to both leptons and quarks, and can thus mediate lepton-quark transitions. As such, they carry both baryon and lepton numbers, and are assigned a fermion number equal to $F = 3B + L$ where B is the baryon number and L is the lepton number of the leptoquark [51]. Given the small amount of existing quark and lepton fields in the SM, the possible types of leptoquarks needed to explain all the couplings can be organised and classified. Such a classification is the Buchmüller-Rückl-Wyler classification which lists 10 leptoquarks divided into 5 scalar leptoquarks (spin-0) and 5 vector leptoquarks (spin-1) [53]. The study from Smirnov looks at the role of vector leptoquarks that originate from the Pati-Salam quark-lepton unification theory [54], which treats lepton number as a fourth colour charge for quarks [55]. Vector leptoquarks lead to enhanced branching fractions for $B^0 \to \ell^\pm\tau^\mp$ as their coupling to both quark and lepton fields allow the decay to proceed by tree-level processes (see Figure 2.4). Furthermore, if composite leptoquarks from the Pati-Salam $SU(4)$ group are considered, an additional increase in the branching fraction is expected as such leptoquarks predominantly couple to tau leptons compared to electrons and muons [52]. Direct searches for leptoquark production are carried out by the ATLAS and CMS experiments at the Large Hadron Collider, with results of their most recent searches found in [56] and [57]. The equation for the branching fraction of $B^0 \to \ell^\pm\tau^\mp$ in the context of a vector leptoquark model as described in [52] is given in Equation 2.7.

$$\mathcal{B}(B^0 \to \ell^\pm\tau^\mp) = \frac{m_B \pi \alpha_{st}^2(M_c) f_B^2 \overline{m}_B^2 (R_B^V)^2}{2 m_V^4 \Gamma_B^{\text{tot}}} \left(1 - \frac{m_\tau^2}{m_B^2}\right)^2 \beta_{B,\ell\tau}^2 \tag{2.7}$$

In the equation, m_V is the mass of the vector leptoquark, m_τ is the tau mass, m_B is the mass of the B^0 meson while $\overline{m}_B^2 = m_B^2/(m_d + m_b)$ where m_d and m_b are the masses of the d and b quarks respectively. Γ_B^{tot} is the total width of the B^0 meson while f_B^2 is the standard B^0 form factor that accounts for the matrix elements of the quark currents. The term $\alpha_{st}^2(M_c)$ is the strong coupling constant at the mass scale M_c where symmetry breaking of the Pati-Salam $SU(4)$ group occurs. The term R_B^V accounts for corrections due to the strong interaction between the vector leptoquark and b and d quark currents, while $\beta_{B,\ell\tau}^2$ is the mixing factor which is related to the mixing matrices that appear in the interaction term between the b and d quarks, and vector leptoquarks, in the Lagrangian for this new physics model. The vector leptoquark model discussed in [52] predicts the branching fractions to be $\mathcal{B}(B^0 \to e^\pm\tau^\mp) = 1.6 \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^\pm\tau^\mp) = 4.4 \times 10^{-9}$.

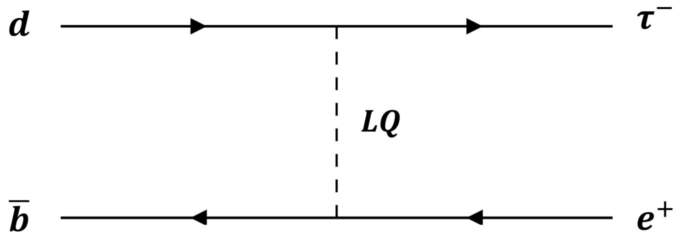


FIGURE 2.4. Feynman diagram for $B^0 \to e^+\tau^-$ based on a vector leptoquark model. The dotted propagator, labelled as *LQ*, indicates the vector leptoquark.

Enhancements in the branching fraction is also achievable by supersymmetric seesaw models as shown in a 2002 paper by Athanasios Dedes, John Ellis and Martti Raidal [58]. Supersymmetric models introduce new particles (*sparticles*) that act as partners to the existing SM particles, with SM fermions acquiring a SUSY boson partner (*bosinos*), and SM bosons partnered with SUSY fermions (*sfermions*). A popular SUSY framework is the minimal supersymmetric SM (MSSM) theory which seeks to consider only the minimal amount of SUSY particles required to maintain phenomenological consistency [59]. A notable aspect of SUSY models is the introduction of additional Higgs boson particles to make two Higgs doublets. These new Higgs particles can act as mediators for CLFV processes in double-penguin processes as shown in Figure 2.5, thereby increasing their branching fractions, particularly for final states involving third generation fermions which have larger Yukawa couplings. In general, SUSY has the advantage of allowing CLFV to occur within the *slepton* sector, making them no longer suppressed by the neutrino mass scale [58]. The Dedes, Ellis and Raidal paper derives the branching fraction equation for $B^0 \to \ell^\pm\tau^\mp$ in the context of a MSSM theory. The equation is given in Equation 2.8.

$$\begin{aligned}\mathcal{B}(B^0 \to \ell\tau) = & \frac{G_F^4 M_W^4}{8\pi^5}|V_{tb}^* V_{ts}|^2 M_B^5 f_B^2 \tau_B \left(\frac{m_b}{m_b + m_d}\right)^2 \\ & \times \sqrt{\left[1 - \frac{(m_\ell + m_\tau)^2}{M_B^2}\right]\left[1 - \frac{(m_\ell - m_\tau)^2}{M_B^2}\right]} \\ & \times \left\{\left(1 - \frac{(m_\ell + m_\tau)^2}{M_B^2}\right)|c_S^{\ell\tau}|^2 + \left(1 - \frac{(m_\ell - m_\tau)^2}{M_B^2}\right)|c_P^{\ell\tau}|^2\right\} \end{aligned} \quad (2.8)$$

In this equation, G_F is the Fermi coupling constant, M_W is the mass of the W boson, V_{ij} are the elements of the CKM matrix while M_B, f_B and τ_B are the mass, decay constant and lifetime of the B^0 meson respectively. The mass of the b and d quarks are given by m_b and m_d respectively while m_τ is the mass of the tau and m_ℓ is the mass of the lepton corresponding to the decay channel $B^0 \to \ell^\pm\tau^\mp$ ($\ell = e, \mu$). The form factors that contain the new physics contributions are in $|c_S^{\ell\tau}|$ and $|c_P^{\ell\tau}|$. An important parameter for SUSY seesaw models is $\tan\beta$ which is the ratio of the values for the vacuum expectation value of the two Higgs doublets [59]. The study in [58] sets $\tan\beta$ set to 60, which is the maximum value that allows for electroweak symmetry breaking. Based on this, the maximum allowable branching fraction for $B^0 \to \ell^\pm\tau^\mp$ was calculated to be $\mathcal{O}(10^{-10})$.

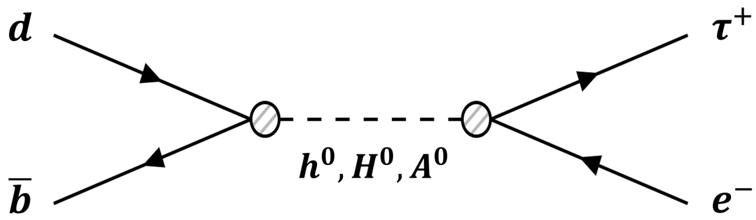


FIGURE 2.5. Feynman diagram for $B^0 \to e^-\tau^+$ within a MSSM theory. The h^0, H^0 and A^0 indicate the new Higgs particles that are part of the two Higgs doublets in SUSY theories such as the MSSM.

2.3.2 Past Experimental Searches

Four prior searches for $B^0 \to \ell^\pm\tau^\mp$ were conducted at various particle colliders, three of which were performed at e^+e^- colliders, while the other was performed at the Large Hadron Collider. As no significant signal was observed for all searches, upper limits on the branching fractions of $B^0 \to \ell^\pm\tau^\mp$ were calculated. A comparison of the upper limits calculated from each study is provided in Table 2.1. Currently, the best branching fraction upper limits for $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ were achieved by the Belle and LHCb searches respectively.

Searching for $B^0 \to \ell^\pm\tau^\mp$ is experimentally challenging as the decay of the τ produces at least one missing neutrino, making it difficult to impose kinematic constraints that help isolate signal events.

TABLE 2.1. Summary of experimental results for past $B^0 \to \ell^\pm\tau^\mp$ searches. The upper limits on the branching fractions for $B^0 \to \ell^\pm\tau^\mp$ are all calculated at the 90% confidence level. No $N_{B\overline{B}}$ number is quoted for LHCb as they do not produce B mesons in pairs, and the cross section for B meson production is not stated to allow for N_B to be quoted.

	CLEO (2004) [60]	**BaBar (2007) [61]**	**LHCb (2019) [15]**	**Belle (2021) [16]**
Sample Size	9.2 fb^{-1}	342 fb^{-1}	3 fb^{-1}	711 fb^{-1}
$\boldsymbol{N_{B\overline{B}}}$	9.6×10^6	378×10^6	–	772×10^6
$\mathcal{B}(B^0 \to e^\pm\tau^\mp)$	$< 1.1 \times 10^{-4}$	$< 2.8 \times 10^{-5}$	–	$< 1.6 \times 10^{-5}$
$\mathcal{B}(B^0 \to \mu^\pm\tau^\mp)$	$< 3.8 \times 10^{-5}$	$< 2.2 \times 10^{-5}$	$< 1.2 \times 10^{-5}$	$< 1.5 \times 10^{-5}$

Although the situation is particularly problematic for leptonic τ decays, which are three-body decays producing two neutrinos, the presence of an additional energetic lepton in the final state is useful for identifying signal and suppressing backgrounds. Alternatively, hadronic τ decays produce only one neutrino thereby providing more kinematic information, but suffer from significant backgrounds from other B decays that generate hadrons. A list of the leptonic and hadronic τ decay modes that were used in three of the prior searches is provided in Table 2.2.

TABLE 2.2. Table of τ^- decay modes considered by past experimental searches for $B^0 \to \ell^\pm\tau^\mp$, along with their branching fractions [41].

	$\boldsymbol{\tau^-}$ **Decay**	**Branching Fraction (%)**
Leptonic	$e^-\overline{\nu}_e\nu_\tau$	17.82 ± 0.04
	$\mu^-\overline{\nu}_\mu\nu_\tau$	17.39 ± 0.04
Hadronic	$\pi^-\nu_\tau$	10.82 ± 0.05
	$\pi^-\pi^0\nu_\tau$	25.49 ± 0.09
	$\pi^-\pi^0\pi^0\nu_\tau$	9.26 ± 0.10
	$\pi^-\pi^-\pi^+\nu_\tau$	9.31 ± 0.05

To combat the lack of kinematic constraints, studies performed by e^+e^- collider experiments utilised the method of hadronic tag reconstruction. At such colliders, a pair of B mesons were produced at the interaction point from an Υ(4S) meson. While one B is reconstructed in the decay mode(s) of interest (called the signal B or B_{sig}), the companion B (referred to as the tag B or B_{tag}) is reconstructed in various hadronic decay modes which produce no neutrinos. The advantage of this is a complete reconstruction of the B_{tag} kinematics given the clean environments of e^+e^- colliders, and the hermetic detectors they use. From the reconstructed B_{tag} 4-vector, the B_{sig} 4-vector can be easily inferred, given that the initial state is known for e^+e^- collisions. A diagram depicting this hadronic tagging method is given in Figure 2.6. In the context of $B^0 \to \ell^{\pm}\tau^{\mp}$ studies, where B_{sig} is reconstructed as $B^0 \to \ell^{\pm}\tau^{\mp}$, hadronic tagging is extremely beneficial as it allows for the B_{sig} rest frame to be recovered. The B daughter lepton can be boosted into the B_{sig} frame, producing a sharply peaked distribution for signal events around the nominal momentum value of 2.34 GeV which can then be used for signal extraction (see Figure 2.7). Furthermore, the combined 4-vector of the missing neutrinos can also be inferred, and selections on the missing energy and direction of missing momentum can be used as background suppression tools. A disadvantage of the hadronic tagging method is the low reconstruction efficiency which is around $\mathcal{O}(0.1\%)$. The impact of this can be mitigated through larger data samples, or to some extent, by using quantities that can achieve good signal-background discrimination with loose selection criteria.

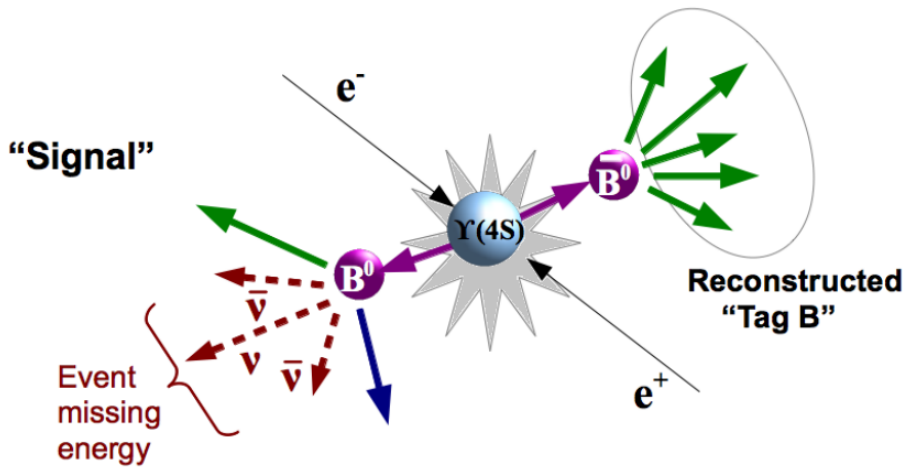


FIGURE 2.6. Diagram of $B\overline{B}$ production in an e^+e^- collider. Decays on the signal side produce neutrinos which are labelled as the missing energy. The green and blue arrows represent the possible hadrons and leptons produced from the B decays [14].

The CLEO experiment which was at the CESR e^+e^- collider first published a search for $B^0 \to \ell^{\pm}\tau^{\mp}$ in 1994 before redoing the search with their full dataset of 9.2 fb^{-1} (9.6×10^6 $B\overline{B}$ pairs) and publishing the result in 2004 [60]. The CLEO study reconstructed the B_{tag} using the hadronic tag method, and considered only the leptonic τ decay modes to achieve a purer sample. The analysis strategy consisted of implementing various signal selections based on the momentum of the B and τ daughter leptons in the B_{sig} frame, the energy of the missing neutrinos, and the quality of the reconstructed B_{tag}. The suppression of background from other B decays and $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) processes were achieved by two separate neural nets. The expected number of background events was estimated from simulated events, and following the application of the selection cuts on data, CLEO observed 39 events, with a total of 49.4 events expected from background. The published upper limits on the branching fractions, calculated at a 90% confidence level, were $\mathcal{B}(B^0 \to e^\pm\tau^\mp) < 1.1 \times 10^{-4}$ and $\mathcal{B}(B^0 \to \mu^\pm\tau^\mp) < 3.8 \times 10^{-5}$.

In 2007, a search was performed using data collected by the BaBar detector which was at the asymmetric e^+e^- PEP-II collider [61]. The search used 342 fb^{-1} of collected data, which corresponded to 378×10^6 $B\overline{B}$ pairs, more than 40 times greater than the data sample collected by CLEO. All the τ decay modes presented in Table 2.2 were utilised, with the specific hadronic decay modes used for the B_{tag} reconstruction listed in [61]. For the BaBar study, the signal was extracted using maximum likelihood fits to the momentum of the B_{sig} daughter lepton boosted into the B_{sig} frame. The fitted distributions of this quantity from the study are provided in Figure 2.7.

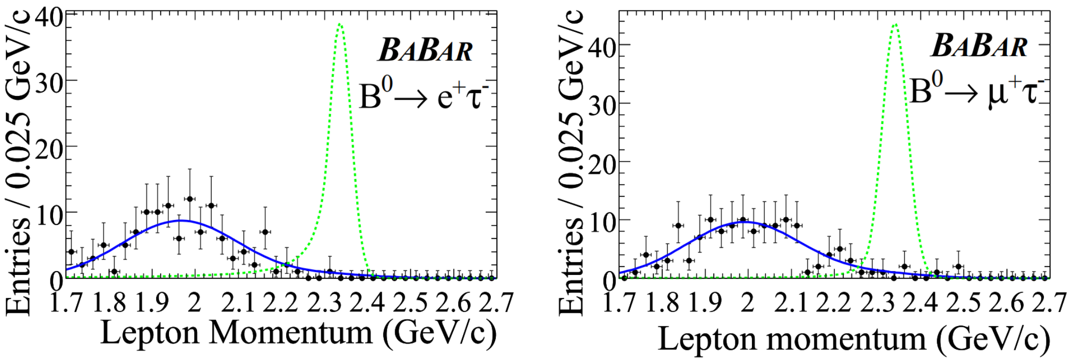


FIGURE 2.7. From the BaBar $B^0 \to \ell^\pm\tau^\mp$ study, distributions of the B_{sig} daughter lepton momentum boosted into the B_{sig} frame following an unbinned maximum likelihood fit to the data (represented with black dots) [61]. The green distribution shows the expected signal shape from simulation, while the blue curve are the simulated background events.

To suppress the background, in addition to requirements on the B_{tag} quality, τ mode-dependent selections were imposed on the direction and magnitude of missing momentum, and the number of particles detected. Additionally, selections were placed on the residual energy measured by the electromagnetic calorimeter E_{extra}, and ΔE_τ which is the difference between the reconstructed and nominal τ energy in the τ rest frame (this frame was determined by using the B daughter lepton 4-vector in the B_{sig} frame). The ΔE_τ quantity is calculated using Equation 2.9, where the sum is over the pion daughters of the τ, the momentum of the neutrino is estimated as $p_\nu \approx |\sum_i \vec{p_\pi}|$ in the B_{sig} frame, and 1.777 is the nominal τ mass.

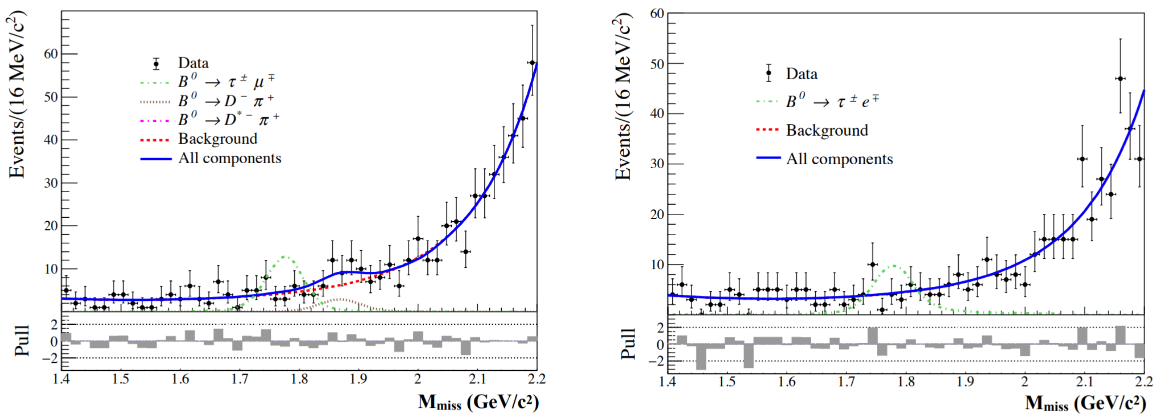
$$\Delta E_\tau = \sum_i E_{\pi_i} + p_\nu - 1.777\ \text{GeV} \tag{2.9}$$

Both E_{extra} and ΔE_τ provided substantial signal-background separation, and as such, these quantities are extensively used in the analysis presented in this thesis. From the fit, the BaBar search found no evidence of any signal, and set 90% confidence level branching fraction upper limits of $\mathcal{B}(B^0 \to e^\pm\tau^\mp) < 2.8 \times 10^{-5}$ and $\mathcal{B}(B^0 \to \mu^\pm\tau^\mp) < 2.2 \times 10^{-5}$.

The best branching fraction upper limit for $B^0 \to e^\pm\tau^\mp$ comes from a 2021 study using data collected by the Belle detector which was at the e^+e^- KEKB collider [16]. In total, 703 fb^{-1} of data was used, which equals around 772×10^6 $B\overline{B}$ pairs, twice the size of the data used by BaBar. In order to combat the decreased reconstruction efficiency due to difficulties in τ reconstruction, this analysis omitted the τ reconstruction, thereby effectively performing an inclusive τ reconstruction. The B_{tag} was reconstructed hadronically using a `NeuroBayes`-based algorithm designed for Belle [62]. The signal was extracted using a maximum likelihood fit to the M_{miss} distribution which is defined according to Equation 2.10 where $(E_{B_{\text{sig}}}, \vec{p}_{B_{\text{sig}}})$ and $(E_\ell, \vec{p}_\ell)$ are the 4-momenta of the B_{sig} and B-daughter lepton respectively calculated in the centre of mass frame.

$$m_{\text{miss}} = \sqrt{(E_{B_{\text{sig}}} - E_\ell)^2 - (\vec{p}_{B_{\text{sig}}} - \vec{p}_\ell)^2} \tag{2.10}$$

The m_{miss} quantity is expected to peak around the τ mass. Figure 2.8 below provides the fitted M_{miss} distribution from the Belle study.



(A) Distribution for $B^0 \to e^\pm\tau^\mp$ events

(B) Distribution for $B^0 \to \mu^\pm\tau^\mp$ events

FIGURE 2.8. Fitted missing mass distributions from the Belle $B^0 \to \ell^\pm\tau^\mp$ study. The data is indicated using black dots, while the green dashed line shows the signal shape from simulation. The dashed red line shows the combinatorial background shape, while the peaking background component for $B^0 \to D^{*-}\pi^+$ is shown in dotted brown. The overall fit result is shown in blue. The subplots underneath each distribution show the residuals divided by the errors.

As can be seen, although the inclusive τ approach yields a higher efficiency, larger backgrounds are observed in the signal extraction region. In particular, for $B^0 \to \mu^\pm\tau^\mp$, peaking backgrounds from $B^0 \to D^-\pi^+$ and $B^0 \to D^{*-}\pi^+$ events caused by the mis-identification of pions as muons are present. Such peaking contributions were studied using a control sample, and were included as additional components in the likelihood fit. The branching fraction upper limits at a 90% confidence level set by the Belle study were $\mathcal{B}(B^0 \to e^\pm\tau^\mp) < 1.6 \times 10^{-5}$ and $\mathcal{B}(B^0 \to \mu^\pm\tau^\mp) < 1.5 \times 10^{-5}$.

The 2019 result by published by the LHCb experiment at the LHC provides the best upper limit for the $B^0 \to \mu^\pm\tau^\mp$ mode [15]. As the LHC is a proton-proton collider, the initial state is unknown, and the simultaneous overlay of multiple proton-proton collisions means a full reconstruction of events is unfeasible. As a result, a different analysis strategy was required compared to the e^+e^- collider searches. To avoid the loss of kinematic information in leptonic τ decays, the τ was reconstructed hadronically. In particular, only the $\tau \to \pi\pi\pi\nu$ mode was reconstructed so that the τ decay vertex could be constrained and determined with high precision. Using this information, and the reconstructed signal B daughter lepton 4-vector, the invariant mass of the signal B meson could be determined. The invariant mass distribution was used to extract the signal, and as no signal was found, an upper limit on the branching fraction at both 90% and 95% confidence levels for $B^0 \to \mu^\pm\tau^\mp$ were set at $< 1.2 \times 10^{-5}$ and $< 1.4 \times 10^{-5}$ respectively.

CHAPTER 3

The Belle II Experiment

The Belle II detector is located in the SuperKEKB collider at the High Energy Accelerator Research Organisation facility (also known as KEK) in Tsukuba, Japan. The Belle II detector is an upgrade of its predecessor Belle which operated between 1999 and 2010 at the KEKB collider. Both KEKB and SuperKEKB were designed to mass produce tens and hundreds of millions of B mesons, respectively, hence why they are referred to as B-factories. Belle II is planning on collecting 50 times the amount of data recorded by Belle, thus improving upon the prior statistical limitations of the Belle experiment [63]. With such data, the Belle II experiment will aim to provide leading measurements of key CKM parameters $|V_{\text{ub}}|$ and $|V_{\text{cb}}|$, accurately measure the amount of CP asymmetry in the B meson system, and precisely determine the CP violation unitarity angles ϕ_1, ϕ_2 and ϕ_3. Tests of lepton flavour universality and searches for lepton flavour violating decays are also a key component of the Belle II physics program, and will require studies into semileptonic, leptonic and radiative B decays [64]. In most cases, the B decay of interest is rare or forbidden in the Standard Model, with expected branching fractions $\mathcal{O}(10^{-6})$ or smaller, thus the motivation for building a B-factory [14][41]. In addition to the copious production of B mesons, the design of the SuperKEKB collider further allows the mass production of charm hadrons and τ leptons, which extends the diversity of the physics program by allowing for studies into $c\bar{c}$ spectroscopy, CP violation in the D meson system, and lepton flavour violation in the τ sector, to name a few examples. In addition, the Belle II experiment conducts dark sector studies such as those involving dark photons, axions, or long-lived particles that could be dark matter candidates [64].

Operations at SuperKEKB began in 2016, with the first collisions occurring in 2018. Since then, the Belle II detector has recorded an integrated luminosity of $\int \mathcal{L} dt = 504$ fb^{-1} of data which is more than half the total data recorded by Belle (988 fb^{-1}), and similar to the total data recorded by BaBar (513 fb^{-1}) [14]. Of this subset, 365 fb^{-1} of data was recorded on the Υ(4S) resonance [65]. From July 2022, Belle II and SuperKEKB entered a long shutdown to undergo a variety of upgrades, the most notable being the upgrade of the pixel detector [66], with an additional two long shutdowns planned during the expected lifetime of the experiment. The goal of these shutdowns is to progressively implement upgrades that will allow Belle II and SuperKEKB to achieve their design performances. Just before the long shutdown commenced, SuperKEKB was able to reach an instantaneous luminosity of 4.7×10^{34} cm^{-2} s^{-1} which is a world record. Since January 2024, collisions have recommenced at SuperKEKB and Belle II has resumed data-taking, with SuperKEKB reaching a peak instantaneous luminosity of 4.08×10^{34} cm^{-2} in its first run period since the long shutdown.

As a collaboration, the Belle II experiment is composed of over 1100 physicists and technical staff originating from 124 institutions from 26 countries, making it among the largest scientific collaborations in the world [64].

3.1 The SuperKEKB Collider

The SuperKEKB collider complex consists of an electron and positron ring, both 3 km in circumference, a positron damping ring, and an electron-positron injector linac. Figure 3.1 shows a schematic diagram of the complex and the location of the Belle II detector which surrounds the interaction point located in the Tsukuba experiment hall.

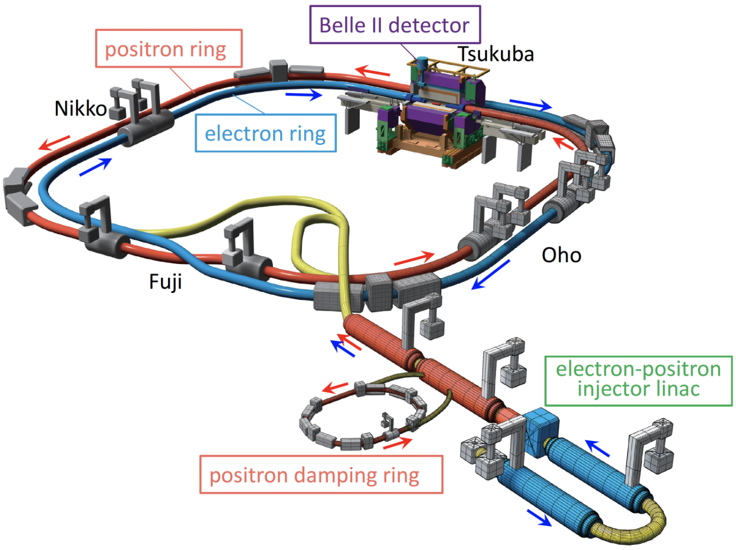


FIGURE 3.1. Schematic diagram of the SuperKEKB collider complex. The four experimental halls, Tsukuba, Nikko, Fuji, and Oho, are also labeled [67].

3.1.1 Linac and Damping Ring

Originally built for use at the Photon Factory at KEK in 1982, the 600 m long electron-positron injector linac has undergone multiple significant upgrades in the past to deal with the requirements of KEKB and SuperKEKB. In order to maximise the instantaneous luminosity of SuperKEKB (see Equation 3.2), the layout of the linac is designed to achieve electron and positron beams that have a high current and minimal transverse beam emittance. In addition, the linac is also capable of performing top-up injections in four storage rings and the positron damping ring simultaneously. Within the linac, a photo-cathode radio frequency (RF) gun is used to source high current and low emittance electron beams. A separate thermionic gun generates high-charge electron bunches that are used to irradiate a tungsten target and produce Bremsstrahlung interactions that result in positron production. Flux concentrators and large-aperture accelerating machinery are placed around the tungsten target to enhance the positron yield. Following production, the positron beams are then sent through the damping ring in order to reduce the beam emittance by a factor of 130.

3.1.2 Beam Energies

SuperKEKB is an asymmetric collider: the positron and electron beam energies are different, specifically 4 and 7 GeV respectively. This difference requires each beam to be circulated around the collider in separate rings, with the ring containing the positrons referred to as the low energy ring or LER, and the electron beam ring labelled as the high energy ring or HER. The LER energy was increased compared to its design for KEKB in order to combat instabilities that arise from high beam emittances from intra-beam scatterings. Smaller beam emittances also motivated the decrease in the HER energy for SuperKEKB. The LER and HER rings are kept separate until they are made to cross over at the interaction point to produce electron-positron collisions. The available energy at the interaction point in the centre of mass frame arising from these beam energies is denoted as $\sqrt{s}$ and is calculated using Equation 3.1

$$\begin{aligned}\sqrt{s} &= \sqrt{(E_{e^-} + E_{e^+})^2 - (E_{e^-} - E_{e^+})^2} \\ &= 10.583 \text{ GeV}\end{aligned} \tag{3.1}$$

which assumes negligible masses for the electrons and positrons, and where E_{e^+} and E_{e^-} are the energies of the positron and electron beams. The beam energies are tuned to correspond to the rest mass of the Υ(4S) meson which decays into a pair of B mesons ($B\overline{B}$) over 96% of the time. In the rest frame of the Υ(4S), each B meson is produced almost at rest as the mass of the Υ(4S) is ≈ 21 MeV higher than the combined mass of the B meson pair, resulting in a small energy release. Due to the asymmetric beam energies, the B mesons produced experience a Lorentz boost of $\beta\gamma \approx 0.287$ in the positive z direction with respect to the laboratory frame. Such boosting allows the decay length of each B meson in the laboratory frame to be adequately long enough so that their decay vertices can be sufficiently separated, and the time difference between the two B decays Δt can be measured. This quantity is important for CP violation studies as CP asymmetry arises in Δt-dependent distributions, and so the asymmetry is measured as a function of Δt. As the time-integrated CP asymmetry is zero, for a symmetric collider where only the total decay time for both B mesons ($t_1 + t_2$) can be measured, no asymmetry can be measurable. The cross-sections of processes that occur from the electron-positron collisions at the interaction point are dependent on the beam energies. For the energies at SuperKEKB, the cross-sections of the most relevant processes that occur are provided in Table 3.1. Continuum events, which refers to the annihilation processes that produce quark pairs ($e^+e^- \to q\bar{q}$ where $q = u, d, s, c$) present the largest background contribution to B physics as their total cross-section is approximately 3.3 times as large as the Υ(4S) cross-section. However, the most dominant process is that of Bhabha scattering which occurs roughly 300 times more frequently than Υ(4S) production, and which the trigger system of Belle II (see Section 3.2.6) plays a vital role in suppressing. Note that the cross-section for Bhabha scattering in Table 3.1 is calculated under an angular cut of $10° \leq \theta \leq 170°$, where θ is the polar angle with respect to the beam pipe, such that the cross-section represents the visible fraction of Bhabha events that lie inside the Belle II detector acceptance.

It should also be noted that the value for the Υ(4S) cross-section is only an approximation. Due to the broadness of the Υ(4S) resonance, small shifts in the $\sqrt{s}$ can lead to large differences in the production cross-section. The result of this is that the number of $B\overline{B}$ pairs produced ($N_{B\overline{B}}$) must be measured from data instead of calculated using $\int \mathcal{L}dt$.

TABLE 3.1. Table comparing cross-sections for the most relevant processes that are produced from collision events at the interaction point of the SuperKEKB collider. [6]. The process highlighted in red is the only one that produces B mesons.

Process	Cross-section (nb)
$e^+e^- \to \Upsilon(4S)$	1.10
$e^+e^- \to u\bar{u}\,(\gamma)$	1.61
$e^+e^- \to d\bar{d}\,(\gamma)$	0.40
$e^+e^- \to s\bar{s}\,(\gamma)$	0.38
$e^+e^- \to c\bar{c}\,(\gamma)$	1.30
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.92
$e^+e^- \to e^+e^-\,(\gamma)$	300
$e^+e^- \to \mu^+\mu^-\,(\gamma)$	1.15
$e^+e^- \to \gamma\gamma\,(\gamma)$	4.99
$e^+e^- \to e^+e^-e^+e^-$	39.7
$e^+e^- \to e^+e^-\mu^+\mu^-$	18.9

3.1.3 Collider Luminosity

The instantaneous luminosity of SuperKEKB can be represented using Equation 3.2 which assumes that the electron and positron beams are approximately flat, and have equal beam sizes in both directions perpendicular to the beam direction, that is $\sigma_{x_1} = \sigma_{x_2}$ and $\sigma_{y_1} = \sigma_{y_2}$ [63].

$$\mathcal{L} = \frac{\gamma_\pm}{2er_e}\left(\frac{I_\pm \xi_{y\pm}}{\beta_{y\pm}}\right)\left(\frac{R_L}{R_{\xi_y}}\right) \tag{3.2}$$

In this equation, the subscript $\pm$ refers to the electron $(-)$ or positron $(+)$ beam, while γ is the Lorentz factor, r_e is the electron radius from classical physics, and e is the elementary electric charge. The instantaneous luminosity achievable is dominated by the total beam current I, the vertical beta function β_y, and the vertical beam-beam parameter ξ_y. This is due to the fraction containing the reduction factors for the luminosity R_L and vertical beam-beam parameter R_{ξ_y} approximately equalling unity. The design value for the vertical beam-beam parameter ξ_y, which represents the tune shift arising from linear beam-beam interactions, is set to 0.09 which equals the value for KEKB [63]. Therefore, maximising the instantaneous luminosity for SuperKEKB involves increasing the beam currents and minimising the β_y function. The β_y function depends on the distance s from the interaction point as shown in Equation 3.3 [68].

$$\beta_y(s) = \beta_y^* \left(1 + \left(\frac{s}{\beta_y^*}\right)^2\right) \tag{3.3}$$

From this equation, it can be seen that the β_y function achieves a minimum value of β_y^* at the interaction point, where $s = 0$. At distances away from the interaction point, the β_y function increases. The beam sizes are related to the β_y function through $\sigma_y = \sqrt{\beta_y(s) \cdot \epsilon_y}$ where ϵ_y is the vertical beam emittance [68]. Therefore, the beam sizes increase approximately linearly away from the interaction point due to growing β_y values. This increase in the beam size is known as the hourglass effect. Examples of the evolving shape of β function is illustrated in Figure 3.2.

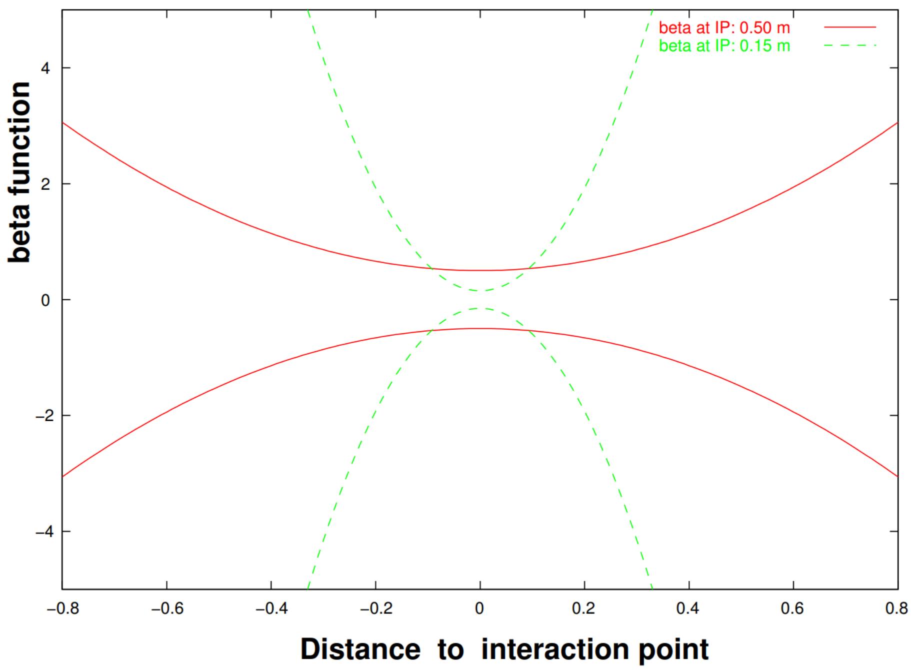
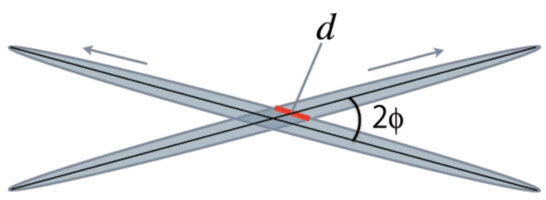


FIGURE 3.2. Visualisation of the $\beta(s)$ function for two different β^* values [68].

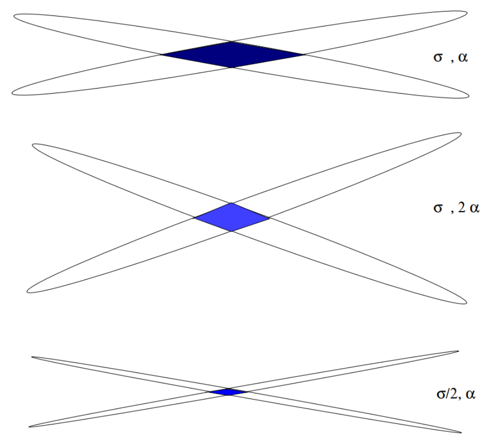
Thus, according to the hourglass effect, collisions should only occur at the interaction point as that is where β_y takes the minimum value of β_y^*, and collisions that take place away from the interaction point lead to a reduced luminosity due to larger β_y values.

To ensure that collisions occur as close to the interaction point as possible, the longitudinal overlap of the beams should be minimised. This would result in a smaller spread of collisions around the interaction point, and instead concentrate them around the point where $\beta_y = \beta_y^*$. The longitudinal overlap d can be approximated using Equation 3.4, where ϕ is the horizontal half crossing angle, and σ_x^* is the horizontal beam width at the interaction point [63]. Visual representations of d and ϕ are provided in Figure 3.3a. Differences in the spread of collisions around the interaction point with changing values of the beam width and crossing angle is demonstrated in the examples shown in Figure 3.3b.

$$d \approx \frac{\sigma_x^*}{\phi} \tag{3.4}$$



(A) Diagram showing the crossing angle 2ϕ and size of the longitudinal overlap region d for the electron and positron beams as they cross at the interaction point [63].



(B) Illustrations of the overlap region for two general beams in blue. Each example changes either the overall beam width σ or the crossing angle, symbolised here as α [68].

FIGURE 3.3. Diagrams of example beam crossings. Note that in the last two diagrams of (B), the value of d is the same though the visual representation depicts them to be different.

SuperKEKB minimises d by increasing the crossing angle compared to the angle used at KEKB (83 mrad compared to 22 mrad for KEKB) [6], and reducing the beam-width using the novel Nano-Beam scheme [69]. The impact of the Nano-Beam scheme is also a reduced β_y^* value that is around 20 times smaller than the value for KEKB [63].

To ensure the beams stay focused at the interaction point, a new magnet system was installed near the interaction point called the QCS system. It consists of 8 quadrupole magnets, 5 of which are superconducting while the other 3 are permanent, in addition to 43 corrector magnets and 4 compensation solenoids that cancel out the magnetic field from the Belle II solenoid. The result of this field cancellation helps minimise the vertical beam emittance near the interaction point. To achieve the design luminosity of $\mathcal{L} = 8 \times 10^{-34}$ cm^{-2} s^{-1} for SuperKEKB, the required values for the beam currents $I_{\pm}$ were chosen given the design value for ξ_y, and the value of β_y^* achievable through the Nano-Beam scheme and QCS system. Table 3.2 details the required $I_{\pm}$ values in addition to the other aforementioned design parameters, and provides a comparison to the KEKB values. A full table of all the design parameters used for both KEKB and SuperKEKB can be found in [63].

TABLE 3.2. Table of values for some fundamental parameters that was achieved by KEKB, compared against the SuperKEKB design values for the same parameters. [63].

Parameter	**KEKB**		**SuperKEKB**	
	LER	HER	LER	HER
ξ_y	0.129	0.090	0.090	0.088
β_y^* (mm)	5.9	5.9	0.27	0.41
I (A)	1.64	1.19	3.60	2.62
Beam energy (GeV)	3.5	8.0	4.0	7.0
$\mathcal{L}$ (10^{35} cm^{-2} s^{-1})	0.211		8	

3.1.4 Beam-Induced Backgrounds

The four main sources of beam-induced backgrounds that impact the performance of the SuperKEKB collider are the following:

(1) Touschek Scattering
This type of scattering is also known as intra-bunch scattering. Following intra-bunch scattering, changes in the momenta of the scattered particles leads to collisions with the beam pipe that create particle showers. The inverse of the beam energy, and the bunch current density, impact the rate of Touschek scattering. Therefore, such background predominantly impacts the LER as it is both a lower energy and higher current than the HER.

(2) Radiative Bhabha Scattering
Bhabha scattering events at the interaction point that radiate photons primarily do so along the z axis which can collide with the magnet system and produce neutrons. These neutrons

present as the dominant source of background for the outermost sub-detector of Belle II which is responsible for K_L^0 and muon detection (described in Section 3.2.5).

(3) Synchrotron Radiation
This type of radiation occurs when charged particles travel along curved paths. Therefore, as the beam particles accelerate around the main ring, they emit synchrotron radiation that can cause severe damage to the inner-most sub-detectors in Belle II. As the radiation power is proportional to the square of the beam energy, the majority of damaging synchrotron radiation comes from the electron beam.

(4) Beam-Gas Scattering
This refers to scattering processes that occur between the beam particles and residual gas particles in the beam pipe, and such processes manifest as either Coulomb scattering or Bremsstrahlung. This results in particle showers being produced within the beam pipe.

To mitigate the impact of beam-induced backgrounds that occur from processes (1), (2) and (4), collimators, metal shielding, and specially-designed magnet systems are relied upon. For synchrotron radiation, a gold coating has been placed on the inner wall of the beam pipe near the interaction point to minimise damage to the Belle II detector.

3.2 The Belle II Detector

The Belle II detector hermetically surrounds the interaction point and is composed of a magnetic spectrometer, electromagnetic calorimeter, and muon detector, all of which are built cylindrically around the interaction point. Surrounding the magnetic spectrometer and electromagnetic calorimeter is a superconducting solenoid that generates a 1.5 T magnetic field parallel to the beam axis. Composed of a niobium and titanium alloy, the superconducting solenoid is cooled using a liquid helium cryogenic system. The detector is not centered around the interaction point, but rather it is asymmetrically positioned so as to reflect the boosted centre of mass system which favours the distribution of final-state-particles towards the forward endcap. The interaction point sets the $z = 0$ position for the coordinate system used for the detector. The axis of the solenoid is used to define the z axis which by extension lies approximately parallel to the electron ring. The x and y axis lie respectively in the horizontal and vertical planes perpendicular to the z axis, with the positive x axis pointing outwards from the collider centre, and the positive y axis pointing upwards. In spherical coordinates, θ represents the polar angle while ϕ represents the azimuthal angle. An introduction to each sub-detector that comprises Belle II is provided below, with more details given in [63] and the references therein. Figure 3.4 provides a schematic of the detector from the top that shows the positioning of each sub-detector. All raw data from the Belle II detector is handled and analysed by the `basf2` software framework [1].

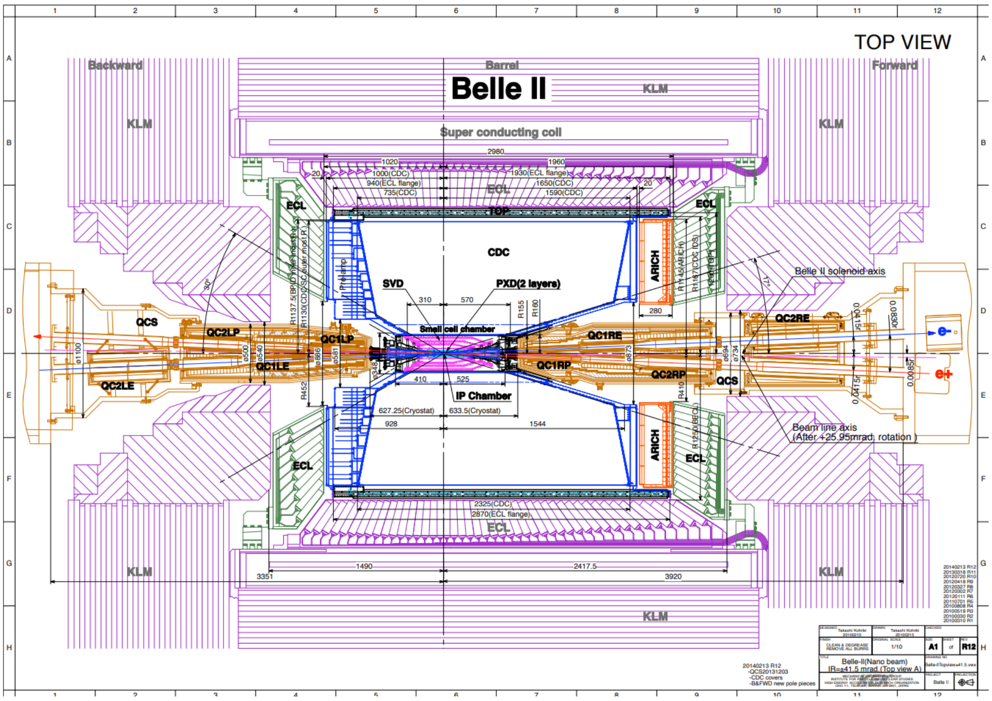


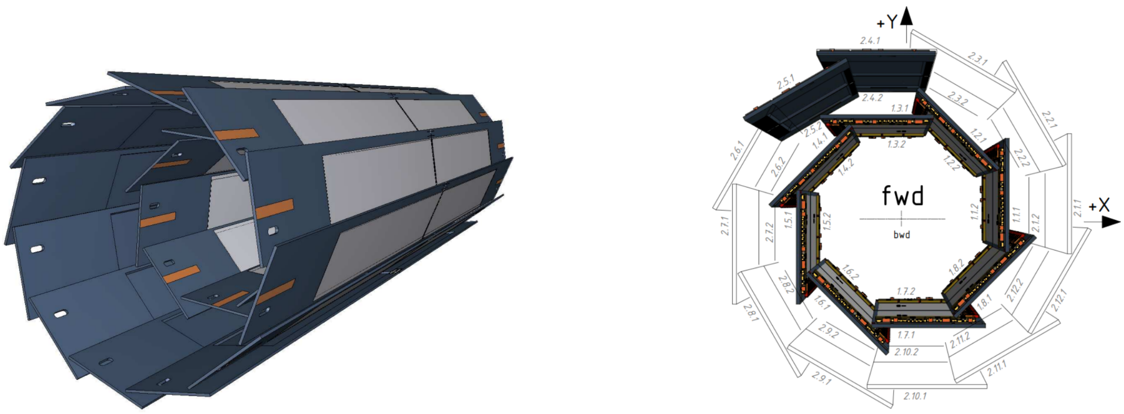
FIGURE 3.4. Schematic showing a top-view cross-section of the Belle II detector and its sub-detectors [6].

3.2.1 VXD

The vertex detector (VXD) performs precise tracking and vertex measurements for B, D and τ decays. In its current setup, the VXD is able to provide resolutions on impact-parameters and D meson decay-times of around 12 μm and 65 fs respectively [70], and has a polar coverage of around $17° < \theta < 150°$ [63].

The VXD is comprised of 6 sensor layers which are divided among two devices: the pixel detector (PXD) which contains 2 layers of pixel sensors, and the silicon vertex detector (SVD) which contains 4 layers of silicon strip sensors. While the Belle detector used a vertex detector made solely of silicon strips, the higher beam background levels arising from increased beam currents as well as a smaller beam pipe, which is 10 mm in radius for Belle II, will cause difficulty for the silicon strip sensors due to their high occupancy. Instead, pixel sensors offer a lower occupancy and are thus used in the layers closest to the beam pipe [63].

The PXD [71] utilises two layers of DEPFET (DEPleted Field Effect Transistor) sensors that are located at radii of 14 mm and 22 mm from the beam axis [63]. Detailed explanations of how DEPFET sensors work are provided in [72]. The inner layer is composed of 8 planar sensor modules while the second layer consists of 12 modules split into two halves (see Figure 3.5a). Currently, two modules are installed in the second layer of the PXD as shown in Figure 3.5b. The active area of each module contains 250 x 768 DEPFET pixels, with the pixel size varying from 50 x 55 μm^2 in the inner layer to 50 x 85 μm^2 in the outer layer. The modules are mounted on an integrated support unit that is attached to the beam pipe. In order to adapt to thermal expansions of the beam pipe, the supporting frame is designed with a gliding mechanism to prevent the bending of any modules. Application Specific Integrated Circuits (ASICs) are used for the PXD readout system, which reads the current from the DEPFET pixels and digitises it through Drain Current Digitiser (DCD) chips. Data Handling Processor (DHP) chips process, compress and buffer the data before performing fast serialisation and sending the data off to the Data Handling Concatenator (DHC) [63].



(A) Geometric arrangement of the module layers, with the light grey areas indicating the DEPFET pixels [63].

(B) Current installation setup of the PXD along the z axis [73].

FIGURE 3.5. Schematics of the PXD from different angles.

The SVD [74] consists of four layers of silicon strip sensors which are mounted at radii 38 – 140 mm from the beam axis. A total of 187 sensors comprise the SVD, each of which vary from 123 x 38 mm^2 to 123 x 58 mm^2 in area. They are mounted in a tilted origami structure similar to the PXD modules (see Figure 3.5b). The shape of the sensors vary, with nearly all sensors being rectangular apart from those in the three outermost layers towards the forward endcap that are trapezoidal in shape. This helps with the limited material space in that region, and also the optimisation of the incident angle of the track. A schematic of the silicon strip sensors are provided in Figure 3.6. As can be seen, the p side consists of long strips that lie parallel to the beam axis with the n side containing short strips that lie along $r - \phi$. As charged particles traverse the SVD and ionise, electron-hole pairs are created with the n (p) side collecting the electrons (holes). Figure 3.6 also demonstrates the effect of tilting the sensors. The electrons and holes created by a particle are deflected by the magnetic field of the solenoid, with the deflection higher for electrons due to their increased mobility. By tilting the sensors, the spread of the electron charge can be reduced, albeit at the cost of increase the spread of the holes. The read out system of the SVD primarily involves APV25 chips that read out data from the sensors, and Flash Analog-to-Digital Converter and Processing (FADC + PROC) modules that perform digitisation [63].

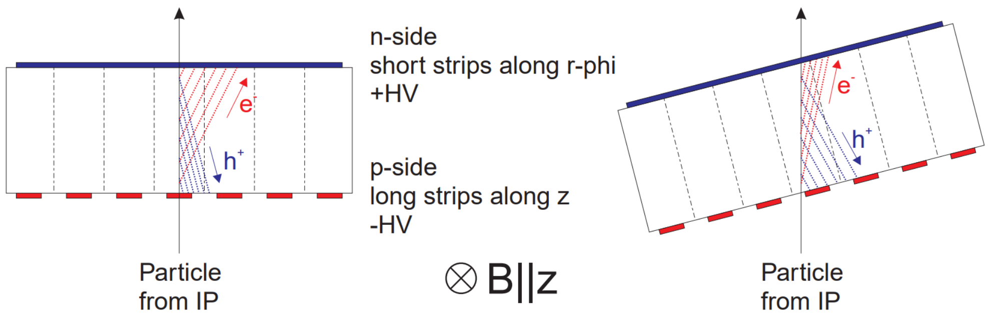


FIGURE 3.6. Diagram of the SVD silicon strip sensors in relation to solenoid field. The blue strip indicates the n side while the red strips indicate the p side. Comparing the two diagrams shows the change in the electron and hole charge spread after the sensors are tilted [63].

3.2.2 CDC

The central drift chamber (CDC) [75] is the main sub-detector responsible for tracking and particle momentum measurements. In addition to this role, it assists with particle identification by measuring energy losses through ionisation (dE/dx), and reliable triggering for charged particles. The CDC has a drift distance resolution of 120 μm [76], a maximum drift time of 350 ns, and the same polar coverage as the VXD [6]. Consisting of a gas chamber with a 50:50 He-C_2H_6 mixture, the CDC has a radius of 160 – 1130 mm from the beam axis and contains a total of 14366 sensor wires that are 30 μm wide. The voltage of the wires is tuned to create a strong electric field so that when a charged particle traverses the gas chamber, electrons produced from ionisation drift towards the sensor wires. An avalanche is created by accelerating the electrons just before they reach the wire, so as to obtain a better measurement. A total of 56 wire layers are installed, with the wires grouped into superlayers depending on their orientation. Wires that are in an axial arrangement (parallel to the solenoid field) are grouped to form a total of 5 axial superlayers. Stereo superlayers, of which there are 4, are arranged at an angle with respect to the axial superlayers in order to provide spatial information in the $x - y$ plane. A diagram of these orientations is provided in Figure 3.7. Together, the axial and stereo orientations allow for a complete reconstruction of 3-dimensional track helices [63]. The read out system incorporates front-end digitiser boards that are attached to the signal and power distributors for the CDC as well as the CDC endplates. The motivation for using such locations is to reduce cross-talk between the CDC channels and other sub-detectors, and to minimise the amount of readout cables [63].

3.2.3 TOP and ARICH

The primary particle identification system at Belle II consists of the Time of Propagation (TOP) counter and Aerogel Ring-Imaging Cherenkov (ARICH) sub-detectors which are located in the barrel and forward endcap respectively. Both sub-detectors detect Cherenkov radiation produced by charged

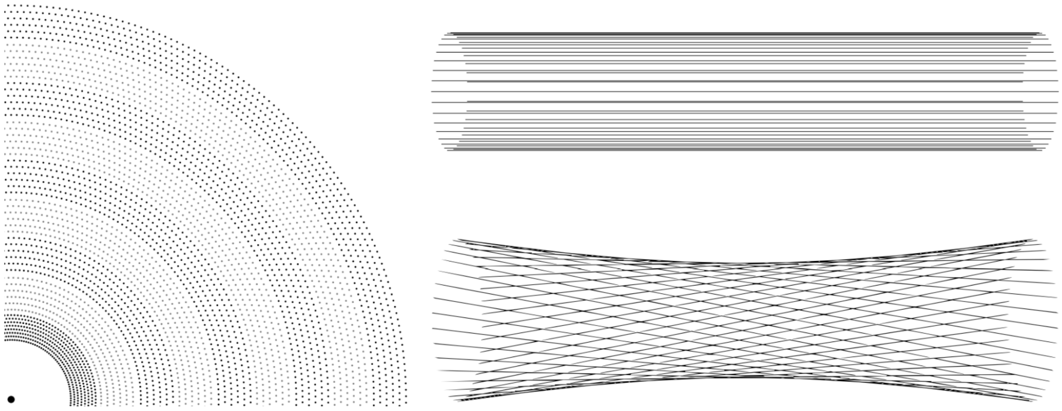


FIGURE 3.7. The left diagram shows a quadrant of the CDC wire layers in the $x - y$ plane. The axial (top) and stereo (bottom) arrangement of the wires is shown on the right [76].

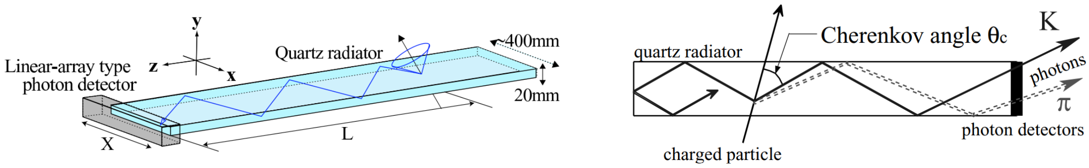
particles, and measure light properties such as the Cherenkov angle θ_c. From the Cherenkov angle, the velocity of the charged particle can be calculated using Equation 3.5 where n is the refractive index of the sub-detector material. By combining this information with the measured momentum of the track, the mass of the particle can be determined.

$$\cos\theta_c = \frac{1}{\beta n} \tag{3.5}$$

Both the TOP and ARICH are capable of providing good kaon-pion separation over most of the kinematic range, and can identify pions, electrons and muons with momenta below 1 GeV/c. With their current hardware and reconstruction software, the TOP counter is capable of providing a 85% kaon identification efficiency with a 10% pion mis-identification rate [77], while the ARICH achieves a 93% separation rate between kaons and pions with a 10% pion mis-identification rate [78].

The TOP counter [79] consists of 16 quartz bar modules that surround the outer wall of the CDC 1.2 m from the beam axis. Figure 3.8a shows a diagram of the modules which are 2.6 x 0.4 m^2 in area and 2 cm thick. A photon detector consisting of a micro-channel plate photomultiplier tube is attached to one end of the quartz bar. Cherenkov light undergoes total internal reflection (see Figure 3.8b) before reaching the photon detector which measures timing and positional information. From this information, a partial reconstruction of the Cherenkov ring can be achieved [63]. Each module contains 4 boardstacks which each contain 16 custom-built 8 channel ASICs that perform waveform sampling readout. From this, timing information is extracted using Standard Control Read-out Data boards (SCRODs) [80].

The ARICH sub-detector [81] consists of 248 silica aerogel Cherenkov radiators arranged in an annular shape with an inner radius of 42 cm and an outer radius of 114.5 cm [63]. The radiators are 2 cm thick and arranged in two layers, with the first (second) layer radiators having a refractive index of

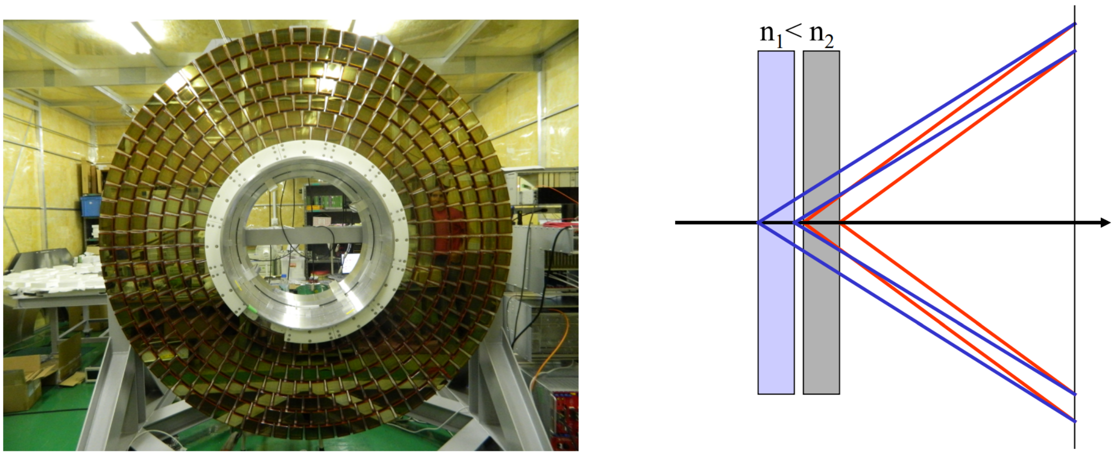


(A) Diagram of a quartz bar module and the various sub-components.

(B) Schematic showing total internal reflection within the quartz module.

FIGURE 3.8. Diagrams of the TOP counter modules and how they are used for particle identification [63].

1.045 (1.055) [6]. The gradually increasing refractive index creates a focusing radiator that improves the photon count without degrading the angular resolution (see Figure 3.9). To mitigate the loss of photons via absorption or Rayleigh scattering, the radiators are required to be highly transparent. Following each radiator is a 20 cm long expansion volume to help achieve the required angular resolution. Hybrid avalanche photon detectors (HAPD) are used for photon detection. They are 73 x 73 mm^2 in size and are equipped with 144 channels. Due to the total number of readout channels being around 80000 for the ARICH, dedicated ASICs are used to process the signals before they are fed to Fast Programmable Gate Arrays (FGPAs) for feature extraction [63].



(A) Front-end view of the HAPD sensors used for photon detection [6].

(B) Cherenkov light rays from a focusing radiator. Credit: Belle II Collaboration.

FIGURE 3.9. Visualisation and schematic of the ARICH components.

3.2.4 ECL

The electromagnetic calorimeter (ECL) is mainly responsible for measuring the energy and angular resolution of photons, assisting with electron identification, and providing online and offline measurements of the instantaneous luminosity. Given that nearly one-third of all B decay products are π^0

mesons which decay to two photons, the calorimeter is required to have high efficiency and resolution for photon reconstruction, and it also must be highly granular in order to separate photons that originate from a π^0. A large acceptance is also key as the total visible energy of a collision event is an important analysis variable for studies involving neutrinos in the final state. The intrinsic energy resolution of the ECL is given by Equation 3.6 where the photon energy E is in units of GeV. For a photon of 100 MeV, σ_E/E is around 2% [63].

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34\%)^2} \tag{3.6}$$

The ECL consists of a barrel and two endcap regions (forward and backward) that altogether provide a polar coverage of $12.4° < \theta < 155.1°$. The barrel section is 3 m long with a radius of 1.25 m while the forward and backward endcaps are situated at $z = 1.96$ m and $z = -1.02$ m respectively. In total, the ECL comprises 8736 thallium-doped caesium iodide CsI(Tl) crystals that are each covered with a 200 μm thick layer of Gore-Tex porous teflon followed by two 25 μm thick layers of aluminium and mylar. Of the 8736 crystals, 6624 are located in the barrel while 2112 are in the endcaps. Within the barrel, where the geometry is simpler, each crystal takes the shape of a truncated pyramid that has an average cross-section of 6 x 6 mm^2 and a length of 30 cm. This length corresponds to around 16.1 radiation lengths for photons at energies typical for Belle II. The support structure for the crystals is kept in a low humidity environment by flushing the structure with dry air, and excess heat from the electronics (around 3 kW in total) is removed by a water cooling system. Attached to the back of each crystal through a 1 mm thick acrylite plate are two 10 x 20 mm^2 Hamamatsu Photonics S2744-08 photodiodes that provide readout for the scintillation light captured by the crystal. Each photodiode is connected to a preamplifier to allow for two independent signals to be output from each crystal. These signals are digitised at external Shaper Digital Signal Processor (ShaperDSP) boards which are placed in 52 crates around the Belle II detector. Waveform fits are done to the digitised signals using FGPAs to extract amplitude and timing information from the original light pulses [63]. Each waveform fit uses 31 points sampled from the digitised signal pulse, with 16 of these points sampled before the pulse peak in order to measure the underlying noise. A fit to obtain information on the the pulse shape is then performed to the sampled points by minimising a set of linear equations whose parameters include the amplitude and time of the pulse.

The performance of the ECL is regularly monitored through a variety of methods. The optical stability can be checked by injecting light pulses through an LED that is attached to the acrylite plate at the back of each crystal. Stability of the electronics system is checked by performing daily test pulse runs which involve injecting artificial inputs into the preamplifiers, processing the signal, and comparing the calculated amplitude and time with the known values. Due to the high occupancy and background experienced by the forward endcap, upgrades to the crystal material are being considered for future long shutdowns, with the primary candidate being pure CsI crystals. The advantage of these is their

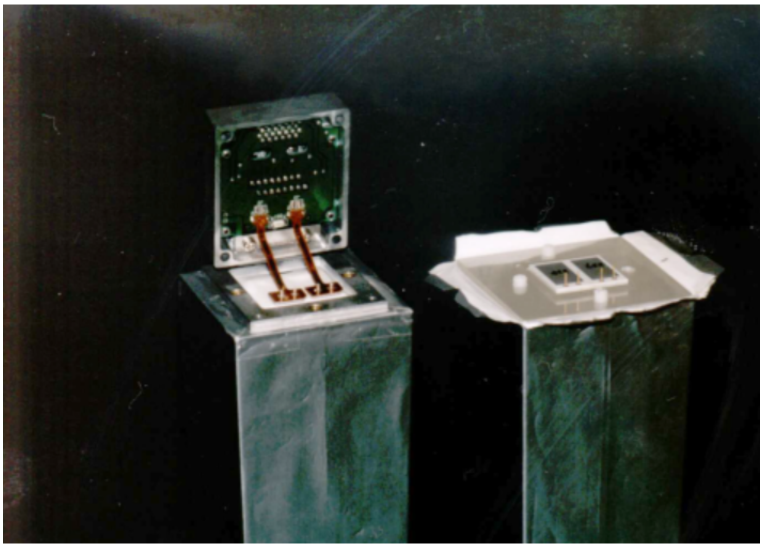


FIGURE 3.10. Picture of the ECL CsI (Tl) crystals encased in the Gore-Tex teflon, aluminium and mylar layers. The crystal on the right shows the two photodiodes attached to the acrylic plate, while the left crystal shows the photodioides attached to the preamplifiers. [63].

significantly smaller decay time of around 30 ns (compared to the currently used Cs(Tl) crystals which have a decay time of 1.3 μs). Such an improvement would help mitigate any degradation in the timing and energy resolution of the ECL that result from increased pileup and radiation damage [63].

3.2.5 KLM

The K_L^0 and muon detector (KLM) [82] is primarily designed to detect muons and K_L^0 mesons, and to serve as the magnetic flux return for the solenoid. Covering a combined polar angle of $20^\circ < \theta < 155^\circ$, the identification efficiency for muons above 1 GeV/c of momenta is 89%, while the K_L^0 reaches a maximum of 80% for a momentum of 3 GeV/c [63]. Like the ECL, the KLM consists of an octagonal barrel and two endcap regions. Within the barrel, there are 15 active detector layers and 14 iron plates, while in the endcaps, there are 14 of each. The iron plates are used to act as the magnetic flux return for the solenoid. In total, the layers provide more than 3.9 interaction lengths, in addition to an existing 0.8 interaction lengths contributed by the ECL. The detector layers in the barrel are glass electrode resistive plate chambers (RPCs) which range from 2.2 x 1.5 m^2 to 2.2 x 2.7 m^2 in area, except for the RPCs near the helium chimney used for the solenoid which are around 0.63 m shorter. The chimney location can be seen in the diagram in Figure 3.11. In the endcaps, RPCs are not used as they are limited by long dead times following electric discharges. This leads to a degradation in the detection efficiency under high beam background. Instead, scintillators make up the active detector layers in the endcaps. Due to the presence of the strong magnetic field and a lack of hardware space, photomultipliers are not used for the scintillators. Instead, plastic scintillation strips embedded with wavelength shifting fibres are used to catch any scintillation light that is produced, with the signal from each fibre read out by a silicon sensor attached to one end of the scintillation strip. In total, 16800 strips are used for the 28 scintillation layers in the endcaps, with each strip having a cross section that averages 8.5 x 40 mm^2 and a length of around 2.8 m. A compromise between the required spatial resolution and the minimisation of readout channels determined the width of each silicon strip. Signals from the silicon strips are read out using a dedicated 16-channel ASIC chip that then passes the signal to an FPGA for digitisation.

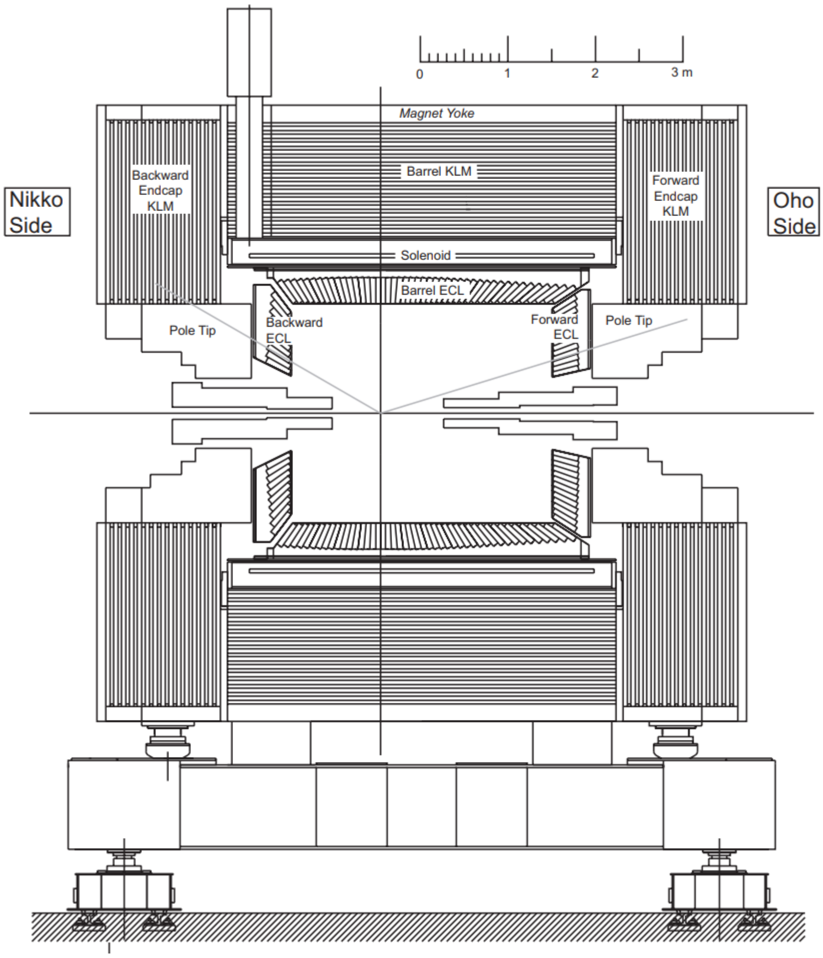


FIGURE 3.11. Schematic showing a cross-section of the mounted Belle II detector from the side. The helium chimney is located between the backward endcap and barrel of the KLM, towards the side of the Nikko experimental hall. The two inserted structures on either side of the interaction point are part of the QCS system. [63].

3.2.6 Trigger and DAQ

Only a fraction of events that occur from collisions at the interaction point are of interest for physics analyses (see Table 3.1). Storing all of the raw data from the detector to storage disks would therefore be wasteful and unnecessary. The purpose of the trigger system is to make decisions about whether a particular event is interesting enough to be triggered and stored. The trigger is required to have a timing precision close to 10 ns, with a timing separation between events of no more than 200 ns [63]. Furthermore, the trigger must be close to 100% efficient for $B\overline{B}$ and continuum events. The trigger system consists of two components: the Level 1 or L1 trigger, and the High-Level Trigger or HLT.

The purpose of the L1 trigger is to utilise information from fast sub-detectors to reject most of the background events. It is completely hardware-based and runs on an array of FPGAs. The primary sub-triggers used by the L1 trigger belong to the CDC and ECL. The CDC sub-trigger yields information on the charged track multiplicity of an event, as well as the reconstructed z coordinate for the primary decay vertex which it can compute in a few microseconds [63]. Similarly, the ECL sub-trigger looks at the cluster multiplicity and the total energy deposited in the ECL. Such low-level information is useful for filtering out distinctive processes such as photon-pair production or Bhabha scattering. Information from these sub-triggers is combined with hit information from the TOP and KLM sub-triggers by the Global Reconstruction Logic (GRL). It is then sent to the Global Decision Logic (GDL) which uses the information to decide whether the event should be triggered [63]. If so, the output of the L1 trigger is passed to the HLT.

The HLT is completely software-based and is responsible for performing an online reconstruction of the event using `basf2` [1]. It utilises information from all sub-detectors apart from the PXD to perform an initial event reconstruction. Once complete, regions of interest are defined by extrapolating reconstructed tracks back to the PXD, and only pixels inside these regions are read out and used to complete the event reconstruction [63]. The HLT uses this online reconstruction to calculate basic quantities on the event topology and particle kinematics in order to make a decision about whether to reject the event as background or transfer it to the offline storage discs.

The role of the data acquisition system (DAQ) is to ensure that all sub-detectors receive trigger signals synchronously and to provide high-speed data links that transfer read-out data from each sub-detector to the HLT system. The maximum acceptance rate of the DAQ is 30 kHz and this limits the maximum rate of the trigger system [63].

3.2.7 Monte Carlo Simulations

Monte Carlo (MC) simulations are used at Belle II to generate a variety of collision events and simulate the detector response to the produced final state particles. MC simulations are an extremely important analysis tool as any given study will first use MC samples to tune selection cuts and develop analysis methods before using real collision data from the detector.

The MC generators used for the Belle II experiment are as follows. The generation of B and D decays into exclusive final states are handled by `EvtGen 1.3` [83], while inclusive decays and light quark hadronisation is performed by the `Pythia 8.2` [84] generator. The generation of τ pair production, and the decay of the τ into exclusive final states, are performed by the `KKMC 4.15` [85] and `TAUOLA` [86] generators respectively. `BABAYAGA.NLO` [87] generates QED background processes that have large cross-sections such as Bhabha scattering and photon-pair production, while `AAFH` [88] generates the $e^+e^- \to e^+e^-e^+e^-$ and $e^+e^- \to e^+e^-\mu^+\mu^-$ decays. Particle interactions with the detector material is modelled using `Geant4` [89]. Lastly, the `Photos` generator is utilised to simulate final state radiative corrections for decays. The same beam parameters are used for all the MC generators mentioned above, and they are stored in a conditions database for easy usage. Explicit values for the beam parameters can be found in [6].

Simulations of the detector response by `Geant4` do not inherently include beam-induced background processes. Instead, such processes are simulated using a dedicated software called SAD which was developed by the KEK accelerator group. The Strategic Accelerator Design (SAD) software [90] simulates the propagation of particles through the SuperKEKB accelerator, and should a beam-induced background process occur, SAD saves the relevant kinematic information before calling upon `Geant4` to simulate the material response [6]. Background overlay files containing these simulated background processes are included in the production of MC samples used for physics analyses to better reflect reality. Different levels of background overlay can be used e.g. a BGx5 MC sample represents a production of MC events with five times the nominal amount of beam-induced background overlay. Background overlay samples can also be obtained from real collision data through the use of a random trigger. However, as this trigger involves a minimum threshold for detection, not all background processes are going to be included.

The MC samples used by Belle II analyses are either run-independent or run-dependent. The latter uses the specific experimental conditions for each data-taking run, including background overlay obtained from random triggers, when generating the simulated events. In this sense, the run-dependent MC will contain simulated events but with a realistic and up-to-date detector response. On the other hand, run-independent MC uses the default experimental conditions, and simulated background overlay, for its production.

3.2.8 Reconstruction

Following data acquisition or simulation, the digitised data output from each sub-detector is first converted into low-level objects such as hits and digits. However, for the information to be usable on an analysis level, higher-level objects and variables are required. The purpose of reconstruction is to reconstruct particles from the collision event in the form of these high-level objects and to calculate variables that are derived from low-level detector information. In the case of simulated data, relations between reconstructed particles and MC-generated particles are accessible during reconstruction, and these MC relations can be saved. To ensure consistency, both raw and simulated data are reconstructed using the same `basf2` versions.

The event reconstruction is never perfect - for example, particles can travel outside detector acceptances or result from beam background processes, and electronic noise in the sub-detectors can degrade energy and timing measurements. In addition to this, the imperfect modeling of the detector material and response in simulated data means some discrepancies between the reconstruction of acquired data and simulated events will always exist. The primary objects that result from event reconstructions are cluster objects made from hits in the ECL, track objects originating from PXD, SVD, and CDC hits, and particle identification likelihood variables.

3.2.8.1 Tracking

Tracking involves reconstructing track objects that represent charged particle trajectories. It is divided into two steps: track finding and track fitting.

Track finding determines suitable track candidates using collections of hits, or clusters, made in the PXD, SVD, and CDC. Suitable clusters are those that combine to form the same trajectory. Track finding begins with CDC clusters, whereby both a global algorithm based on Legendre transforms, and a localised cellular automation algorithm is used to find suitable CDC clusters [91]. These clusters are fit using a deterministic annealing filter (DAF) from the `GENFIT2` package to form CDC-only tracks [92]. A visualisation of this is given in Figure 3.12. Following this, a combined Kalman filter (CKF) is used to find suitable SVD clusters that are combined with the CDC-only tracks through an additional DAF fit [93]. The result is then extrapolated to the PXD using a CKF. Subsequently, track fitting using a DAF is performed using a helix parametrisation for the particle trajectories. The goal of track fitting is to estimate the kinematic properties of the particle such that its position and momentum can be accurately determined. This is achieved by using a single particle hypothesis (kaon, pion, or proton) for the fit.

For a V^0 particle such as a K_S^0 that produces two charged tracks at some distance from the interaction point, a special tracking algorithm is used after the charged track reconstruction. This algorithm first finds all oppositely charged pairs of tracks and extrapolates them back to the first hit made by either track. Pairs of tracks that fail this extrapolation are rejected. Track combinations that succeed have their vertex fitted using the RAVE package [94], and track pairs that have a failed vertex fit are removed from the list. The last step is to impose quality cuts using the χ^2 from the vertex fit, and, if the vertex lies inside the beam pipe, a cut on the mass difference of the track using the nominal mass for the track hypothesis used [6].

3.2.8.2 Clustering

Clustering involves grouping individual crystals into connected regions from which cluster objects are made. A connected region (CR) is a contiguous set of crystals that passes particular crystal energy

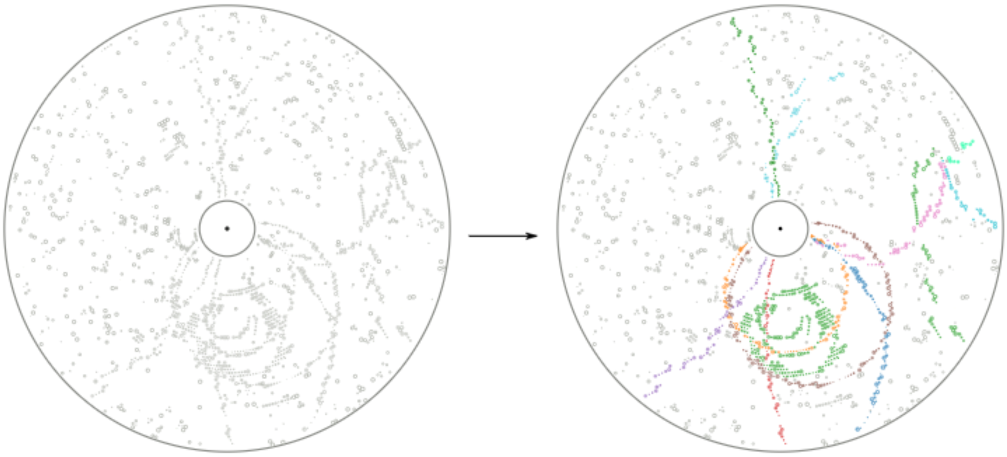


FIGURE 3.12. Hits in the CDC shown in an $x-y$ cross-section of the sub-detector. The left-hand plot shows the hits prior to the track finding, while the right-hand plot shows the suitable clusters that have been found using the track finding in different colors. Hits that remain grey are rejected as background. Credit: Belle II Collaboration.

cuts whose values are tuned for robustness against electronic noise and beam background. The average energy deposited in a single crystal by beam background is 0.5-1 MeV in the ECL barrel and forward endcap, and up to 2 MeV for the backward endcap, while electronic noise contributes around 0.35 MeV.

The steps involved in the clustering process are described below [6]:

(1) Determination of the seed crystal and CR

Crystals with energy greater than 10 MeV are designated as seed crystals. The surrounding 8 crystals whose energies are above 0.5 MeV are added to the seed crystal to form a CR. If any of the neighboring 8 crystals have energies greater than 10 MeV, they become seed crystals themselves, and this step repeats and grows the CR. If a crystal is shared between two CRs, the CRs are merged.

(2) Splitting the CR

The CR is split into n clusters where n is the number of local maxima in the CR. The local maximum is defined as the crystal whose energy is greater than the surrounding 8 crystals.

(3) Forming clusters and energy splitting

If only one local maximum is found in the CR, the cluster comprises the 20 crystals surrounding the local maximum such that it makes a 5x5 grid with the corners removed.

If more than one local maximum exists, the energies of crystals in the CR are divided among the clusters using weights. The weights for up to 20 crystals surrounding the local maximum are calculated using e^{R/R_w} where R is the distance from the crystal to the cluster centroid (initially the local maximum) and $R_w = 1.43$ cm. For each crystal, the ratio between the weight from a given local maximum to the sum of all weights for the crystal is used to determine the fraction of the crystal energy assigned to each local maximum.

This process is iterative, with the centroid position updated each iteration, and the process stopping once the centroids are stable.

The above procedure is used as-is for CRs that have a photon or electron hypothesis. For CRs with a hadron hypothesis, more than 20 neighbouring crystals are used in the third step. Once cluster objects are formed, cluster properties or shower-shape quantities are calculated using the individual crystals that compose the cluster. These are calculated following energy corrections applied due to leakage and other calibrations [6]. The result of the clustering algorithm can be visualised in Figure 3.13.

Following both clustering and tracking, clusters are matched to track objects using one of two methods depending on the track momentum and polar angle. For tracks with transverse momentum greater than 0.3 GeV/c^2, or those pointing towards the forward endcap, track-cluster matching is performed by comparing the polar angle distance between the particle shower and track. Conversely, for tracks with less than 0.3 GeV/c^2 transverse momentum, or those pointing towards the gaps between the ECL barrel and endcaps, the matching is done using the ECL hits obtained from extrapolating the trajectory of the track through the ECL. The track-cluster matching is only performed for clusters that have an energy greater than 50 MeV, and a timing that lies inside the window defined by 99% of real photons with the same energy as the cluster. These requirements reduce the computation time by ensuring that track-cluster matching is not performed for clusters that are likely to be a product of beam background processes [6].

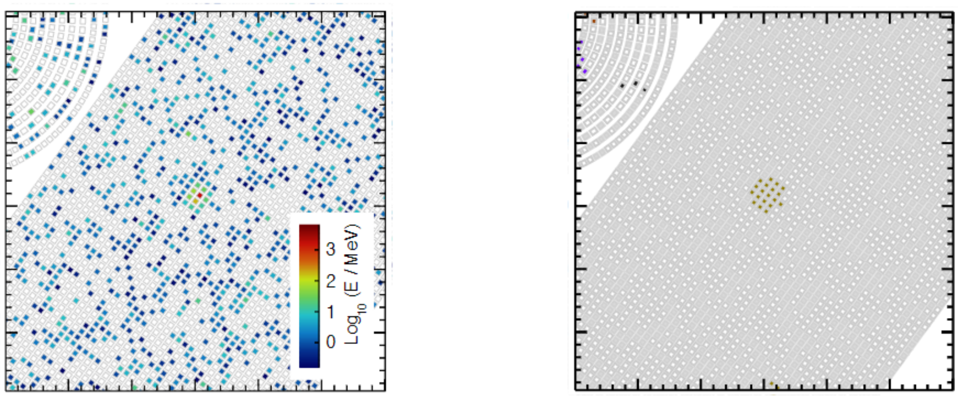


FIGURE 3.13. Hits in the ECL shown in the $\theta - \phi$ plane. The left-hand plot shows the activated crystals from a true photon in the centre, surrounded by crystal deposits from beam-background processes. The right-hand plot shows the result of the clustering algorithm, which reconstructs the photon cluster and rejects all other ECL hits. Credit: Belle II Collaboration.

3.2.8.3 Particle Identification

Particle identification is performed for standard charged particles ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$ and $d^{\pm}$) and standard neutral particles (γ and K_L^0). Specific sub-detector information is combined to form an identity likelihood for a given particle, with the type of information used depending on the particle hypothesis.

For charged hadrons, a significant value is the ionisation loss dE/dx which is measured independently using the SVD and CDC. Additionally, for K/π separation, low-level hit information on Cherenkov rings and angles from the TOP and ARICH sub-detectors are used. The longitudinal profile and transverse scattering probabilities associated with KLM hits are used as inputs for the muon hypothesis likelihood, while electrons use shower-shape information from clusters reconstructed in the ECL. For each sub-detector, the likelihood for a given particle hypothesis i can be combined to form an overall likelihood as shown in Equation 3.7.

$$\mathcal{L}_i = \prod_{\text{det}} \mathcal{L}_i^{\text{det}} \tag{3.7}$$

From this, a global particle identification quantity (PID) for a given particle hypothesis i can be derived using Equation 3.8, where j covers all the standard charged particles including particle i.

$$\text{PID}_i = \frac{\mathcal{L}_i}{\sum\limits_j \mathcal{L}_j} \tag{3.8}$$

Alternative to the global PID is a binary PID quantity that compares the likelihoods between two specific hypotheses i and k, and is calculated using Equation 3.9.

$$\text{PID}(i;k) = \frac{\mathcal{L}_i}{\mathcal{L}_i + \mathcal{L}_k} \tag{3.9}$$

Photons are identified using ECL clusters that are not matched to any tracks. Various kinematic, timing and shower-shape information derived from the crystals in the cluster are used to form a photon likelihood. Shower-shape quantities are particularly useful as electromagnetic showers tend to be more symmetrical and compact in distribution as opposed to hadronic showers which are asymmetrical and broad.

For K_L^0 identification, both ECL and KLM information is used to build two separate multivariate classifiers. For the KLM-based classifier, 19 features are used, with the most significant ones being the number of innermost layers that are hit (hadronic showers have a wider radius and thus a smaller number of layers) and the hit timing (most hits from beam background occur before or after the collision time). The ECL-based classifier has 38 features, and it largely relies on shower-shape quantities and the distance between the ECL cluster and the nearest track (neutral clusters do not have any nearby tracks).

CHAPTER 4

Analysis Techniques

This chapter introduces the main analysis techniques that will be relevant for subsequent chapters. It consists of technical information on the workings of binary classifiers and evaluations of classifier performance, with specific discussions on decision trees and gradient-boosted decision trees. Such information is relevant for the E_{extra} cleanup classifiers introduced in Chapter 5. Additionally, a discussion of statistical fits and calculations of branching fraction upper limits will be presented. These concepts pertain to the $B^0 \to \ell^\pm \tau^\mp$ analysis that is presented in Chapter 6.

4.1 Binary Classifiers

A binary classifier is an algorithm designed to learn the functional relationship between vectors of input variables $\mathbf{x} = \{\mathbf{x_1}, ..., \mathbf{x_n}\}$ called features and corresponding output variables $\mathbf{y} = (y_1, ..., y_n)$ where y_i represents one of two classes (usually 0 or 1). In this context, a data point or data sample $i \in [1, n]$ is defined as a corresponding pair of features and output $(\mathbf{x}_i, y_i)$. The relationship between $\mathbf{x}$ and y can be expressed generally using Equation 4.1 where ϵ represents the error term, and f is the functional relationship which defines the mapping $\mathbf{x} \to y$.

$$y = f(\mathbf{x}) + \epsilon \tag{4.1}$$

The exact mapping learned by the classifier is dependent on the classifier's internal parameters, whose values must be learned by fitting to a set of example data. This process of fitting is referred to as training. The optimal parameters are determined by minimising a pre-defined loss function evaluated over the training data. The loss function measures the difference between the output predicted by the classifier $\hat{y}$ and the true class labels y. These functions are therefore representative of the accuracy of the classifier, and are useful for measuring the performance of trained classifiers on unseen data, also referred to as test data. The binary cross-entropy loss function is the most common choice for binary classifiers and it is discussed further in Section 4.1.3 in the context of evaluating classifier performance. A classifier makes a prediction by calculating the probability of a data sample belonging to a particular class (conventionally class 1). In order to convert this to a class label, the probability is compared to a decision threshold, typically 0.5, and if the probability is higher, then the sample is labelled as class 1.

Classifiers have additional configuration parameters whose values must be determined prior to training. Such parameters are called hyperparameters and they dictate the properties of the classifier architecture and training process. Tuning using a dedicated data set independent of the training and test data, called the validation data, is used to obtain the optimal hyperparameter values, and this process is described in Section 4.1.4.

Although there are countless different types of binary classifiers available, the remainder of this section will focus on two specific types: decision trees and gradient boosted decision trees.

4.1.1 Decision Trees

Decision trees are one of the most common and popular binary classifiers due to their simplicity and interpretability. A typical decision tree is shown in Figure 4.1. Based on a divide-and-conquer approach, a decision tree consists of layers of nodes that together perform a series of consecutive cuts to classify data. All data samples start at the root and traverse through the tree depending on which cuts the data sample satisfies. Classification is done when the data sample reaches a terminal node (also called a leaf), at which point the probability of the data sample belonging to class 1 is equal to the proportion of class 1 samples that reached the same terminal node during training.

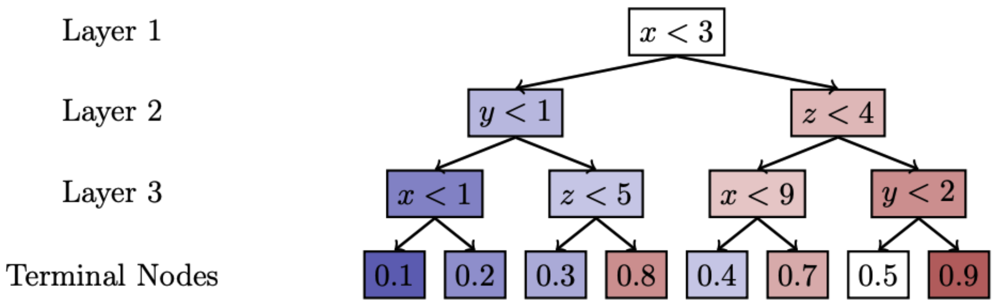


FIGURE 4.1. Visualisation of a typical three layer decision tree. All data samples start at the root node which is the single node in Layer 1. The numbers denoted on the terminal nodes indicates the proportion of class 1 data samples that have reached the terminal node after traversing through the tree. [95].

The choice of cuts is determined during training using metrics that use a purity-based measure called entropy. The entropy of a data set S represents the homogeneity of the data and is calculated using Equation 4.2 where P_i is the proportion of data samples that belong to class i [96].

$$H(S) = -\sum_i P_i \cdot \log_2(P_i) \tag{4.2}$$

A common entropy-based metric used during training is the information gain which is calculated for a given cut using Equation 4.3 [96]. It measures the decrease in entropy of the data set if a given cut is applied, and by comparing the information gain across a series of potential cuts, the most optimal cut for a specific node can be selected. The implication of this is that the root node of any decision tree contains the cut that yields the highest information gain for the training data.

$$\mathrm{Gain}(S \mid \mathrm{cut}) = H(S) - H(S \mid \mathrm{cut}) \tag{4.3}$$

The complexity of a decision tree is largely governed by a single hyperparameter called the maximum depth which is the number of layers in the tree. Simple decision trees that are built with a small maximum depth (around 2-3) are called weak learners. The output of a single decision tree is largely dominated by statistical fluctuations present in the training data. However, this can be overcome if weak learners are used in combination and trained as an ensemble. Provided the individual accuracy of each weak learner is better than a random guess, ensembles of weak learners are significantly more robust and accurate than a standalone decision tree [96]. The most common type of such an ensemble is the gradient boosted decision tree.

4.1.2 Gradient Boosted Decision Trees

A boosted decision tree refers to an ensemble which is built sequentially from individual weak learners that are trained on the errors of the prior learner. Gradient boosted decision trees iteratively add weak learners to the ensemble using gradient descent so as to minimise the overall loss function of the ensemble. Such ensembles can be used for both regression and classification tasks. For a binary classification problem, the loss function for a single weak learner F_j in a gradient boosted decision tree with J learners is defined using Equation 4.4. The gradient of this function is given in Equation 4.5 [96].

$$L(y, F_j(\mathbf{x})) = -\sum_{i}^{N} y_i \log(F_j(\mathbf{x}_i)) + (1 - y_i)\log(1 - F_j(\mathbf{x}_i)) \tag{4.4}$$

$$\mathbf{g}_j = \frac{\partial \sum_i^n L(y_i, F_{j-1}(\mathbf{x}_i))}{\partial F_{j-1}(\mathbf{x}_i)} \tag{4.5}$$

The equation for the update of the gradient boosted decision tree is given by Equation 4.6. Here $h_j(\mathbf{x})$ is a weak learner that has been trained on the gradient $\mathbf{g}_j$, while α_j is the learning rate or shrinkage parameter. The value of α can be set as a constant for all training iterations, or can be derived for each iteration using the output of $h_j(\mathbf{x})$ and the loss function for F_{j-1} [96]. Predictions on unseen data can then be made by the ensemble using Equation 4.7 [96].

$$F_j(\mathbf{x}) = F_{j-1}(\mathbf{x}) - \alpha_j h_j(\mathbf{x}) \tag{4.6}$$

$$\hat{y} = F_J(\mathbf{x}) = \sum_{j=1}^{J} \alpha_j h_j(\mathbf{x}) \tag{4.7}$$

Gradient boosted decision trees are one of the best performing classifiers provided their hyperparameters are properly tuned. Some of the more important hyperparameters include the number of weak learners, the maximum depth of each learner, the learning rate, and the fraction of features to consider when building each weak learner. The performance of a gradient boosted decision tree can be improved by using only a subset of the training data when training each weak learner. This is known as a stochastic gradient boosted decision tree, and they are faster to train and more robust against statistical fluctuations in the training data. The ratio of the training data used for each iteration is an additional hyperparameter for these classifiers.

For the purposes of the E_{extra} study presented in Chapter 5, a special type of stochastic gradient boosted decision tree is used called a `FastBDT` [95]. This classifier is specifically cache-optimised and discretises continuous features to allow for faster training.

4.1.3 Evaluating Classifiers

For binary classifiers, the true positive (TP) and true negative (TN) number equals the amount of correct class 1 and 0 predictions, respectively. Complementary to this are the false positive (FP) and false negative (FN) values which are the number of incorrect class 1 and 0 predictions. Using these four numbers, the true positive rate (TPR) and false positive rate (FPR) can be defined according to Equations 4.8 and 4.9

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4.8}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{4.9}$$

The TPR and FPR are used to create one of the most widely used classifier evaluation methods which is the receiver operating characteristic (ROC) curve. An ROC curve plots the TPR and FPR (or variations of them such as TPR and 1 - FPR) under different decision thresholds ranging from 0 to 1. The area under the ROC curve, referred to as the AUC, is used as the evaluation metric. For a random classification, the AUC equals 0.5, while a perfect classification yields an AUC of exactly 1.

The AUC is one of the metrics used in Chapter 5, the other being the binary cross-entropy loss function, also called the log-loss. It is defined in Equation 4.10, where y is the true label of a given sample, and p is the probability of the given sample being in class 1 as determined by the classifier [97]. The smaller the log-loss, the better the predictions made by the model. For instances where the training data is split equally between class 0 and 1, the baseline log-loss value is 0.693 [97].

$$L(y, p) = -y\log(p) - (1-y)\log(1-p) \tag{4.10}$$

Evaluating the classifier performance between training and unseen data is a key method of determining whether a classifier underfits or overfits. When a classifier is unable to learn the complete mapping between features and classes, often due to a lack of complexity or training, the classifier is said to be underfitting and have a high bias. Conversely, an overfitted classifier is one which has "memorised" the training data by focusing on learning the statistical fluctuations as opposed to the general relationship. Such classifiers are said to have a loss of generality or high variance, and it is generally a result of too much fine-tuning, training or complexity. Overfitting is determined to occur if there is as a significant decrease in performance for the unseen data compared to the training data. On the other hand, if the performance for both the training and unseen data is poor, underfitting is likely occurring.

4.1.4 Hyperparameter Optimisation

Most methods of hyperparameter optimisation involve performing a grid search. This is where the range of hyperparameter values to be tested are stored in an N-dimensional grid, where N is the number of hyperparameters to be tuned. This grid is then searched using the chosen optimisation algorithm, so that the best combination of hyperparameter values can be found. Depending on the classifier training time, the grid search can either be exhaustive, or limited to a random selection of combinations.

One common and straightforward optimisation algorithm, that is often paired with grid search, is holdout. Holdout involves dividing all of the data into three subsets: one for training, one for validation, and one for testing. The chosen ratios for the data split is largely dependent on the complexity of the classifier - the more complex a classifier, the more training data required for the classifier to adequately learn. The holdout method is as follows. For each hyperparameter combination in the grid, a classifier is trained on the training subset, and tested using the validation subset. By comparing the classifier performance for the validation data between the various classifiers, the combination of hyperparameter values that yielded the best classifier is designated as the most optimal choice. Typically, what follows is a final retraining of the best classifier with the training and validation subsets combined, before the last test is done using the testing subset.

When the available data is not large enough to allow for a split where a sufficient amount of samples are used for training, an alternative method called cross-validation can be used. A detailed explanation of cross-validation can be found in [96].

4.2 Maximum Likelihood Fits

Binned maximum likelihood (ML) fits are used in this analysis to derive the estimate for one of the background components. ML fits first involve forming a model to describe the data based on a given probability density function (PDF) $f(x;\boldsymbol{\theta})$ where x is the data and $\boldsymbol{\theta}$ is the vector of PDF parameters that describe its shape. Assuming each data point is statistically independent, the corresponding likelihood function is described by Equation 4.11 where n is the total number of data points. In the context of particle physics where each data point represents an event, n is equivalent to the total event yield.

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) \tag{4.11}$$

The ML estimates for the parameters $\boldsymbol{\theta}$ are then determined when fitting the model to the data by maximising the logarithm of the likelihood function. Effectively, this means finding the parameter values that solve Equation 4.12. The logarithm of the likelihood function is used as it is better behaved and maintains the position of the ML parameter estimates.

$$\frac{\partial \ln\mathcal{L}}{\partial\theta_i} = 0 \tag{4.12}$$

The likelihood function in Equation 4.11 is extended in the context of particle physics as n is not fixed but instead is drawn from a Poisson distribution with mean μ. This extended likelihood function is given in Equation 4.13 which is equivalent to Equation 4.11 but with the Poisson function added.

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{\mu^n}{n!}e^{-\mu}\prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) \tag{4.13}$$

For unbinned data, the ML fit proceeds as described above. However, for binned data where the data becomes the vector $\boldsymbol{n} = (n_1, ..., n_N)$ where N is the number of bins, the data has the expectations $\boldsymbol{\mu} = E[\boldsymbol{n}]$ and probabilities $f(\boldsymbol{n};\boldsymbol{\mu})$. For expectations that depend on $\boldsymbol{\theta}$, the likelihood function can be maximised by fitting to the contents of each bin.

For this analysis, all binned ML fits are performed using `RooFit` [98] which uses an algorithm that allows for the output of all PDF parameters and normalisations along with their relative uncertainties. The uncertainty is derived from the statistical uncertainty of the sample used to perform the fit.

4.3 Calculating Upper Limits

In instances where no significant signal is observed, as was the case with previous $B^0 \to \ell^{\pm}\tau^{\mp}$ searches, an upper limit on the branching fraction is calculated. Though there are many methods to calculate upper limits, the one employed in this analysis is based on a confidence interval derived using the Rolke-Lopez method. This method was introduced by Wolfgang Rolke and Angel López in their 2000 paper [99] as an alternative to the Feldman-Cousins interval. A key aspect of the Rolke-Lopez interval method is the treatment of the signal efficiency and background yield error as systematic uncertainties by profiling unlike the Feldman-Cousins interval which has no treatment of systematic uncertainties. This approach provides more robust intervals with better coverage for analyses featuring very low background, where knowledge of the background rate is difficult to ascertain precisely. The method is designed to provide a two-sided interval for small signals and an upper limit for larger background samples.

A description of the upper limit method as derived from [99] is provided below. Given an observed number of events x in the signal extraction region and y background events in the background region, the joint Poisson probability distribution of observing (x, y) given the signal μ and background b rates is given in Equation 4.14 below.

$$P_{\mu,b}(X = x, Y = y) = \frac{(\mu + b)^x}{x!} e^{-(\mu+b)} \cdot \frac{b^y}{y!} e^{-b} \tag{4.14}$$

The confidence interval is then obtained as follows. Starting with the null hypothesis $H_0 : \mu = \mu_0, b = b_0$ and a set of k observations for (u_i, v_i), the first step is to form a corresponding joint probability distribution. The Rolke-Lopez method discards all probabilities below 10^{-6}. The probabilities for each observation are sorted from highest to lowest, and partial sums of the probability are calculated starting with the highest up to and including the probability for the k_{th} observation. This is done until the desired $(1 - \alpha)$ limit is reached where α is the confidence level. If the observation (x, y) is in the list of (u_i, v_i) pairings that contribute to the aforementioned partial sum, H_0 is accepted, otherwise it is rejected. To form the confidence region based on this, this method is reversed and for a given observation (x, y), all the (μ, b) pairings that accept H_0 for a given $(1 - \alpha)$ limit are combined to form the confidence region. To include the systematic uncertainties in the upper limit calculation, a profile likelihood is built into the likelihood ratio based on the derivatives of the log-likelihoods. In our case, which involves a single Poisson yield equalling the signal plus background, a Gaussian background rate measurement and uncertainty, and a Gaussian signal efficiency measurement and uncertainty, the system of equations that arise from this can be solved analytically.

The Rolke-Lopez method is implemented within `ROOT` as the `TRolke` class [98]. The particular instance of the class used for this analysis is the one which assumes the uncertainty on the signal efficiency and background are Gaussian. For the signal efficiency, this is a reasonable assumption as the uncertainty on the signal efficiency is dominated by the statistical uncertainty of the large control channel study yield on data. For the background measurement, the uncertainty on the yield is derived from the integral of the 2D fit within the signal extraction region. The parabolic shape of the negative log-likelihood of the fit demonstrated that the uncertainties were Gaussian-like albeit asymmetric. Some correction to the Rolke-Lopez method will be investigated to manage this asymmetry.

CHAPTER 5

E_{extra} Cleanup Classifiers

The quantity E_{extra} refers to the total energy sum of electromagnetic clusters unaccounted for following the reconstruction of a collision event. It is an important quantity for a multitude of analyses at Belle II due to its strong signal-background separation power which makes it useful for signal extraction and background suppression. In the context of this $B^0 \to \ell^{\pm}\tau^{\mp}$ analysis, it is employed as one of the signal extraction variables which is detailed in Chapter 6. For correctly reconstructed signal events, E_{extra} peaks at 0 GeV, while for background events, the E_{extra} distribution broadens and peaks at higher values [14]. To increase the separation power of E_{extra}, and thus its impact as an analysis variable, ensuring a strong peak at 0 GeV for the signal distribution is imperative. In practice this is difficult as contributions to E_{extra} arise from energy deposits in the electromagnetic calorimeter created by beam background processes or calorimeter energy deposits that are split into multiple clusters during reconstruction, also referred to as fake photons. Such contributions will affect properly reconstructed signal events, making the distribution appear more background-like i.e. a broad nonzero peak. At the current SuperKEKB luminosity, fake photons make a larger contribution to E_{extra} than beam background clusters. However, as beam currents increase at SuperKEKB, beam background will continue to worsen, and so each contribution must be treated as important.

The aim for this E_{extra} study was to develop a tool that identifies beam background clusters and fake photons such that the E_{extra} distribution for signal events can be "cleaned" from such contributions. A multivariate analysis approach was chosen for the identification of fake photons and beam background clusters to allow for the information from various cluster-related variables to be used in combination. The classifier chosen is the `FastBDT` whose architecture is explained in Section 4.1.2. Two classifiers were developed to target beam background clusters and fake photons separately, the benefit of this being that each classifier could have its feature selection and training optimised for each contribution.

5.1 Data Samples

Signal, beam background, and fake photons are all selected from photon candidates reconstructed from run-dependent MC generated $B^0\overline{B}^0$ events. All considered photon candidates are required have a minimum cluster energy of 0.05 GeV and a polar angle between 17° and 150°. Additionally, the sum of crystal weights in the photon cluster is required to be more than 1.5. For non-overlapping clusters, this sum equals the total number of crystals in the cluster. For overlapping clusters, a division of the crystal energy occurs between the clusters leading to a non-integer value of the crystal weight sum.

Clusters are classified using information from the MC generators. The MC generators retain information about the energy deposited by individual particles in individual crystals, and then calorimeter clusters. This allows clusters whose energy is predominantly from a single particle to be *matched* to that particle in the simulation. Signal photons are defined using a cut which ensures the photon has been correctly reconstructed from the collision event. Signal photons that result from secondary decays simulated by `Geant4` are specifically vetoed using their corresponding `Geant4` process ID. To select beam background and fake photon clusters, the following procedure is used:

(1) Any cluster that is matched to a particle descended from the fundamental e^+e^- collision (e.g. B^- or D^- decay daughters, $e^+e^- \to q\bar{q}$ fragmentation products, initial- and final-state radiation) is rejected.
(2) For clusters passing this step, the total energy in the cluster due to particles descended from the e^+e^- collision is summed.
(3) Clusters where this sum $\geq$ 0.053 GeV are classified as fake photons.
(4) Clusters where this sum $\leq$ 0.025 GeV are classified as beam background clusters.

The thresholds in steps 3 and 4 were determined by studying distributions from various MC generated samples, with and without hadronic tracks, and with and without overlaid beam background.

In the case of both the beam background and fake photon classifiers, class 1 refers to signal photons, while class 0 signifies either beam background clusters or fake photons. The number of photons that remain for each simulated sample following the aforementioned cuts is provided in Table 5.1. A requirement of equal class sizes for each classifier was implemented to limit issues that may result from a class imbalance [100]. Holdout was implemented so that feature selection and hyperparameter tuning could be performed without biasing the test data. The ratio of the split between the training, validation and test data is 60:20:20 as reflected in Table 5.1.

5.2 Classifier Features

This section outlines the pool of features that were considered for both the beam background cluster and fake photon classifiers. These features consist of cluster-based and cluster-track distance variables that demonstrated some level of separation between signal photons and beam background clusters or fake photons. Feature selection was performed to determine, for each respective classifier, which of the following features are most significant. A descriptive list of the features are provided below:

(1) <u>*clusterE*</u> - the estimated energy of the photon which produced the cluster

TABLE 5.1. Number of signal photons and beam background clusters, or fake photons, used to train each classifier. The total number of photons present in each sample is presented, alongside the split for holdout.

(A) Beam background classifier

	Signal	**Beam Background**
Training	58,800	58,800
Validation	19,600	19,600
Test	19,600	19,600
Total	98,000	98,000

(B) Fake photon classifier

	Signal	**Fake**
Training	252,000	252,000
Validation	84,000	84,000
Test	84,000	84,000
Total	420,000	420,000

The clusters created by beam background are expected to predominantly have lower energies as they are the result of particle showers produced from scattering processes. For signal photons that are produced from the collision event, higher energies are achievable as they can be produced at any stage along the decay chain of the $\Upsilon(4S)$. Fake photons exhibit a trend in between as they are clusters produced from the collision event, but contain smaller proportions of the total energy deposited as they often result from such clusters being incorrectly divided. Distributions of this feature are presented in Figure 5.1a.

(2) *clusterLAT* - the lateral energy distribution of the cluster

$$S = \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^{n} w_i E_i r_i^2}$$

where $r_0 \approx 5$ cm, and for crystals sorted in decreasing order of energy (with E_0 the highest), the w_i and E_i are the weight and energy respectively of crystal i, and r_i is the distance from crystal i to a projection of the cluster onto a plane perpendicular to the photon trajectory [101]

Both this quantity and *clusterE1E9* contain information on the energy distribution, thus the same logic for *clusterE1E9* applies. Distributions of this feature are presented in Figure 5.1c.

(3) *clusterTheta* - the polar angle of the cluster

This variable is not highly discriminating between signal photons and fake photons as both result from particles produced at the interaction point, and thus predominantly have polar angles corresponding to the barrel region. However, as beam background processes occur at all points along the beam pipe, with some disproportionately producing particles at smaller polar angles or low momentum particles, beam background clusters are expected to have a higher occupancy in the endcaps, resulting in a polar angle skewed to the extremes. Distributions of this figure are presented in Figure 5.1d.

(4) *clusterZernikeMVA* - returns the output of a separate classifier that uses eleven of the Zernike moments of a cluster to identify the shape of energy distributions as belonging to hadronic particles (class 0) or photons (class 1)

Photons produce electromagnetic showers that are largely symmetric around the incident photon trajectory, whereas fake photons result from non-symmetric energy distributions. Beam background clusters are expected to exhibit the same behaviour as fake photons but to a lesser extent. Distributions of this feature are presented in Figure 5.1e.

(5) *clusterTiming* - the difference in timing $t - t_0$ between the time of the collision at the interaction point t_0 and the cluster time t

Beam background clusters can be produced any time there are electrons and positrons traveling through the beam pipe. Therefore, the timing distribution for such clusters would be broad, highly uniform and independent of the collision time t_0. However, for signal photons and fake photons, their timing distributions are expected to peak close to 0 as a result of them being produced shortly following collision events. Distributions of this feature are presented in Figure 5.1f.

(6) *minC2TDist* - the distance (in cm) between the cluster and its closest particle track (this is measured using the point of closest approach for the particle track extrapolated through the electromagnetic calorimeter)

This variable depends on the relative positions of clusters and particle tracks in the event. It is a key variable for fake photons as they correspond to clusters that are nearby the cluster matched to the original particle. Therefore the variable is expected to peak at low values. Both signal photons and beam background clusters are expected to have broader distributions peaked at higher values as the locations of their clusters are (to the lowest order) independent of the positions of other particles. Distributions of this feature are presented in Figure 5.1g.

(7) *clusterPulseShapeDiscriminationMVA* - returns the output of a classifier that identifies clusters as either electromagnetic showers (class 1) or hadronic showers (class 0) [102]

The response of this classifier is driven by the information obtained from fits to the scintillation pulse shape. Distributions for this feature are presented in Figure 5.1h.

5.3 Beam Background Classifier

In this section, the feature selection and fitting procedure for the beam background cluster classifier is presented. The data sample selection process for signal photons and beam background clusters is outlined in Section 5.1.

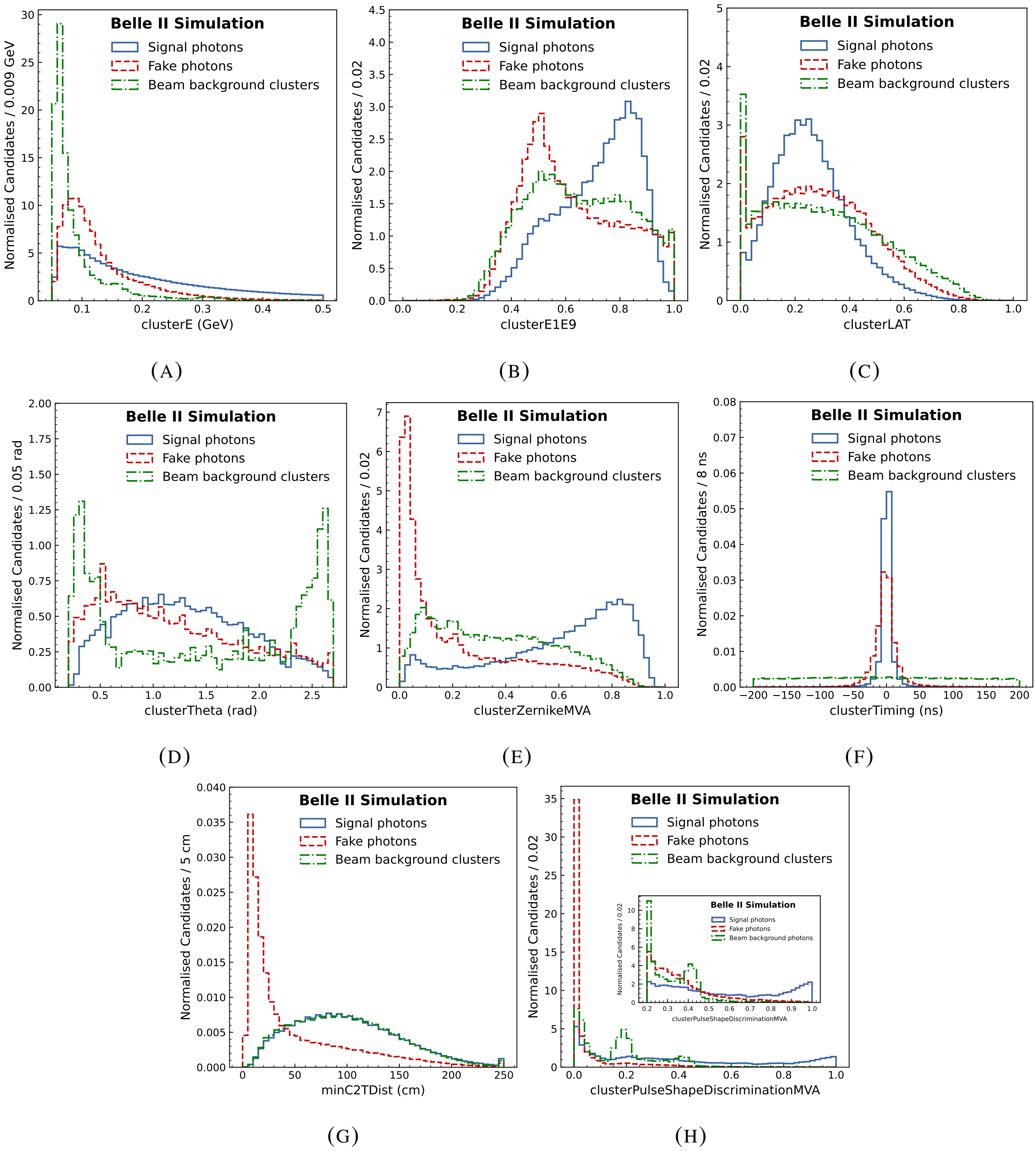


FIGURE 5.1. Distributions of the features for signal photons, fake photons and beam background clusters. All distributions are normalised to 1. An inset for *clusterPulseShapeDiscriminationMVA* is given to show the distributions in the tail region [0.2, 1].

5.3.1 Feature Selection

Internal to the `FastBDT` are calculations of the feature importance scores. These scores equal the sum of the information gain for each feature across all weak learners in the ensemble, and will be primarily used to perform the feature selection. A dedicated `FastBDT` classifier is built for the feature selection using the default hyperparameter values provided in Table 5.2. Using these default values to perform feature selection is not expected to have a significant impact on the importance scores as these values are sensible choices within the problem context, and will not severely impact the classifier performance [95].

TABLE 5.2. Default hyperparameter values for the `FastBDT` classifier. This set of hyperparameter values were used for feature selection.

Hyperparameter	**Value**
Number of trees	100
Maximum depth	3
Shrinkage	0.1
Data sample ratio	0.5

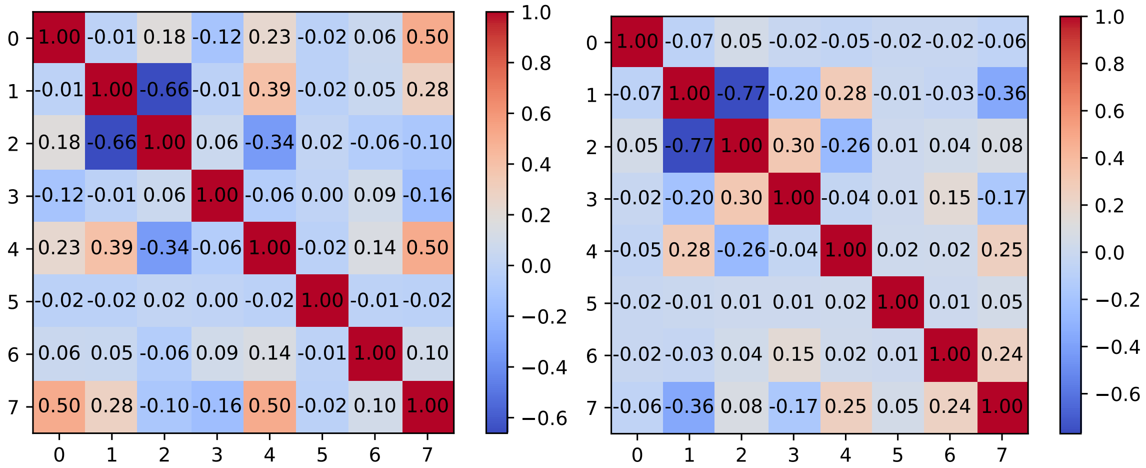
This classifier was trained on the training simulation samples. The importance score of each feature was obtained, and compared alongside a correlation matrix of the features to check for multicollinearity, and determine the validity of the importance scores. The importance scores are given in Table 5.3 and the correlation matrix is provided in Figure 5.2.

From the importance scores, it can be seen that the *minC2TDist* feature is effectively redundant for the beam background classifier as it has close to no information gain. The features *clusterLAT* and *clusterE1E9* also have low scores and from Figure 5.2, it can be seen that they have the highest multicollinearity between them for both signal photons and beam background clusters, so *clusterE1E9* is removed. For *clusterLAT*, in addition to the low importance score, the separation in the signal photon and beam background cluster distributions is minimal compared to the other remaining features, so *clusterLAT* is also removed.

TABLE 5.3. Importance scores for the features described in Section 5.2 sorted in descending order of importance. These are calculated for the beam background classifier fit using the default hyperparameter values.

Index	**Features**	**Internal Importance**
7	*clusterPulseShapeDiscriminationMVA*	0.4920
5	*clusterTiming*	0.4701
0	*clusterE*	0.0142
3	*clusterTheta*	0.0113
4	*clusterZernikeMVA*	0.0081
2	*clusterLAT*	0.0031
1	*clusterE1E9*	0.0010
6	*minC2TDist*	0.0001

Thus, the features that are selected for the beam background classifier are: *clusterE*, *clusterTheta*, *clusterZernikeMVA*, *clusterTiming* and *clusterPulseShapeDiscriminationMVA*.



(A) Signal photons (B) Beam background photons

FIGURE 5.2. Correlations between the features described in Section 5.2. The features are indexed according to Table 5.3. The three most important features are indexed as follows: *clusterPulseShapeDiscriminationMVA* (7), *clusterTiming* (5), *clusterE* (0).

5.3.2 Hyperparameter Tuning

The hyperparameters that were tuned were the number of trees and the maximum depth as the `FastBDT` classifier shows great variability in performance over a range of values for these hyperparameters [95]. The default values for all other hyperparameters, namely the sub-sample ratio and shrinkage, were used.

With the feature selection complete, the hyperparameter tuning is performed for the number of trees and the maximum depth of each weak learner. A simple grid search was performed for the following values:

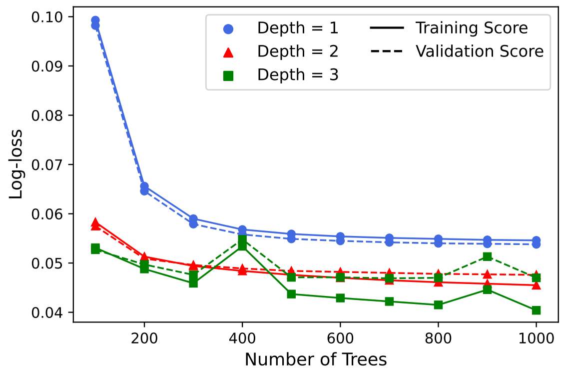
- Number of trees: 100 $\rightarrow$ 1000 in steps of 100
- Maximum depth: 1 $\rightarrow$ 6 in steps of 1

The log-loss values for each grid search pairing is compared between the training and validation data sets to check for overfitting. To mitigate the effect of random fluctuations introduced by the sub-sampling of training data, each grid search pairing was repeated 5 times, with the average log-loss results given in Table 5.4. The results are plotted in Figure 5.3. The ratio between the training and validation log-loss scores are provided in Table 5.5 and Figure 5.4 as a further check for overfitting. From Figure 5.3, for depths greater than 3, clear instability and overfitting is observed as there are increasingly large differences between the training and validation scores for most numbers of trees. This is supported by the ratio values which show steep decreases below one indicating increasingly better performances for the training data as opposed to the validation data. Such results demonstrate that such a maximum depth on this particular dataset yields individual weak learners that are too complex, and this issue is exacerbated when the classifier complexity increases with higher numbers of trees. Stability is seen for depths of 1 and 2, with the classifier performance improving rapidly up to 300 trees, before slowing down to the point of diminishing returns as more trees are added to the ensemble. A similar observation is made for a depth of 3, however beyond 400 trees, clear overfitting is seen to occur as shown by the steady decrease in the ratios in Figure 5.4.

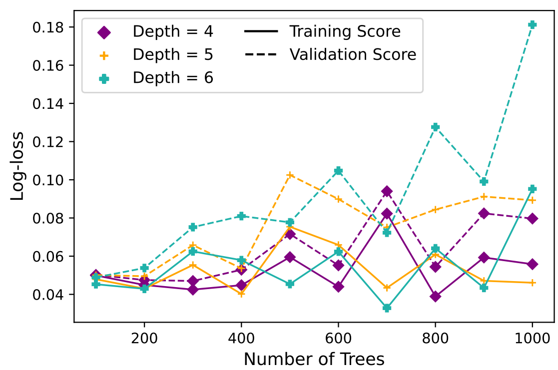
Decreasing the shrinkage parameter to values on the order of 10^{-2} to combat overfitting across various hyperparameter settings led to either no or minimal improvements, with shrinkage values below 0.05 not tested due to concerns over training and application times. The parameter values that provided the best log-loss values and exhibiting little signs of overfitting were (100, 3) and (300, 2) for the number of trees and maximum depth respectively, with the ratios and log-loss scores comparable between the two pairings. Although the best ratio and log-loss score is achieved for the (300, 2) pairing, due to the increase in application times for an ensemble of 300 trees [95], and to reduce the complexity of the classifier, the chosen values for the hyperparameters are

- Number of trees = 100
- Maximum depth = 3

The slight decrease in performance with the (100, 3) choice is not a concern as an effect of the log-loss score is that small differences are accentuated due to the log function. Therefore, the actual impact on the AUC score is minimal.



(A)



(B)

FIGURE 5.3. Plots of the log-loss scores presented in Table 5.4. Scores for depths 1-3 are shown in (A) while scores for depths 4-6 are in (B).

5.3.3 Training

A final training of the beam background classifier was conducted on the training and validation simulation samples combined. The classifier was evaluated with the test simulation samples. The ROC

TABLE 5.4. Log-loss values for the training and validation data sets calculated during the grid search for the beam background classifier.

(A) Training data

Max Depth	Number Of Trees									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
1	0.0993	0.0656	0.0590	0.0568	0.0559	0.0554	0.0551	0.0549	0.0547	0.0546
2	0.0583	0.0513	0.0494	0.0484	0.0476	0.0470	0.0465	0.0461	0.0458	0.0455
3	0.0531	0.0488	0.0459	0.0534	0.0437	0.0429	0.0422	0.0415	0.0446	0.0404
4	0.0499	0.0449	0.0425	0.0448	0.0595	0.0440	0.0823	0.0389	0.0593	0.0558
5	0.0479	0.0430	0.0554	0.0402	0.0754	0.0659	0.0434	0.0609	0.0471	0.0461
6	0.0453	0.0429	0.0625	0.0579	0.0454	0.0624	0.0328	0.0641	0.0434	0.0952

(B) Validation data

Max Depth	Number Of Trees									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
1	0.0982	0.0646	0.0579	0.0558	0.0549	0.0545	0.0542	0.0540	0.0539	0.0538
2	0.0575	0.0509	0.0496	0.0489	0.0484	0.0482	0.0480	0.0478	0.0477	0.0476
3	0.0527	0.0497	0.0476	0.0548	0.0471	0.0471	0.0469	0.0470	0.0513	0.0470
4	0.0502	0.0475	0.0470	0.0529	0.0716	0.0552	0.0940	0.0544	0.0824	0.0797
5	0.0501	0.0495	0.0658	0.0536	0.1025	0.0899	0.0751	0.0844	0.0912	0.0893
6	0.0489	0.0538	0.0752	0.0810	0.0777	0.1048	0.0723	0.1276	0.0991	0.1812

TABLE 5.5. Ratio of the beam background classifier training log-loss score to the validation log-loss score as documented in Table 5.4 for each corresponding parameter pairing.

Max Depth	Number Of Trees									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
1	1.0112	1.0155	1.0190	1.0179	1.0182	1.0165	1.0166	1.01666	1.0148	1.0149
2	1.0139	1.0079	0.9960	0.9898	0.9834	0.9751	0.9688	0.9644	0.9602	0.9559
3	1.0076	0.9819	0.9643	0.9745	0.9278	0.9108	0.8998	0.8829	0.8694	0.8596
4	0.9940	0.9453	0.9043	0.8469	0.8310	0.7971	0.8755	0.7151	0.7197	0.7001
5	0.9561	0.8687	0.8419	0.7500	0.7356	0.7330	0.5779	0.7216	0.5164	0.5162
6	0.9264	0.7974	0.8311	0.7148	0.5843	0.5954	0.4537	0.5024	0.4379	0.5254

curve and the corresponding AUC value, and distributions for the classifier output probability of class 1 is presented in Figure 5.5.

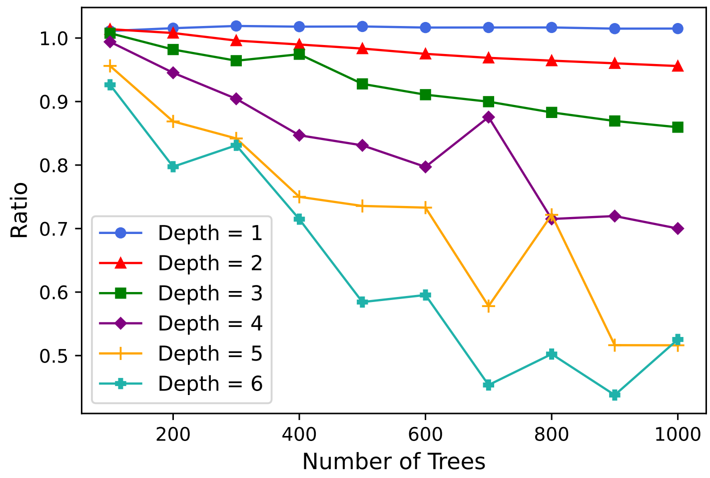
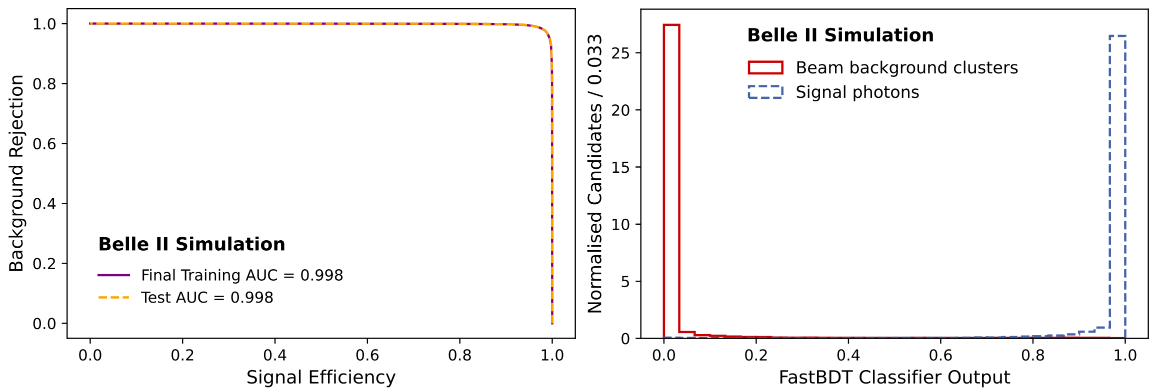


FIGURE 5.4. Plot of the ratios presented in Table 5.5.

The log-loss values for the training and test data sets was 0.0557 and 0.0560 respectively, which aligns with the results from Table 5.4.



(A)

(B)

FIGURE 5.5. The ROC curve with AUC (A) and classifier output probability distribution for class 1 (B) on the test data sets following a final training on the combined training and validation data sets. The distribution in (B) is normalised to 1.

The small log-loss value and high AUC value reflect the near perfect separation achieved by the classifier. Strong, narrow peaks at low and high classifier output values for beam background clusters and signal photons respectively demonstrates the high likelihood with which the classifier predicts the class of a given cluster. Such behaviour is expected when observing the well separated distributions for these classifier features in Figure 5.1, as well as the high importance scores of *clusterTiming*, *clusterPulseShapeDiscriminationMVA* and *clusterE* in Table 5.3 which show a large degradation on the ROC performance should these features be removed. These results clearly demonstrate the substantial information contained in these features, in particular, their distinguishable behaviours for each class which was discussed in Section 5.2. Moreover, these features can be seen to be largely uncorrelated with each other for beam background clusters when observing the matrices in Figure 5.2 indicating that the information content of the features is maximised - each feature contains new and distinct information on the timing and energy shape of the cluster allowing for a parameter space that yields high separation between the classes.

5.4 Fake Photon Classifier

This section details the feature selection and training procedure for the fake photon classifier. The simulation sample selection process is provided in Section 5.1.

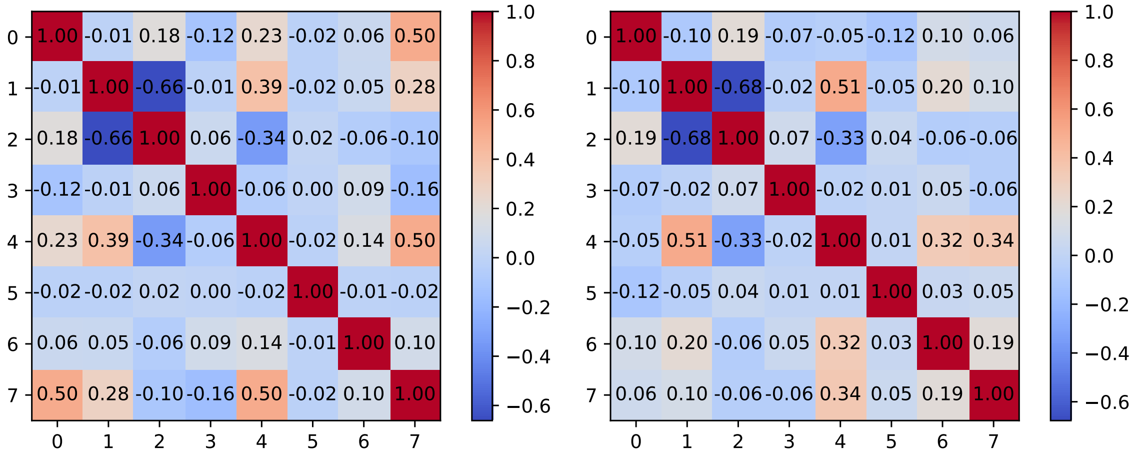
5.4.1 Feature Selection

To perform feature selection, a `FastBDT` classifier was built using the default hyperparameter values stated in Section 5.3.1. This classifier was fitted on the training data. The importance ranking of each feature was obtained, and compared alongside a correlation matrix of the features as before. The importance rankings are given in Table 5.6 and the correlation matrix is provided in Figure 5.6.

TABLE 5.6. Importance scores for the features described in Section 5.2 sorted in descending order of importance. These are calculated for the fake photon classifier fit using the default hyperparameter values.

Index	**Features**	**Importance**
7	*clusterPulseShapeDiscriminationMVA*	0.3768
6	*minC2TDist*	0.1883
4	*clusterZernikeMVA*	0.1882
0	*clusterE*	0.1098
3	*clusterTheta*	0.0543
5	*clusterTiming*	0.0413
2	*clusterLAT*	0.0305
1	*clusterE1E9*	0.0107

The feature with the lowest importance is *clusterE1E9* and it is highly correlated with *clusterLAT*, thus the decision to remove *clusterE1E9* was made. As there likely exists multicollinearity between *clusterE1E9* and *clusterLAT*, it is unclear how much of the *clusterLAT* importance score was compensated by *clusterE1E9*, but as *clusterE1E9* had a low importance score, it is likely there was little to none. Removing *clusterE1E9* would also help reduce multicollinearity between it and *clusterZernikeMVA* for fake photons.



(A) Signal photons

(B) Fake photons

FIGURE 5.6. Correlations between the features described in Section 5.2. Features are indexed according to Table 5.6. The three most important features are indexed as follows: *clusterPulseShapeDiscriminationMVA* (7), *minC2TDist* (6), *clusterZernikeMVA* (4).

Therefore, the final features chosen for the fake photon classifier are: *clusterE*, *clusterLAT*, *clusterTheta*, *clusterZernikeMVA*, *clusterTiming*, *minC2TDist* and *clusterPulseShapeDiscriminationMVA*.

5.4.2 Hyperparameter Tuning

Hyperparameter tuning for the number of trees and maximum depth were performed using a simple grid search over the same test values as before:

- Number of trees: $100 \rightarrow 1000$ in steps of 100
- Maximum tree depth: $1 \rightarrow 6$ in steps of 1

The log-loss values for each grid search pairing is compared between the training and validation data sets to check for overfitting. To mitigate the effect of random fluctuations introduced by the sub-sampling of training data, each grid search pairing was repeated 5 times, with the average log-loss results given in Table 5.7. The results are plotted in Figure 5.7. The ratio between the training and validation log-loss scores are provided in Table 5.8 and Figure 5.8 as before. As was the case with the beam background classifier, depths of 4 or more show clear signs of overfitting for all numbers of trees, with the overfitting more egregious than what was observed for the beam background classifier. This is particularly evident in the ratio plots in Figure 5.8 where the decrease is significantly steeper. This could be due to the complicated and varied mechanisms that produce fake photons, which can drive a loss of generality when attempting to build a classifier that works for all examples of fake photons. For maximum depths less than 4, the training is more stable, and a maximum depth of 3 is clearly the best performing. Beyond 400 trees, the ratio in Table 5.8 drops below 0.99 so the decision was to choose (300, 3), for the number of trees and maximum depth respectively, as the optimal set of hyperparameters. As with the beam background classifier, the shrinkage parameter was decreased for the value pairings where overfitting was observed, with only minimal improvements seen. Hence, to summarise, the hyperparameters chosen for the fake photon classifier are:

- Number of trees = 300
- Maximum depth = 3

Although there is a slight difference between the training and validation scores that might indicate some overfitting, a check using the AUC scores (which can be seen in Figure 5.9a) demonstrates that overfitting is small.

TABLE 5.7. Log-loss values for the training and validation data sets calculated during the grid search for the fake photon classifier.

(A) Training data

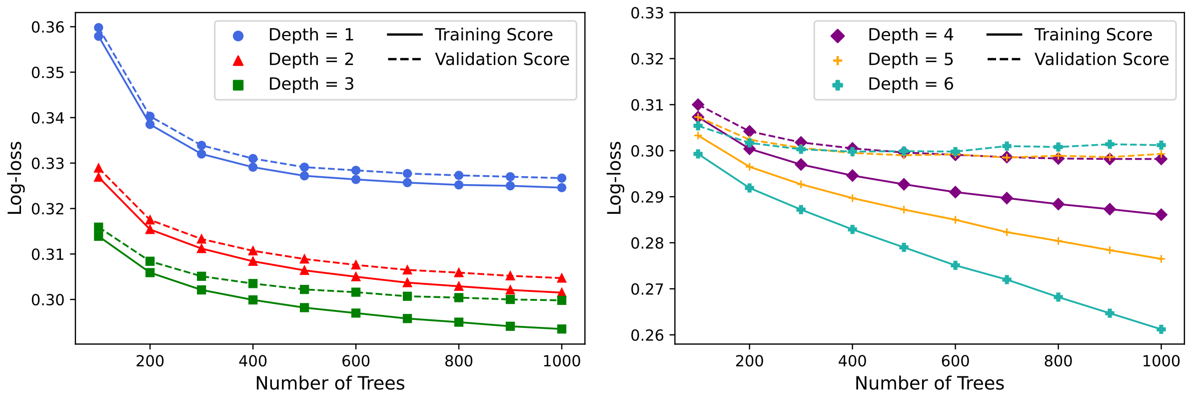
Max Depth	**Number Of Trees**									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
1	0.3579	0.3385	0.3320	0.3291	0.3272	0.3264	0.3257	0.3252	0.3250	0.3246
2	0.3269	0.3154	0.3112	0.3084	0.3064	0.3050	0.3037	0.3029	0.3021	0.3015
3	0.3139	0.3059	0.3021	0.2999	0.2982	0.2970	0.2958	0.2950	0.2941	0.2935
4	0.3073	0.3004	0.2970	0.2946	0.2927	0.2910	0.2897	0.2884	0.2873	0.2861
5	0.3033	0.2965	0.2927	0.2897	0.2872	0.2850	0.2823	0.2804	0.2784	0.2765
6	0.2993	0.2919	0.2872	0.2829	0.2790	0.2751	0.2720	0.2682	0.2647	0.2612

(B) Validation data

Max Depth	**Number Of Trees**									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
1	0.3598	0.3403	0.3339	0.3310	0.3291	0.3284	0.3277	0.3273	0.3270	0.3267
2	0.3289	0.3175	0.3133	0.3107	0.3089	0.3076	0.3065	0.3059	0.3052	0.3047
3	0.3159	0.3084	0.3051	0.3035	0.3022	0.3016	0.3007	0.3004	0.3000	0.2998
4	0.3100	0.3042	0.3018	0.3005	0.2996	0.2991	0.2986	0.2983	0.2982	0.2982
5	0.3073	0.3024	0.3006	0.2995	0.2990	0.2992	0.2985	0.2989	0.2986	0.2993
6	0.3054	0.3017	0.3003	0.2998	0.2999	0.2998	0.3010	0.3008	0.3014	0.3012

5.4.3 Training

A final training of the fake photon classifier was conducted on the training and validation data sets combined. The classifier was tested on the test simulation data, with the ROC curve and AUC value, and distributions for the classifier output probability of class 1, presented in Figure 5.9.



(A) (B)

FIGURE 5.7. Plots of the log-loss scores presented in Table 5.7. Scores for depths 1-3 are shown in (A) while scores for depths 4-6 are in (B).

TABLE 5.8. Ratio of the fake photon classifier training log-loss score to the validation log-loss score as documented in Table 5.7 for each corresponding parameter pairing.

Max Depth	**Number Of Trees**									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
1	0.9947	0.9947	0.9943	0.9943	0.9942	0.9939	0.9939	0.9936	0.9939	0.9936
2	0.9939	0.9934	0.9933	0.9926	0.9919	0.9915	0.9909	0.9902	0.9898	0.9895
3	0.9937	0.9919	0.9902	0.9881	0.9868	0.9847	0.9837	0.9820	0.9803	0.9790
4	0.9913	0.9875	0.9841	0.9804	0.9770	0.9729	0.9702	0.9668	0.9634	0.9594
5	0.9869	0.9805	0.9737	0.9673	0.9605	0.9525	0.9457	0.9381	0.9324	0.9238
6	0.9800	0.9675	0.9564	0.9436	0.9303	0.9176	0.9037	0.8916	0.8782	0.8672

The log-loss for the training and test set were 0.290 and 0.291 respectively which aligns with the values in Table 5.7.

Strong peaks for signal photons and fake photons can be seen in the output distribution, demonstrating that for the majority of examples, the classifier is relatively certain about its labelling of the classes. However, the long tails of the distributions along with the lower AUC value (compared to the beam background classifier) demonstrate cases that the classifier is struggling to classify. Presumably, these are cases where the fake photon cluster is distant from the cluster matched to the original ionising particle, and a large, symmetric energy deposition is made, causing the fake photon to appear more signal-like. Alternatively, signal photons that coincide with or occur nearby tracks could easily be mistaken as fake photons given the importance that *minC2TDist* has for the classifier. Ultimately, the impact of such specific edge cases, and the cause of the long tails in the output distributions, will require further investigations.

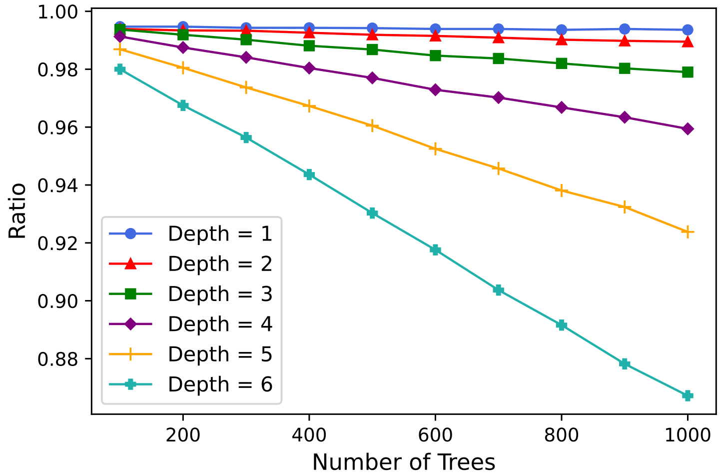


FIGURE 5.8. Plot of the ratios presented in Table 5.8.

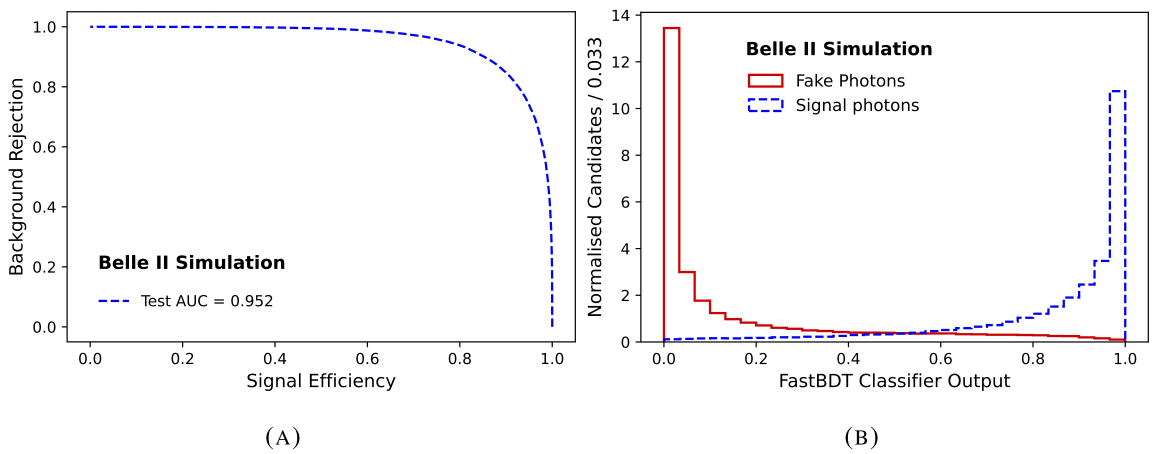


FIGURE 5.9. The ROC curve with AUC (A) and classifier output probability distribution for class 1 (B) on the test data set following a final training on the combined training and validation data sets. The distribution in (B) is normalised to 1.

5.5 Data-MC Agreement

As the classifiers are trained using MC-generated events but will be applied to data, the data-MC agreement of the used features alongside with the classifier outputs is checked to determine if using these classifiers will introduce any bias. To ensure valid comparisons, the latest run-dependent MC samples are used alongside the latest re-processing of the data. Furthermore, the following corrections are applied:

- *Data - Photon Energy Bias*

This is applied to data in order to correct for the mis-modelling of leakage effects of the photon energy in the ECL. The corrections are applied using the `correctEnergyBias` module from `basf2` [1].

- *MC - Photon Efficiency Corrections*
  This is applied to MC in order to correct for the difference in photon reconstruction efficiency between data and MC. The corrections consider the momentum, polar angle and azimuthal angle of the photons, and is applied using `basf2` [1].

Figure 5.10 shows the data-MC agreement of the features used in the classifiers. Overall, good agreement is observed between data and MC aside from the *clusterTiming* and *clusterPulseShapeDiscriminationMVA* features. However, these disagreements are to be expected. For *clusterTiming*, the disagreement lies in the long tails of the distribution which is only populated by beam background clusters. The many processes that can create a beam background cluster are difficult to model in MC and so disagreements are expected. As for *clusterPulseShapeDiscriminationMVA*, most of the disagreement is driven by the unavailability of specific corrections in the latest reprocessing of the data. Despite these disagreements, due to the importance of these two features for the classifiers, they remain in the set of features used.

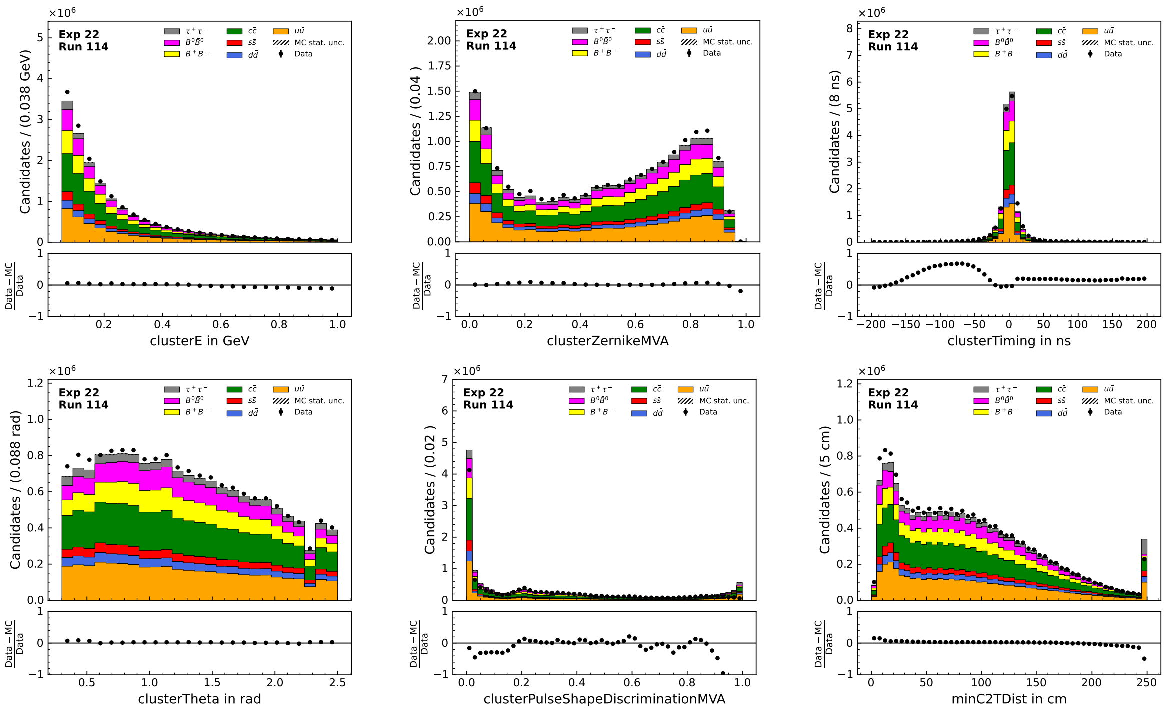


FIGURE 5.10. Data-MC agreement plots for the features used in the beam background and fake photon classifiers. The data originates from experiment 22 run 114.

A further check was done on the data-MC agreement of the classifier outputs in order to determine the impact of any disagreements for the features. The results are provided in Figure 5.11 and indicate an overall good agreement at the peaks of either end of the distributions where the majority of photons lie. The disagreement between the peaks is small for the fake photon classifier, but more pronounced for the beam background classifier. However, this is not a concern as beam background photons are typically low energy, and so the contribution to any disagreements within E_{extra} from this would be negligible.

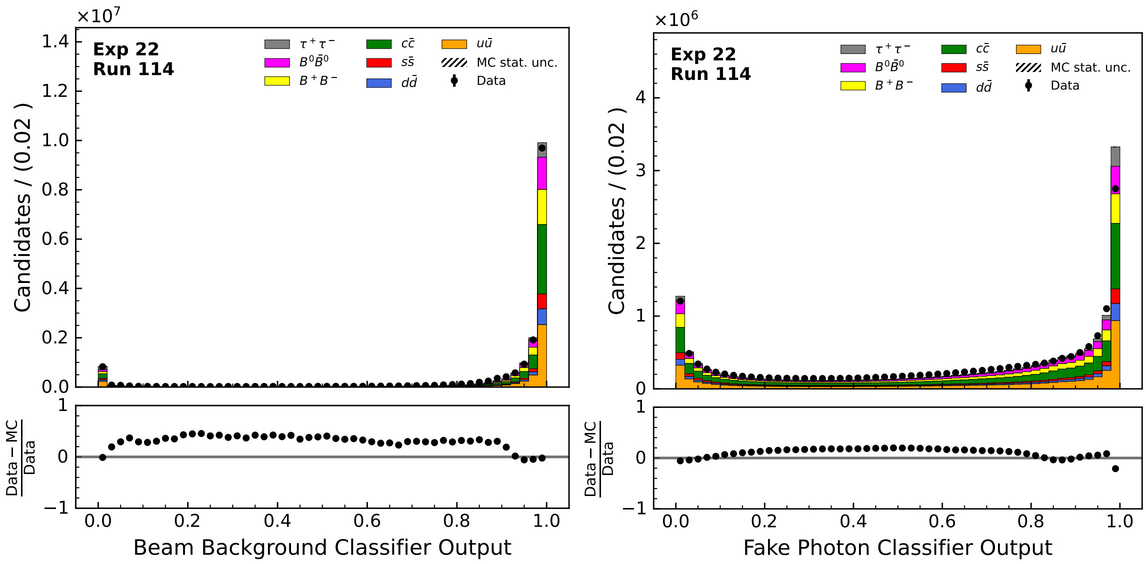


FIGURE 5.11. Data-MC agreement plots for the outputs of each classifier. The data originates from experiment 22 run 114.

CHAPTER 6

Search For $B^0 \to \ell^\pm \tau^\mp$

Following on from the context on relevant analysis techniques provided in Chapter 4, and the introduction of E_{extra} quantity and the cleanup classifiers given in Chapter 5, this chapter contains the crux of the thesis which is the search for $B^0 \to \ell^\pm \tau^\mp$.

Given that the aim of this analysis is to provide an analysis strategy for future Belle II searches which improves upon the sensitivity of prior analysis approaches, the chapter begins by focusing on the analysis strategy employed in this study and comparing it to the more relevant B-factory studies of BaBar and Belle. After this, the datasets used are introduced in Section 6.2 following which emphasis is given on discussing the selection criteria, cut flow, and efficiency in Section 6.3. Data-MC agreement checks are shown in Section 6.4. Detailed background breakdowns and discussions involving background estimation strategies are provided in Section 6.5. The last two sections, Sections 6.6 and 6.7 respectively, detail the control channel study performed for the signal efficiency estimate and the results including MC estimates for the branching fraction upper limits.

6.1 Analysis Strategy

Like previous B-factory studies, the B_{tag} is reconstructed hadronically. However, for this study, the latest `FastBDT`-based Belle II tagging algorithm called the Full Event Interpretation (FEI) was used [103]. The FEI yields an increased B_{tag} reconstruction efficiency compared to the tagging algorithm of Belle [62] and the discrete B_{tag} reconstruction of BaBar. A key point of difference between all $B^0 \to \ell^\pm \tau^\mp$ searches, including this one, is the treatment of the τ reconstruction. In this analysis, the τ modes that are reconstructed are provided in Table 6.1, and they total just under 97% of the total τ branching fraction. Unlike all prior studies, the analysis presented here considers leptonic and hadronic τ modes separately for the signal extraction procedure (hereafter referred to as the *leptonic partition* and *hadronic partition*). In the current version of the analysis, only the *leptonic partition* is used to derive the branching fraction upper limits. However, certain results regarding the *hadronic partition* stemming from initial studies are also presented. The *leptonic partition* uses p_ℓ^B as one dimension for signal extraction. This is the B daughter lepton (hereafter referred to as the signal lepton) momentum in the B_{sig} rest frame, and was used in the BaBar analysis as the sole quantity defining their signal region. To gain signal-background discrimination power, an additional dimension is added for both sets of modes. For the *leptonic partition*, the residual energy in the calorimeter E_{extra} is employed. The equivalent of this variable was used in the BaBar analysis as a clean-up cut for the individual τ modes. For the *hadronic partition*, the quantity ΔE_τ, as defined by Equation 2.9, was introduced as a clean-up cut for the hadronic τ modes in the BaBar analysis, is used.

Table 6.2 provides a comparison of the main differences between our analysis strategy and the strategies of the BaBar and Belle analyses. For the *leptonic partition*, the *signal extraction region* is defined as $p_\ell^B \in [2.24, 2.4] \times E_{\text{extra}} \in [0, 0.3]$; these values were optimised using the procedure described in Appendix A. The larger *fitting region* of $p_\ell^B \in [2, 2.4] \times E_{\text{extra}} \in [0, 1.5]$ is utilised to study the shape and composition of background events, and is ultimately fit to in order to obtain a background estimate for the *signal extraction region*. No such regions for the *hadronic partition* has been defined yet. Due to the mono-energetic signal lepton and the *signal extraction region* covering the kinematic endpoint of p_ℓ^B, very little background is left in this region. As a result, proper background estimations are a key aspect of the analysis. It should be noted that alternate fitting variables for p_ℓ^B and E_{extra} for the leptonic τ modes were investigated. The results of this are provided in Appendix B.

The upper limit on the branching fraction is obtained using the following quantities calculated within the *signal extraction region*: number of background events and its uncertainty; signal efficiency and its uncertainty; number of observed events. These quantities are combined to obtain a confidence interval using the Rolke-Lopez method as described in Section 4.3, from which the branching fraction upper limit is calculated.

TABLE 6.1. Table of reconstructed τ modes reconstructed with their branching fractions. The h denotes either a π or K.

	τ **Decay**	**Branching Fraction (%)**
Leptonic	$e^-\overline{\nu}_e\nu_\tau$	17.82 ± 0.04
	$\mu^-\overline{\nu}_\mu\nu_\tau$	17.39 ± 0.04
Hadronic	$h^-\nu_\tau$	11.51 ± 0.05
	$h^-\pi^0\nu_\tau$	25.93 ± 0.09
	$h^-\pi^0\pi^0\nu_\tau$	9.48 ± 0.10
	$h^-h^-h^+\nu_\tau$	9.80 ± 0.05
	$h^-h^-h^+\pi^0\nu_\tau$	4.76 ± 0.05
Total		96.69

6.2 Datasets

The data intended to be used for the final upper limit calculation corresponds to the entirety of the Belle II dataset collected between 2019-2022. Data-MC agreement checks were done using a 21.035

TABLE 6.2. Comparison between our analysis method and the BaBar and Belle analyses.

	BaBar	**Belle**	**Our Analysis**
B_{tag} *Reconstruction*	Hadronic - specific modes	Hadronic - FR	Hadronic - FEI
τ *Reconstruction*	$\ell\overline{\nu}\nu$, $\pi\nu$, $\pi\pi^0\nu, \pi\pi^0\pi^0\nu$, $\pi\pi\pi\nu$	Inclusive τ reconstruction	$\ell\overline{\nu}\nu$, $h\nu$, $h\pi^0\nu, h\pi^0\pi^0\nu$, $hhh\nu$, $hhh\pi^0\nu$
	Signal extracted from fit to all τ modes combined		τ modes partitioned into leptonic and hadronic modes for signal extraction
Signal Fit Region	$p_\ell^B \in [2.2, 2.4]$	$m_{\text{miss}} \in [1.4, 2.2]$	Leptonic τ $p_\ell^B \in [2, 2.4] \times E_{\text{extra}} \in [0, 1.5]$

fb^{-1} subset of this aforementioned data. MC samples produced using the official FEI skim for signal and background were used to derive and validate the analysis procedure. The latest run-independent MC samples with the nominal level of background overlay (BGx1) were used. The only exceptions to these were the specially produced MC samples used to estimate the *two true lepton* component of the background. This is explained in Section 6.5.2. The signal MC samples consisted of 3 million events for each $\ell = e, \mu$ mode. The background samples studied and presented here consist of $B^0\overline{B}^0$, B^+B^- and $q\overline{q}$ ($q = u, d, s, c$) components. The total MC used for the $B^0\overline{B}^0$ and B^+B^- components amounted to 26.8 ab^{-1} each (2.8 ab^{-1} from the official production, and 24 ab^{-1} from the specially produced sample), while the sample sizes for the $q\overline{q}$ events consisted of 25 ab^{-1} (1.0 ab^{-1} from the official production, and 24 ab^{-1} from the specially produced sample). For $\tau^+\tau^-$ events, 1.0 ab^{-1} of the official production was used to verify that none survive the selection criteria.

It is important to note the entirety of the MC samples currently used will be re-produced to make run-dependent versions, in which case the study presented here will be redone. Due to time and resource issues within the collaboration, run-dependent MC was not available during large portions of the time over which this work was conducted. As such, run-independent MC was the only option. Switching to run-dependent MC is not expected to cause major changes in the results as this study does not involve precision measurements.

6.3 Selection Criteria

A complete list of selection cuts is supplied in Table 6.3. Distributions of certain selection variables prior to any key signal-side selection cuts is given in Figures 6.1 and 6.2 for the $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ modes respectively. Additionally, distributions for the N_{tracks} in the rest of event (ROE) variable is given in Figure 6.3 to explain the difference in cut efficiency between the *leptonic* and *hadronic partitions*. Here, the ROE refers to all tracks and clusters that remain following the Υ(4S) reconstruction. Tables 6.4, 6.6, 6.7, 6.8, 6.9 and 6.10 represent cut-flow tables for signal and background events. Table 6.5 presents the cut-flow table for signal events in the *leptonic partition*. A combined cut-flow table for each mode are in Tables 6.11 and 6.13 for the $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ modes respectively. Tables 6.12 and 6.14 respectively provide the combined cut-flow table using the *leptonic partition*. All combined cut-flow tables were derived from a smaller MC sample than the entirety that was available.

TABLE 6.3. List of cuts applied following the reconstruction of the B_{tag} using the FEI.

Particle	**Cut Applied**
B_{tag}	$M_{\text{bc}} > 5.27$ GeV/c^2, $-0.15 < \Delta E < 0.1$ GeV, `sigProb` > 0.001
Signal lepton	$1.45 < p_{\text{cms}} < 2.55$ GeV/c BDT electron ID (BDT muon ID) > 0.9 for $\ell = e$ (μ)
τ daughters	$\lvert z_0\rvert < 2$ cm, $\lvert d_0\rvert < 0.5$ cm, $p_T > 0.1$ GeV/c BDT electron ID (BDT muon ID) > 0.9 for $\ell = e$ (μ) for leptonic modes
π^0 daughters	`clusterNHits` > 1.5 and $0.2967 <$ `clusterTheta` < 2.6180 rad $E > 0.08$ GeV in forward, $E > 0.055$ GeV in barrel, $E > 0.06$ GeV in backward `beamBackgroundSuppression` > 0.4
π^0	$0.120 < M_{\text{inv}} < 0.145$ GeV/c^2
ROE tracks	$\lvert z_0\rvert < 4$ cm, $\lvert d_0\rvert < 2$ cm, $p_{\text{T}} > 0.1$ GeV/c, N_{tracks} in ROE = 0
ROE clusters	$E > 0.055$ GeV, `inCDCAcceptance` $= 1$ `beamBackgroundSuppression` > 0.4, `fakePhotonSuppression` > 0.2
Event	B_{tag} `cosTBTO` < 0.9, $-0.2 < U_{\text{CM}} = E_{\text{miss}_{\text{CM}}} - \lvert\vec{p}_{\text{miss}_{\text{CM}}}\rvert < 1.5$ GeV

Reconstruction of a $B^0 \to \ell^\pm\tau^\mp$ event begins with the hadronic reconstruction of the B_{tag}, which is achieved through the FEI. The standard FEI skim cuts are as follows:

- Number of clean tracks in the event is $>= 3$. A clean track is defined as having $\lvert z_0\rvert < 2$ cm and $\lvert d_0\rvert < 0.5$ cm, where z_0 is the z coordinate and d_0 is the distance in $r - \phi$ plane to the point of closest approach to the interaction point. A minimum transverse momentum of $p_\mathrm{T} > 0.1$ GeV is required as well.

- Number of clean unmatched ECL clusters in the event is $>= 3$. A clean cluster is defined as having a minimum energy of $E > 0.1$ GeV, and a polar angle in the range of 0.296706 $< \theta <$ 2.61799 which corresponds to the CDC acceptance.
- The total measured energy of the event is > 4 GeV in the CMS frame.

Focusing on each set of the remaining cuts individually:

- B_tag *Candidates*
  A vertex fit is performed, with all candidates failing the fit removed. The signal probability, which is the output of the FEI `FastBDT` that describes the quality of the reconstructed B_tag, was kept at the skim-level cut of `sigProb` $>$ 0.001, with no signal probability cut applied for the $B^0 \to J/\psi K_S^0$ mode. Only the top 5 B_tag candidates according to the highest signal probability are kept for offline reconstruction. The cuts on the beam-constrained mass $M_\mathrm{bc} > 5.27$ GeV/c^2 and energy difference $-0.15 < \Delta E < 0.1$ GeV are required to ensure that any corresponding FEI corrections are valid. Equations 6.1 and 6.2 define the M_bc and ΔE quantities respectively.

$$M_\mathrm{bc} = \sqrt{E_\mathrm{CM}^2 - |\vec{p}_{B_\mathrm{tag}}|^2} \tag{6.1}$$

$$\Delta E = E_{B_\mathrm{tag}} - E_\mathrm{CM} \tag{6.2}$$

- *Signal Lepton and* τ *Lepton Daughters*
  The signal lepton is required to be the candidate with the highest momentum in the recoil frame of the top B_tag candidate. In some instances, due to poor reconstruction of the B_tag kinematics, the daughter lepton of the τ has a higher momentum and is thus chosen as the signal lepton. This cross-feed between the signal and τ daughter lepton occurs 1.67% (1.64%) of the time for the $\ell = e$ (μ) mode. Such events are still considered as signal. As performed in the BaBar analysis, to reduce clear background events, a loose p_cms cut on the signal lepton momentum in the CMS frame is applied with the BaBar cut value used. The same track-selection cuts on z_0, d_0 and p_T used in the FEI skim are applied for the τ daughters. The BDT-based electron and muon ID variables are used for the lepton ID cuts as they provide lower fake rates and were better behaved when modelling the *fake lepton* component of background (see Section 6.5.3). The standard threshold of 0.9 was chosen for the lepton ID.
- π^0 *Candidates and Photon Daughters*
  Values for the cuts on the invariant mass M_inv of the π^0, and quality selection cuts on the photon daughters, are unchanged from the standard list used in the collaboration. To remove photons that originate from beam background processes, a cut on the beam background classifier introduced in Chapter 5 of `beamBackgroundSuppression` > 0.4 was applied.

- *ROE Tracks*
  The standard skim-level ROE mask cuts are applied, with the additional cut on the number of ROE tracks = 0 to help remove background events. A distribution comparing the number of ROE tracks for signal and background is provided in Figure 6.3. Though there are signal events with one or more tracks left in the ROE, it can be seen that most of these cases are for *hadronic partition*. Therefore, the efficiency loss for the *leptonic partition* is minimal. This is represented in Table 6.5.
- *ROE Clusters*
  The standard skim-level cuts on the clusters being within the CDC acceptance and having a minimum energy are applied. Cuts on the `beamBackgroundSuppression` and `fakePhotonSuppression` are applied to clean-up the ROE clusters from beam background contributions and fake photons. The former represents the output of the fake photon classifier introduced in Chapter 5. The specific cut values chosen for each MVA correspond to the recommended starting values by the Belle II Performance group. Optimisation of these variables, and of the ROE cuts in general, remains a point of further work for this analysis when it is redone using run-dependent MC. However, an initial study into the optimal set of ROE cuts is presented in Appendix C.
- *Event Cuts*
  To suppress continuum events, the cut of `cosTBTO` < 0.9 is used. This quantity is defined as the cosine of the angle between the thrust axes of the B_{tag} and the ROE. The cut value of 0.9 is chosen as the currently available FEI corrections assume a value of 0.9 is used. To further suppress background, the quantity U_{CM} is used. This variable was originally used in various CLEO and BaBar analyses with missing energy, and is defined as $U_{\text{CM}} = E_{\text{miss}_{\text{CM}}} - |\vec{p}_{\text{miss}_{\text{CM}}}|$ where $E_{\text{miss}_{\text{CM}}}$ and $|\vec{p}_{\text{miss}_{\text{CM}}}|$ are the total missing energy and momentum of the event in the CMS frame. Distributions comparing this for signal and background is given in Figures 6.1 and 6.2 for the $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ modes respectively. In particular, a comparison between the *leptonic* and *hadronic partitions* is presented as they display notably different behaviours.

Bremsstrahlung corrections are applied to the signal lepton for the $\ell = e$ mode. Standard cuts on the photon energy for this are used: $E > 0.075$ GeV if in the backwards endcap, $E > 0.055$ GeV if in the barrel, or $E > 0.1$ GeV if in the forwards endcap. The best five Υ(4S) candidates ranked according to the lowest $|U_{\text{CM}}|$ are saved for offline analysis. Following all selection and *fitting region* cuts, the distribution of candidate multiplicity for signal and background are given in Figure 6.4. The candidate multiplicity distribution for $\tau^+\tau^-$ events is not included as no events survive after the selection cuts. Table 6.15 details the average number of candidates per event for signal and background.

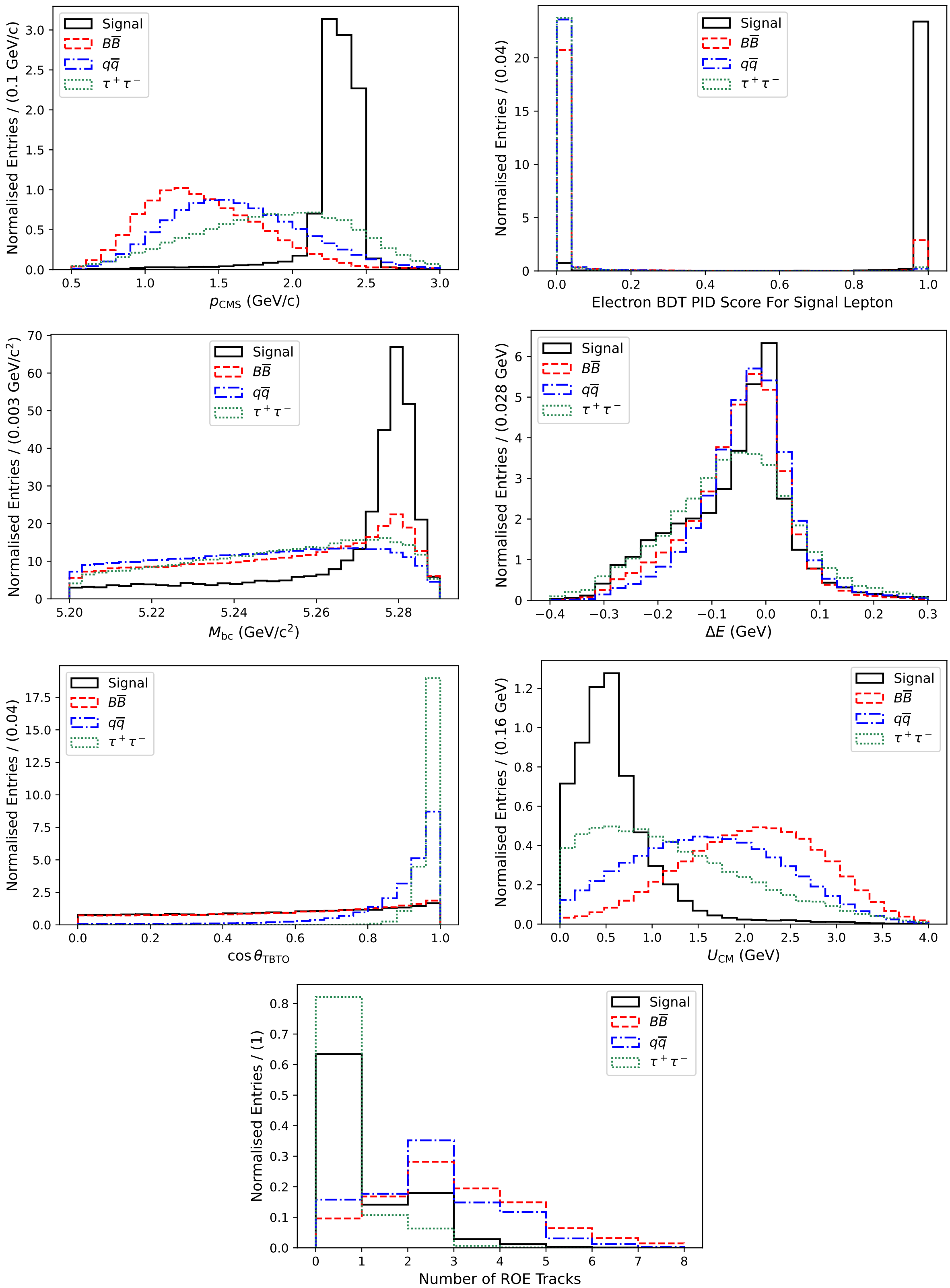


FIGURE 6.1. Distributions of various selection variables after reconstruction of the tag side and the application of FEI skim cuts, the B_{tag} vertex fit, and track (z_0, d_0, p_T) and π^0 selection cuts. Only the $B^0 \to e^{\pm}\tau^{\mp}$ mode is shown here.

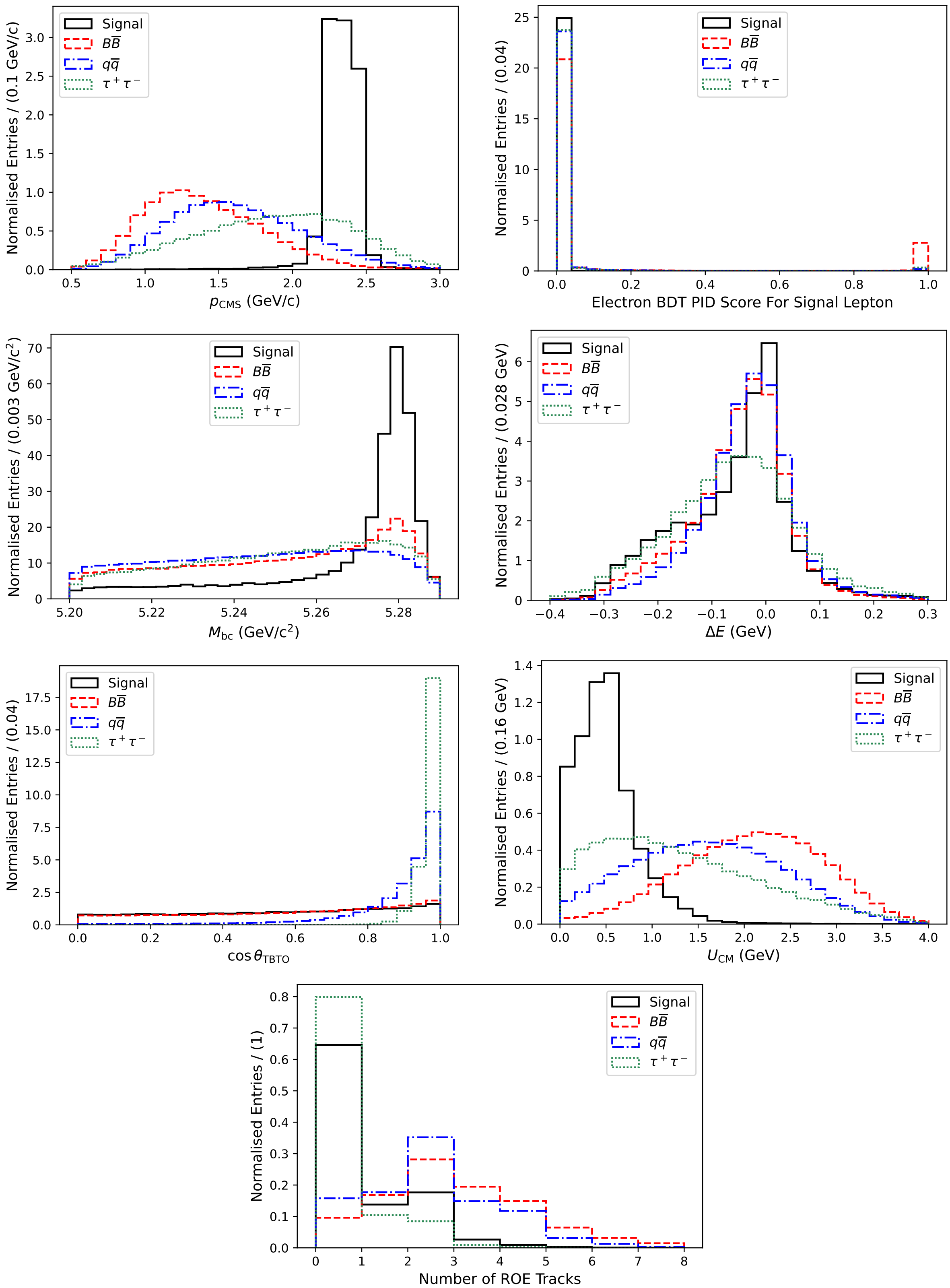
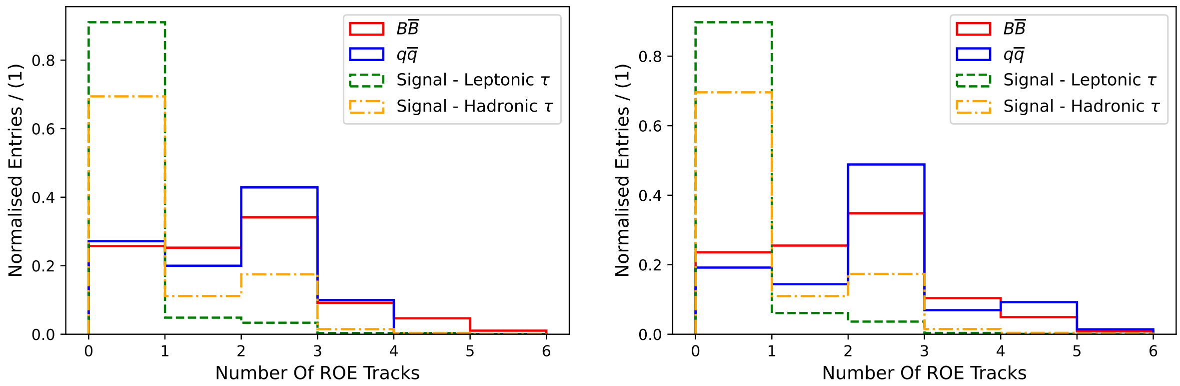


FIGURE 6.2. Distributions of various selection variables after reconstruction of the tag side and the application of FEI skim cuts, the B_{tag} vertex fit, and track (z_0, d_0, p_T) and π^0 selection cuts. Only the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode is shown here.



(A) $B^0 \to e\tau$ mode (B) $B^0 \to \mu\tau$ mode

FIGURE 6.3. Number of ROE tracks for background and continuum for each mode with the signal divided into the *leptonic* and *hadronic partitions*. This is after all other cuts have been applied.

6.3.1 Efficiencies and Cut Flow Tables

The following set of tables represent cut flow tables for signal and background events, with some representing just the *leptonic partition*. When considering all events, the efficiency is defined according to Equation 6.3, where n_{gen} is the total number of signal events generated (see Section 6.2) and n_{rec} is the total number of reconstructed signal events. When considering events in the *leptonic partition*, the definition of efficiency is as follows. For signal events, the n_{gen} in Equation 6.3 is scaled according to the leptonic τ branching fraction (35.21% according to PDG) while for background, the specific n_{gen} definition is provided in the description of the relevant tables. The tables include vetoes for the J/ψ*-like decay* background component which are described in detail in Section 6.5.4.

$$\epsilon = \frac{n_{\text{rec}}}{n_{\text{gen}}} \tag{6.3}$$

Distributions of the *fitting region* following all selection cuts and best candidate selection are given in Figure 6.10 for both modes. Tables comparing just the final efficiencies and event counts within the *fitting* and *signal extraction regions* for each sample are provided in Section 6.3.4.

TABLE 6.4. Cut flow table for reconstructed signal events. The **Total** column indicates the percentage of events that remain out of the total generated events, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	Total (%)	Kept (%)	Total (%)	Kept (%)
	$B^0 \to e^\pm\tau^\mp$ Mode		$B^0 \to \mu^\pm\tau^\mp$ Mode	
Generated	100	–	100	–
FEI skim	3.0532	3.053	3.7458	3.746
B_{tag} vertex fit	1.8894	61.883	2.3703	63.279
Track + π^0 selections	0.8124	42.998	0.9171	38.691
Signal lepton p_{cms}	0.7767	95.610	0.9042	98.593
Signal lepton PID	0.7738	99.627	0.7777	86.010
τ Daughter PID	0.7738	100	0.7777	100
M_{bc}	0.5845	75.537	0.5932	76.276
ΔE	0.4809	82.276	0.4859	81.917
$\cos\theta_{\text{TBTO}}$	0.4262	88.612	0.4288	88.236
U_{CM}	0.4221	99.054	0.4261	99.386
N_{tracks}	0.3358	79.548	0.3390	79.553
J/ψ veto	0.3279	97.647	0.3322	97.984
ψ(2S) veto	0.3235	98.648	0.3277	98.645
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	0.3103	95.940	0.3140	95.829
$p_\ell^B \in [2, 2.4]$	0.2958	95.317	0.3025	96.327

TABLE 6.5. Cut flow table for reconstructed signal events in the *leptonic partition*. The **Total** column indicates the percentage of events that remain out of the total generated leptonic τ events, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	**Total (%)**	**Kept (%)**	**Total (%)**	**Kept (%)**
	$B^0 \to e^{\pm}\tau^{\mp}$ Mode		$B^0 \to \mu^{\pm}\tau^{\mp}$ Mode	
Generated	100	–	100	–
FEI skim	1.7215	1.722	1.7116	1.712
B_{tag} vertex fit	1.0708	62.202	1.1231	65.617
Track + π^0 selections	0.6396	59.730	0.7041	62.693
Signal lepton p_{cms}	0.6185	96.698	0.6989	99.261
Signal lepton PID	0.6102	98.658	0.6100	87.280
Leptonic τ candidate (step 1 - candidate exists)	0.6102	100	0.6100	100
Leptonic τ candidate (step 2 - passes PID cuts)	0.4033	66.093	0.3937	64.541
M_{bc}	0.3725	92.363	0.3636	92.355
ΔE	0.3078	82.631	0.2921	80.336
$\cos\theta_{\text{TBTO}}$	0.2744	89.149	0.2562	87.770
U_{CM}	0.2702	98.469	0.2542	99.219
N_{tracks} in ROE	0.2572	95.189	0.2413	94.925
J/ψ veto	0.2479	96.384	0.2346	97.223
ψ(2S) veto	0.2425	97.822	0.2297	97.911
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	0.2390	98.557	0.2269	98.781
$p_{\ell}^{B} \in [2, 2.4]$	0.2283	95.523	0.2200	96.959

TABLE 6.6. Cut flow table for $B^0\overline{B}^0$ background events. The **Total** column indicates the percentage of events that remain out of the total generated, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	**Total (%)**	**Kept (%)**	**Total (%)**	**Kept (%)**
	$B^0 \to e^\pm\tau^\mp$ Mode		$B^0 \to \mu^\pm\tau^\mp$ Mode	
Generated	100	–	100	–
FEI skim	8.0980	8.098	8.0980	8.098
B_{tag} vertex fit	4.5577	56.282	4.5577	56.282
Track + π^0 selections	0.8157	17.897	0.8123	17.823
Signal lepton p_{cms}	0.3513	43.067	0.3511	43.223
Signal lepton PID	0.0925	26.331	0.0931	26.517
τ Daughter PID	0.0925	100	0.0931	100
M_{bc}	0.0583	63.027	0.0592	63.588
ΔE	0.0492	84.391	0.0489	82.601
$\cos\theta_{\text{TBTO}}$	0.0431	87.602	0.0428	87.526
U_{CM}	0.0318	73.782	0.0320	74.766
N_{tracks}	0.0102	32.076	0.0095	29.688
J/ψ veto	9.9178×10^{-3}	97.233	9.3425×10^{-3}	98.342
ψ(2S) veto	9.8904×10^{-3}	99.724	9.3151×10^{-3}	99.707
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	9.0959×10^{-3}	91.967	8.4932×10^{-3}	91.177
$p_\ell^B \in [2, 2.4]$	1.5616×10^{-3}	17.168	1.2603×10^{-3}	14.839

TABLE 6.7. Cut flow table for B^+B^- background events. The **Total** column indicates the percentage of events that remain out of the total generated, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	**Total (%)**	**Kept (%)**	**Total (%)**	**Kept (%)**
	$B^0 \to e^\pm\tau^\mp$ Mode		$B^0 \to \mu^\pm\tau^\mp$ Mode	
Generated	100	–	100	–
FEI skim	7.6015	7.602	7.6536	7.654
B_{tag} vertex fit	4.0781	53.649	4.1061	53.649
Track + π^0 selections	0.6502	15.944	0.6483	15.789
Signal lepton p_{cms}	0.2584	39.742	0.2585	39.874
Signal lepton PID	0.0516	19.969	0.0560	21.663
τ Daughter PID	0.0516	100	0.0560	100
M_{bc}	0.0185	35.853	0.0196	35.000
ΔE	0.0144	77.838	0.0156	79.592
$\cos\theta_{\text{TBTO}}$	0.0122	84.722	0.0131	83.974
U_{CM}	0.0090	73.770	0.0096	73.282
N_{tracks}	0.0028	31.111	3.1156×10^{-3}	32.454
J/ψ veto	2.7347×10^{-3}	97.668	3.0612×10^{-3}	98.254
ψ(2S) veto	2.7347×10^{-3}	100	3.0612×10^{-3}	100
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	2.4898×10^{-3}	91.045	2.8027×10^{-3}	91.556
$p_\ell^B \in [2, 2.4]$	1.6327×10^{-4}	6.558	3.2653×10^{-4}	11.651

TABLE 6.8. Cut flow table for $u\overline{u}$, $d\overline{d}$ and $s\overline{s}$ background events combined. The **Total** column indicates the percentage of events that remain out of the total generated, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	**Total (%)**	**Kept (%)**	**Total (%)**	**Kept (%)**
	$B^0 \to e^\pm\tau^\mp$ Mode		$B^0 \to \mu^\pm\tau^\mp$ Mode	
Generated	100	–	100	–
FEI skim	14.0823	14.082	14.0823	14.082
B_{tag} vertex fit	10.5863	10.586	10.5050	10.505
Track + π^0 selections	2.6041	24.599	2.6037	25.907
Signal lepton p_{cms}	1.7023	65.370	1.7025	65.388
Signal lepton PID	0.0185	1.087	0.0486	2.855
τ Daughter PID	0.0185	100	0.0486	100
M_{bc}	0.0045	24.324	0.0119	24.486
ΔE	0.0037	82.222	0.0101	84.874
$\cos\theta_{\text{TBTO}}$	0.0011	29.730	0.0032	31.683
U_{CM}	0.0008	72.727	0.0025	78.125
N_{tracks}	1.7799×10^{-4}	22.249	6.0938×10^{-4}	24.375
J/ψ veto	1.7391×10^{-4}	97.708	6.0122×10^{-4}	98.661
ψ(2S) veto	1.7188×10^{-4}	98.833	5.9511×10^{-4}	98.984
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	1.3247×10^{-4}	77.071	4.4497×10^{-4}	74.771
$p_\ell^B \in [2, 2.4]$	1.2228×10^{-5}	92.308	1.1413×10^{-4}	25.649

TABLE 6.9. Cut flow table for $c\bar{c}$ background events. The **Total** column indicates the percentage of events that remain out of the total generated, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	Total (%)	Kept (%)	Total (%)	Kept (%)
	$B^0 \to e^{\pm}\tau^{\mp}$ Mode		$B^0 \to \mu^{\pm}\tau^{\mp}$ Mode	
Generated	100	–	100	–
FEI skim	10.7597	10.760	10.7597	10.760
B_{tag} vertex fit	5.7843	53.759	5.7740	53.663
Track + π^0 selections	1.1501	19.883	1.1499	19.915
Signal lepton p_{cms}	0.5966	51.874	0.5968	51.900
Signal lepton PID	0.0138	2.313	0.0225	3.770
τ Daughter PID	0.0138	100	0.0225	100
M_{bc}	0.0031	22.464	0.0049	21.778
ΔE	0.0026	83.871	0.0040	81.633
$\cos\theta_{\text{TBTO}}$	0.0009	34.615	0.0013	32.500
U_{CM}	0.0005	55.556	0.0009	69.231
N_{tracks}	1.4509×10^{-4}	29.018	2.2545×10^{-4}	37.575
J/ψ veto	1.4063×10^{-4}	96.926	2.2545×10^{-4}	100
ψ(2S) veto	1.4063×10^{-4}	100	2.2545×10^{-4}	100
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	1.2054×10^{-4}	85.714	1.7411×10^{-4}	77.228
$p_{\ell}^{B} \in [2, 2.4]$	1.1161×10^{-5}	92.592	4.0179×10^{-5}	23.077

TABLE 6.10. Cut flow table for $\tau^+\tau^-$ background events. The **Total** column indicates the percentage of events that remain out of the total generated, while the **Kept** column shows the percentage that remains with respect to the previous line.

Cut	Total (%)	Kept (%)	Total (%)	Kept (%)
	$B^0 \to e^\pm\tau^\mp$ Mode		$B^0 \to \mu^\pm\tau^\mp$ Mode	
Generated	100	–	100	–
FEI skim	0.1365	0.137	0.1365	0.137
B_{tag} vertex fit	0.0770	56.410	0.0770	56.410
Track + π^0 selections	0.0126	16.364	0.0125	16.233
Signal lepton p_{cms}	0.0101	80.159	0.0101	80.800
Signal lepton PID	0.0002	1.980	0.0004	3.960
τ Daughter PID	0.0002	100	0.0004	100
M_{bc}	4.4070×10^{-5}	22.035	0.0001	25.000
ΔE	2.8292×10^{-5}	64.198	6.8009×10^{-5}	68.009
$\cos\theta_{\text{TBTO}}$	$< 5.4406 \times 10^{-7}$	0	2.1763×10^{-6}	3.200
U_{CM}	$< 5.4406 \times 10^{-7}$	–	5.4406×10^{-7}	24.816
N_{tracks} in ROE	$< 5.4406 \times 10^{-7}$	–	5.4406×10^{-7}	100
J/ψ veto	$< 5.4406 \times 10^{-7}$	–	5.4406×10^{-7}	100
ψ(2S) veto	$< 5.4406 \times 10^{-7}$	–	5.4406×10^{-7}	100
Fitting Regions				
$E_{\text{extra}} \in [0, 1.5]$	$< 5.4406 \times 10^{-7}$	–	5.4406×10^{-7}	100
$p_\ell^B \in [2, 2.4]$	$< 5.4406 \times 10^{-7}$	–	$< 5.4406 \times 10^{-7}$	0

6.3.2 Best Candidate Selection

Following all the selection and *fitting region* cuts in Table 6.3, the candidate multiplicity for each event is given in the distributions in Figure 6.4. The average number of candidates for each sample is presented in Table 6.15. Best candidate selection for this analysis serves two purposes: (1) it selects one candidate per event; (2) it separates events into either the *leptonic* or *hadronic partition* depending on the reconstructed mode of the τ for the chosen candidate. Therefore, the strategy chosen for this analysis is to choose the Υ(4S) candidate using the mode and quality of the reconstructed τ. In an event where a leptonic reconstruction of the τ exists, the corresponding Υ(4S) candidate is chosen as leptonic

TABLE 6.11. Combined cut-flow table for signal and background events reconstructed in the $B^0 \to e^\pm\tau^\mp$ mode. The values are the percentage of events that remain out of the total generated.

Cuts	**Signal**	$\boldsymbol{B^0\overline{B}^0}$	$\boldsymbol{B^+B^-}$	$\boldsymbol{uds}$	$\boldsymbol{c\bar{c}}$	$\boldsymbol{\tau^+\tau^-}$
Generated	100	100	100	100	100	100
FEI skim	3.0532	8.0980	7.6015	14.082	10.760	0.1365
B_{tag} vertex fit	1.8894	4.5577	4.0781	10.586	5.7843	0.0770
Track + π^0 selections	0.8124	0.8157	0.6502	2.6041	1.1501	0.0126
Signal lepton p_{cms}	0.7767	0.3513	0.2584	1.7023	0.5966	0.0101
Signal lepton PID	0.7738	0.0925	0.0516	0.0185	0.0138	0.0002
τ Daughter PID	0.7738	0.0925	0.0516	0.0185	0.0138	0.0002
M_{bc}	0.5845	0.0583	0.0185	0.0045	0.0031	4.4070×10^{-5}
ΔE	0.4809	0.0492	0.0144	0.0037	0.0026	2.8292×10^{-5}
$\cos\theta_{\mathrm{TBTO}}$	0.4262	0.0431	0.0122	0.0011	0.0009	$< 5.4406 \times 10^{-7}$
U_{CM}	0.4221	0.0318	0.0090	0.0008	0.0005	$< 5.4406 \times 10^{-7}$
N_{tracks} in ROE	0.3358	0.0102	0.0028	0.0002	0.0001	$< 5.4406 \times 10^{-7}$
J/ψ veto	0.3279	0.0099	0.0027	0.0002	0.0001	$< 5.4406 \times 10^{-7}$
ψ(2S) veto	0.3235	0.0099	0.0027	0.0002	0.0001	$< 5.4406 \times 10^{-7}$
Fitting Regions						
$E_{\mathrm{extra}} \in [0, 1.5]$	0.3103	0.0091	0.0025	0.0001	0.0001	$< 5.4406 \times 10^{-7}$
$p_\ell^B \in [2, 2.4]$	0.2958	0.0016	0.0002	1.2228×10^{-5}	1.1161×10^{-5}	$< 5.4406 \times 10^{-7}$

TABLE 6.12. Combined cut-flow table for signal and background events reconstructed in the $B^0 \to e^\pm\tau^\mp$ mode for the *leptonic partition*. The percentages are calculated using the total generated leptonic τ events for signal, while it is the total generated events for the background.

Cuts	Signal	$B^0\overline{B}^0$	B^+B^-	uds	$c\bar{c}$	$\tau^+\tau^-$
Generated	100	100	100	100	100	100
FEI skim	1.7215	8.0980	7.6015	14.082	10.760	0.1365
$B_{\rm tag}$ vertex fit	1.0708	4.5577	4.0781	10.586	5.7843	0.0770
Track + π^0 selections	0.6396	0.8157	0.6502	2.6041	1.1501	0.0126
Signal lepton $p_{\rm cms}$	0.6185	0.3513	0.2584	1.7023	0.5966	0.0101
Signal lepton PID	0.6102	0.0925	0.0516	0.0185	0.0138	0.0002
Leptonic τ candidate (step 1 - candidate exists)	0.6102	–	–	–	–	–
Leptonic τ candidate (step 2 - passes PID cuts)	0.4033	0.0119	0.0047	0.0021	0.0011	3.5364×10^{-5}
$M_{\rm bc}$	0.3725	0.0085	0.0016	0.0005	0.0003	7.0729×10^{-6}
ΔE	0.3078	0.0073	0.0013	0.0004	0.0002	3.8085×10^{-6}
$\cos\theta_{\rm TBTO}$	0.2744	0.0064	0.0001	0.0001	7.5893×10^{-5}	$< 5.4406 \times 10^{-7}$
$U_{\rm CM}$	0.2702	0.0036	0.0011	0.0001	2.9018×10^{-5}	$< 5.4406 \times 10^{-7}$
$N_{\rm tracks}$ in ROE	0.2572	0.0013	0.0001	5.4348×10^{-6}	6.6964×10^{-5}	$< 5.4406 \times 10^{-7}$
J/ψ veto	0.2479	0.0013	0.0001	5.4348×10^{-6}	6.6964×10^{-5}	$< 5.4406 \times 10^{-7}$
ψ(2S) veto	0.2425	0.0013	0.0001	5.4348×10^{-6}	6.6964×10^{-5}	$< 5.4406 \times 10^{-7}$
Fitting Regions						
$E_{\rm extra} \in [0, 1.5]$	0.2390	0.0012	0.0001	$< 6.7935 \times 10^{-7}$	6.6964×10^{-5}	$< 5.4406 \times 10^{-7}$
$p_\ell^B \in [2, 2.4]$	0.2283	0.0002	1.9710×10^{-5}	$< 6.7935 \times 10^{-7}$	$< 2.2321 \times 10^{-5}$	$< 5.4406 \times 10^{-7}$

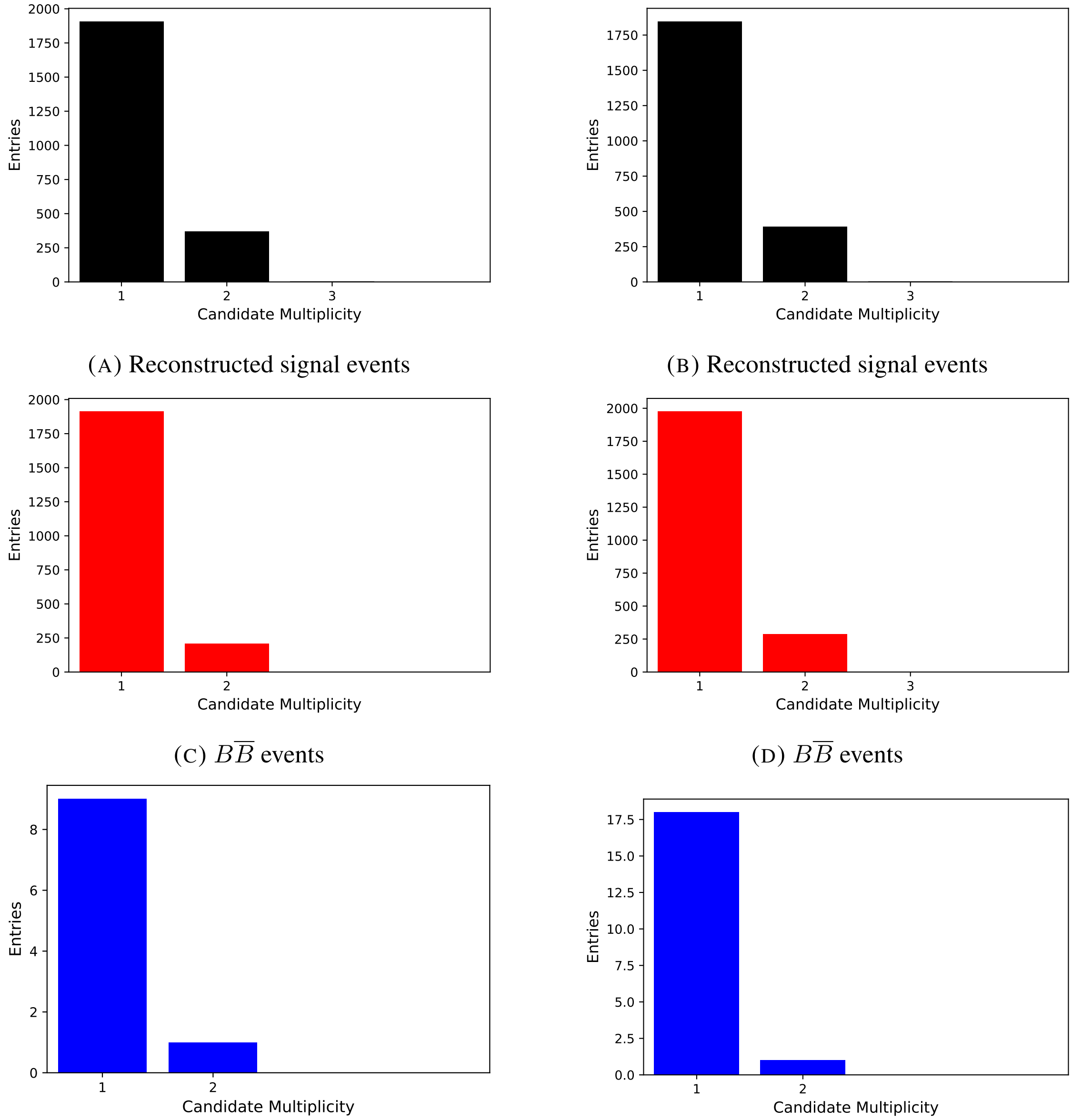
TABLE 6.13. Combined cut-flow table for signal and background events reconstructed in the $B^0 \to \mu^\pm \tau^\mp$ mode. The values are the percentage of events that remain out of the total generated. The last line provides the raw number of events the remain from the original MC sample.

Cuts	**Signal**	$B^0\overline{B}^0$	B^+B^-	uds	$c\bar{c}$	$\tau^+\tau^-$
Generated	100	100	100	100	100	100
FEI skim	3.7458	8.0980	7.6536	14.082	10.760	0.1365
B_{tag} vertex fit	2.3703	4.5577	4.1061	10.505	5.7740	0.0770
Track + π^0 selections	0.9171	0.8123	0.6483	2.6037	1.1499	0.0125
Signal lepton p_{cms}	0.9042	0.3511	0.2585	1.7025	0.5968	0.0101
Signal lepton PID	0.7777	0.0931	0.0560	0.0486	0.0225	0.0004
τ Daughter PID	0.7777	0.0931	0.0560	0.0486	0.0225	0.0004
M_{bc}	0.5932	0.0592	0.0196	0.0119	0.0049	0.0001
ΔE	0.4859	0.0489	0.0156	0.0101	0.0040	6.8009×10^{-5}
$\cos\theta_{\text{TBTO}}$	0.4288	0.0428	0.0131	0.0032	0.0013	2.1763×10^{-6}
U_{CM}	0.4261	0.0320	0.0096	0.0025	0.0009	5.4406×10^{-7}
N_{tracks} in ROE	0.3390	0.0095	0.0031	0.0006	0.0002	5.4406×10^{-7}
J/ψ veto	0.3322	0.0093	0.0031	0.0006	0.0002	5.4406×10^{-7}
ψ(2S) veto	0.3277	0.0093	0.0031	0.0006	0.0002	5.4406×10^{-7}
Fitting Regions						
$E_{\text{extra}} \in [0, 1.5]$	0.3140	0.0085	0.0028	0.0004	0.0002	5.4406×10^{-7}
$p_\ell^B \in [2, 2.4]$	0.3025	0.0016	0.0003	0.0001	4.0179×10^{-5}	$< 5.4406 \times 10^{-7}$

TABLE 6.14. Combined cut-flow table for signal and background events reconstructed in the $B^0 \to \mu^\pm\tau^\mp$ mode for the *leptonic partition*. The values are the percentage of events that remain out of the total generated leptonic τ events for signal, while it is the total generated events for the background. As no events with a true leptonic τ candidate exist for background, those entries are left blank in the table.

Cuts	**Signal**	$\boldsymbol{B^0\overline{B}^0}$	$\boldsymbol{B^+B^-}$	$\boldsymbol{uds}$	$\boldsymbol{c\bar{c}}$	$\boldsymbol{\tau^+\tau^-}$
Generated	100	100	100	100	100	100
FEI skim	1.7116	8.0980	7.6536	14.082	10.760	0.1365
B_{tag} vertex fit	1.1231	4.5577	4.1061	10.505	5.7740	0.0770
Track + π^0 selections	0.7041	0.8123	0.6483	2.6037	1.1499	0.0101
Signal lepton p_{cms}	0.6989	0.3511	0.2585	1.7025	0.5968	0.0004
Signal lepton PID	0.6100	0.0931	0.0560	0.0486	0.0225	0.0004
Leptonic τ candidate (step 1 - candidate exists)	0.6100	–	–	–	–	–
Leptonic τ candidate (step 2 - passes PID cuts)	0.3937	0.0121	0.0050	0.0027	0.0016	1.5234×10^{-5}
M_{bc}	0.3636	0.0084	0.0017	0.0007	0.0004	4.3526×10^{-6}
ΔE	0.2921	0.0068	0.0012	0.0006	0.0003	2.1763×10^{-6}
$\cos\theta_{\text{TBTO}}$	0.2562	0.0058	0.0010	0.0002	0.0001	$< 5.4406 \times 10^{-7}$
U_{CM}	0.2542	0.0033	0.0005	0.0001	2.4554×10^{-5}	$< 5.4406 \times 10^{-7}$
N_{tracks} in ROE	0.2413	0.0010	0.0001	5.4348×10^{-6}	4.4643×10^{-6}	$< 5.4406 \times 10^{-7}$
J/ψ veto	0.2346	0.0009	0.0001	5.4348×10^{-6}	4.4643×10^{-6}	$< 5.4406 \times 10^{-7}$
ψ(2S) veto	0.2297	0.0009	0.0001	5.4348×10^{-6}	4.4643×10^{-6}	$< 5.4406 \times 10^{-7}$
Fitting Regions						
$E_{\text{extra}} \in [0, 1.5]$	0.2269	0.0009	0.0001	2.0380×10^{-6}	2.2321×10^{-6}	$< 5.4406 \times 10^{-7}$
$p_\ell^B \in [2, 2.4]$	0.2200	0.0002	1.8652×10^{-5}	2.0380×10^{-6}	$< 2.2321 \times 10^{-6}$	$< 5.4406 \times 10^{-7}$

τ modes are easy to reconstruct well with particle ID requirements. Meanwhile, hadronic τ decays typically suffer from multiple points of mis-reconstruction and cross-feed, therefore often resulting in poorer quality $\Upsilon(4S)$ candidates. This idea lends itself to the use of a flowchart to select the best $\Upsilon(4S)$ candidate per event. A diagram of the flowchart is given in Figure 6.5.



(A) Reconstructed signal events

(B) Reconstructed signal events

(C) $B\overline{B}$ events

(D) $B\overline{B}$ events

(E) $q\overline{q}$ events

(F) $q\overline{q}$ events

FIGURE 6.4. Distribution of candidate multiplicities for reconstructed signal and background events. The $B^0 \to e^{\pm}\tau^{\mp}$ mode is shown on the left while the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode is on the right.

TABLE 6.15. Average number of candidates per event for reconstructed signal and background following all the selection cuts in Table 6.3.

MC Sample	**Multiplicity**	
	$B^0 \to e^\pm\tau^\mp$ Mode	$B^0 \to \mu^\pm\tau^\mp$ Mode
Reconstructed Signal	1.16	1.18
$B\overline{B}$	1.09	1.13
$q\overline{q}$	1.10	1.05

Further details about the logic for each of the flowchart steps is given below:

- The leptonic τ modes are the cleanest to reconstruct so those are prioritised first. A threshold on the lepton ID is used to determine if a clear τ daughter lepton is present. If multiple candidates from a given event pass the lepton ID requirement, the candidate selected is the one with the highest lepton ID.
- Among the hadronic τ modes, the 3-prong decay, including the mode with the additional π^0, is focused on next due to the high track multiplicity which makes these modes harder to fake. If the τ is reconstructed in either of these modes, the corresponding Υ(4S) candidate is chosen. If multiple candidates have a τ reconstructed in either mode, then the candidate with the lowest E_{extra} is kept.
- For a similar reasoning as the previous point, except with π^0 multiplicities instead, the $h\pi^0\pi^0$ mode is focused on next. If the τ is reconstructed in this mode, the corresponding Υ(4S) candidate is chosen, and if multiple candidates are chosen, the best one is selected using E_{extra}.
- The $h\pi^0$ mode is often mis-reconstructed from $h\pi^0\pi^0$ events where a π^0 is not reconstructed. The resulting residual energy in the calorimeter yields higher E_{extra} values, hence the use of the $E_{\text{extra}} < 0.4$ GeV threshold. If multiple Υ(4S) candidates yield an E_{extra} lower than 0.4 GeV, the best candidate for the event is chosen using E_{extra}.
- For all events that remain (in other words, no best candidate selection has taken place yet), the Υ(4S) candidates in each event are ranked according to E_{extra} with the lowest one selected.

The performance of the flowchart in classifying events is shown in the cross-feed tables in Section 6.3.3. Errors in the reconstruction that limit the classification performance are also discussed there. A check was done to determine if this best candidate selection method led to any sculpting of the background shape. This was achieved by comparing the above flowchart-based method to a method based on randomly selecting a candidate from each event. A check on the shape of the *fitting region* along each dimension was performed, with the distributions provided in Figures 6.6, 6.7, 6.8 and 6.9. No major differences in shape between the two methods were observed apart from a few cases. For the $q\overline{q}$ events in the *leptonic partition*, changes in shape are consistent with the effect of low statistics, and the

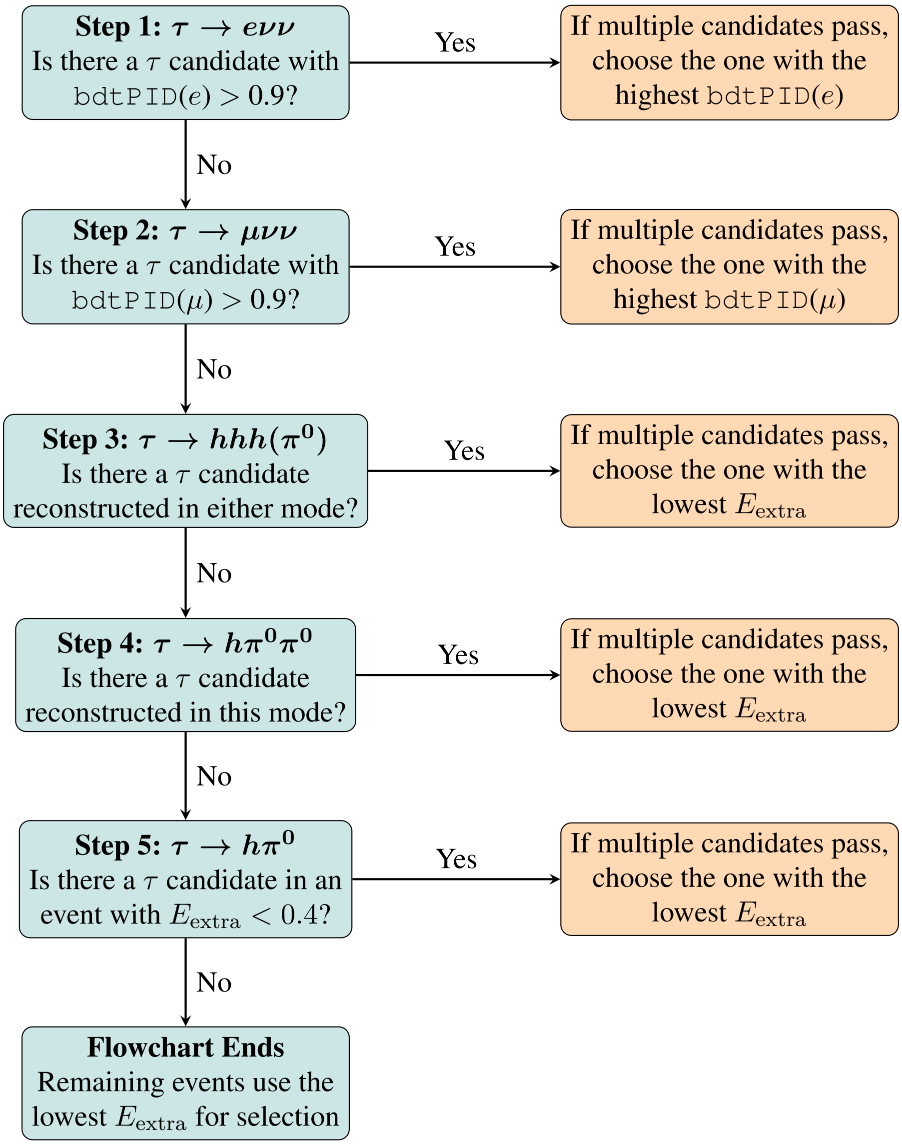


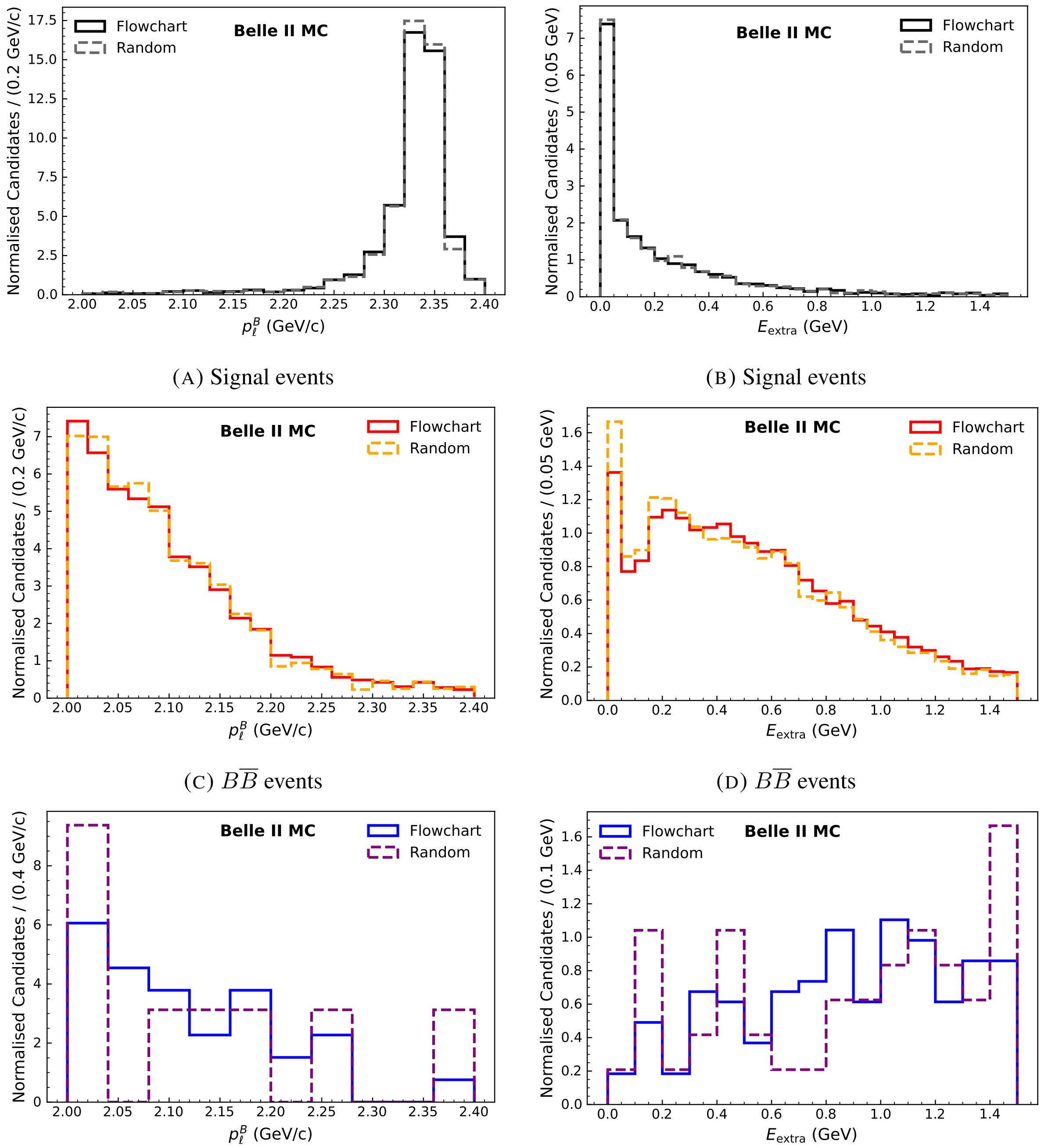
FIGURE 6.5. Flowchart depicting the stages of the best candidate selection process. In Step 3, the considered τ modes are $\tau \to hhh$ and $\tau \to hhh\pi^0$.

overall trend of the distribution is preserved. For reconstructed signal events in the *hadronic partition*, an increase in the peaked component is seen, likely attributed to the better quality candidates chosen.

In its current form, the best candidate selection is performed after the *fitting region* cuts are imposed. A common alternative is to instead perform the best candidate selection prior to applying such cuts. Using this alternative setup, checks were performed on the candidate multiplicity and background count. Regarding the average candidate multiplicity, the results were the following: 2.32 for reconstructed signal, 3.07 for $B\overline{B}$ and 2.87 for $q\overline{q}$ events for the $B^0 \to e^{\pm}\tau^{\mp}$ mode; 2.29 for reconstructed signal, 3.06 for $B\overline{B}$ and 2.88 for $q\overline{q}$ events for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode. Comparing this with the multiplicities of the current approach as shown in Table 6.15 demonstrates that imposing the *fitting region* cuts prior to the best candidate selection is favourable. A further check was done to see if the alternative approach led to significant changes in the background count. However, the results showed a decrease in the total background of only 1 event for each mode. Given this minor change, and the large improvement in candidate multiplicity, it was concluded that the alternative approach should not be used.

It should also be noted that the flowchart went through various stages of development before reaching its final form as presented in Figure 6.5. The most notable developments were the following:

- Instead of using E_{extra} to decide between multiple events that may pass any given step from *Step 3* onwards, initially ΔE_τ was used. However, as ΔE_τ is considered as one of the potential dimensions for the signal extraction for *hadronic partition*, this switch was made in order to avoid any sculpting of the background.
- The $\tau \to hhh\pi^0$ mode was added to the analysis while the flowchart was being developed. Therefore, this mode was not considered in the original flowchart, and when added to the analysis, it was added to *Step 3* as $\tau \to hhh\pi^0$ had the same distinct signature of three charged tracks as $\tau \to hhh$.
- As a result of adding the $\tau \to hhh\pi^0$ mode to the flowchart, the variable used to select among multiple events that may pass *Step 3* was changed from the χ^2 probability of the vertex fit to the B_{tag} to E_{extra} to accommodate the possibility of choosing between multiple events that have $\tau \to hhh\pi^0$ candidates.
- To be consistent with the selection cuts, the flowchart uses the BDT-based PID quantities in *Step 1* and *Step 2*. Earlier iterations of the flowchart used other LID cuts, for example when the analysis selection cuts used the likelihood-based PID quantities, with the flowchart matching this. Likelihood-based binary PID quantities were also investigated and it was found to provide little to no difference in terms of best candidate selection and τ mode cross-feed.



(A) Signal events

(B) Signal events

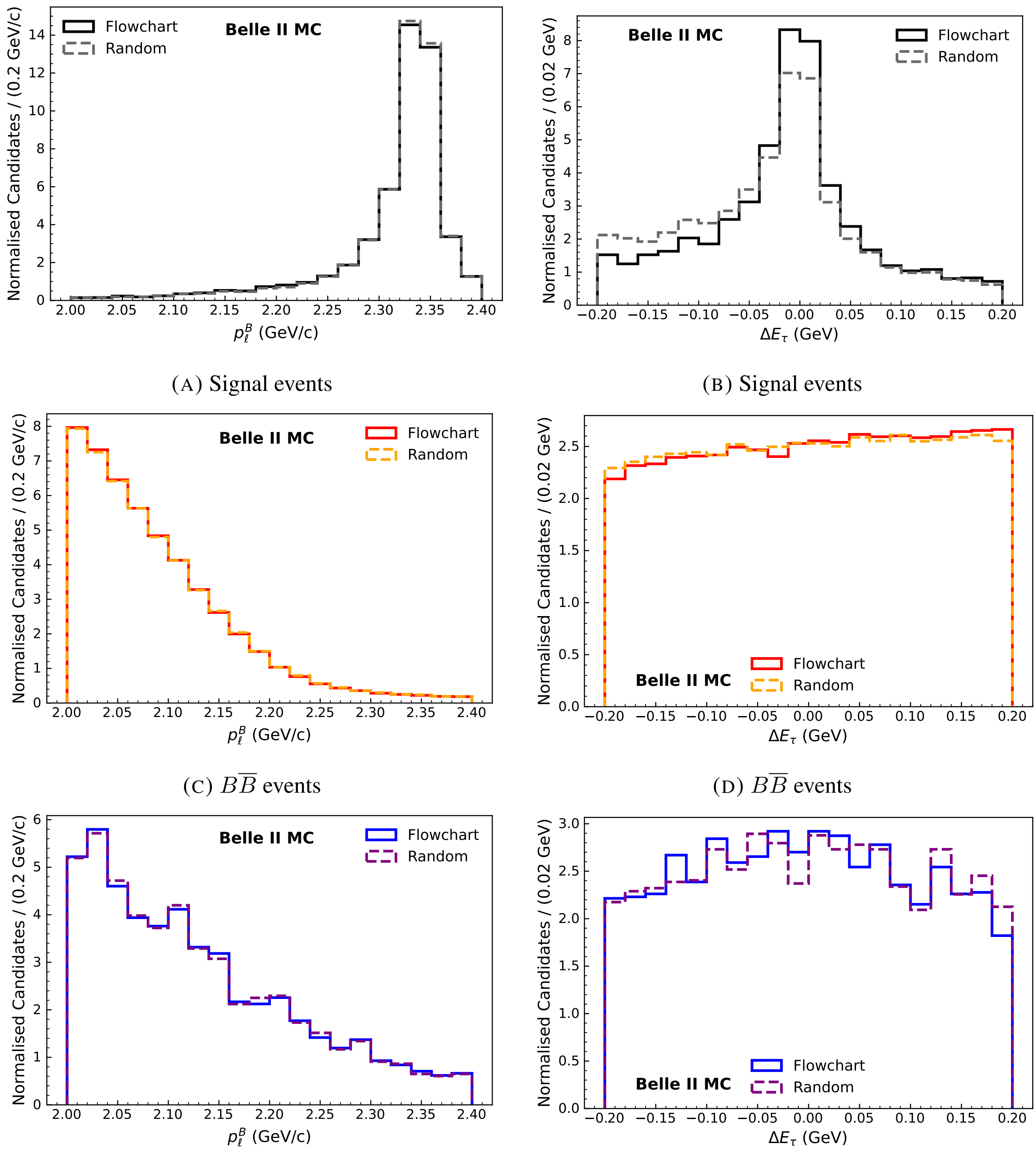
(C) $B\overline{B}$ events

(D) $B\overline{B}$ events

(E) $q\overline{q}$ events

(F) $q\overline{q}$ events

FIGURE 6.6. Distributions of the variables p_ℓ^B and $E_{\rm extra}$ that define the *fitting region* for the *leptonic partition*. Signal and background reconstructed in the $B^0 \to e^\pm\tau^\mp$ mode is presented here for the *leptonic partition*. Each distribution compares a random best candidate selection with the flowchart method outlined above.



(A) Signal events

(B) Signal events

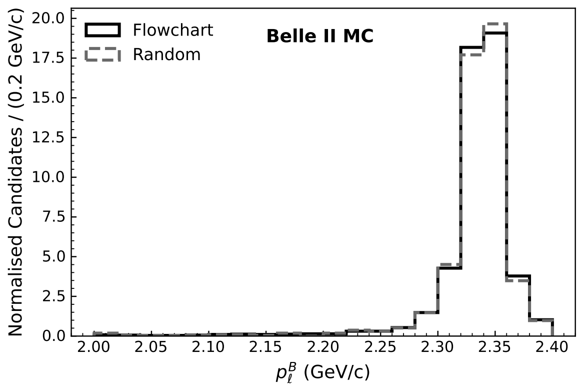
(C) $B\overline{B}$ events

(D) $B\overline{B}$ events

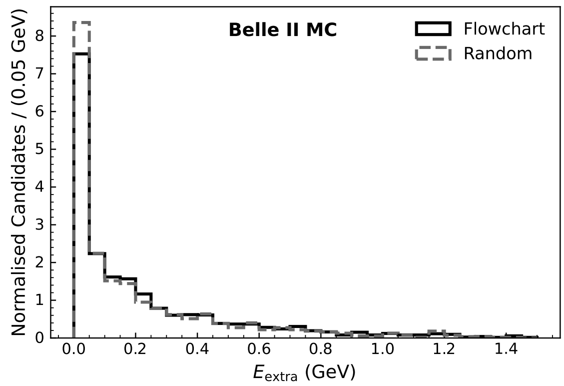
(E) $q\overline{q}$ events

(F) $q\overline{q}$ events

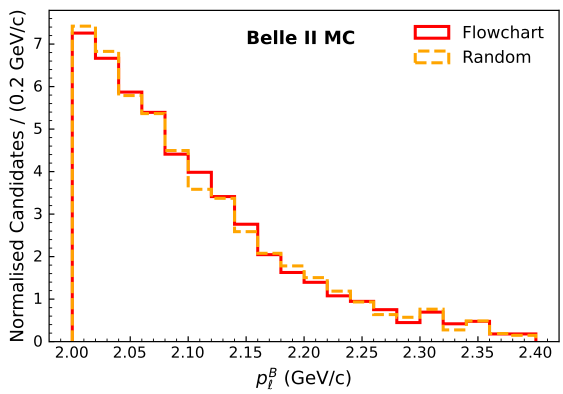
FIGURE 6.7. Distributions of the variables p_ℓ^B and ΔE_τ that define the *fitting region* for the *hadronic partition*. Signal and background reconstructed in the $B^0 \to e^\pm\tau^\mp$ mode is presented here for the *hadronic partition*. Each distribution compares a random best candidate selection with the flowchart method outlined above.



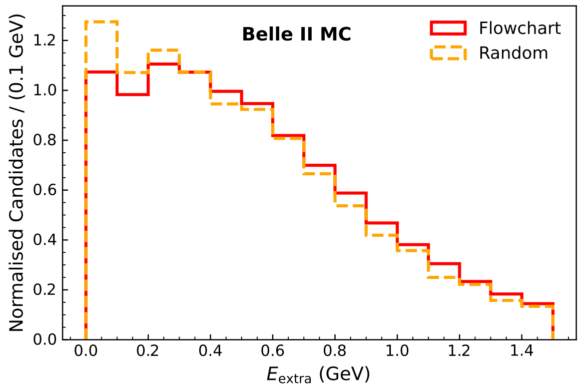
(A) Signal events



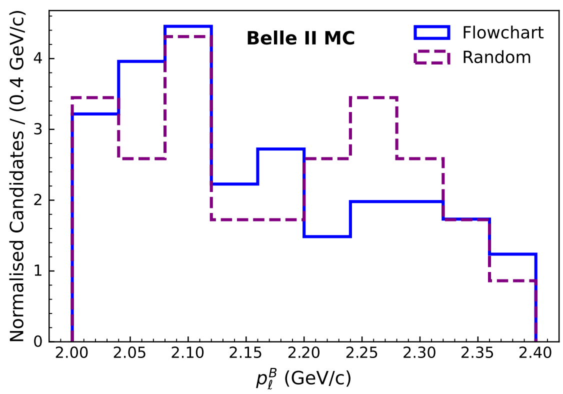
(B) Signal events



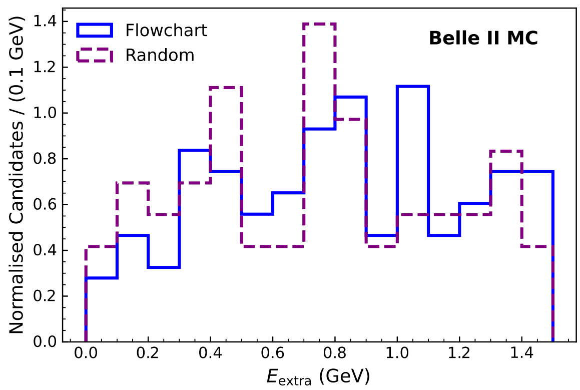
(C) $B\overline{B}$ events



(D) $B\overline{B}$ events

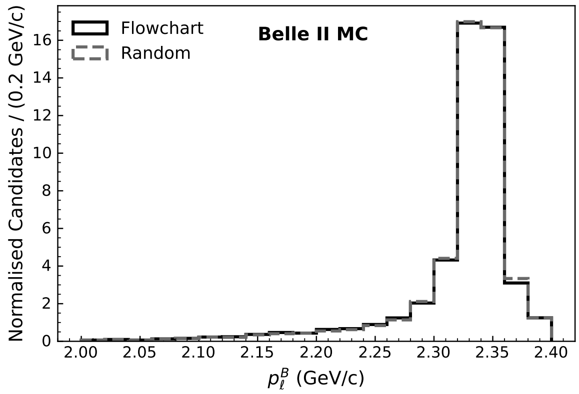


(E) $q\overline{q}$ events

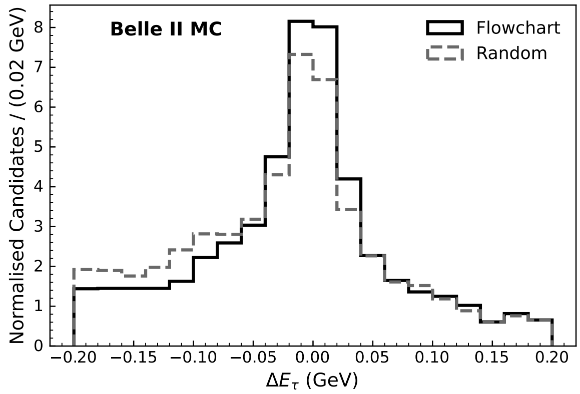


(F) $q\overline{q}$ events

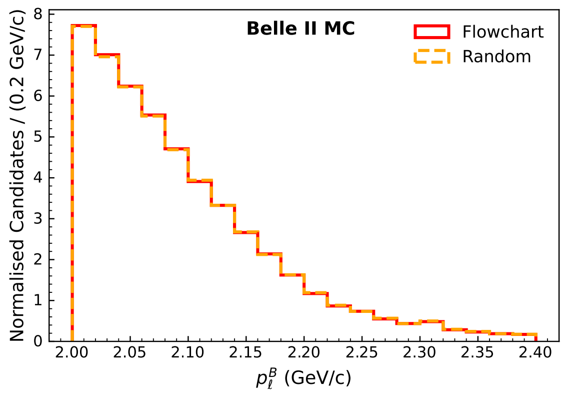
FIGURE 6.8. Distributions of the variables p_ℓ^B and $E_{\rm extra}$ that define the *fitting region* for the *leptonic partition*. Signal and background reconstructed in the $B^0 \to \mu^\pm\tau^\mp$ mode is presented here for the *leptonic partition*. Each distribution compares a random best candidate selection with the flowchart method outlined above.



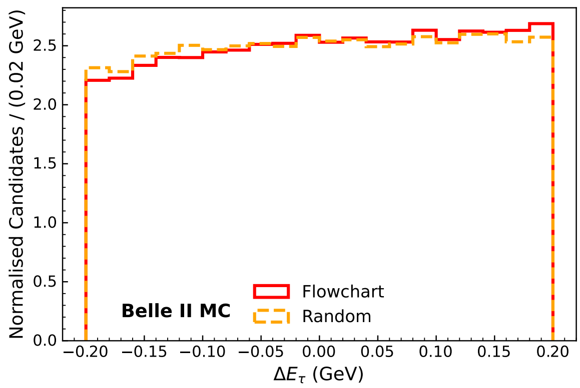
(A) Signal events



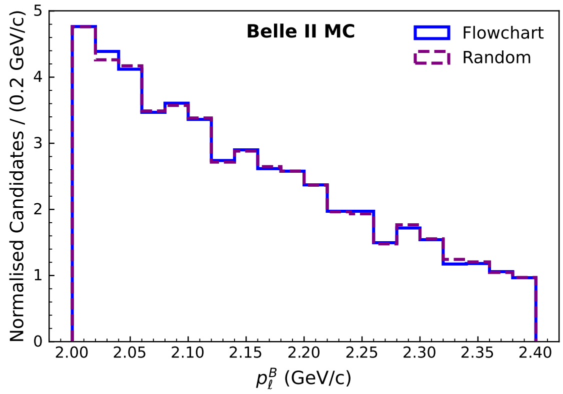
(B) Signal events



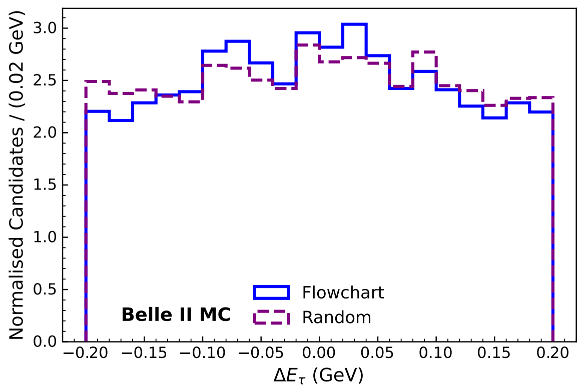
(C) $B\overline{B}$ events



(D) $B\overline{B}$ events



(E) $q\overline{q}$ events



(F) $q\overline{q}$ events

FIGURE 6.9. Distributions of the variables p_ℓ^B and ΔE_τ that define the *fitting region* for the *hadronic partition*. Signal and background reconstructed in the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode is presented here for the *hadronic partition*. Each distribution compares a random best candidate selection with the flowchart method outlined above.

6.3.3 τ Tau Mode Cross-Feed

The cross-feed between the τ modes is investigated following the selection cuts and best candidate selection. Here, the cross-feed is defined as the crossovers between the different generated and reconstructed τ modes. The reconstructed τ modes was obtained using the decay mode ID. This is the ID number designated to specific decay modes when reconstructing composite particles within `basf2` so that any given reconstructed decay mode can be later identified. The generated τ modes was determined using the PDG code of the particles in the decay chain on the generator level. Table 6.16 and 6.17 present results on the τ mode cross-feed (Table 6.18 contains the same information as the latter table but represented as a fraction). The numbers are derived from all reconstructed signal events. As seen from Table 6.17a and 6.17b, significant cross-feed between the π^0 modes arises from either the loss or incorrect addition of a π^0 to the τ reconstruction. The cross-feed between the h and hhh modes is also noticeable which is expected as the loss of a single track, either due to low p_T or a polar angle falling outside of the detector acceptance, means that the single track is the most suitable reconstruction mode for such a case. The cross-feed between the leptonic and hadronic modes as a whole demonstrates the benefit of our best candidate selection approach. Of the reconstructed signal in the *leptonic partition*, almost all are truly generated as either of the leptonic τ modes. On the other hand, there are around 12-13% of reconstructed signal events that have been incorrectly placed in the *hadronic partition*. This represents an efficiency loss when it comes to calculating the upper limit for the *leptonic partition*. Aside from the noticeable impact of muons that fail the PID cut, the main factor that contributes to the efficiency loss are null values for the BDT-based PID variables. As the hadronic τ modes have no PID requirements, events with these null cases remain as hadronically reconstructed τ candidates. Table 6.19 quantifies this impact, along with the contribution to the efficiency loss from leptons that fail the PID cut. A further discussion on these null cases is given in Appendix D. Though the likelihood-based PID values have significantly fewer null values, they showed larger mis-modelling of the *fake lepton* background component.

TABLE 6.16. The τ cross-feed grouped into leptonic and hadronic modes. The percentages are calculated from the fraction of remaining signal events following selection cuts and best candidate selection.

Generated Mode	**Reconstructed Mode**	
	Leptonic	*Hadronic*
Leptonic	20.73%	12.96%
Hadronic	1.20%	65.11%

(A) Results for $B^0 \to e^\pm \tau^\mp$

Generated Mode	**Reconstructed Mode**	
	Leptonic	*Hadronic*
Leptonic	19.50%	12.59%
Hadronic	1.34%	66.57%

(B) Results for $B^0 \to \mu^\pm \tau^\mp$

TABLE 6.17. The τ cross-feed among the different hadronic modes. These are absolute numbers derived from the reconstructed signal sample following selection cuts and best candidate selection.

Generated Mode	Reconstructed Mode h	$h\pi^0$	$h\pi^0\pi^0$	hhh	$hhh\pi^0$
h	978	201	19	3	1
$h\pi^0$	891	1706	507	36	20
$h\pi^0\pi^0$	112	425	522	11	15
hhh	221	147	72	689	134
$hhh\pi^0$	66	126	88	141	319

(A) Results for $B^0 \to e^\pm\tau^\mp$

Generated Mode	Reconstructed Mode h	$h\pi^0$	$h\pi^0\pi^0$	hhh	$hhh\pi^0$
h	1149	210	10	2	2
$h\pi^0$	894	1765	498	28	15
$h\pi^0\pi^0$	109	390	526	11	19
hhh	268	152	69	710	97
$hhh\pi^0$	54	130	74	154	304

(B) Results for $B^0 \to \mu^\pm\tau^\mp$

TABLE 6.18. The τ cross-feed among the different hadronic modes. The numbers represent fractions of the generated mode that are present in each reconstructed mode.

Generated Mode	Reconstructed Mode h	$h\pi^0$	$h\pi^0\pi^0$	hhh	$hhh\pi^0$
h	0.81	0.17	0.02	$<$ 0.01	$<$ 0.01
$h\pi^0$	0.28	0.54	0.16	0.01	0.01
$h\pi^0\pi^0$	0.10	0.39	0.48	0.01	0.01
hhh	0.17	0.12	0.06	0.55	0.11
$hhh\pi^0$	0.09	0.17	0.12	0.19	0.43

(A) Results for $B^0 \to e^\pm\tau^\mp$

Generated Mode	Reconstructed Mode h	$h\pi^0$	$h\pi^0\pi^0$	hhh	$hhh\pi^0$
h	0.84	0.15	0.01	$<$ 0.01	$<$ 0.01
$h\pi^0$	0.28	0.55	0.16	0.01	0.01
$h\pi^0\pi^0$	0.10	0.37	0.50	0.01	0.01
hhh	0.21	0.12	0.05	0.55	0.07
$hhh\pi^0$	0.08	0.18	0.10	0.22	0.42

(B) Results for $B^0 \to \mu^\pm\tau^\mp$

TABLE 6.19. Table of percentages that quantify the contributions for the efficiency losses for the *leptonic partition*. In each instance, the total is considered separately for each $\tau \to \ell\nu\nu$ mode. The percentages are calculated following selection cuts and best candidate selection, and the percentages are independent across the two categories.

Contribution	$B^0 \to e^\pm\tau^\mp$ Mode		$B^0 \to \mu^\pm\tau^\mp$ Mode	
	$\tau \to e\nu\nu$	$\tau \to \mu\nu\nu$	$\tau \to e\nu\nu$	$\tau \to \mu\nu\nu$
PID Is Null				
% of generated events	11.43	8.53	9.86	6.97
% of total hadronic cross-feed	53.60	18.72	50.79	15.07
PID Is Below Cut Value				
% of generated events	9.32	36.81	8.87	39.09
% of total hadronic cross-feed	43.73	80.70	45.71	84.47

Note that the hadronic cross-feed cases unexplained by either a failure to pass the PID cut threshold or a null PID value originate from instances where the leptonically-reconstructed candidate is removed by one of the selection cuts following the τ daughter lepton PID cut, and so only the hadronic candidates for that event reach the best candidate selection stage and remain in the final signal sample.

6.3.4 Summary

The 2D *fitting region*, and their 1D projections, for the *leptonic partition* is in Figure 6.10 for both modes. Within each sub-figure, the *signal extraction region* is demarcated. Table 6.20 compares the final efficiencies and event counts within the *fitting* and *signal extraction regions* for each mode within the *leptonic partition*. For these, the efficiency was calculated according to Equation 6.3.

TABLE 6.20. Table of final efficiencies for events reconstructed in the *leptonic partition*. The *fitting* and *signal extraction regions* are denoted as **FR** and **SER** respectively. The number of background events presented is scaled to the data luminosity. The n_{gen} for the signal efficiency is equal to the total size of the signal MC sample.

MC Sample	**ϵ in FR (%)**	**ϵ in SER (%)**	**Events in FR**	**Events in SER**
		$B^0 \to e^\pm \tau^\mp$ Mode		
Signal	0.0834	0.0576	2502	1729
$B\overline{B}^0$	0.0002	1.5456×10^{-6}	302.1	2.8
B^+B^-	1.9710×10^{-5}	1.9842×10^{-7}	38.5	0.4
$q\overline{q}$	3.5670×10^{-7}	$< 2.7439 \times 10^{-8}$	1.7	0.0
		$B^0 \to \mu^\pm \tau^\mp$ Mode		
Signal	0.0808	0.0591	2423	1772
$B\overline{B}^0$	0.0002	4.9179×10^{-6}	359.9	9.1
B^+B^-	1.8652×10^{-5}	4.6299×10^{-7}	36.5	0.9
$q\overline{q}$	7.4084×10^{-7}	2.7349×10^{-8}	9.8	0.4

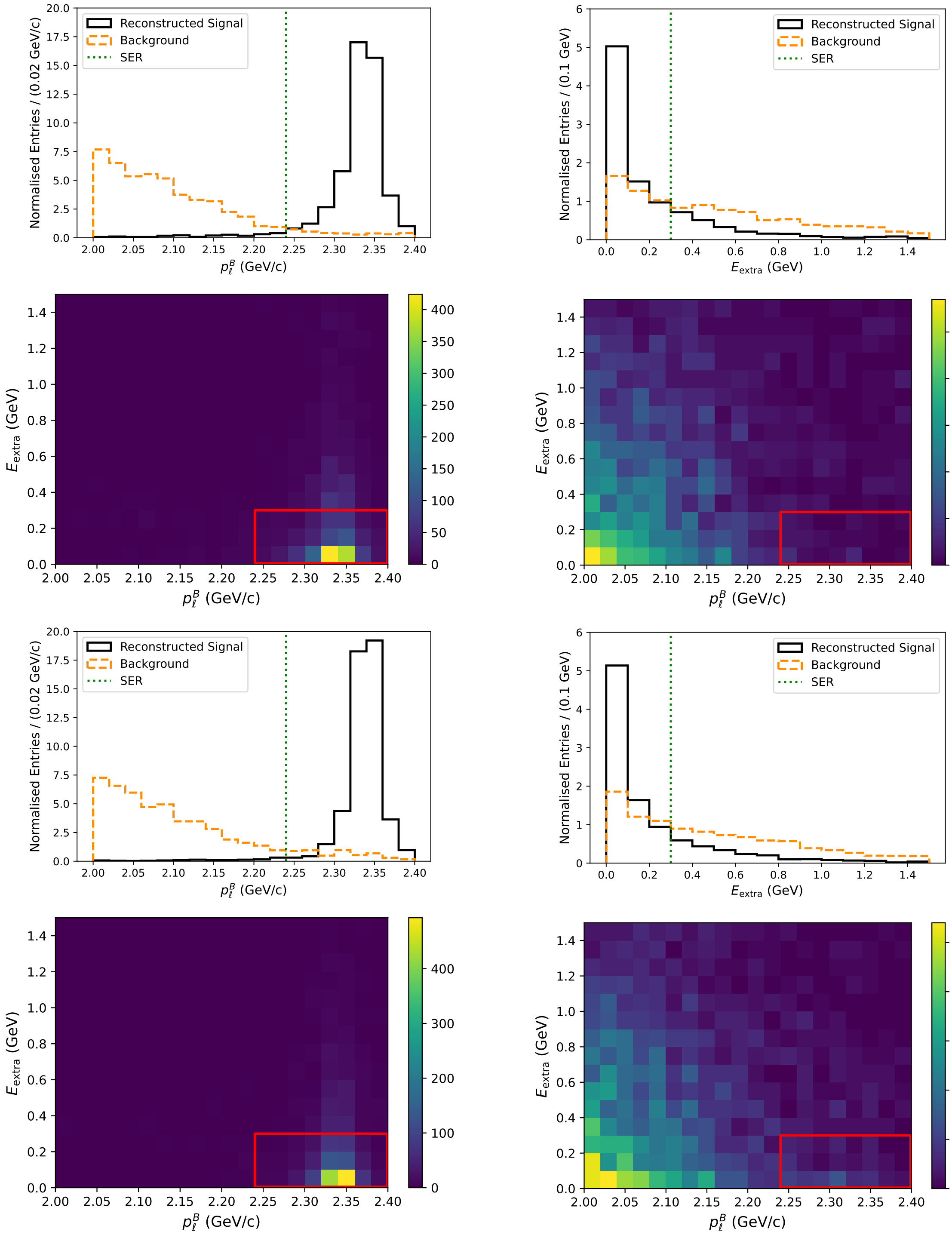


FIGURE 6.10. 1D projections and 2D plots of the *fitting region* for the $B^0 \to e^\pm\tau^\mp$ (top four plots) and $B^0 \to \mu^\pm\tau^\mp$ (bottom four plots) mode. In each case, the left 2D plot is for reconstructed signal, while the right plot is for background. The *signal extraction region* is demarcated on the 1D projections in green and on the 2D plots in red. The background component refers to $B^0\overline{B}^0$, B^+B^- and $q\overline{q}$ samples combined.

6.4 Data and MC Agreement

Agreement between data and MC was checked using the distributions of quantities used to perform cuts after reconstruction. Checks on the sidebands for the *signal extraction region* are left until permission from the working group is obtained. The datasets used for these checks are described in Section 6.2. The corrections applied to either data or MC are as follows:

- *Data - Photon Energy Bias*
  Corrections to the energy of photons used for the ROE and π^0 reconstruction are applied using the `basf2` module `correctEnergyBias` [1].
- *Data - Track Momentum Scaling*
  The momentum of tracks used to reconstruct the Υ(4S) candidate is re-scaled using the flat scale factor of 0.99987. It is applied using the `basf2` module `scaleTrackMomenta` [1].
- *MC - Lepton ID Corrections*
  Corrections to the efficiency of lepton ID cuts are applied to the leptonic τ modes and the signal lepton. The `PIDvar` package [3] is a python package used to apply the corrections following the `basf2` reconstruction. The package allows the corrections to be applied in the recommended (p, θ, q) bins, where $q \in [-1, 1]$ is the lepton charge, and to save the corresponding uncertainties for systematic uncertainty calculations. Coverage distributions of the corrections that are produced by the package are provided in Appendix E. Figures E.1 and E.2 show an overall good coverage of the leptons that are corrected.
- *MC - Hadronic FEI Corrections*
  The corrections are applied using the reconstructed decay mode ID of the B_{tag}. A best candidate selection based on the B_{tag} with the highest signal probability is applied for the data-MC agreement checks. Furthermore, a selection of $-0.15 < \Delta E < 0.1$ GeV/c and $M_{\text{bc}} > 5.27$ GeV/c^2 is applied for the B_{tag} as the corrections are only valid with these selections.

Aside from the M_{bc} and ΔE selections mentioned for the FEI corrections above, no other selection cuts past the τ daughter lepton PID within the cut flow are applied to the samples for this check. The data-MC agreement check in Figures 6.11 and 6.12 is performed for key variables used as cuts to suppress background, and for distributions related to the FEI reconstruction. Overall, a reasonable though imperfect agreement is observed across all distributions. There are no consistent deviations between the data and MC samples that might be cause for concern. Before it was decided to use only run-independent MC for this iteration of the analysis, checks on the data-MC agreement were performed using run-dependent MC samples. For this check, an improved agreement between data and MC was observed. However, as in the case of run-independent MC, the agreement was still not perfect, though the deviations did not show any particular trend.

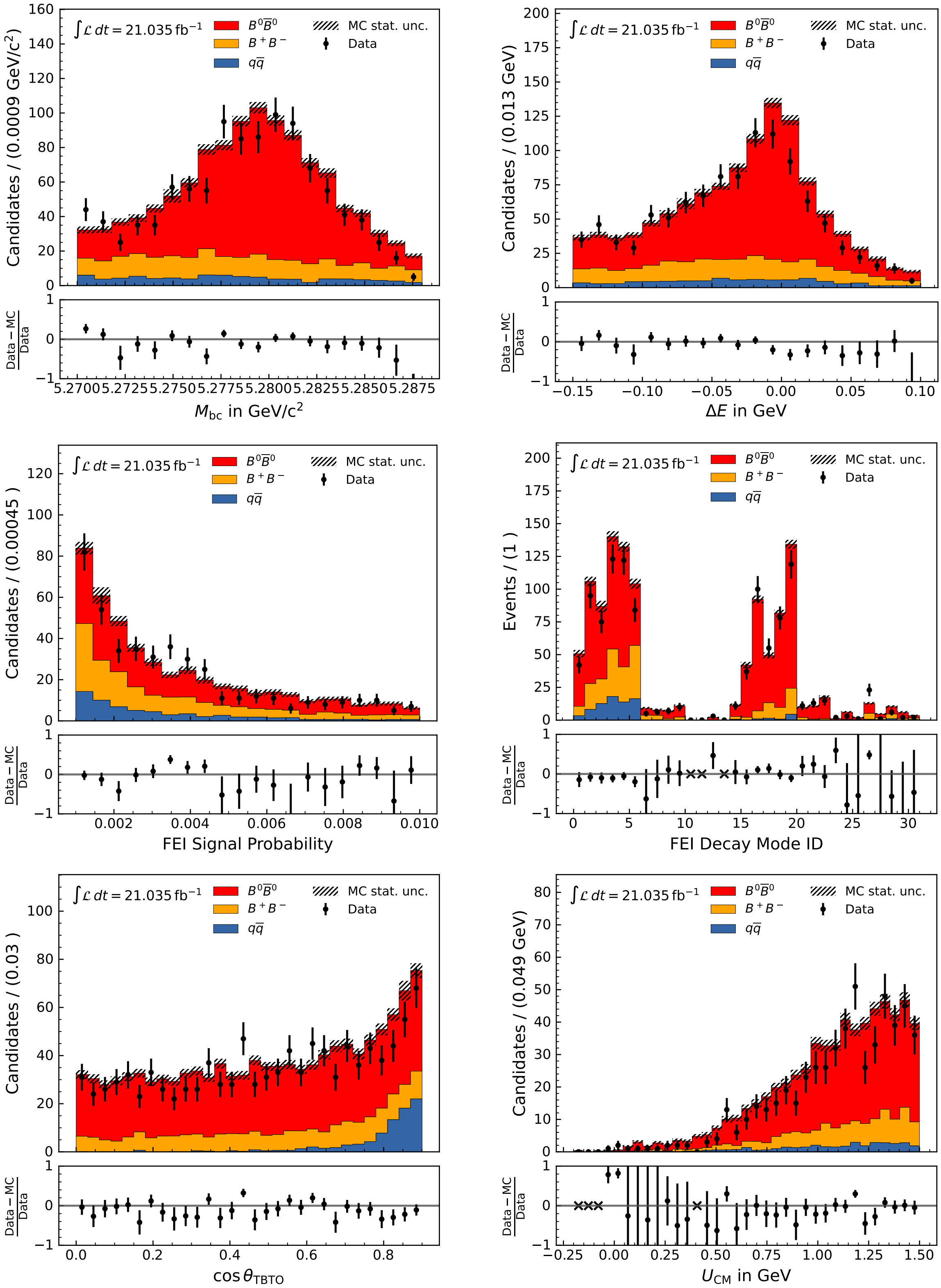


FIGURE 6.11. Data-MC agreement plots for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. Only events in the *leptonic partition* is shown here. The black crosses on the pull distribution at the bottom of each distribution represent bins where no data is present.

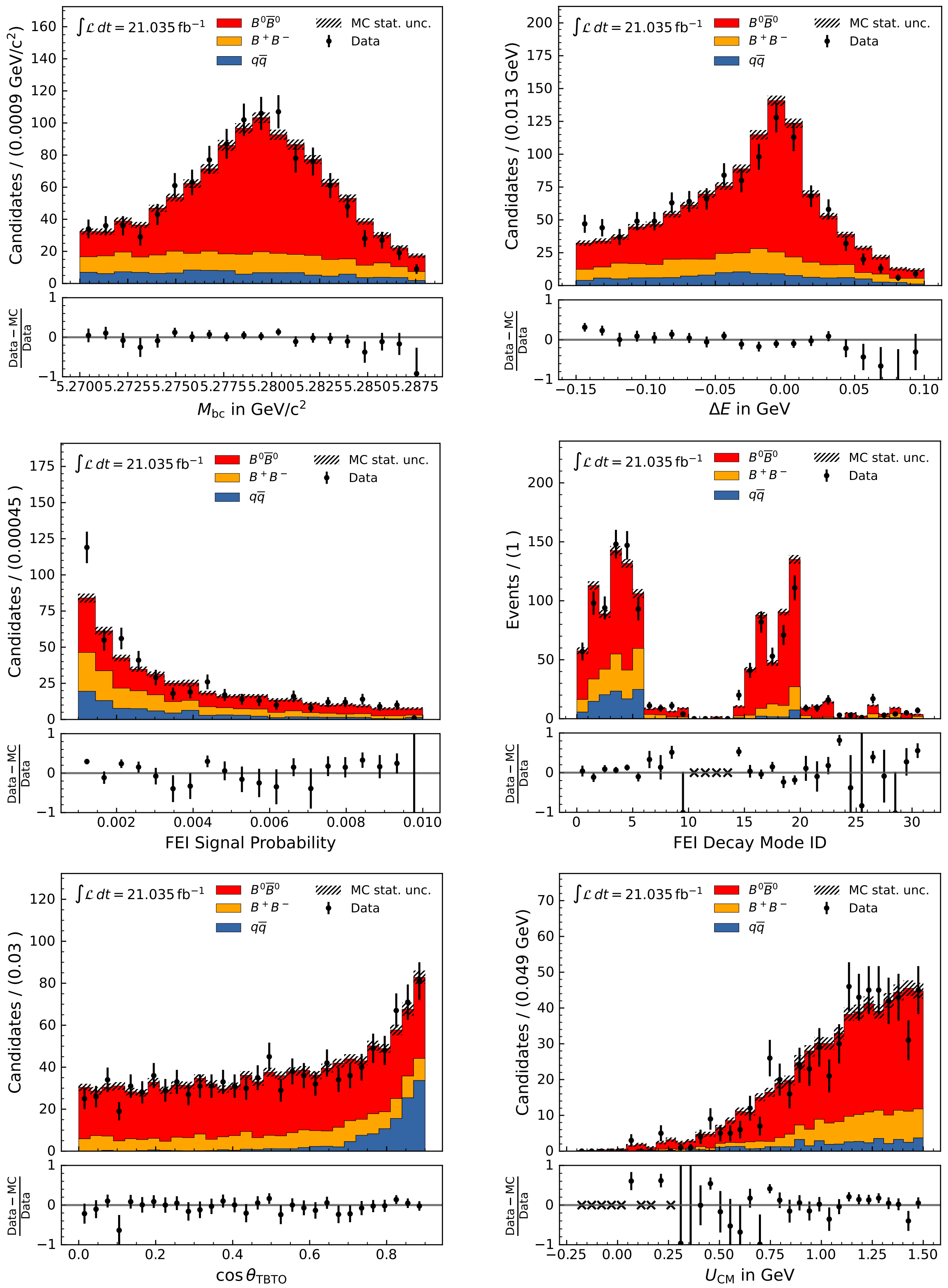


FIGURE 6.12. Data-MC agreement plots for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode. Only events in the *leptonic partition* is shown here. The black crosses on the pull distribution at the bottom of each distribution represent bins where no data is present.

6.5 Background Study For Leptonic τ Modes

Detailed background studies were performed for the leptonic τ modes as they are distinct from the hadronic τ modes in terms of their background level and sensitivity. As the leptonic τ modes have very low backgrounds, a detailed breakdown of the background events that remain in the *signal extraction region* is both imperative and feasible. It would allow systematic division of the background into constituents, which would permit for a targeted and improved background estimation.

6.5.1 Component Breakdown

The background that remains in the signal region is dominated by $B^0\overline{B}^0$ events. Nevertheless, we apply the same background estimation strategies for B^+B^- and $q\overline{q}$ events so that we obtain the estimated background from all constituents. The three background components present are introduced below, with their amounts in the *fitting region* and *signal extraction region* presented in Tables 6.21 and 6.22 respectively. The components are:

(1) *Two True Leptons (e.g. $B^0 \to [D \to K_L^0\ell\nu]\ell\nu$)*
Events involving two semileptonic decays, such as those of a B and D meson, produce two leptons that can mimic the signal and τ daughter lepton for our analysis. In particular, around 67% of such decays across both modes produce the leptons alongside a K_L^0, and as a reliable K_L^0 veto or identification method has yet to be developed within Belle II, this subset of decays are effectively irreducible backgrounds. Decays with two true leptons are the dominant background for the $B^0 \to e^\pm\tau^\mp$ mode. As only 14 events from all the available MC survive in this region, an MC sample larger than the existing 2.8 ab^{-1} one is required to obtain a better understanding of the shape and its uncertainty. To achieve this with the current computing resources, an `EvtGen` skimming module was implemented. This is described in Section 6.5.2.

(2) *Fake Leptons (e.g. $B^0 \to \pi\ell\nu$)*
Semileptonic B and D decays that produce a hadron, such as a π, in the final state alongside a lepton can mimic our signal if the hadron is mis-identified as either an e or μ. This component is the leading background for the $B^0 \to \mu^\pm\tau^\mp$ mode. The aim is to estimate this background straight from data so as to limit the impact of any potential data-MC disagreements. The method with which this is achieved is described in Section 6.5.3.

(3) *J/ψ-Like Decays*
This component is the same as the first component but is separated into its own category as it is vetoed. This component specifically consists of semileptonic B decays that produce a J/ψ or ψ(2S) alongside a K_L^0. Therefore, like the *two true lepton* component, they are effectively irreducible without a K_L^0 veto or identification method. A veto based on the reconstructed mass of the J/ψ or $\psi(2S)$ lepton daughter pair is used to remove the background from J/ψ contributions.

TABLE 6.21. Amounts of each background component for the different background MC samples for the *fitting region*. The numbers are scaled to the integrated luminosity of the data. Only the events in the *leptonic partition* are shown here.

MC Sample	Two True Leptons	Fake Leptons	J/ψ-Like Decays
		$B^0 \to e^\pm\tau^\mp$ Mode	
$B\overline{B}^0$	263.4	20.3	18.5
B^+B^-	29.5	6.5	2.3
$q\overline{q}$	2.5	2.1	0.0
		$B^0 \to \mu^\pm\tau^\mp$ Mode	
$B\overline{B}^0$	308.3	48.2	3.1
B^+B^-	27.2	9.1	0.1
$q\overline{q}$	2.9	6.9	0.0

TABLE 6.22. Amounts of each background component for the different background MC samples for the *signal extraction region*. The numbers are scaled to the integrated luminosity of the data. Only the events in the *leptonic partition* are shown here.

MC Sample	Two True Leptons	Fake Leptons	J/ψ-Like Decays
		$B^0 \to e^\pm\tau^\mp$ Mode	
$B\overline{B}^0$	1.8	0.3	0.8
B^+B^-	0.0	0.1	0.1
$q\overline{q}$	0.0	0.0	0.0
		$B^0 \to \mu^\pm\tau^\mp$ Mode	
$B\overline{B}^0$	2.7	6.2	0.1
B^+B^-	0.1	0.9	0.0
$q\overline{q}$	0	0.4	0.0

6.5.2 EvtGen Skimming Module

As demonstrated by the Tables 6.21 and 6.22, the *two true lepton* background component is the most dominant background in the *signal extraction region* for the *leptonic partition*. Given the amount of statistics available from the centrally produced (official) MC samples, an adequate understanding of the shape of this background is difficult to obtain. Knowing the shape is key in determining which fitting function to use for this component and to ascertain if correlated terms are required in the fit to account for the relationship between p_ℓ^B and E_{extra}. Therefore, more MC is needed.

The generation of large MC samples is mostly restricted by the `Geant4` simulation stage which heavily consumes the available time and space. Therefore, the idea behind this module is to skim events from `EvtGen` prior to `Geant4` so that only a skimmed subset are simulated. The aim of the skim is to retain background events with a kinematic and topological signature that mimics the background in the $p_\ell^B \in [2.0, 2.2]$ GeV/c $\times$ $E_{\text{extra}} \in [0, 0.5]$ GeV sideband. The module is written in `C++` and is called `GeneratorPreselectionLeptonic` in `basf2` [1]. The following cuts were used to define the skim. All quantities were calculated using generator-level information in the CMS frame. Distributions of background events in this particular sideband for each quantity are provided in Figures 6.13 and 6.14. Distributions that demonstrate what proportion of generated events are retained by the skim are in Figures 6.15 and 6.16. These were obtained using a small subset of 182,500 $B^0\overline{B}^0$ events. Other background types were not used as they are sub-leading and thus would have little impact on the shape.

Multiplicity Cuts

- The presence of one electron or one muon with a CMS momentum in the range $1.9 < p_{\text{CM}} < 2.3$ GeV/c. This allows for a plausible signal lepton candidate. Note that setting this upper limit on p_{CM} is a conceptual mistake in hindsight as it trimmed the very end of the p_ℓ^B distribution tail. The upper limit on this cut will be removed when these samples are re-produced for run-dependent MC.
- The presence of at least one lepton (electron or muon) with a CM momentum $0.11 < p_{\text{CM}} < 1.72$ GeV/c. This allows for plausible τ lepton candidates. More than one candidate is kept to account for losses due to detection efficiencies that occur during simulation and reconstruction.

Kinematic and Topology Cuts

Given that events with multiple τ lepton candidates are selected by the above skim cut, the following quantities that are lepton-based were calculated for each {signal lepton candidate, τ lepton candidate} pairing in the event. If at least one pairing passed the skim cuts, the event was flagged as being suitable for retention.

- The absolute-value difference in the z coordinate of the production vertex between the signal lepton and τ lepton Δz_{prod} being $-0.085 < \Delta z_{\text{prod}} < 0.033$ cm. This helps cut events where the signal lepton and τ lepton candidates are produced from semileptonic decays occurring in both the signal and tag sides.
- Requirements on the projection of the signal lepton momentum onto the τ lepton momentum, and the angle between the signal and τ leptons, of $-1.63 < \text{projection} < 0.136$ and $-1 <$

$\cos\theta < 0.33$ respectively

$$\cos\theta = \frac{p_\ell^{\rm CM}\cdot p_{\tau_\ell}^{\rm CM}}{|p_\ell^{\rm CM}||p_{\tau_\ell}^{\rm CM}|} \qquad \text{projection} = \frac{p_\ell^{\rm CM}\cdot p_{\tau_\ell}^{\rm CM}}{|p_\ell^{\rm CM}|}$$

For similar reasons to the prior cut, there is some relationship between the angular distribution of the signal and τ lepton.

- The event-based cut of $0 < U < 1$ GeV where $U = E_{\rm miss} - p_{\rm miss}$. The missing 4-vector of the event on the generator-level solely arises from the neutrinos. Therefore, this cut effectively isolates background events with neutrinos whose 4-vector resembles what is expected from signal events. Note that this is different to the *reconstructed* $U_{\rm CM}$ quantity which is an event cut (see Table 6.3). The *reconstructed* $U_{\rm CM}$ quantity is broader, with a longer tail to large $U_{\rm CM}$ values, due to contributions from reconstruction errors.

The reduction factor is defined as the number of events generated by `EvtGen` divided by the number of events being retained and therefore simulated. A table of the reduction factors for each background component is given in Table 6.23. As event generation is a random process, on any given use of the module, slightly different reduction factors may be obtained. Therefore, the reduction factors provided here are nominal. These reduction factors were based on tests using 10,000 generated events for each component, aside from the $u\overline{u}$, $d\overline{d}$ and $s\overline{s}$ backgrounds which are based on 100,000 generated events. Using these reduction factors, the maximum amount of MC that was feasible to produce was ascertained by the Belle II Data Production group, with the agreed upon value being 24 ab^{-1}. Therefore, 24 ab^{-1} samples of $B\overline{B}$ and $q\overline{q}$ events were generated to increase the statistics for the background estimation.

TABLE 6.23. Reduction factors for the different background components separated by signal mode. For the $u\overline{u}$, $d\overline{d}$ and $s\overline{s}$ components, no events remained after generation, thus a lower bound is placed on the factor.

Background Component	**Signal Mode**	
	$B^0 \to e^\pm\tau^\mp$ Mode	$B^0 \to \mu^\pm\tau^\mp$ Mode
$B^0\overline{B}^0$	128.2	108.7
B^+B^-	185.2	140.0
$u\overline{u}$, $d\overline{d}$, $s\overline{s}$	$> 10^5$	$> 10^5$
$c\overline{c}$	769.2	555.6

The skim cuts were derived from an early sample of background events that fall into the sideband. This early sample did not use the FEI but rather reconstructed the $B_{\rm tag}$ hadronically in the specific decay modes used by BaBar [61]. Due to this, the skim cuts are not entirely inclusive of the sideband background events that result from the FEI-skimmed samples. A check is performed on the shape of sideband background events to determine if the events removed by the skim cuts lead to any sculpting or bias in the shape that may impact the background estimation. Figure 6.17 compares the shape for the sideband background events, and only minimal changes to the shape are observed. A further check was performed using the EvtGen-skimmed samples for each background type to ensure the shape and normalisation are correct by comparing it to the $B^0\overline{B}^0$ and B^+B^- centrally-produced MC samples. Figure 6.18 contains the distributions comparing the shape, while Tables 6.24 and 6.25 contain information on the normalisation comparison for the $B^0 \to e^{\pm}\tau^{\mp}$ and $B^0 \to \mu^{\pm}\tau^{\mp}$ modes respectively. The normalisation results indicate an issue with the skim retention as the actual sizes of the EvtGen samples significantly higher than what is expected. As the EvtGen sample uses skim cuts to reject events, if the normalisations are different, the expectation is for the EvtGen sample normalisation to be lower. It is unclear what is causing this inflated skim size, and so further investigations are required prior to unboxing. In total, with the existing centrally-produced generic MC samples, the total amount of MC used for this background estimation is the following: 26.8 ab^{-1} for $B^0\overline{B}^0$ and B^+B^- each, and 25 ab^{-1} for each $q\overline{q}$ type.

TABLE 6.24. Normalisation comparisons between the centrally-produced MC sample and the EvtGen skimmed sample for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. The **Central Production** value indicates the amount of events in the centrally-produced MC samples. The **Expected 24 ab^{-1}** value comes from rescaling the centrally-produced numbers, while the **Actual 24 ab^{-1}** value comes from the EvtGen skimmed MC samples.

Background Component	**Central Production**	**Expected 24 ab^{-1}**	**Actual 24 ab^{-1}**
$B^0\overline{B}^0$	1107	9489	10704
B^+B^-	90	771	898
$q\overline{q}$	7	60	58

TABLE 6.25. Normalisation comparisons between the centrally-produced MC sample and the EvtGen skimmed sample for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode. The **Central Production** value indicates the amount of events in the centrally-produced MC samples. The **Expected 24 ab^{-1}** value comes from rescaling the centrally-produced numbers, while the **Actual 24 ab^{-1}** value comes from the EvtGen skimmed MC samples.

Background Component	**Central Production**	**Expected 24 ab^{-1}**	**Actual 24 ab^{-1}**
$B^0\overline{B}^0$	1274	10920	14075
B^+B^-	74	634	917
$q\overline{q}$	6	51	62

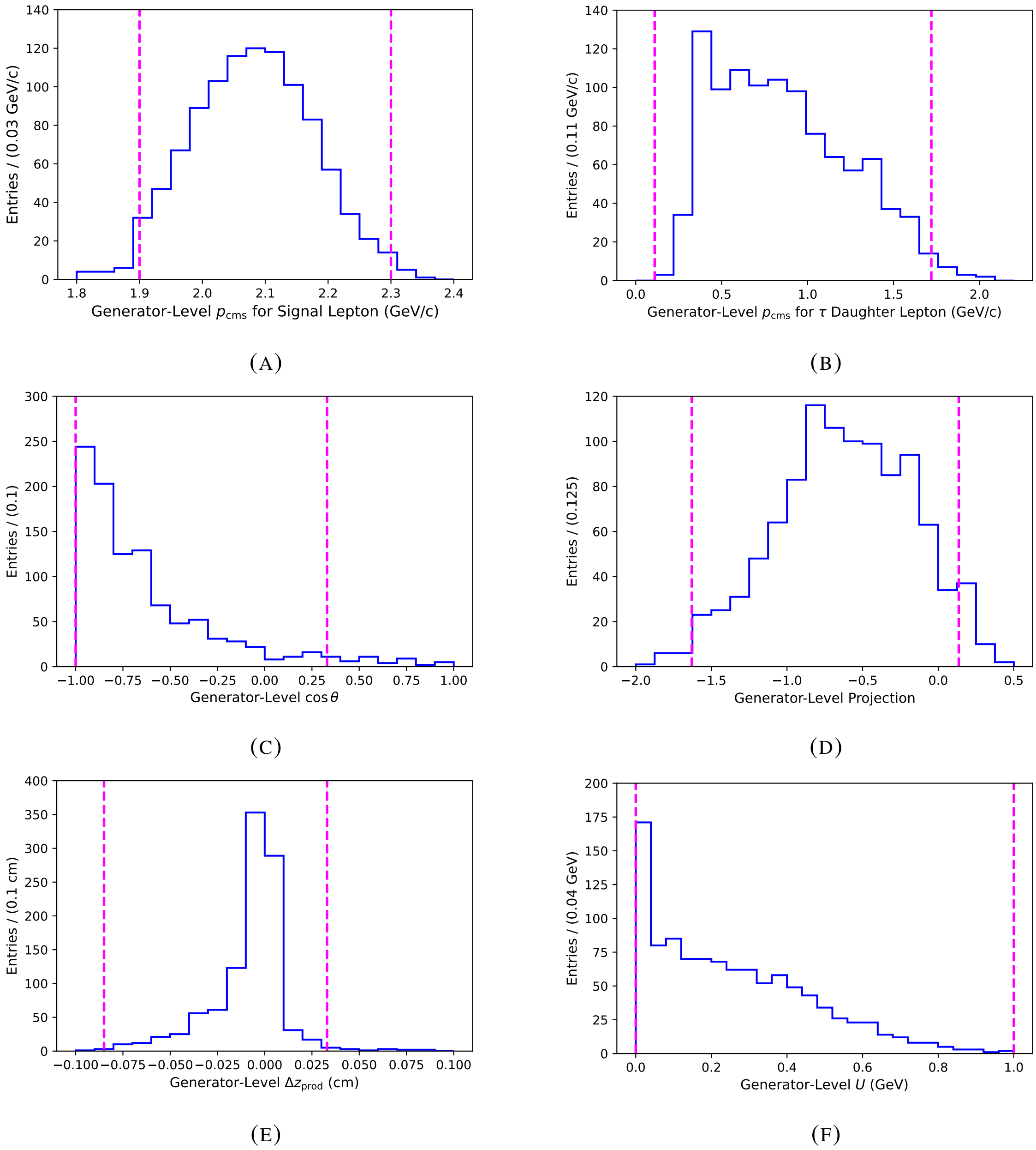


FIGURE 6.13. Distributions for quantities used in the EvtGen skim consisting of background events, from centrally-produced MC samples, in the $p_\ell^B \in [2, 2.2] \times E_{\text{extra}} \in [0, 0.5]$ sideband. The cut values chosen for the EvtGen skim are displayed using the magenta dashed lines. Only the $B^0 \to e^\pm\tau^\mp$ mode is displayed here.

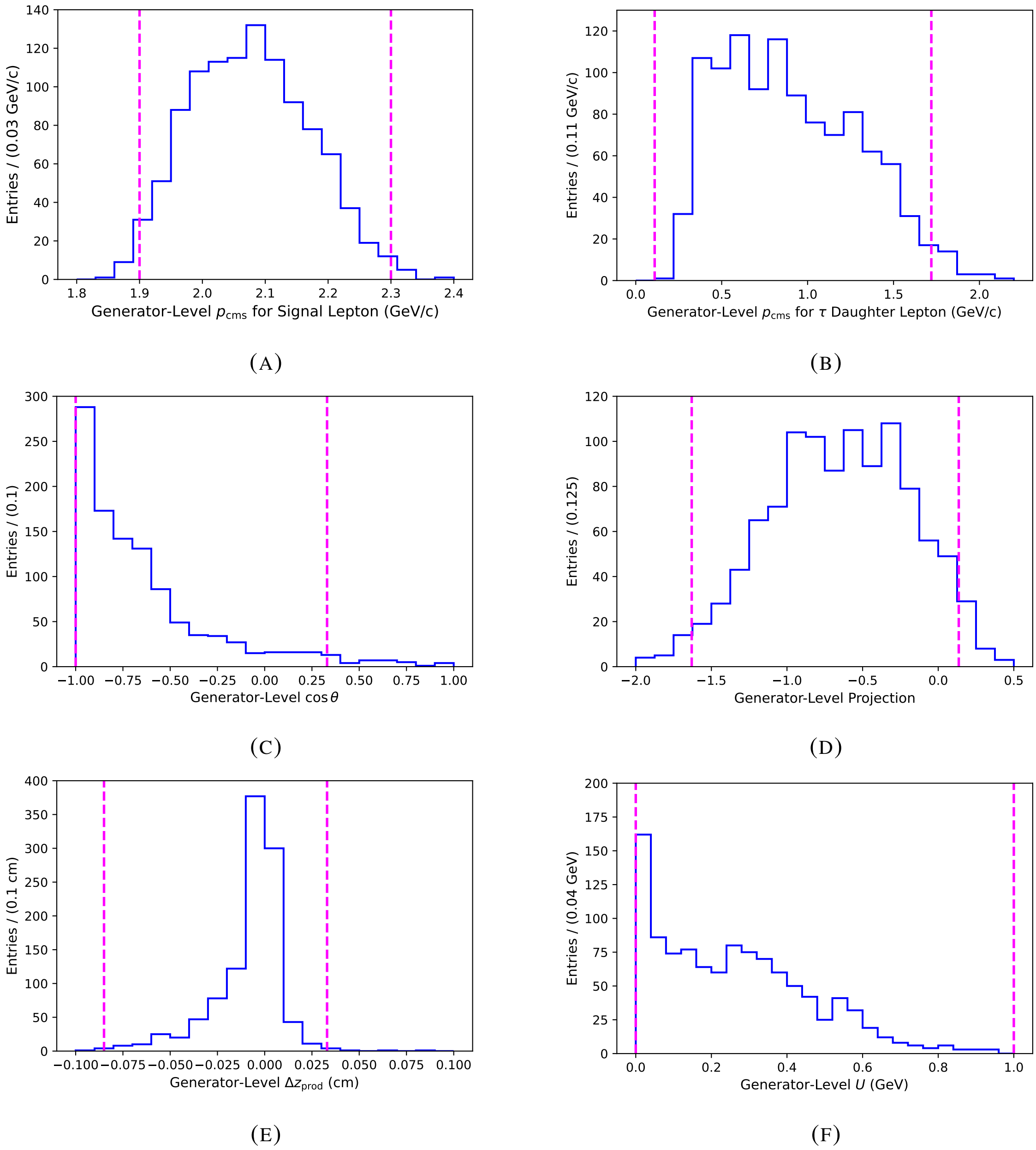


FIGURE 6.14. Distributions for quantities used in the `EvtGen` skim consisting of background events, from centrally-produced MC samples, in the $p_\ell^B \in [2, 2.2] \times E_{\text{extra}} \in [0, 0.5]$ sideband. The cut values chosen for the `EvtGen` skim are displayed using the magenta dashed lines. Only the $B^0 \to \mu^\pm\tau^\mp$ mode is displayed here.

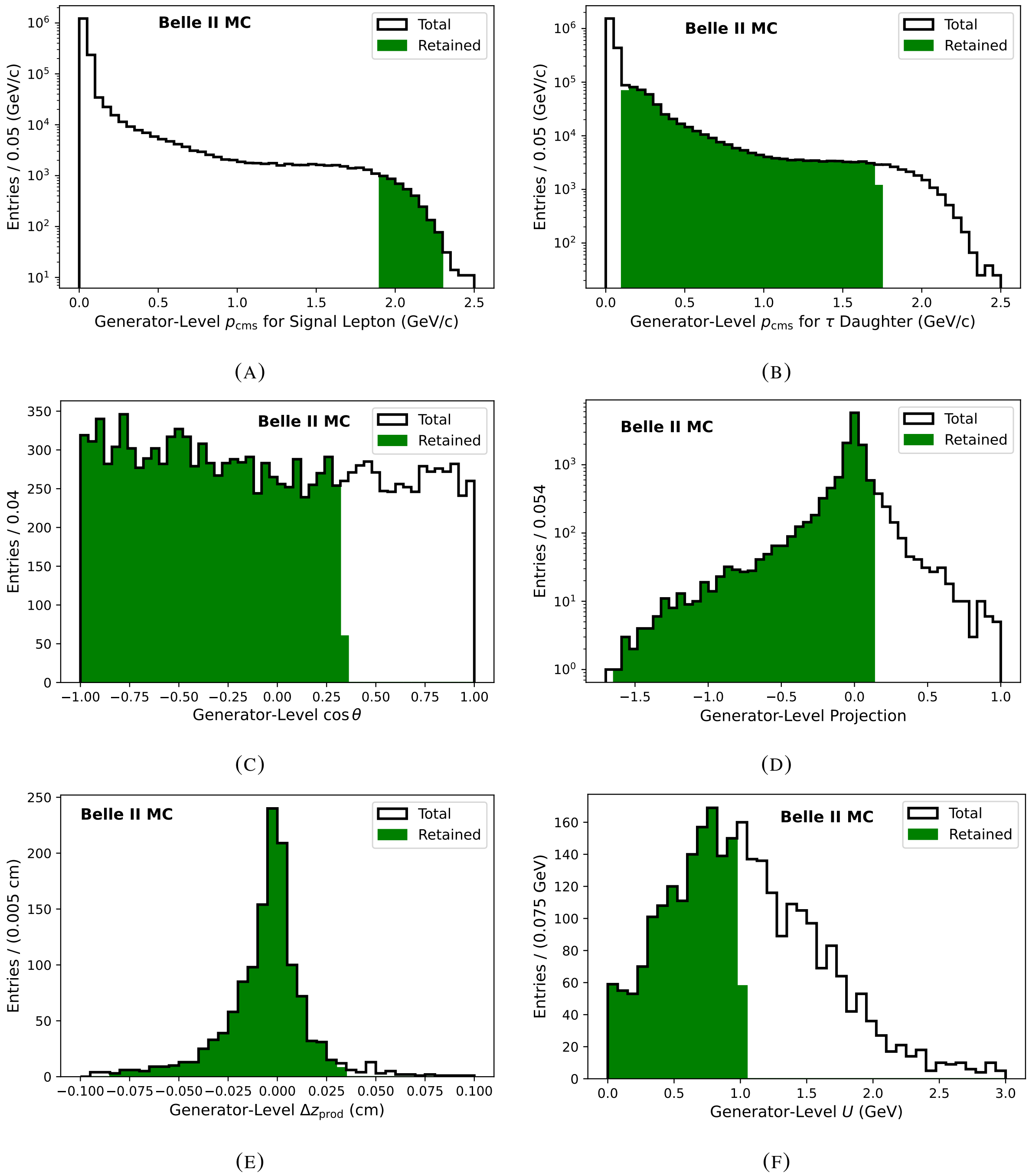


FIGURE 6.15. Distributions of the EvtGen skim cut variables for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. The black outline for (a) and (b) indicate the distribution of all $B^0\overline{B}^0$ events that contain at least two real leptons, while the black outline for the remaining subfigures are for $B^0\overline{B}^0$ events that contain at least two real leptons passing the required p_{cms} cuts for the signal and τ daughter lepton. For all instances, the signal lepton is required to be an electron. The green regions represent the subset of events retained by the corresponding skim cut. Each distribution is independent of each other such that the final subset of events retained by the EvtGen skim corresponds to the intersection of all the green regions.

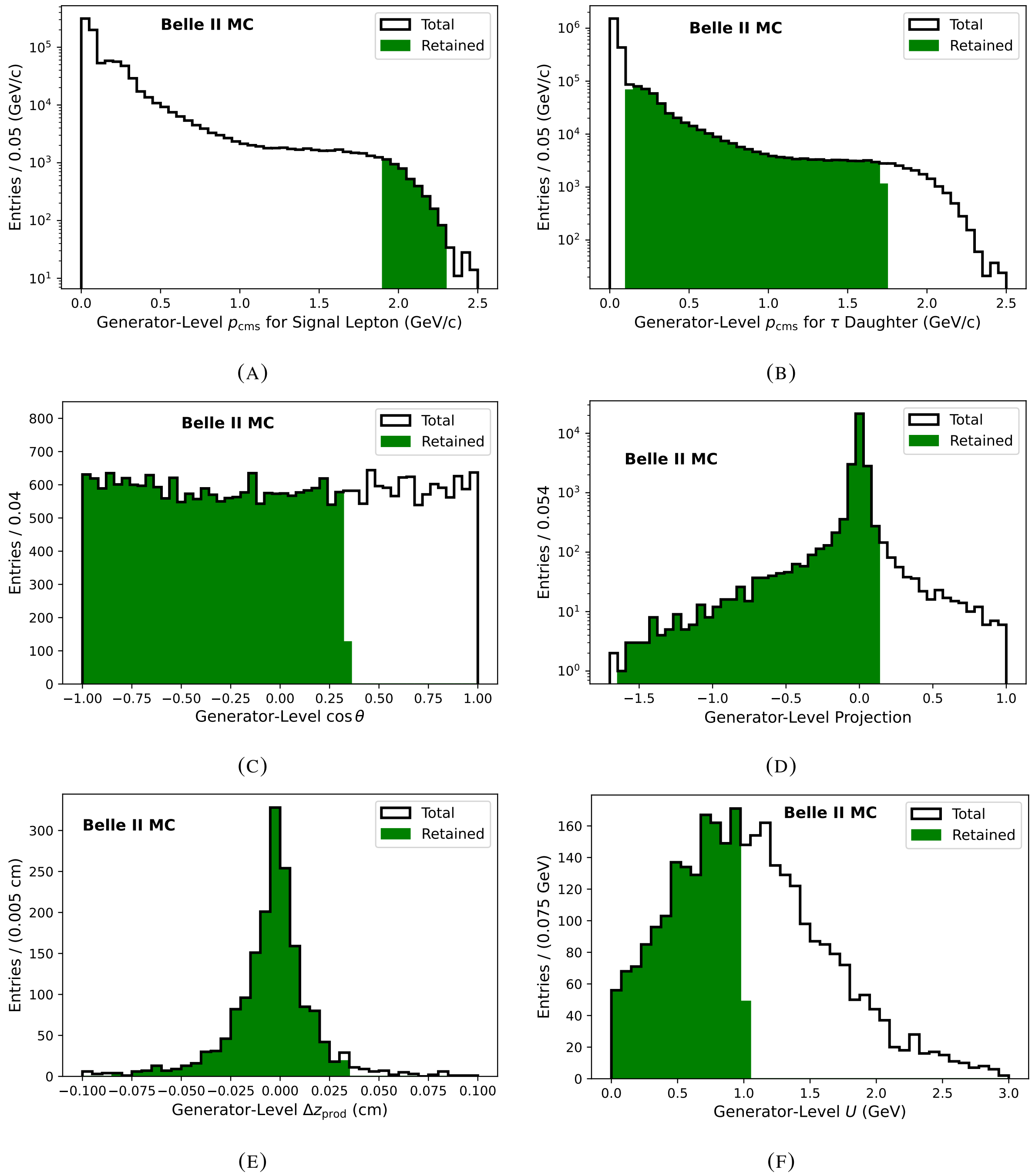


FIGURE 6.16. Distributions of the EvtGen skim cut variables for the $B^0 \to \mu^\pm\tau^\mp$ mode. The black outline for (a) and (b) indicate the distribution of all $B^0\overline{B}^0$ events that contain at least two real leptons, while the black outline for the remaining subfigures are for $B^0\overline{B}^0$ events that contain at least two real leptons passing the required p_{cms} cuts for the signal and τ daughter lepton. For all instances, the signal lepton is required to be an muon. The green regions represent the subset of events retained by the corresponding skim cut. Each distribution is independent of each other such that the final subset of events retained by the EvtGen skim corresponds to the intersection of all the green regions.

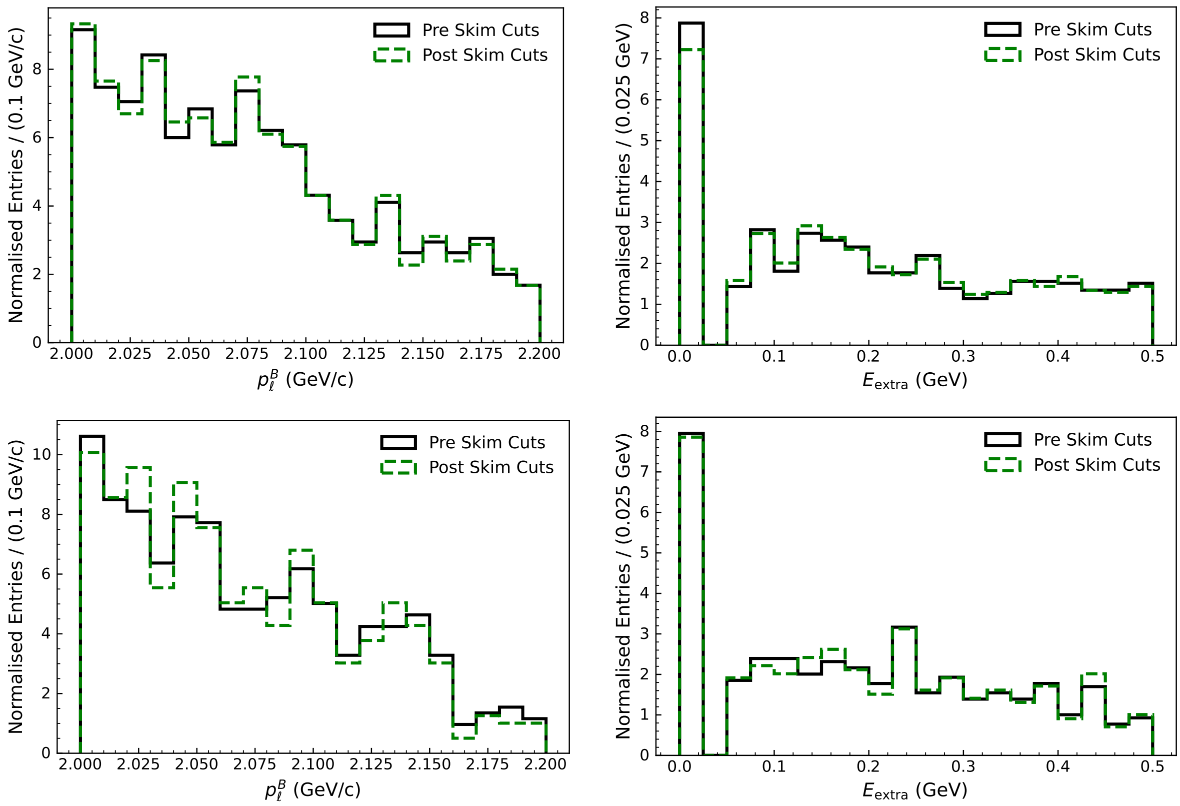


FIGURE 6.17. Normalised distributions comparing the sideband background events before and after the EvtGen skim cuts are applied. Distributions for the $B^0 \to e^\pm\tau^\mp$ modes are on top, while the distributions for $B^0 \to \mu^\pm\tau^\mp$ modes are on the bottom.

6.5.3 Fake Lepton Estimation

The *fake lepton* background component can cause estimation issues due to the data-MC disagreement that often exists with lepton ID. This is key for the $B^0 \to \mu^\pm\tau^\mp$ mode which has this component as the leading background. To minimise the impact of any data-MC disagreements on the reliability of this background estimate, a data-driven approach will be used for this estimation. The procedure for this data-driven approach is as follows:

(1) Run the analysis with one of the lepton ID cuts removed. (For the remaining steps, the lepton track in question refers to the one with their PID cut removed).
The lepton ID cuts removed depend on which fake rate will be estimated. For the estimation of fake signal leptons, the signal lepton PID cut is removed, while for the estimation of fake τ daughter leptons, each of their PID cuts are removed individually.

(2) For each event, the likely identity of the lepton candidate is determined by comparing the likelihood-based PID quantities, with the highest PID being used to assign the identity of

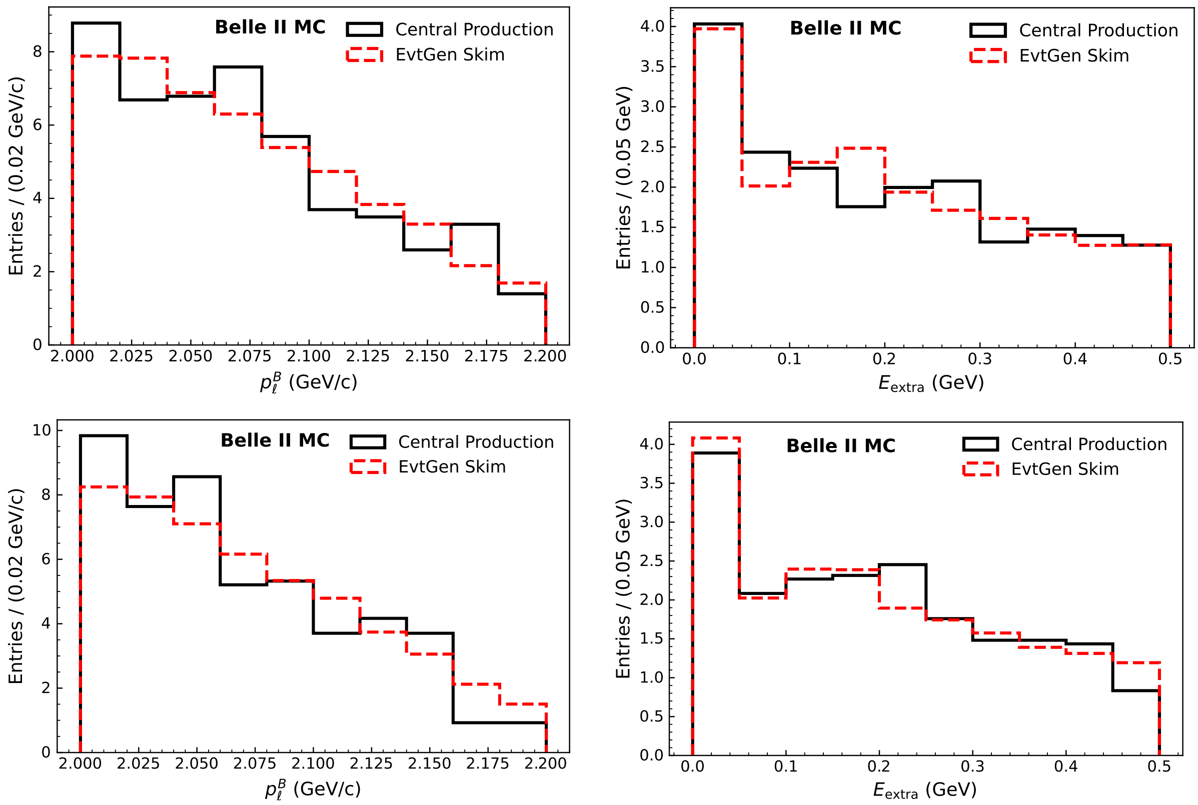


FIGURE 6.18. Normalised distributions comparing the shapes of the sideband background events from the centrally-produced MC samples and the `EvtGen` skimmed MC samples. The $B^0 \to e^\pm \tau^\mp$ mode is on the top while the $B^0 \to \mu^\pm \tau^\mp$ mode is on the bottom.

the lepton candidate. Tracks that are likely to be pions proceed to the next step as they will contribute to the *fake lepton* background.

(3) The tracks are separated by charge, with each charge binned in $p_{\mathrm{lab}} \times \theta_{\mathrm{lab}}$. The binnings are derived from those used by the performance group studies to obtain the fake rates from data:
- $\pi \to e$
  $p_{\mathrm{lab}} \in [0.2, 0.4, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0]$ GeV/c
  $\theta_{\mathrm{lab}} \in [0.22, 0.56, 1.13, 1.57, 1.88, 2.23, 2.71]$ rad
- $\pi \to \mu$
  $p_{\mathrm{lab}} \in [0.2, 0.4, 0.5, 0.7, 1.0, 1.5, 2.0, 2.5, 3.0]$ GeV/c
  $\theta_{\mathrm{lab}} \in [0.4, 0.64, 0.82, 1.16, 1.46, 1.78, 2.13, 2.22, 2.6]$ rad

(4) The fake rates are multiplied by the corresponding entries binned in the $p_{\mathrm{lab}} \times \theta_{\mathrm{lab}}$ for each track charge subset. The diagram in Figure 6.19 depicts this step. A 2D heatmap of the fake rates across the $p_{\mathrm{lab}} \times \theta_{\mathrm{lab}}$ parameter space for each $\pi \to e$ and $\pi \to \mu$ for each track charge

is given in Figure 6.20. Bins of $p_{\text{lab}} \times \theta_{\text{lab}}$ for which no fake rate exists have a fake rate of zero and are filled in red in these figures.

(5) The total amount of fakes, and thus the corresponding background estimate, is derived by integrating across the entire $p_{\text{lab}} \times \theta_{\text{lab}}$ distribution. Figure 6.21 depicts this step.

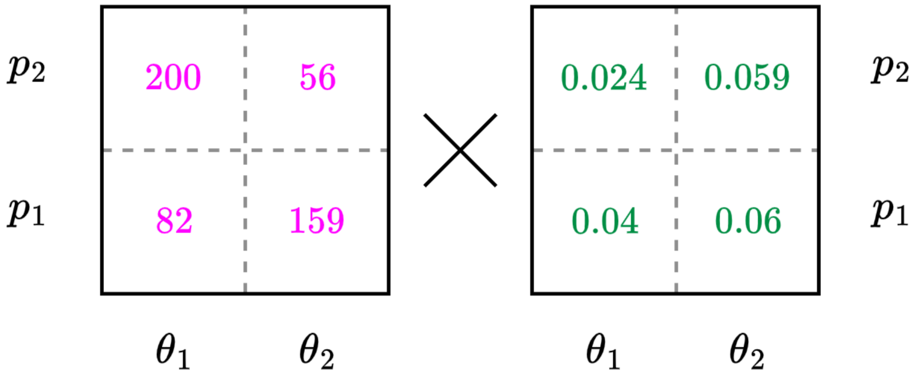


FIGURE 6.19. Diagram depicting step 4 of the *fake lepton* estimation process. The pink values indicate the entries for the lepton candidates, while the green values are the corresponding fake rates for those bins.

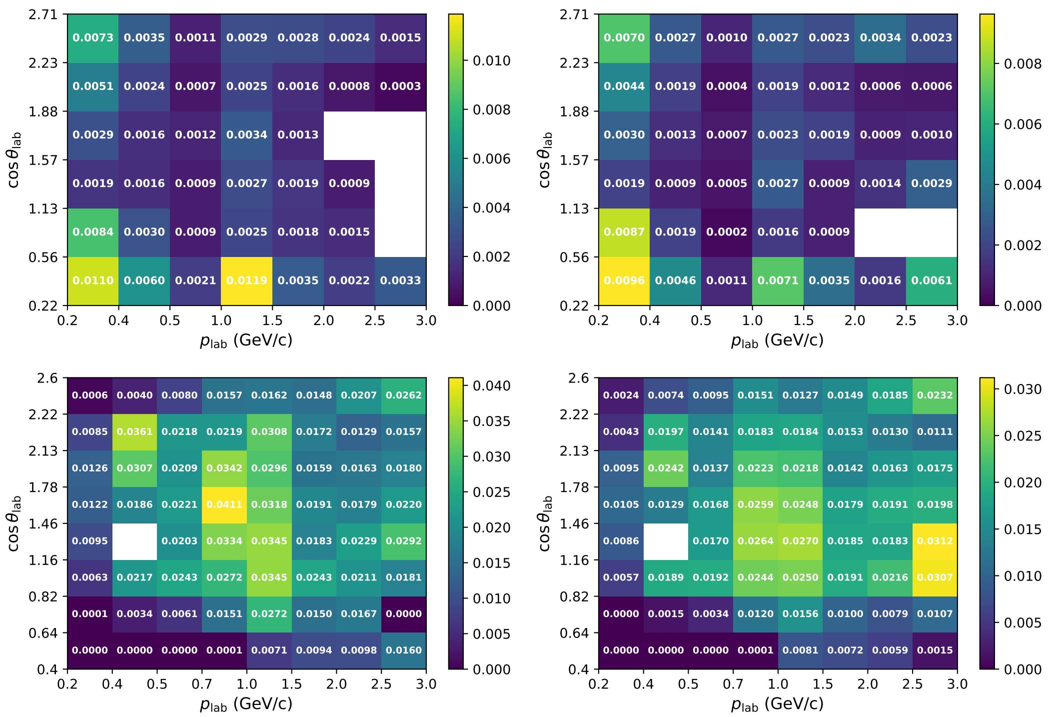


FIGURE 6.20. Fake rates for the electron ID (top) and muon ID (bottom) fake rates for positive (left) and negative (right) charged tracks. The bins that are white indicate no fake rate exists for this bin.

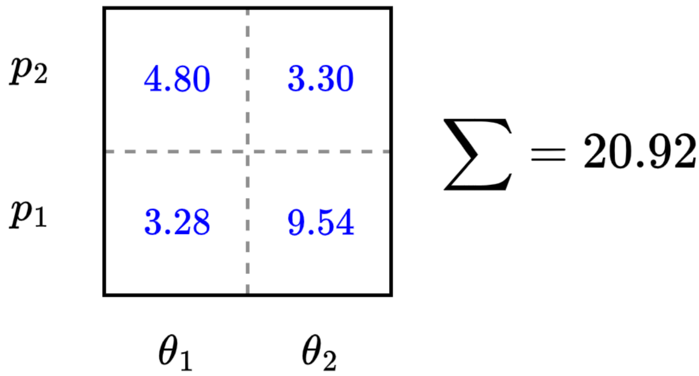


FIGURE 6.21. Diagram depicting step 5 of the *fake lepton* estimation process. The blue values indicate the result of the product between the corresponding entry and fake rate. The sum represents the integration across the region and therefore the estimated *fake lepton* component.

To determine if this method provides an accurate estimation of the *fake lepton* background component, a closure test using MC was performed. The MC samples used to calculate the expected fake rate for this closure test involved a 1 ab^{-1} subset of $B^0\overline{B}^0$ and B^+B^- background events. The closure test calculates the expected fake rates given the steps above, except for step 2 where the generator-level PDG code of the lepton candidate is used to determine if it is a pion. The expected rate is then compared to the rate observed for this analysis using the entire 2.8 ab^{-1} samples of $B^0\overline{B}^0$ and B^+B^- events. In the case of the closure test only, bins of $(p_{\rm lab}, \theta_{\rm lab})$ for which no fake rate exists are accounted for by subtracting the entry count of the observed sample in the corresponding bin from the total. The results of the closure test are shown in Figures 6.22 and 6.23 for $B^0\overline{B}^0$ and B^+B^- backgrounds respectively. The $q\overline{q}$ sample was not included in the closure test as only a fraction of an event remained for the the *fake lepton* background component from MC studies as shown in Table 6.22. In general, good agreement is observed across most distributions. In a few bins where the agreement is met due to the uncertainty of the expected rate, it is assumed that moving to a higher statistics MC sample to derive these expected rates would yield some disagreements. However, due to the large errors on the observed fake rate counts that are expected from the data, these disagreements will be small. To cover the disagreements that exist in certain bins, a systematic uncertainty will be applied.

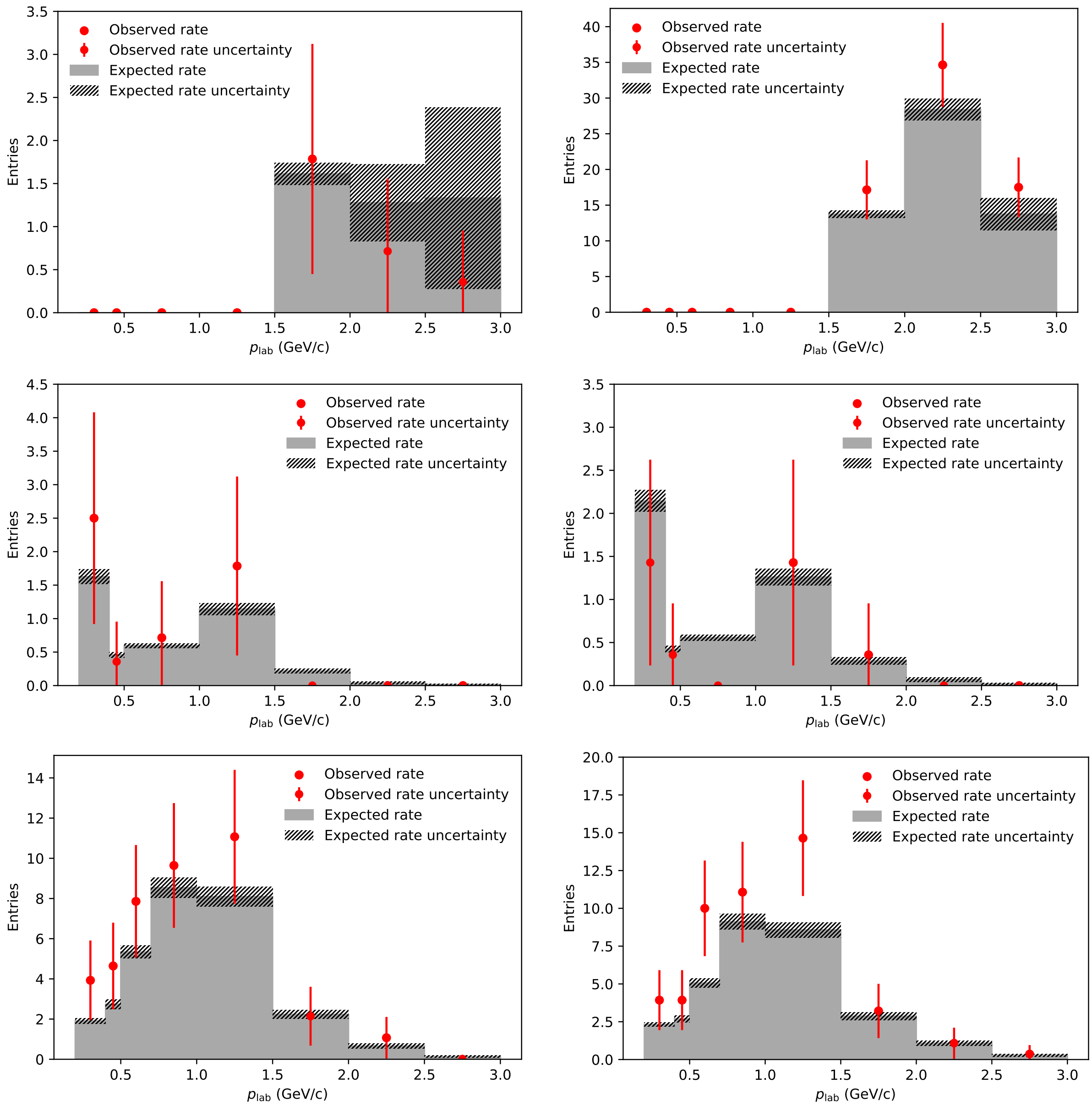


FIGURE 6.22. Distributions comparing the observed and expected fake rates for $B^0\overline{B}^0$ events. The steps used to derive these rates are outlined in Section 6.5.3. The fake rates for the signal lepton and $\tau \to e\nu\bar{\nu}$ and $\tau \to \mu\nu\bar{\nu}$ daughter leptons are shown on the top, middle and bottom respectively. Results for the $B^0 \to e^\pm\tau^\mp$ are on the left while results for the $B^0 \to \mu^\pm\tau^\mp$ are on the right.

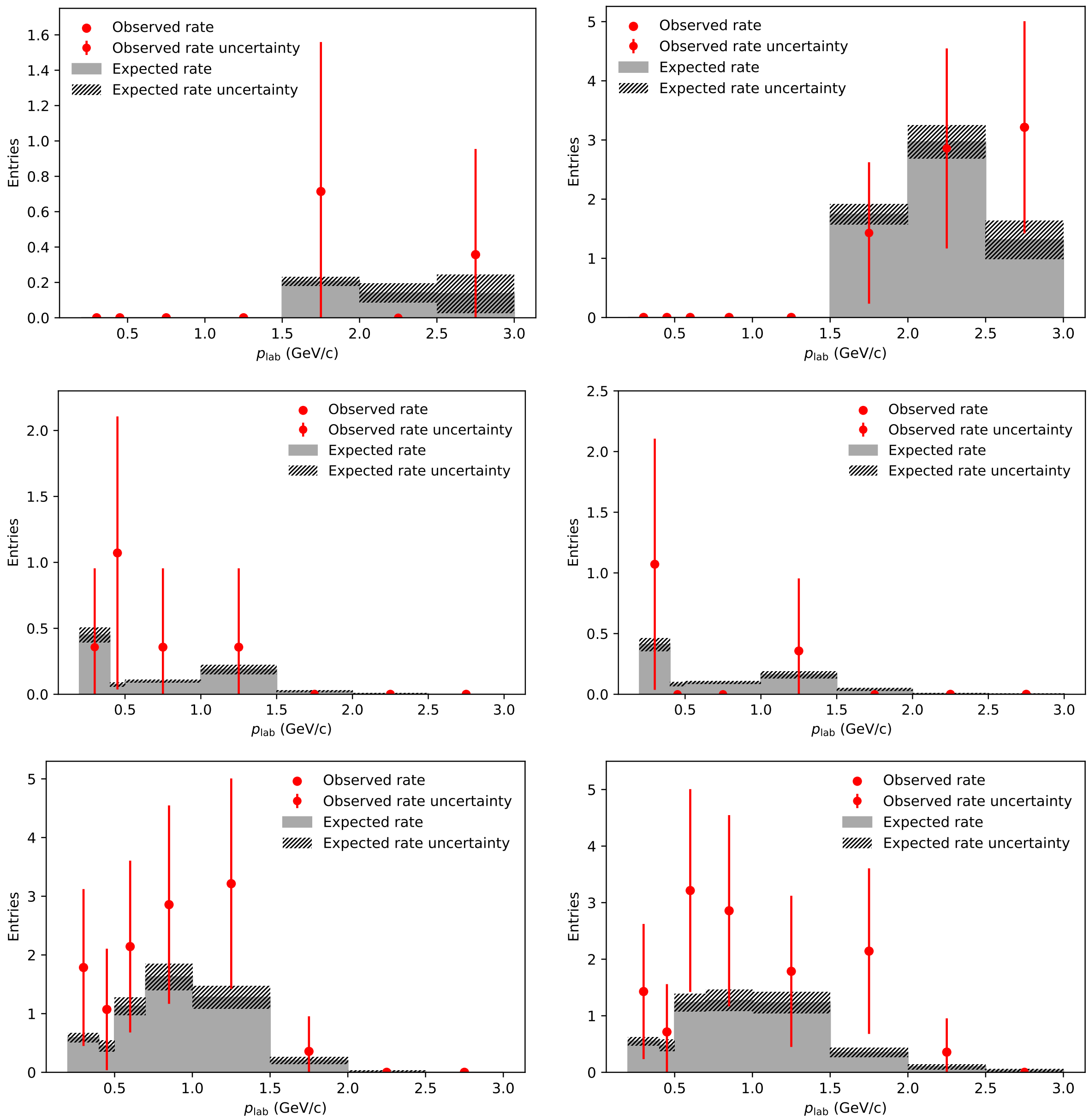


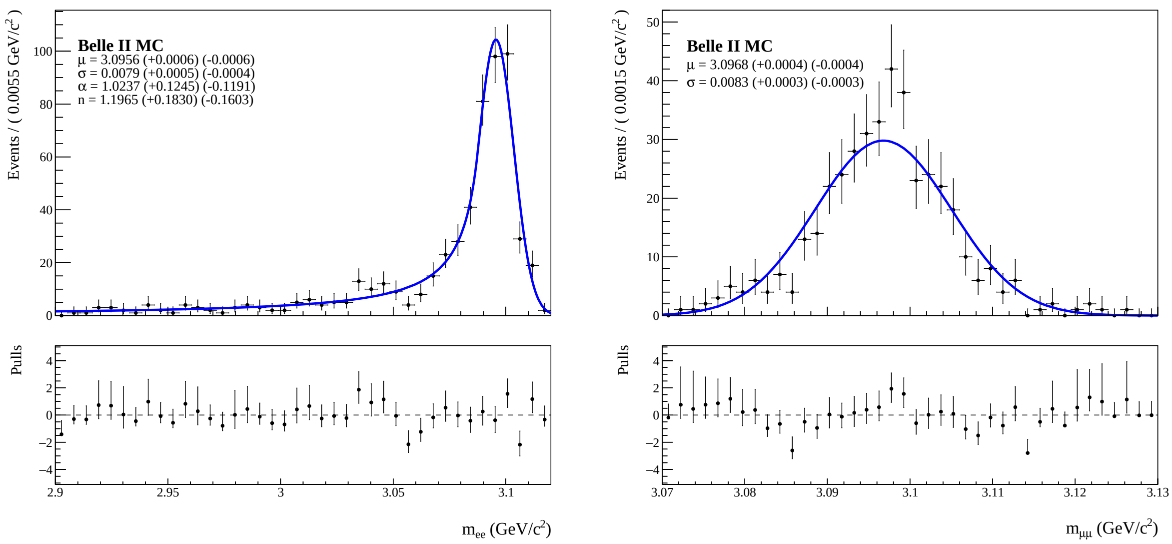
FIGURE 6.23. Distributions comparing the observed and expected fake rates for B^+B^- events. The steps used to derive these rates are outlined in Section 6.5.3. The fake rates for the signal lepton and $\tau \to e\nu\bar{\nu}$ and $\tau \to \mu\nu\bar{\nu}$ daughter leptons are shown on the top, middle and bottom respectively. Results for the $B^0 \to e^{\pm}\tau^{\mp}$ are on the left while results for the $B^0 \to \mu^{\pm}\tau^{\mp}$ are on the right.

6.5.4 J/ψ and ψ(2S) Veto

To remove the background contribution from B decays involving a J/ψ or ψ(2S), a veto is introduced based on the invariant mass of the daughter leptons $m_{\ell\ell}$. This is defined in Equation 6.4 where E_ℓ and $\vec{p_\ell}$ is the CMS energy and momentum of each lepton.

$$m_{\ell\ell} = \sqrt{\left(\sum E_\ell\right)^2 - \left(\left|\sum \vec{p_\ell}\right|\right)^2} \tag{6.4}$$

For the J/ψ veto, an extended maximum likelihood fit using `RooFit` [98] is performed to inform the choice of the veto window. For both the $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ modes, the events that are fit to comprise all the background events within the *fitting region* that have the signal and τ daughter lepton originating from a J/ψ decay. To increase the statistics, the events are derived from the `EvtGen` skim 24 ab^{-1} MC samples. Due to the bremsstrahlung tail for the $B^0 \to e^\pm\tau^\mp$ mode, a Crystal Ball function is used for the fit, while for the $B^0 \to \mu^\pm\tau^\mp$ mode, a simple Gaussian is sufficient. In both instances, the normalisation and shape parameters are allowed to float in the fit. The fits are shown in Figure 6.24 with the fitted parameters shown on the plots.



(A) $B^0 \to e^\pm\tau^\mp$ mode (B) $B^0 \to \mu^\pm\tau^\mp$ mode

FIGURE 6.24. Fits performed to the $m_{\ell\ell}$ for each mode for the J/ψ veto. The parameters of each fit function are displayed on the plots.

The J/ψ window for each mode was chosen so as to achieve roughly the same level of background reduction in either case. The chosen windows are:

- $[\mu - 5\sigma, \mu + 4\sigma] = [3.056, 3.127]$ GeV/c^2 for the $B^0 \to e^\pm\tau^\mp$ mode where $\mu = 3.0956$ and $\sigma = 0.0079$ are the fitted parameter values

Considering the 24 ab^{-1} sample for the largest background which are $B^0\overline{B}^0$ events, this veto removes 89% of the J/ψ background from the sample. Considering only the *leptonic partition*, the efficiency of signal within the *signal extraction region* changes from 0.0613% to 0.0592%.

- $[\mu - 2\sigma, \mu + 2\sigma] = [3.080, 3.113]$ GeV/c^2 for the $B^0 \to \mu^\pm\tau^\mp$ mode where $\mu = 3.0968$ and $\sigma = 0.0083$ are the fitted parameter values
  Considering the 24 ab^{-1} sample for the largest background which are $B^0\overline{B}^0$ events, this veto removes 93% of the J/ψ background from the sample. Considering only the *leptonic partition*, the efficiency of signal within the *signal extraction region* changes from 0.0621% to 0.0603%.

Note that an asymmetric window for the $B^0 \to e^\pm\tau^\mp$ mode was chosen to reflect the shape of the m_{ee} distribution. For the ψ(2S) veto, even with the 24 ab^{-1} statistics, not enough events are available for a fit. Instead, veto windows are derived using $\mu = 3.68609$ GeV/c^2 which is the current PDG value for the mass of the ψ(2S), with the σ in either case being equal to the width of the fitted J/ψ peaks for each respective mode. That is, it is assumed that the shape of the ψ(2S) peak would be approximately the same as that of the J/ψ peak, and in the case of the $B^0 \to e^\pm\tau^\mp$ mode, it would also be asymmetric. With this in mind, the chosen windows for the ψ(2S) peaks are:

- $[\mu - 4\sigma, \mu + 3\sigma] = [3.654, 3.710]$ GeV/c^2 for the $B^0 \to e^\pm\tau^\mp$ mode
  Considering the 24 ab^{-1} sample for the largest background which are $B^0\overline{B}^0$ events, this veto removes 50% of the ψ(2S) background from the sample. This 50% represents one of two events that lie in the *signal extraction region*. Considering only the *leptonic partition*, and following the application of the J/ψ veto, the efficiency of signal within the *signal extraction region* changes from 0.0592% to 0.0578%.
- $[\mu - 2\sigma, \mu + 2\sigma] = [3.669, 3.703]$ GeV/c^2 for the $B^0 \to e^\pm\tau^\mp$ mode
  Considering the 24 ab^{-1} sample for the largest background which are $B^0\overline{B}^0$ events, this veto removes all of the ψ(2S) background from the sample. Considering only the *leptonic partition*, and following the application of the J/ψ veto, the efficiency of signal within the *signal extraction region* changes from 0.0603% to 0.0592%.

6.6 Control Channel Study

In order to check the signal efficiency calculation, and determine what data-MC correction needs to be applied, a control channel study is performed. Finding a suitable control channel was difficult as it needed to fulfill the requirements of a visible two-body B decay with leptonic B granddaughters. Ultimately, it was decided to use the $B^0 \to D^-\pi^+$ decay, with the π^+, hereafter referred to as the signal pion, fulfilling the role of the signal lepton. By reconstructing the D^- in various semileptonic modes, the role of the τ can also be mimicked.

The list of reconstructed D^- modes is presented Table 6.26. As $\mathcal{B}(B^0 \to D^-\pi^+) = (2.51 \pm 0.08) \times 10^{-3}$, and the total branching fraction of reconstructed D^- modes is 13.72% (see Table 6.26), the effective branching fraction is $\mathcal{B}_{\text{eff}}(B^0 \to D^-\pi^+) = 0.34 \times 10^{-3}$.

No dedicated skims of $B^0 \to D^-\pi^+$ with an exclusive semileptonic D^- reconstruction exist. Therefore, the MC study of this control channel is performed using the existing 2.8 ab^{-1} MC sample of $B^0\overline{B}^0$ events to obtain the $B^0 \to D^-\pi^+$ events. The 2.8 ab^{-1} and 1 ab^{-1} MC samples of B^+B^- and $q\overline{q}$ events respectively are also used in the study.

TABLE 6.26. List of semileptonic D^- modes reconstructed for the control channel study. The **Effective Branching Fraction** of each mode includes the K_S^0 decay branching fraction for the relevant modes. All branching fraction numbers were obtained from the PDG.

D^- Decay Mode	Effective Branching Fraction (%)
$K_S^0(\to \pi^+\pi^-)\, e^-\overline{\nu}_e$	3.02
$K_S^0(\to \pi^+\pi^-)\, \mu^-\overline{\nu}_\mu$	3.03
$K^+\pi^-e^-\overline{\nu}_e$	4.02
$K^+\pi^-\mu^-\overline{\nu}_\mu$	3.65
Total	13.72

6.6.1 Analysis Setup

The analysis setup for the control mode is designed to mimic all possible aspects of the original $B^0 \to \ell^\pm\tau^\mp$ analysis. As with the original analysis, reconstruction of a $B^0 \to D^-\pi^+$ event begins with the hadronic reconstruction of the B_{tag}, which is achieved through the FEI. The standard skim cuts (outlined in Section 6.3) are applied along with all the same B_{tag} cuts as mentioned in Table 6.3.

For the signal-side, the applications of the selection cuts from the original analysis are modified to fit the control mode. Therefore, the selections for the signal-side of the control mode are explicitly outlined below:

- *Signal Pion*
  The signal pion is required to be the candidate with the highest momentum in the recoil frame of the top B_{tag} candidate. Instead of imposing a lepton ID cut, a likelihood-based global pion ID cut is imposed. The standard threshold of 0.3 was chosen.
- K_S^0
  All candidates formed from oppositely-charged pion tracks are required to pass a vertex fit and an invariant mass cut of $0.450 < M_{\text{inv}} < 0.550$ GeV/c^2.

- *D Daughters*
  All tracks have the same track-selection cuts for z_0, d_0 and p_T as those used in the FEI skim. For the leptonic daughters, BDT-based lepton ID cuts are used with the standard threshold of 0.9 chosen.

Lastly, all the ROE and event selections from Table 6.3 remain. To summarise the selections used, and to highlight the new selections that were not used in the original $B^0 \to \ell^\pm\tau^\mp$ analysis, Table 6.27 is presented. A combined cut-flow table for the reconstructed events is in Table 6.28. Due to issues with the grid computing system, the efficiency numbers for the combined $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ samples could not be obtained in time. These aforementioned samples were two orders of magnitude lower in efficiency than the $c\bar{c}$ sample for the *leptonic partition* in the $B^0 \to \ell^\pm\tau^\mp$ analysis, so they are a very minor contribution, if any. The best candidate selection procedure from Figure 6.5 is modified to remove the hadronic τ mode candidate selection, and instead to only focus on selecting the event based on the lepton with the same method as the original analysis to ensure the control mode is as representative, and therefore valid, as possible. The same *fitting* and *signal extraction regions* are chosen, namely $p_\ell^B \in [2, 2.4] \times E_{\mathrm{extra}} \in [0, 1.5]$ for the former and $p_\ell^B \in [2.24, 2.4] \times E_{\mathrm{extra}} \in [0, 0.3]$ for the latter. Projection distributions of the *fitting region* for the reconstructed $B^0\overline{B}^0$ events are given in Figure 6.28. The categories shown here comprise multiple sub-categories that were initially kept separate before being combined into the broad categories shown in the figure. Notable aspects of this grouping are explained below:

- $D^-(\to X\ell\nu)\,\pi^+$
  Not only does this category contain the correctly reconstructed $D^-\pi^+$ events, but also the events where the signal-side was a correctly reconstructed $B^0 \to D^-\pi^+$ candidate but the tag-side was incorrectly reconstructed. This component displayed the same shape as correctly reconstructed $D^-\pi^+$ events but with a wider peak. This is shown in Figure 6.25, with a decomposition of the incorrect tag-side modes shown in Figure 6.26. The decomposition shows that either high track or π^0 multiplicity modes contribute the largest which is unsurprising due to the increased difficulty in reconstruction. In the *signal extraction region* there are some smaller contributions from $D^-(\to \rho/\omega\ell\nu)\,\pi^+$ events that were combined with correctly reconstructed $D^-\pi^+$ events. Incorrectly reconstructed $D^-\pi^+$ events that arose due to cross-feed between the reconstructed D^- modes were also absorbed into the $D^-(\to X\ell\nu)\,\pi^+$ category.
- $D^{*-}X_{\mathrm{had}}$ *and* D^-X_{had}
  A large component in the tail of the distribution originated from $D^{(*)-}\rho^+$ decays. This was split into the D^- and D^{*-} modes and combined with other smaller $D^{*-}X_{\mathrm{had}}$ and D^-X_{had} components to form these broader categories.

As the $D^-(\to X\ell\nu)\,\pi^+$ and $D^-(\to X_{\mathrm{had}})\,\pi^+$ components in Figure 6.28 both peak in the same location and are similar in shape, the “signal" distribution for the control mode is considered to be both

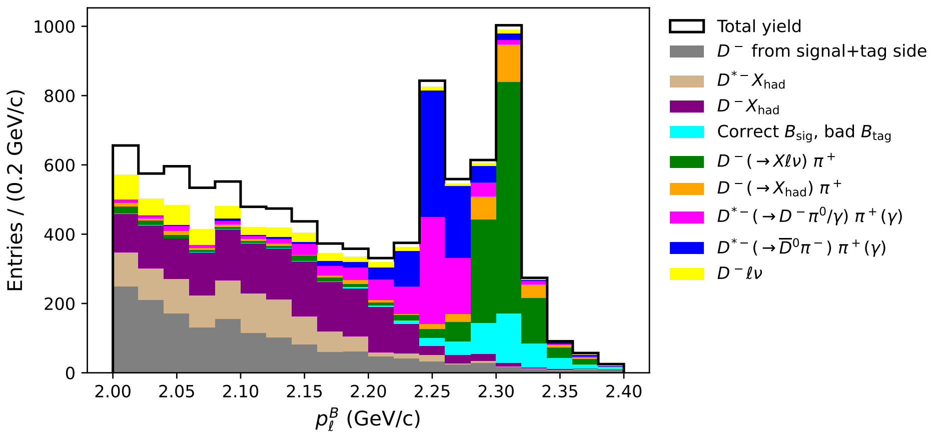


FIGURE 6.25. Projections of the p_ℓ^B *fitting region* for $B^0\overline{B}^0$ events reconstructed in the control mode. In each instance, X or $X_{\rm had}$ refers to any collection of hadronic particles while simply D without a charge or sign indicates any D meson. The category "D^- *from signal+tag side*" means that the D^- candidate is composed of one or more tracks from the tag-side. For all other categories, all the tracks forming the $D^{(*)-}$ candidate come from the signal-side. For the case of π^0/γ, this refers to both the $D^{*-} \to D^-\pi^0$ and $D^{*-} \to D^-\gamma$ decays.

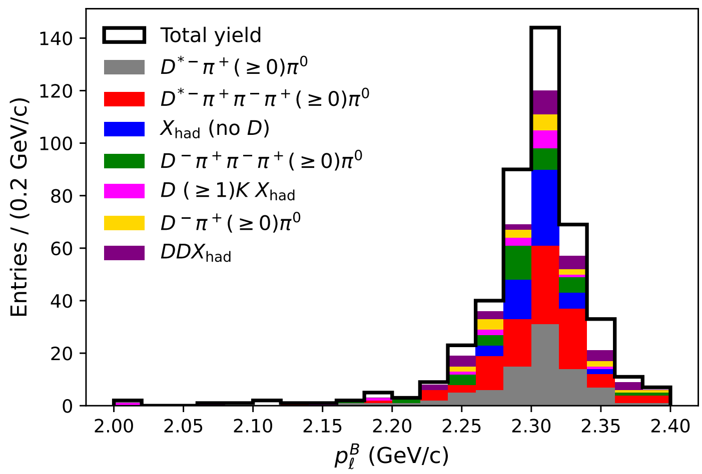


FIGURE 6.26. Projections of the p_ℓ^B *fitting region* for $B^0\overline{B}^0$ events where the signal-side is correctly reconstructed as $D^-\pi^+$ but the tag-side is incorrectly reconstructed. In each instance, X or $X_{\rm had}$ refers to any collection of hadronic particles while simply D without a charge or sign indicates any D meson. The distribution is decomposed according to the generator-level tag-side modes.

of these categories combined. As they are both ultimately $B^0 \to D^-\pi^+$ events, such a definition of "signal" is conceptually reasonable. Figure 6.29 compares the shape of signal for the control mode,

TABLE 6.27. List of cuts applied for the control channel mode following the reconstruction of the B_{tag} using the FEI. Cuts that are new compared to the original $B^0 \to \ell^\pm\tau^\mp$ analysis are shown in magenta.

Particle	**Cut Applied**
B_{tag}	$M_{\text{bc}} > 5.27$ GeV/c^2, $-0.15 < \Delta E < 0.1$ GeV, `sigProb` > 0.001
Signal pion	$1.45 < p_{\text{cms}} < 2.55$ GeV/c, **likelihood pion ID > 0.3**
K_S^0	**$0.450 < M_{\text{inv}} < 0.550$ GeV/c^2**
D daughters	$\lvert z_0\rvert < 2$ cm, $\lvert d_0\rvert < 0.5$ cm, $p_T > 0.1$ GeV/c BDT electron ID (BDT muon ID) > 0.9 for $\ell = e$ (μ) for leptons
ROE tracks	$\lvert z_0\rvert < 4$ cm, $\lvert d_0\rvert < 2$ cm, $p_{\text{T}} > 0.1$ GeV/c, N_{tracks} in ROE = 0
ROE clusters	$E > 0.055$ GeV, `inCDCAcceptance` $= 1$ `beamBackgroundSuppression` > 0.4, `fakePhotonSuppression` > 0.2
Event	B_{tag} `cosTBTO` < 0.9, $-0.2 < U_{\text{CM}} = E_{\text{miss}_{\text{CM}}} - \lvert\vec{p}_{\text{miss}_{\text{CM}}}\rvert < 1.5$ GeV

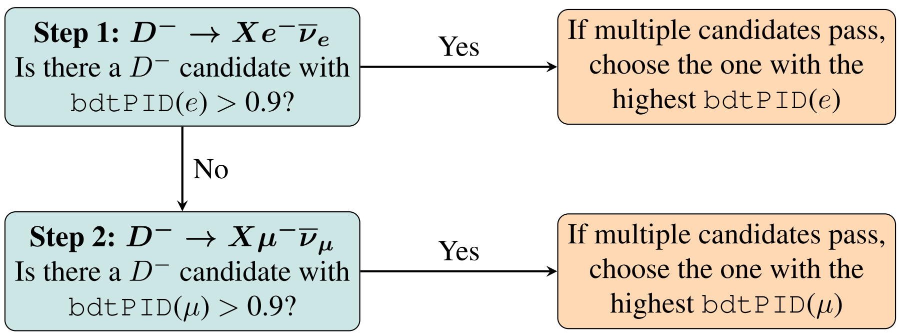


FIGURE 6.27. Flowchart depicting the stages of the best candidate selection process for the control mode. The X refers to either the K_S^0 or $K^+\pi^-$ pair.

alongside all other background samples including B^+B^- and $q\overline{q}$ events. The $D^{*-}(\to D^-\pi^0/\gamma)\,\pi^+(\gamma)$ and $D^{*-}(\to \overline{D}^0\pi^-)\,\pi^+(\gamma)$ components are kept separate as they are peaking in p_ℓ^B. From this picture, it is clear for the next steps to be a $D^{*-} \to \overline{D}^0\pi^-$ veto, followed by a three component fit to the following components: signal, D^{*-} backgrounds, and all other backgrounds. A veto on the $D^{*-} \to D^-\pi^0$ background was not considered as it is likely that a suitable veto would sculpt the signal shape. This is because such events are ultimately signal but with an additional π^0 produced in the event. Furthermore, the ultimate goal of this control channel study is to obtain the $B^0 \to D^-\pi^+$ yield, and this can be achieved by keeping the $D^{*-} \to D^-\pi^0$ peaking component but properly modelling it during the fit so that its shape does not interfere with that of the signal.

TABLE 6.28. Combined cut-flow table for the $D^-(\to X\ell\nu)\,\pi^+$ and $D^-(\to X_{\rm had})\,\pi^+$ components, which comprise the signal for the control mode, and the background components. The numbers represent the percentage of events that remain out of the total generated events. For the signal components, the FEI skim and $B_{\rm tag}$ vertex fit efficiency is assumed to be the same as for $B^0\overline{B}^0$ events overall which is why they are wrapped in brackets.

Cuts	$\boldsymbol{D^-(\to X\ell\nu)\,\pi^+}$	$\boldsymbol{D^-(\to X_{\rm had})\,\pi^+}$	$\boldsymbol{B^+B^-}$	$\boldsymbol{c\bar{c}}$
Generated	100	100	100	100
FEI skim	(8.589)	(8.589)	7.995	10.746
$B_{\rm tag}$ vertex fit	(4.826)	(4.826)	4.294	5.765
Track + π^0 selections	1.519	2.335	2.693	3.876
Signal lepton $p_{\rm cms}$	0.797	1.071	0.239	0.594
Signal lepton PID	0.669	0.887	0.114	0.321
D lepton PID	0.416	0.056	0.034	0.025
$M_{\rm bc}$	0.346	0.044	0.004	0.006
ΔE	0.280	0.041	0.003	0.005
$\cos\theta_{\rm TBTO}$	0.248	0.037	0.002	0.001
$U_{\rm CM}$	0.244	0.036	0.002	0.001
$N_{\rm tracks}$ in ROE	0.204	0.024	0.001	5.936×10^{-4}
Fitting Regions				
$E_{\rm extra}\in[0,1.5]$	0.189	0.019	1.296×10^{-4}	5.644×10^{-5}
$p_\ell^B\in[2,2.4]$	0.155	0.016	1.156×10^{-4}	5.022×10^{-5}

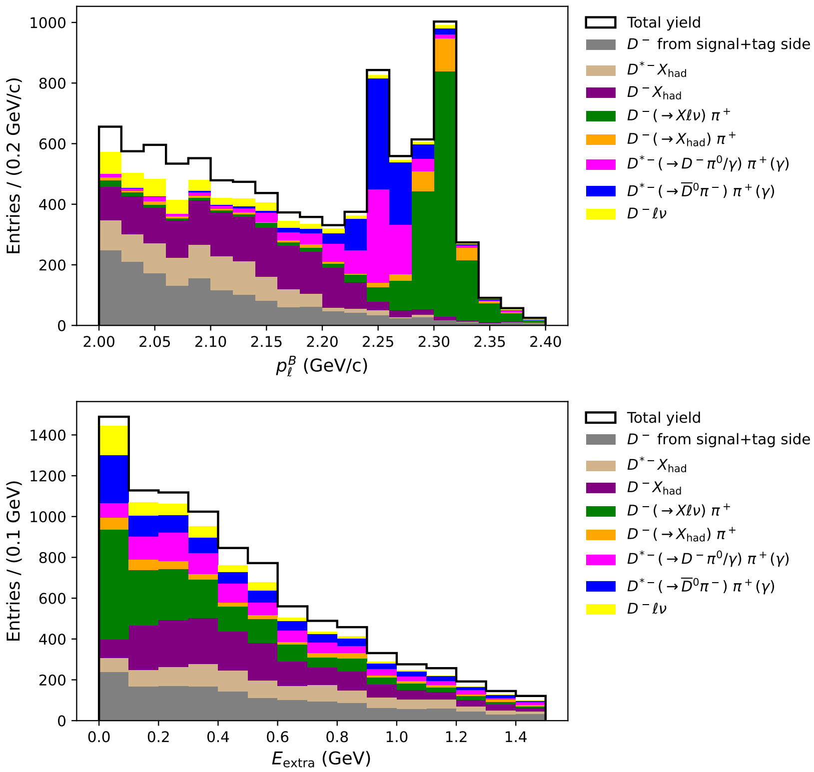


FIGURE 6.28. Projections of the *fitting region* for $B^0\overline{B}^0$ events reconstructed in the control mode. In each instance, X or $X_{\rm had}$ refers to any collection of hadronic particles. The category "D^- *from signal+tag side*" means that the D^- candidate is composed of one or more tracks from the tag-side. For all other categories, all the tracks forming the $D^{(*)-}$ candidate come from the signal-side. For the case of π^0/γ, this refers to both the $D^{*-} \to D^-\pi^0$ and $D^{*-} \to D^-\gamma$ decays.

6.6.2 $D^{*-} \to \overline{D}^0\pi^-$ Veto

The veto for the $D^{*-}(\to \overline{D}^0\pi^-)\,\pi^+(\gamma)$ component is based on recovering the reconstructed D^{*-} and D^0 masses from the recoil particles. These masses are obtained using Equation 6.5, where the summation over i in the equation changes depending on the recoil mass to be calculated. In the instance of $M_{\rm recoil}(B_{\rm tag}, \pi^+_{\rm signal}, \pi^-)$, which recovers the D^0 mass, the summation is over the reconstructed tag-side B, the signal pion, and the pion daughter of the reconstructed D which is oppositely charged to the signal pion. For $M_{\rm recoil}(B_{\rm tag}, \pi^+_{\rm signal})$, which recovers the D^{*-} mass, the summation is just over the reconstructed tag-side B and the signal pion.

$$M_{\rm recoil}(B_{\rm tag}, \pi^+_{\rm signal}, \pi^-),\; M_{\rm recoil}(B_{\rm tag}, \pi^+_{\rm signal}) = \sqrt{\left(\sum_i E_i\right)^2 - \left|\sum_i \vec{p}_i\right|^2} \tag{6.5}$$

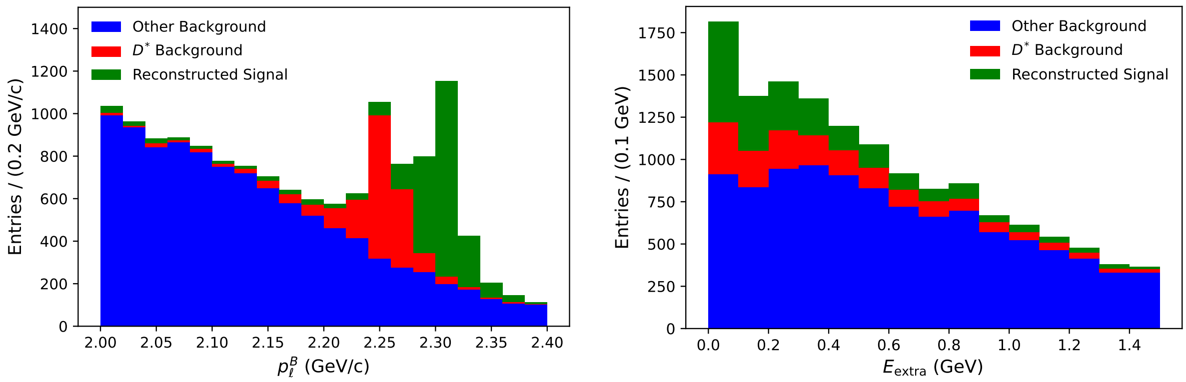


FIGURE 6.29. Projections of the *fitting region* for events reconstructed in the control mode. The *Reconstructed Signal* component contains both the $D^-(\to X\ell\nu)\ \pi^+$ and $D^-(\to X_{\text{had}})\ \pi^+$ components. The $D^{*-}(\to D^-\pi^0/\gamma)\ \pi^+(\gamma)$ and $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ components comprise the D^* *Background* category. All other categories from $B^0\overline{B}^0$ events, in addition to B^+B^- and $q\overline{q}$ events, are labelled as *Other Background*.

Distributions of each recoil mass for the different $B^0\overline{B}^0$ event categories are given in Figure 6.30. In these distributions, the $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ component peaks at the expected locations which are the D^{*-} and D^0 masses. Initially, the difference between these recoil masses defined by $M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}}, \pi^-) - M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}})$ was tested as a veto. However, there was little separation between the signal and $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ components for this quantity so it was not used. Instead, these two quantities were considered simultaneously, and by observing the 2D plots in Figure 6.31 it can be seen that a simple cut on both quantities is effective at removing the $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ background. Following on from this, the veto implemented is: $1.90 < M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}}) < 2.15$ GeV/c ^{2} and $1.74 < M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}}, \pi^-) < 2.05$ GeV/c ^{2}. Events that fall inside this window are removed. The efficiency of the veto for various event categories is provided in Table 6.29. It can be seen that the veto keeps most of signal events while removing a majority of the target $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ background component. This demonstrates the effectiveness of the veto. However, by observing the shape of the $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ component in Figure 6.31, it can be seen that an improved veto is attainable by having a non-linear window within which events are removed. This idea has yet to be explored. Following the veto, the resulting projections of the *fitting region* is given in Figure 6.32. It can be seen that the D^* background peak is reduced, and that this is due to the reduction in the $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ component.

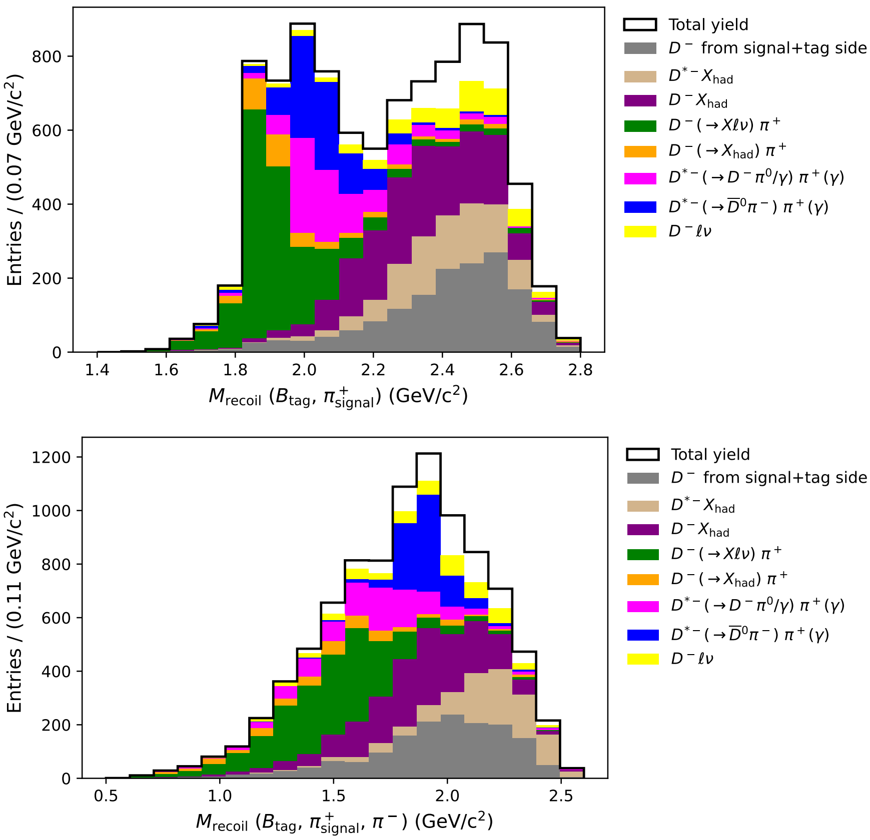
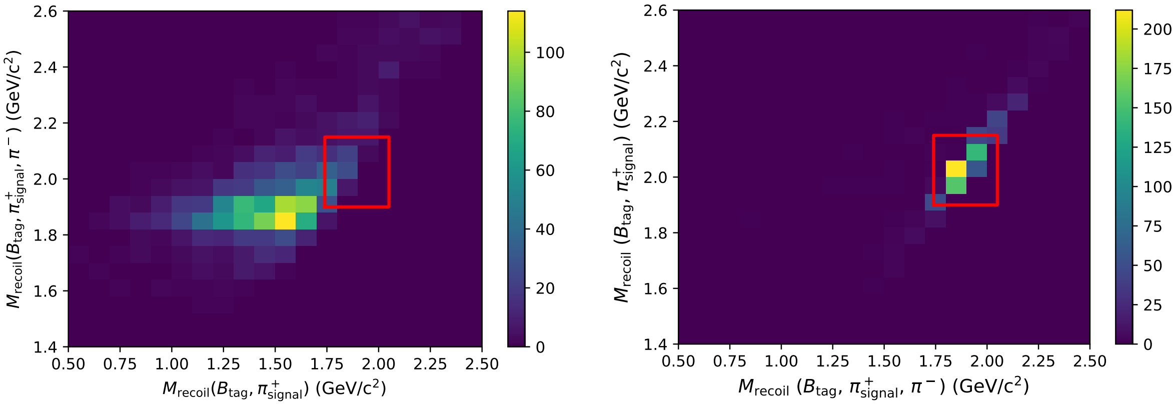


FIGURE 6.30. Distributions of the $M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}})$ and $M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}}, \pi^-)$ variables for the different $B^0\overline{B}^0$ event categories.



(A) $D^-(\to X\ell\nu)\,\pi^+$ (B) $D^{*-}(\to \overline{D}^0\pi^-)\,\pi^+(\gamma)$

FIGURE 6.31. Distributions of the $M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}}, \pi^-) \times M_{\text{recoil}}(B_{\text{tag}}, \pi^+_{\text{signal}})$ parameter space for $D^-(\to X\ell\nu)\,\pi^+$ and $D^{*-}(\to \overline{D}^0\pi^-)\,\pi^+(\gamma)$ events. The red lines and box indicates the veto window, with all events inside the red box being removed.

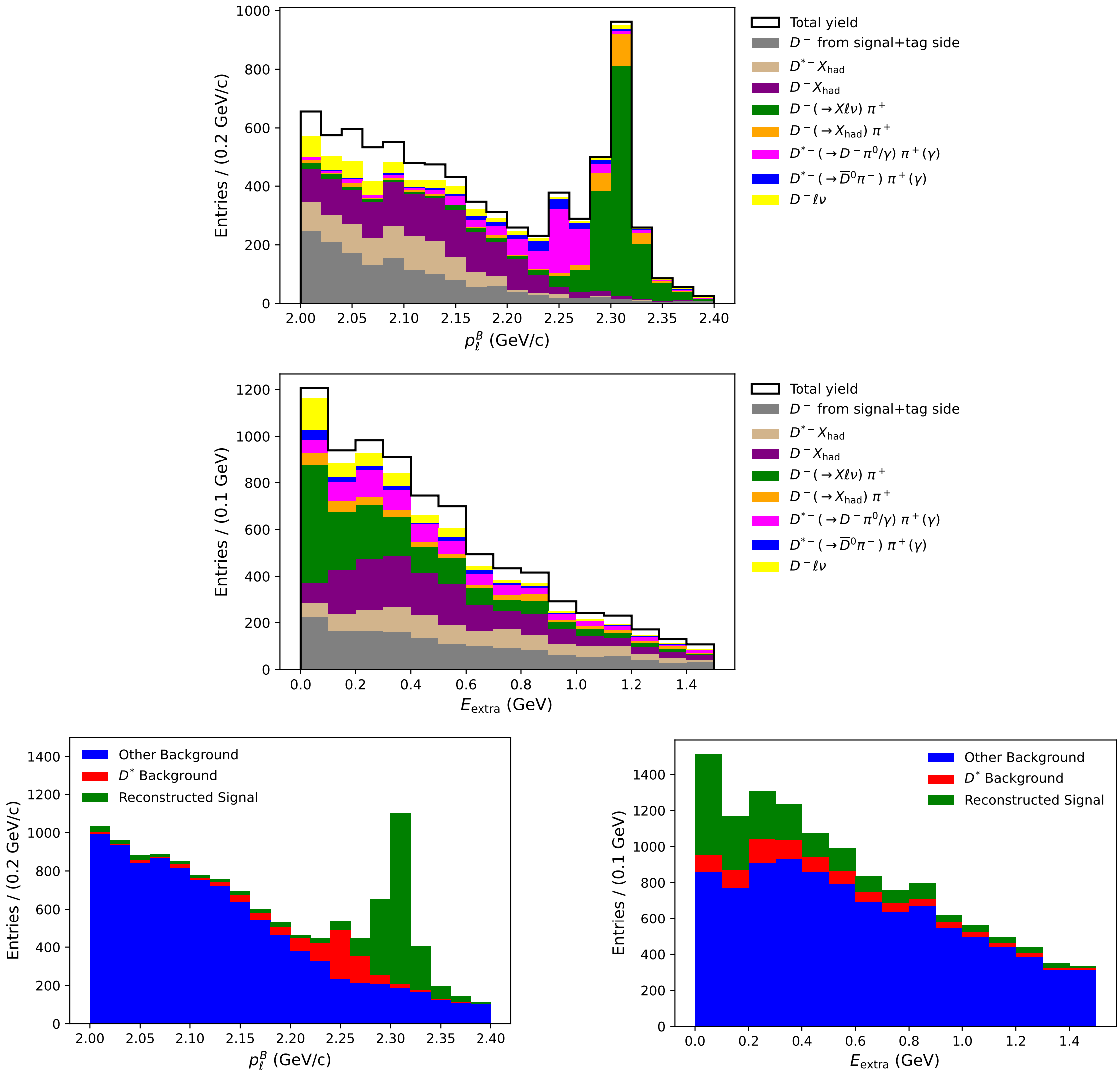


FIGURE 6.32. Projections of the *fitting region* for events reconstructed in the control mode following the $D^{*-} \to D^0\pi^-$ veto. The top two plots contain just the $B^0\overline{B}^0$ event categories. In the bottom left and right plots, the *Reconstructed Signal* component contains both the $D^-(\to X\ell\nu)\ \pi^+$ and $D^-(\to X_{\rm had})\ \pi^+$ components, while the $D^{*-}(\to D^-\pi^0/\gamma)\ \pi^+(\gamma)$ and $D^{*-}(\to \overline{D}^0\pi^-)\ \pi^+(\gamma)$ components comprise the D^* *Background* category. In these two plots, all other categories from $B^0\overline{B}^0$ events, in addition to B^+B^- and $q\overline{q}$ events, are labelled as *Other Background*.

TABLE 6.29. Efficiency of the $D^{*-} \to D^0\pi^-$ veto for different event categories.

Event Category	**Veto Efficiency (%)**
Signal	93.11
$D^{*-}(\to D^-\pi^0/\gamma)\,\pi^+(\gamma)$	78.93
$D^{*-}(\to \overline{D}^0\pi^-)\,\pi^+(\gamma)$	22.62
Other $B^0\overline{B}^0$	95.50
B^+B^-	93.17
$q\overline{q}$	92.37

6.6.3 Fitting and Results

To obtain the $B^0 \to D^-\pi^+$ yield, a 2D binned extended maximum likelihood fit to the combined distributions in Figure 6.32 is performed. This is performed using the `RooFit` framework [98]. The fit has the following three components: reconstructed signal; D^* background; other background. To obtain the shape of each component, individual fits to each are first performed during which all parameters of the PDF are floated. After the shape is ascertained, a combined fit is performed with all components included at which point the shapes of each PDF is fixed, and only the normalisations are allowed to float. For each component, the PDF shapes are outlined below:

- *Signal*
  The p_ℓ^B is fit using a Johnson function from the `RooJohnson` class of `RooFit` while E_{extra} is fit using an exponential from the `RooExponential` class of `RooFit`. The Johnson function is given in Equation 6.6 where μ and λ represent the mean and width respectively of the Gaussian component, δ represents the strength of this component in the overall function, and γ controls the tail distortions so as to achieve an asymmetry similar to that of a double-sided Crystal Ball function. A Johnson function has the advantage of having two less parameters than a double-sided Crystal Ball function.

$$F_J(x) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+\left(\frac{x-\mu}{\lambda}\right)^2}} \exp\left[-\frac{1}{2}\left(\gamma + \delta \sinh^{-1}\left(\frac{x-\mu}{\lambda}\right)^2\right)\right] \tag{6.6}$$

The signal PDF is given in Equation 6.7 where $F_{\rm J}$ indicates the Johnson function. The fit was restricted to the range $p_\ell^B \in [2.2, 2.4] \times E_{\rm extra} \in [0, 1.5]$ in order to help the fit converge. All signal events that fall outside this range were included in the *Other Background* component in the fit. The shape parameters obtained from this fit are used for the signal PDF that is fit over the entire *fitting region*. Projections of the fit are provided in Figure 6.33.

$$\mathrm{PDF}_{\rm Signal}(x, y) = \mathrm{PDF}_{p_\ell^B}(x) \times \mathrm{PDF}_{E_{\rm extra}}(y) = N \cdot F_{\rm J}(x) \cdot \ \exp(-cy) \tag{6.7}$$

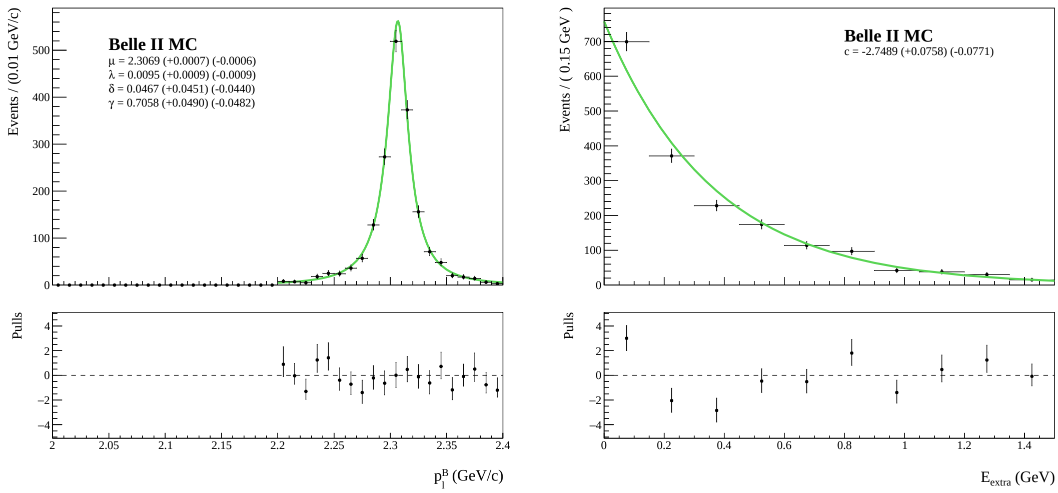


FIGURE 6.33. Projections of the 2D fit to an MC sample of the *Signal* component. The fitted shape parameters are displayed on each projection for the corresponding fit function.

- D^* *Background*

  The p_ℓ^B is fit using a Johnson function from the `RooJohnson` class of `RooFit` while $E_{\rm extra}$ is fit using a double Gaussian with both Gaussians coming from the `RooGaussian` class of `RooFit`. A double Gaussian is needed to account for the peaked component in the $E_{\rm extra}$ distribution at around 0.3 GeV which arises from photons in the ROE originating from $D^{*-} \to D^-\pi^0$ decays (see Appendix F). The PDF is given in Equation 6.8 where $F_{\rm J}$ indicates the Johnson function whose form is given in Equation 6.6, and f parametrises the amount of each Gaussian component. Projections of the fit along each dimension are provided in Figure 6.34.

$$\begin{aligned}\mathrm{PDF}_{D^*}(x, y) &= \mathrm{PDF}_{p_\ell^B}(x) \times \mathrm{PDF}_{E_{\rm extra}}(y) \\ &= N \cdot F_{\rm J}(x) \cdot \left[f\exp\left(\frac{-(y-\mu_1)^2}{2\sigma_1^2}\right) + (1-f)\exp\left(\frac{-(y-\mu_2)^2}{2\sigma_2^2}\right)\right]\end{aligned} \tag{6.8}$$

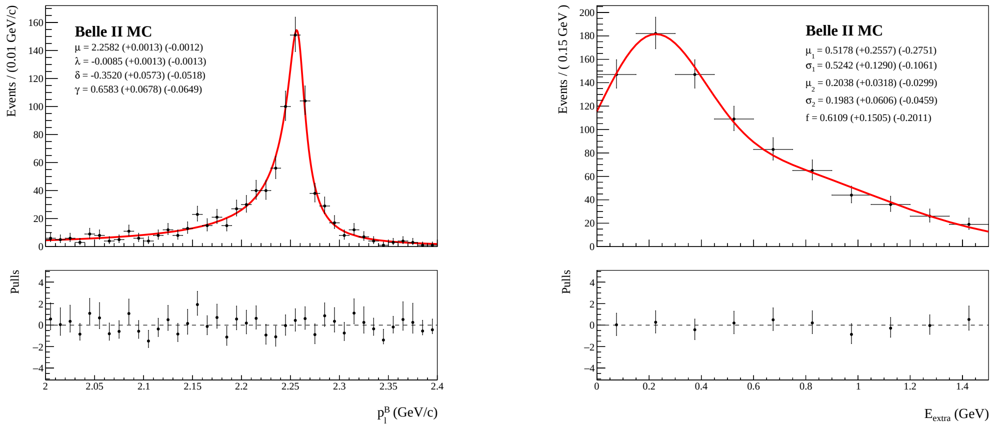


FIGURE 6.34. Projections of the 2D fit to an MC sample of the D^* *Background* component. The fitted shape parameters are displayed on each projection for the corresponding fit function.

- *Other Background*
  Both the p_ℓ^B and E_{extra} dimensions are fitted using Chebychev polynomials from the `RooChebychev` class of of `RooFit`. A fourth-order polynomial is used to model p_ℓ^B while a third-order polynomial is used for E_{extra}. The PDF is given in Equation 6.9. Projections of the fit along each dimension are provided in Figure 6.35.

$$\begin{aligned} \text{PDF}_{\text{other}}(x,y) &= \text{PDF}_{p_\ell^B}(x) \times \text{PDF}_{E_{\text{extra}}}(y) \\ &= N \cdot \left(1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4\right) \cdot \left(1 + b_1 y + b_2 y^2 + b_3 y^3\right) \end{aligned} \tag{6.9}$$

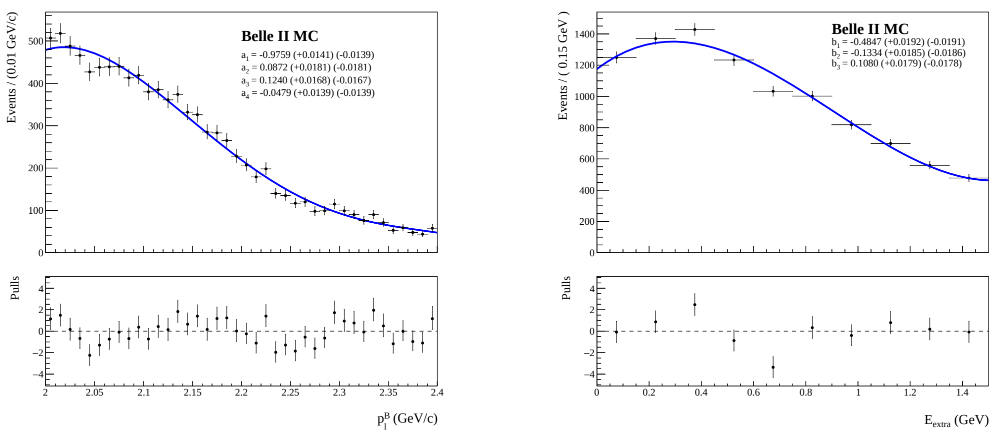


FIGURE 6.35. Projections of the 2D fit to an MC sample of the *Other Background* component. The fitted shape parameters are displayed on each projection for the corresponding fit function.

The different components are combined into a single fit according to Equation 6.10. The shape parameters for each PDF is fixed while their respective normalisation is floated. Projections of the combined fit along each dimension are provided in Figure 6.36.

$$\mathrm{PDF}_{\mathrm{total}}(x,y) = \mathrm{PDF}_{\mathrm{Signal}}(x,y) + \mathrm{PDF}_{D^*}(x,y) + \mathrm{PDF}_{\mathrm{other}}(x,y) \tag{6.10}$$

The fitted yields for each component are as follows, with the true yields from MC provided in the brackets for reference:

- *Signal* = 1877.39 ± 53.61 ($N_{\mathrm{MC}}^{\mathrm{signal}} = 1808$)
- D^* *background* = 773.54 ± 64.26 ($N_{\mathrm{MC}}^{D^*} = 858$)
- *Other background* = 9885.50 ± 114.66 ($N_{\mathrm{MC}}^{\mathrm{other}} = 9870$)

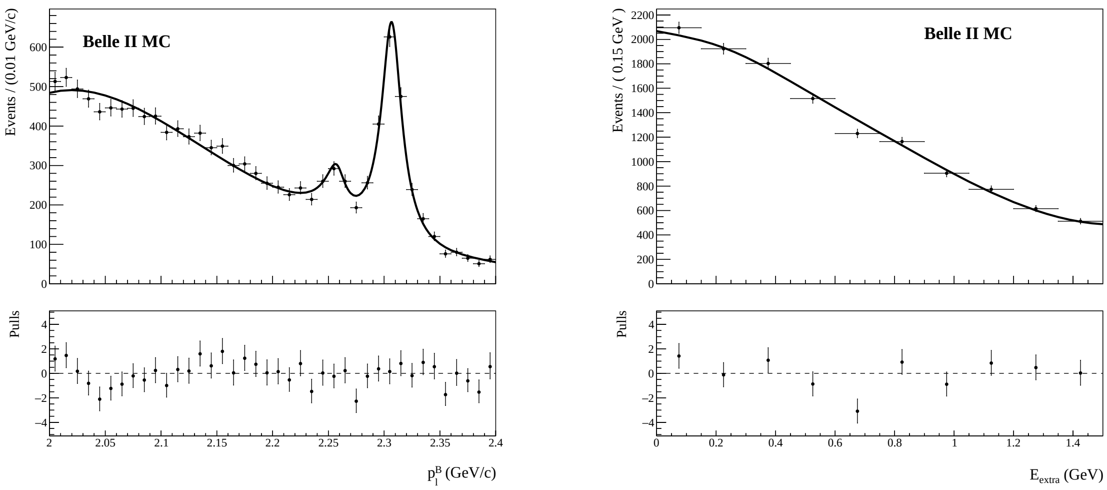


FIGURE 6.36. Projections of the 2D fit to the MC sample containing all the components across the entire *fitting region*.

It is clear that while the fitted yield for the *Other background* component agrees with the expected yield from MC, both the *Signal* and D^* *background* components agree only within 1.3σ. This is likely due to their close peaks in p_ℓ^B which leads to some D^* *background* events being absorbed into the *Signal* component during the fit. In order to determine if these deviations in the fitted yields were a result of some bias in the fit or statistical fluctuations in the specific MC sample used, a toy ensemble test was conducted. The results are presented in Section 6.6.4 below.

6.6.4 Toy Ensemble Study

To determine if the fit contains any bias, a toy ensemble test was conducted. The goal of the test was to obtain pull plots for the fitted yields for each of the three components. The steps used to perform the toy ensemble study are outlined below:

(1) Using the shape parameters obtained from Section 6.6.3, a PDF for each of the three components is defined. These are used to generate the data for each toy.

(2) For each of the 1000 toys in the ensemble test, the following steps are performed:

  (a) Three different datasets are generated separately for each component using their respective fixed PDF from *Step 1*. The size of the dataset is randomly drawn from a Poisson distribution with means equal to their respective expected yields from MC. For example, a Poisson distribution with a mean equal to 1808 is used for the *Signal* component.

  (b) A new PDF is defined for each component and is fit to the corresponding generated dataset from *Step 2a*. Due to the complexity of the 2D PDF for the *Signal* and D^* *background* components, these fits were performed in stages.

  For the *Signal* component, the first iteration involved fixing only the μ parameter from the Johnson PDF for p_ℓ^B to the value obtained from the fit in Section 6.6.3, with all other shape parameters in Equation 6.7 floated. In the second iteration, all the shape parameters that were floated in the first fit were fixed to their fitted values, with only the μ parameter from the Johnson PDF for p_ℓ^B floated. The μ parameter was then fixed to the value from this second fit.

  For the D^* *background*, the first iteration involved fixing the μ parameter from the Johnson PDF for p_ℓ^B and the Gaussian fraction f for the double Gaussian PDF for E_{extra} to their values from the fit in Section 6.6.3, with all other shape parameters in Equation 6.8 floated. In the second iteration, all the shape parameters that were floated in the first fit were fixed to their fitted values, with only the μ parameter from the Johnson PDF for p_ℓ^B and the Gaussian fraction f for the double Gaussian PDF for E_{extra} floated. The μ and f parameters were then fixed to the value from this second fit.

  (c) The PDF of each component obtained in *Step 2b* were combined and fit to the combined generated data from *Step 2a*. In this fit, only the normalisation of each PDF component was floating. The fitted normalisation was extracted from the fit and saved.

  (d) The pull for each of the three fitted yields were calculated separately using Equation 6.11 if the fitted yield was higher than the expected yields from the original MC sample, or Equation 6.12 if the fitted yield was lower. For example, if the fitted yield for the *Signal* component was 1850, then Equation 6.11 would be used to calculate the pull for this toy. In these equations, $\sigma_{\text{fitted}}^{\text{low}}$ and $\sigma_{\text{fitted}}^{\text{high}}$ respectively represent the lower and upper error on the fitted yield. Such a definition of the pull allows for a more Gaussian distribution with less skew in the pull distribution.

$$N_{\text{pull}} = \frac{N_{\text{fitted}} - N_{\text{true}}}{|\sigma_{\text{fitted}}^{\text{low}}|} \tag{6.11}$$

$$N_{\text{pull}} = \frac{N_{\text{fitted}} - N_{\text{true}}}{\sigma_{\text{fitted}}^{\text{high}}} \tag{6.12}$$

(3) When the pull for fitted yields of each of the 1000 toys has been calculated, a Gaussian fit using `RooFit` is performed to the three pull distributions. The parameters of the Gaussian fit and their errors are saved in each case in order to determine if there is any bias in the yields.

This ensemble test was performed using expected yields from the 2.8 (1.0) ab^{-1} MC sample for $B\overline{B}$ events ($q\overline{q}$ events), and performed for the real case of 365 fb^{-1}. The 365 fb^{-1} expected yields used for the ensemble test are as follows: $N_{\mathrm{MC}}^{\mathrm{signal}} = 234$; $N_{\mathrm{MC}}^{D^*} = 111$; $N_{\mathrm{MC}}^{\mathrm{other}} = 2064$. The pull plot distributions are shown in Figure 6.37 for both sets of expected yields. None of the 1000 toy fits failed for any of the components for either set of expected yields used for the ensemble setup.

For an unbiased fit, a Gaussian fit on the pulls would have a mean consistent with 0 and a standard deviation consistent with 1 within their respective errors. In all cases, the pull plots indicate that the fit is biased for the D^* and *Other background* components, and also that the spread of yields is less than expected, especially for the D^* *background.* The situation improves to some degree when looking at the 365 fb^{-1} case. It is not yet currently understood what is causing this irregular behaviour in the pull plots. In order to determine if it is an issue with how the fit is implemented, a pull plot on the combined fitted yield, that is $N_{\mathrm{MC}}^{\mathrm{combined}} = N_{\mathrm{MC}}^{\mathrm{signal}} + N_{\mathrm{MC}}^{D^*} + N_{\mathrm{MC}}^{\mathrm{other}}$, was calculated. The corresponding error is calculated according to Equation 6.13 with all terms derived from the fit result. For a properly implemented fit, this pull plot should be well-behaved. Figure 6.38 presents this pull plot, and indeed the Gaussian fit yields a mean consistent with 0 and a variance consistent with 1. As this demonstrates that there is no issue with fit implementation, it suggests that some correlations between the different components is likely to exist and therefore skew the shape of the pulls. Investigating this possibility remains a point of further work for this study and is mentioned in Section 7.1.

$$\begin{aligned}\Sigma^2 = {} & \sigma^2_{\mathrm{signal}} + \sigma^2_{D^*} + \sigma^2_{\mathrm{other}} + \\ & 2\left[\mathrm{Cov}(N_{\mathrm{MC}}^{\mathrm{signal}}, N_{\mathrm{MC}}^{D^*}) + \mathrm{Cov}(N_{\mathrm{MC}}^{\mathrm{signal}}, N_{\mathrm{MC}}^{\mathrm{other}}) + \mathrm{Cov}(N_{\mathrm{MC}}^{\mathrm{other}}, N_{\mathrm{MC}}^{D^*})\right] \qquad (6.13)\end{aligned}$$

6.6.5 Linearity Study

Because the branching fractions of $B^0 \to D^-\pi^+$ and the reconstructed D modes have an uncertainty attached, the true yield of $B^0 \to D^-\pi^+$ can vary across some range. A linearity study is therefore conducted to determine if the fit maintains the correct behaviour for a range of different *Signal* component yields. Given the uncertainty on $\mathcal{B}(B^0 \to D^-\pi^+)$ and $\mathcal{B}(D^- \to X\ell\nu)$ is $\approx 3\%$ and $\approx 19\%$ respectively, the linearity test is performed across the range of $[0.75 \times N_{\mathrm{MC}}^{\mathrm{signal}}, 1.25 \times N_{\mathrm{MC}}^{\mathrm{signal}}]$ in steps of $0.05 \times N_{\mathrm{MC}}^{\mathrm{signal}}$, where $N_{\mathrm{MC}}^{\mathrm{signal}} = 234$.

The linearity test is conducted by performing almost all the same steps as the ensemble toy test in Section 6.6.4 for each $N_{\mathrm{MC}}^{\mathrm{signal}}$ value in the testing range. The only difference is after *Step 2c*, where a Gaussian fit using `RooFit` is performed on the fitted yield values rather than on the pulls. This is so that the averaged fitted yield for a given $N_{\mathrm{MC}}^{\mathrm{signal}}$ value can be derived from the 1000 toys. The average fitted yield from the Gaussian fit is then plotted against the corresponding $N_{\mathrm{MC}}^{\mathrm{signal}}$ value this is being tested. Each of the Gaussian fits that were performed to obtain the average fitted yields is provided in Figure 6.41 below. A linear fit is then performed on these points to obtain the gradient and intercept of the line. An ideal linear fit behaviour shows as a gradient consistent with 1 and an intercept consistent with 0 within their respective errors. A plot of the points and the linear fit is provided in Figure 6.39, with a table of values indicating the range of $N_{\mathrm{MC}}^{\mathrm{signal}}$ values tested, and their corresponding average fitted yield from the 1000 toys, presented in Table 6.30. From the gradient of the fitted line, it can be seen that the behaviour of the fit is essentially linear and therefore well-behaved under different signal yields.

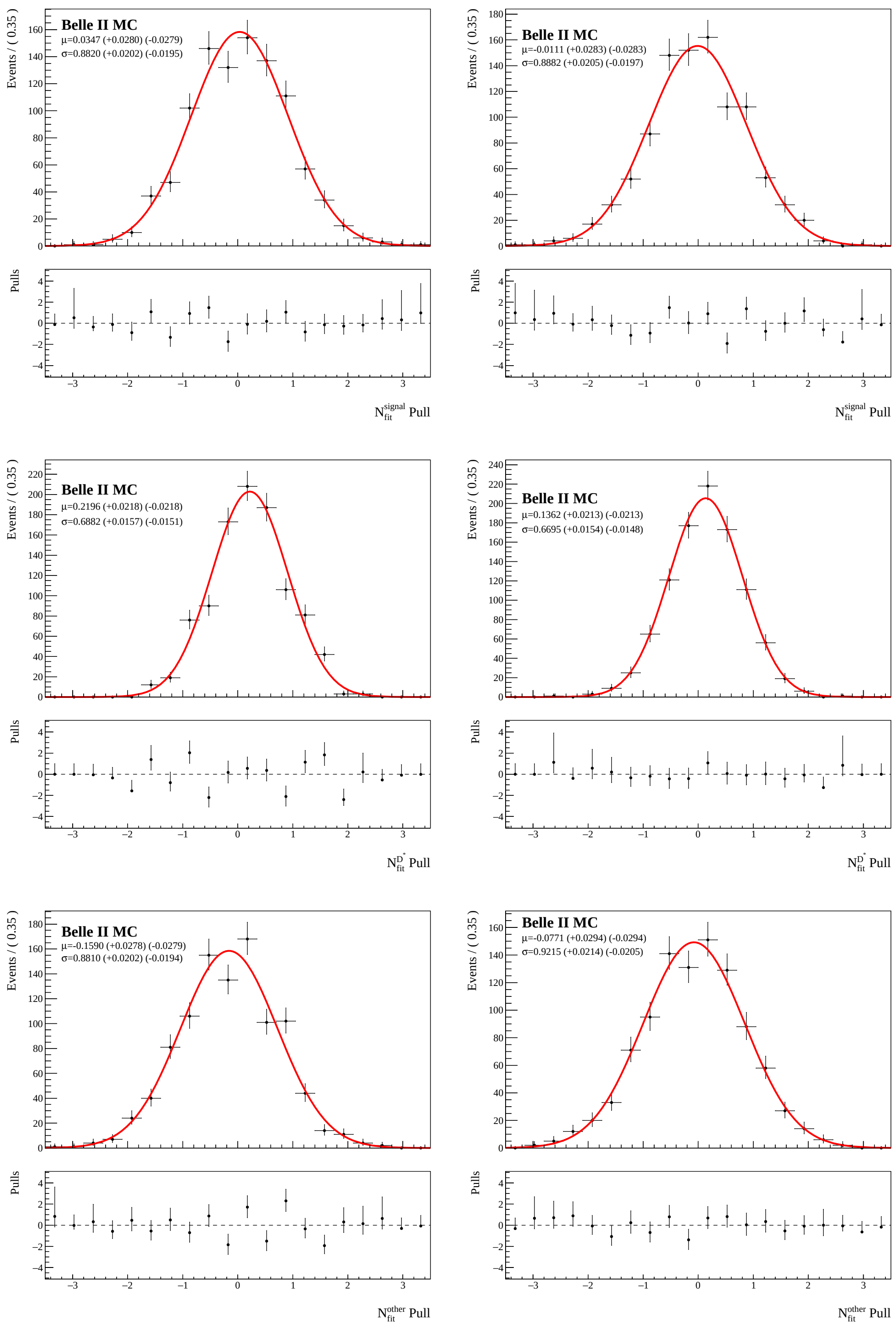


FIGURE 6.37. Pull plot distributions for the fitted yields of the *Signal* component $N_{\text{fit}}^{\text{signal}}$ (top), D^* *background* component $N_{\text{fit}}^{D^*}$ (middle) and *Other background* component $N_{\text{fit}}^{\text{other}}$ (bottom). The pulls were calculated according to the steps outlined in Section 6.6.4. The distributions on the left and right respectively represent the pulls from toys based on the 2.8 (1.0) ab^{-1} and 365 fb^{-1} expected yields.

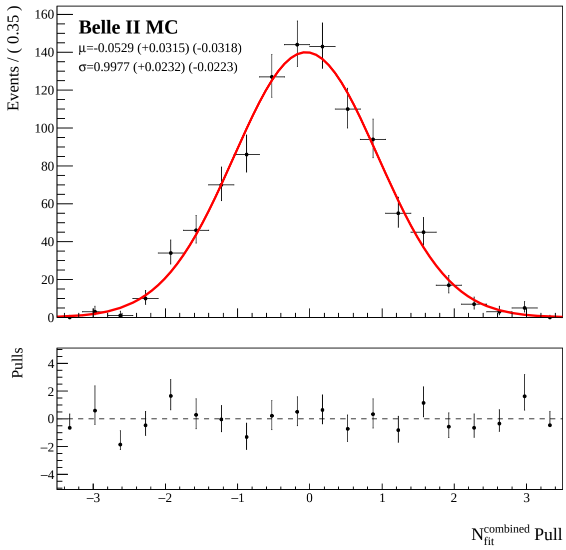


FIGURE 6.38. Pull plot distributions for the combined fitted yields of the *Signal* component, D^* *background* component and *Other background* component. The pulls were calculated according to the steps outlined in Section 6.6.4.

TABLE 6.30. Results of the linearity test for the *Signal* component. The x column indicates the amount by which the $N_{\mathrm{MC}}^{\mathrm{signal}}$ is scaled while the **Expected Yield** equals $x \times N_{\mathrm{MC}}^{\mathrm{signal}}$. The **Fitted Yield** comes from the Gaussian fit to the fitted yields of the 1000 toys as described above.

x	**Expected Yield**	**Fitted Yield**
0.75	176	176.29 ± 0.51
0.80	187	187.16 ± 0.51
0.85	199	199.26 ± 0.53
0.90	211	210.14 ± 0.53
0.95	222	222.23 ± 0.53
1.00	234	233.87 ± 0.53
1.05	246	246.33 ± 0.55
1.10	257	257.12 ± 0.56
1.15	269	269.02 ± 0.58
1.20	281	280.56 ± 0.61
1.25	293	292.99 ± 0.59

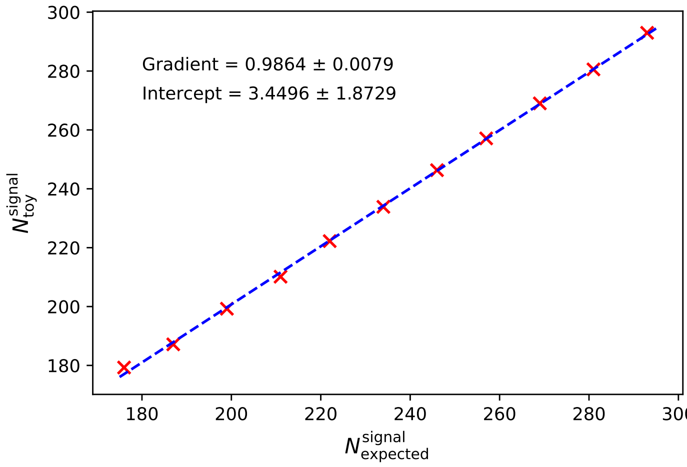


FIGURE 6.39. Result of a linear fit to the fitted yields from the toys ($N_{\text{toy}}^{\text{signal}}$) versus the expected yield ($N_{\text{expected}}^{\text{signal}}$) for the *Signal* component. Table 6.30 provides the values plotted here. The linear fit is represented by the dashed blue line, with the gradient and intercept of the fitted line displayed on the plot.

6.7 Results

The upper limit calculation utilises a hybrid cut-and-count approach that includes a fit to obtain the estimate of the *two true lepton* component. An alternate strategy that used MC template fits to obtain the signal and background yields was tried but resulted in problems with negative PDF values, biases and failed fits. In the future, when the data collected reaches the ab^{-1} scale and the expected number of observed events is more than a few, MC template fits would be the most appropriate method to use for the results extraction. However, given the current situation, the aforementioned hybrid cut-and-count approach is most reliable. The outline of this method is as follows:

(1) Obtain the background counts and background uncertainty of each component:
   (a) For the *two true lepton* component, estimates are obtained using the template derived from the 24 ab^{-1} MC sample. Fits and fitter checks are provided in Section 6.7.1.
   (b) For the *fake lepton* component, the method outlined in Section 6.5.3 is applied to data to obtain an estimated count.
   (c) For the J/ψ*-like* component, the estimated count that remains following the J/ψ and $\psi(2S)$ veto as described in Section 6.5.4 is used.

(2) The estimated signal efficiency and its uncertainty is obtained using MC studies and the control channel study described in Section 6.6.

(3) The number of observed events in the *signal extraction region* is obtained.

(4) The 90% confidence interval from these inputs is calculated using the Rolke-Lopez interval method which is described in Section 4.3.

(5) The expected upper limit on the branching fraction is then calculated according to Equation 6.14 where $n_{\text{sig}}^{\text{high}}$ comes from the $[0, n_{\text{sig}}^{\text{high}}]$ confidence interval, $N_{B\overline{B}}$ is the number of $B\overline{B}$ pairs and f_{00} is the $\Upsilon(4S) \to B^0\overline{B}^0$ branching fraction.

$$\mathcal{B}_{\text{UL}}(B^0 \to \ell^\pm\tau^\mp) = \frac{n_{\text{sig}}^{\text{high}}}{2 \times N_{B\overline{B}} \times f_{00}} \tag{6.14}$$

This method is currently applied to MC in order to obtain the MC expectation of the upper limit, and similar steps will be applied to data during the box opening procedure. The difference for data will be the background count which will be derived from a single fit to the entire *fitting region*, excluding the *signal extraction region,* as the different background components cannot be distinguished and measured separately within data. Currently, the fit is performed across the entire *fitting region*, so prior to the data fit, the stability of fitting to the *fitting region* excluding the *signal extraction region* will need to be verified using MC. The intended box opening procedure is to be done in stages in order to confirm the background estimates prior to opening the box within the *signal extraction region*. An example of the box opening strategy is shown in Figure 6.40 wherein the first stage allows for a preliminary check of the background estimate while the next two stages allows for more robust checks along each dimension of the *fitting region*. The last stage is where the entire *fitting region* including the *signal extraction region* is unboxed and the final measurement from data takes place. A separate box opening procedure for the *fake lepton* background estimate will also take place simultaneously using data sidebands defined by the removal of specific PID cuts.

6.7.1 Background Fit Results

The *two true lepton* component is estimated using a 2D binned extended maximum likelihood fit within the *fitting region*. A Gaussian is used to model the p_ℓ^B component while a second-order Chebychev polynomial is used for E_{extra}. The 2D PDF used for the fit is provided in Equation 6.15 where μ and σ are the mean and width of the Gaussian PDF, a and b are the polynomial constants, and N is the normalisation. The `RooFit` framework is used for the fit, with a Gaussian and Chebychev polynomial used to model each component [98].

$$\text{PDF}_{\text{bkg}}(x, y) = \text{PDF}_{p_\ell^B}(x) \times \text{PDF}_{E_{\text{extra}}}(y) = N \cdot \left[\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \times (ay + by^2 + 1)\right] \tag{6.15}$$

To obtain an estimate, a fit is done to the *two true lepton* template obtained from the 24 ab^{-1} `EvtGen` skim sample. MC corrections, as outlined in Section 6.4, are applied to the template before the fit. Following the fit, the PDF is integrated across the *signal extraction region* to obtain a background estimate. This is then scaled down to obtain an estimate that is in accordance with 365 fb^{-1} and therefore suitable for the upper limit calculation. Projections of the fit to the *two true lepton* template for both $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ modes are provided in Figure 6.42 below.

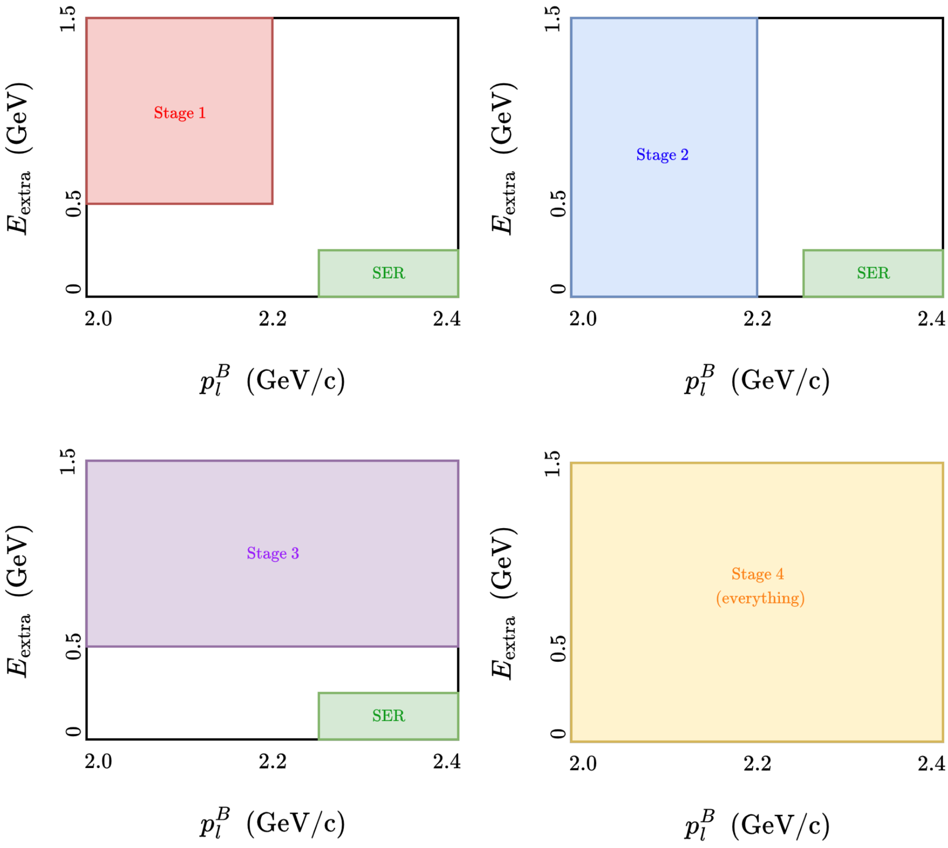


FIGURE 6.40. Proposed stages of the unboxing procedure within the *fitting region* in comparison to the *signal extraction region* which is denoted as *SER*.

In general, reasonable agreement is observed within the *signal extraction* bins for both quantities. Table 6.31 shows the background estimates obtained from the fits. In order to determine the reliability of the *two true lepton* background estimate obtained from the fit, both toy ensemble and linearity tests are performed. Beginning with the toy ensemble test, the procedure used to perform this, for both 24 ab^{-1} and 365 fb^{-1} integrated luminosities, is as follows:

(1) Using the shape parameters from the fit presented in Figure 6.42, a PDF for the *two true lepton* component is defined. These are used to generate the data for each toy.

(2) For each of the 1000 toys in the ensemble test, the following steps are performed:

   (a) A dataset is generated using the fixed PDF from *Step 1*. The size of the dataset is randomly drawn from a Poisson distribution with mean equal to the expected yields for the *two true lepton* component from MC. The 24 ab^{-1} expected yields used were $N_{\text{MC}} = 17836$ and 22412 for the $B^0 \to e^{\pm}\tau^{\mp}$ and $B^0 \to \mu^{\pm}\tau^{\mp}$ modes respectively. The 365 fb^{-1} expected yields were derived by scaling these down.

   (b) A new PDF is defined and fit to the generated dataset from *Step 2a*. The fitted normalisation is extracted to calculate the pull.

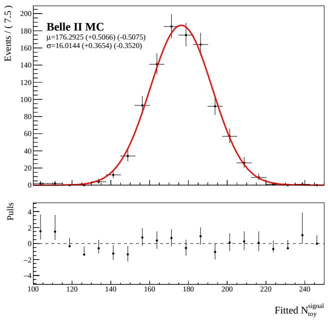
(c) As with the control channel study, the pull was calculated Equation 6.11 if the fitted yield was higher than the expected yields from the original MC sample, or Equation 6.12 if the fitted yield was lower.

(3) When the pull for fitted yields for each of the 1000 toys has been calculated, a Gaussian fit using `RooFit` is performed to the pull distribution. The parameters of the Gaussian fit and their errors are saved in each case in order to determine if there is any bias in the yields.

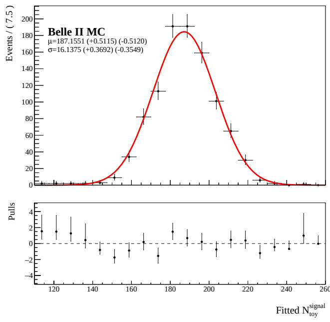
Figure 6.43 shows the fitted pull distributions for each mode and for each integrated luminosity tested. The results show a near-ideal pull, with effectively no bias on the background estimate, and a reasonable spread of fitted yields. Following the toy ensemble test, a linearity test is also performed to check the stability of the fit for different background yields. The range of values tested (N_{toy}) is $[N_{\text{MC}} - 3\sigma, N_{\text{MC}} + 3\sigma]$ in steps of σ, where N_{MC} and σ are the background estimates and their Poisson errors. For the $B^0 \to e^\pm\tau^\mp$ mode this is equal to 269 $\pm$ 16, while for the $B^0 \to \mu^\pm\tau^\mp$ mode this is equal to 338 $\pm$ 18. The linearity test is only performed for the 365 fb^{-1} integrated luminosity case. The same procedure as described in Section 6.6.5 is used. Each of the Gaussian fits that were performed to obtain the average fitted yields is provided in Figure 6.44 and 6.45 for the $B^0 \to e^\pm\tau^\mp$ and $B^0 \to \mu^\pm\tau^\mp$ modes respectively. A plot of the points and the linear fit is provided in Figure 6.46, with tables indicating the values tested provided in Table 6.32 and 6.33 respectively. The gradient and intercept of the fitted lines agrees perfectly with an ideal linear fit, which indicates that the fit is stable for different background yields.

TABLE 6.31. Background estimates for the *two true lepton* component for each mode obtained from fits using the 24 ab^{-1} `EvtGen` skim sample. These are calculated within the *signal extraction* region. The $\int \mathcal{L}dt$ column denotes the corresponding integrated luminosity. Estimates for 365 fb^{-1} are obtained by scaling the numbers for 24 ab^{-1}.

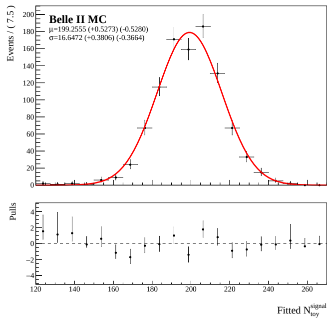
	$\int \mathcal{L}dt$ (ab^{-1})	**Background Estimate**
$B^0 \to e^\pm\tau^\mp$	24	108.79 $\pm$ 4.55
	0.362	1.64 $\pm$ 0.07
$B^0 \to \mu^\pm\tau^\mp$	24	141.71 $\pm$ 5.30
	0.362	2.14 $\pm$ 0.08



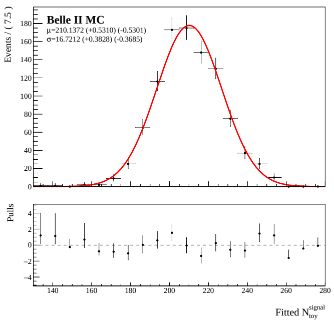
(A) $x = 0.75$



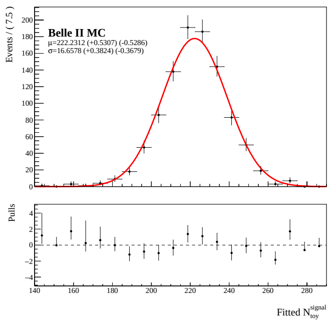
(B) $x = 0.80$



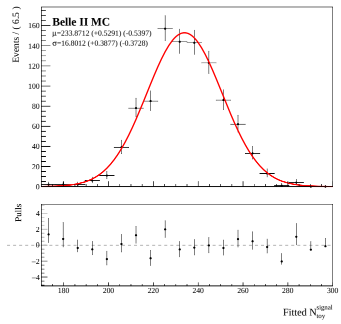
(C) $x = 0.85$



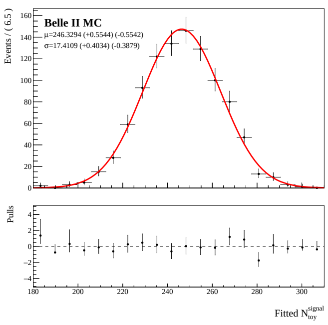
(D) $x = 0.90$



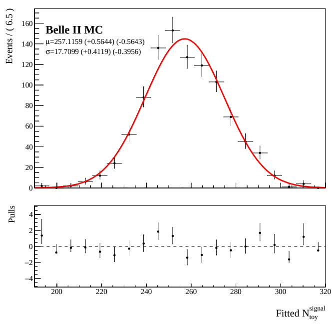
(E) $x = 0.95$



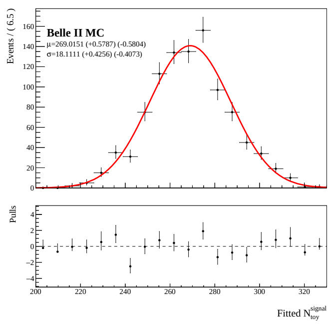
(F) $x = 1.00$



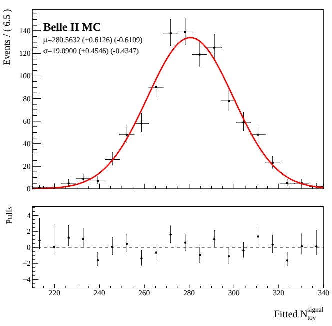
(G) $x = 1.05$



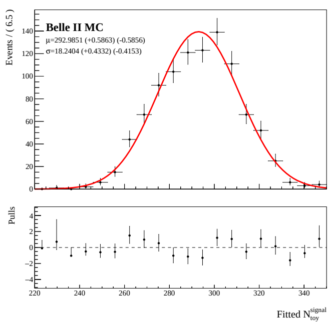
(H) $x = 1.10$



(I) $x = 1.15$



(J) $x = 1.20$



(K) $x = 1.25$

FIGURE 6.41. Distributions of the individual Gaussian fit to each set of toys used in the linearity test. The caption of each sub-figure correspond to the values in Table 6.30 and indicate which expected yield (equal to $x \times N_{\mathrm{MC}}^{\mathrm{signal}}$) is tested.

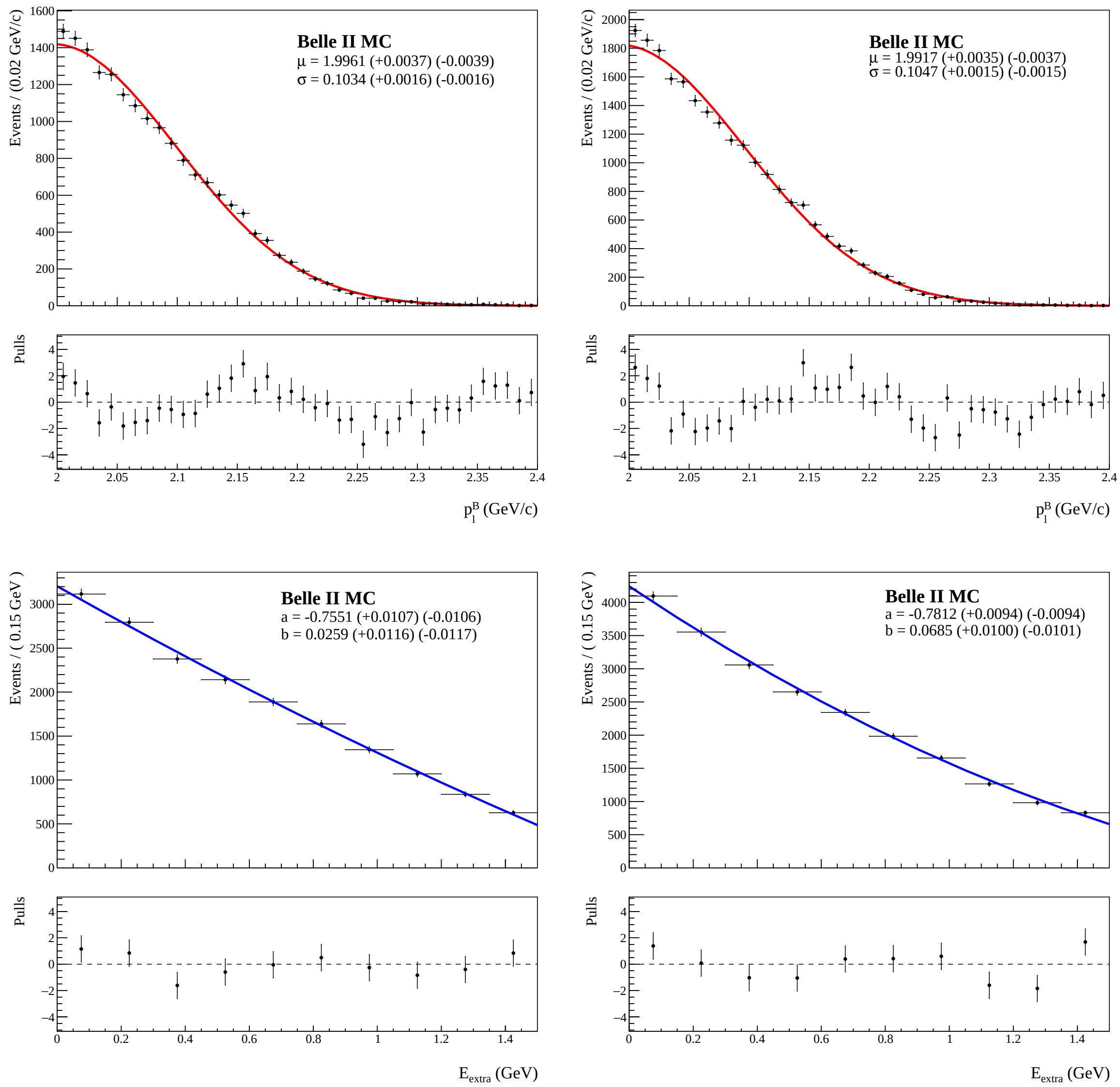


FIGURE 6.42. Fit projections to the *two true lepton* template for the $B^0 \to e^{\pm}\tau^{\mp}$ mode (left) and $B^0 \to \mu^{\pm}\tau^{\mp}$ mode (right). The fitted values of the PDF shape parameters are provided on each sub-figure.

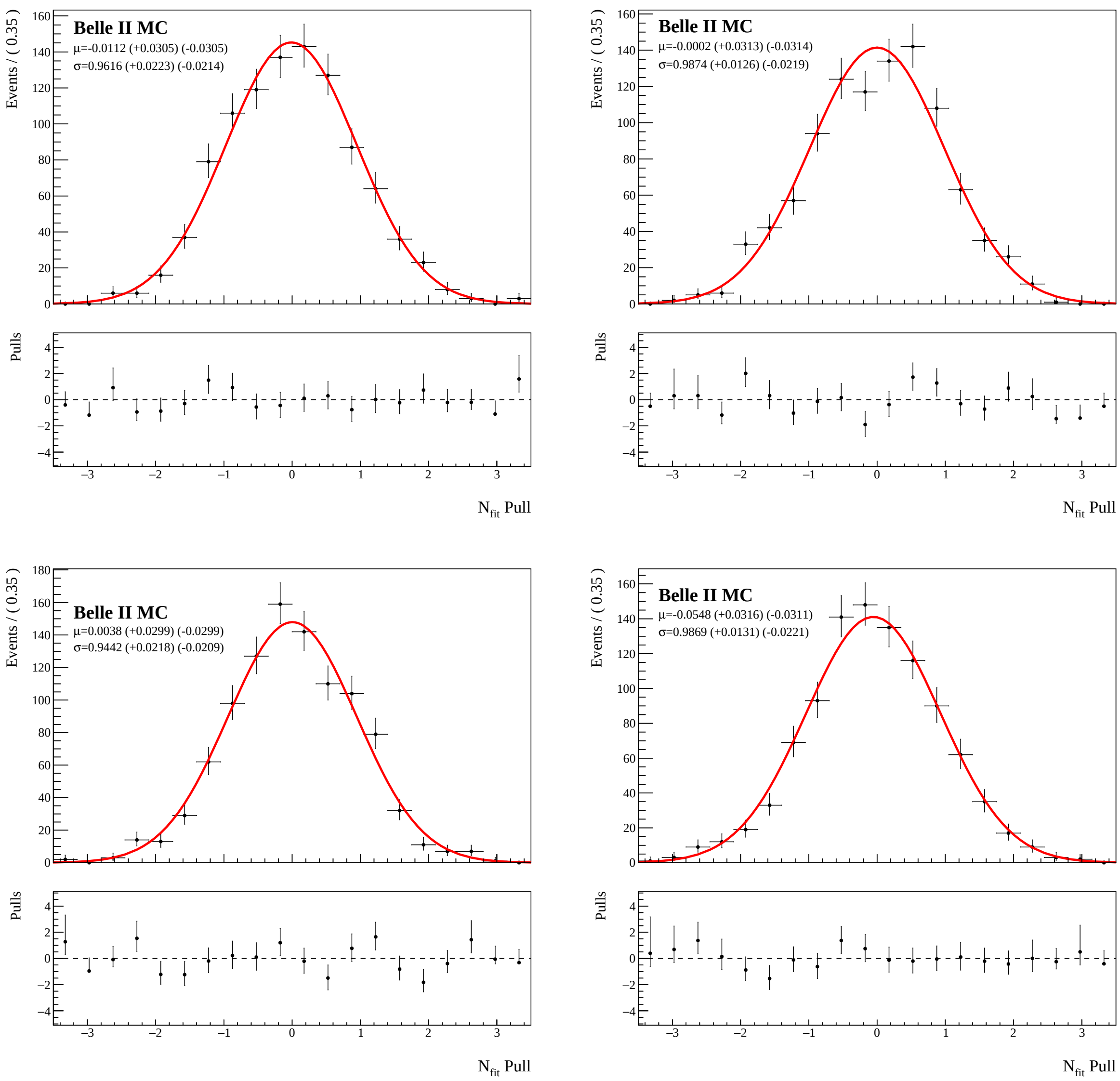


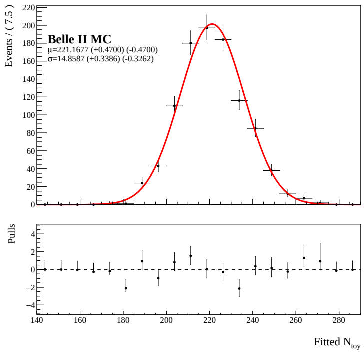
FIGURE 6.43. Pull plots for the *two true lepton* fitted yields of the toy ensemble test described above. The distributions on the left and right respectively represent the pulls from toys based on the 24 ab^{-1} and 365 fb^{-1} expected yields. The $B^0 \to e^{\pm}\tau^{\mp}$ mode is shown in the top plots while the $B^0 \to \mu^{\pm}\tau^{\mp}$ is on the bottom plots. The parameters of the Gaussian fit to the pulls are displayed on each sub-figure.

TABLE 6.32. Results of the linearity test for the *two true lepton* component background estimate for the $B^0 \to e^\pm\tau^\mp$ mode. The $\boldsymbol{x}$ column indicates the amount by which the N_{MC} is changed while the **Expected Yield** equals $N_{\mathrm{MC}} + \boldsymbol{x}\sigma$. Here, σ is the uncertainty on the background estimate from Table 6.31. The **Fitted Yield** comes from the Gaussian fit to the fitted yields of the 1000 toys as described above.

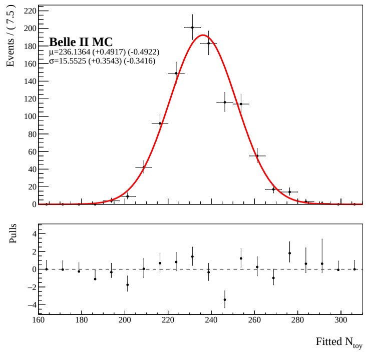
$\boldsymbol{x}$	**Expected Yield**	**Fitted Yield**
-3	221	221.17 ± 0.47
-2	237	236.14 ± 0.49
-1	253	251.87 ± 0.50
0	269	269.08 ± 0.52
+1	285	284.66 ± 0.54
+2	301	301.30 ± 0.51
+3	317	316.19 ± 0.55

TABLE 6.33. Results of the linearity test for the *two true lepton* component background estimate for the $B^0 \to \mu^\pm\tau^\mp$ mode. The $\boldsymbol{x}$ column indicates the amount by which the N_{MC} is changed while the **Expected Yield** equals $N_{\mathrm{MC}} + \boldsymbol{x}\sigma$. Here, σ is the uncertainty on the background estimate from Table 6.31. The **Fitted Yield** comes from the Gaussian fit to the fitted yields of the 1000 toys as described above.

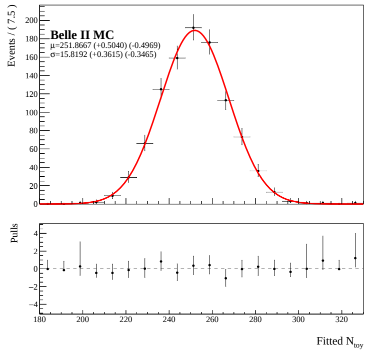
$\boldsymbol{x}$	**Expected Yield**	**Fitted Yield**
-3	284	284.47 ± 0.54
-2	302	300.76 ± 0.55
-1	320	320.15 ± 0.57
0	338	337.91 ± 0.57
+1	356	356.40 ± 0.64
+2	374	373.08 ± 0.60
+3	392	393.35 ± 0.61



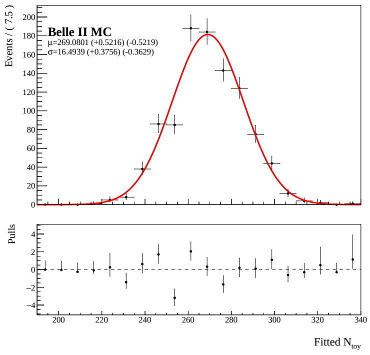
(A) $\boldsymbol{x} = -3$



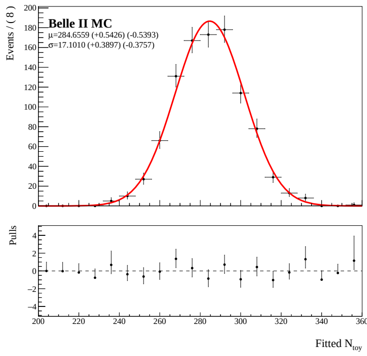
(B) $\boldsymbol{x} = -2$



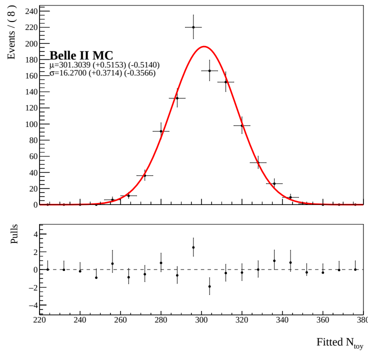
(C) $\boldsymbol{x} = -1$



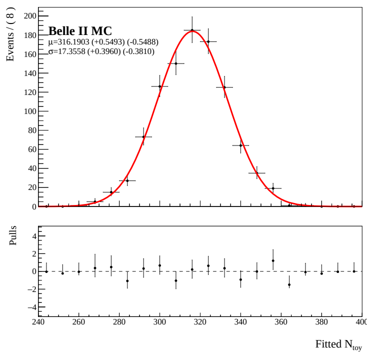
(D) $\boldsymbol{x} = 0$



(E) $\boldsymbol{x} = 1$

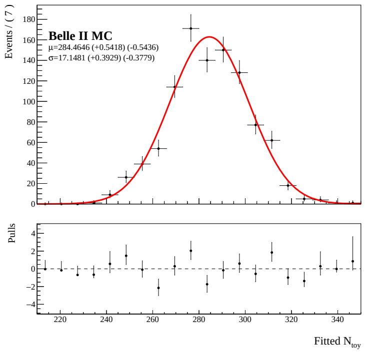


(F) $\boldsymbol{x} = 2$

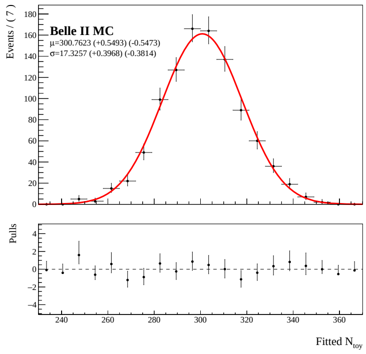


(G) $\boldsymbol{x} = 3$

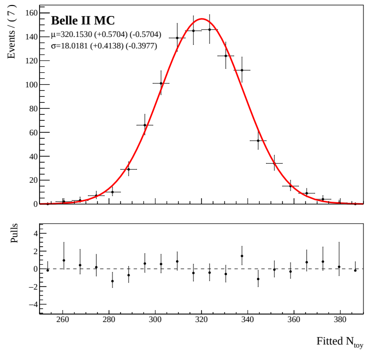
FIGURE 6.44. Distributions of the individual Gaussian fit to each set of toys used in the linearity test for Section 6.7.1 for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. The caption of each sub-figure correspond to the values in Table 6.32 and indicates which expected yield (equal to $N_{\mathrm{MC}} + \boldsymbol{x} \times \sigma$) is tested.



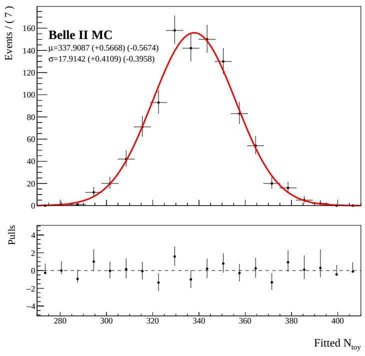
(A) $\boldsymbol{x} = -3$



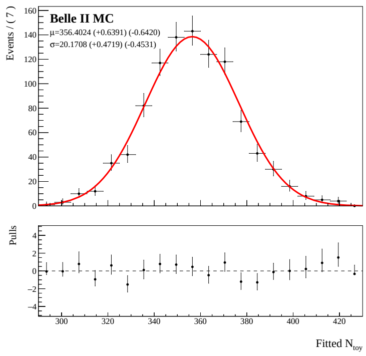
(B) $\boldsymbol{x} = -2$



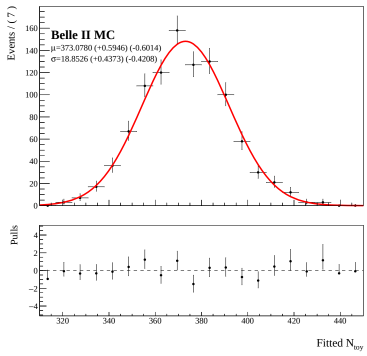
(C) $\boldsymbol{x} = -1$



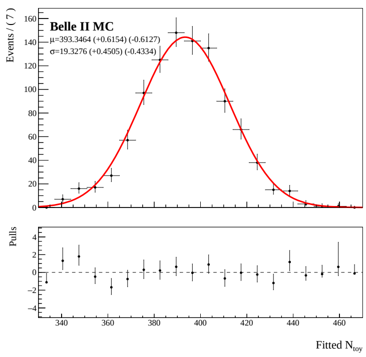
(D) $\boldsymbol{x} = 0$



(E) $\boldsymbol{x} = 1$



(F) $\boldsymbol{x} = 2$



(G) $\boldsymbol{x} = 3$

FIGURE 6.45. Distributions of the individual Gaussian fit to each set of toys used in the linearity test for Section 6.7.1 for the $B^0 \to \mu^\pm\tau^\mp$ mode. The caption of each sub-figure correspond to the values in Table 6.33 and indicates which expected yield (equal to $N_{\mathrm{MC}} + \boldsymbol{x} \times \sigma$) is tested.

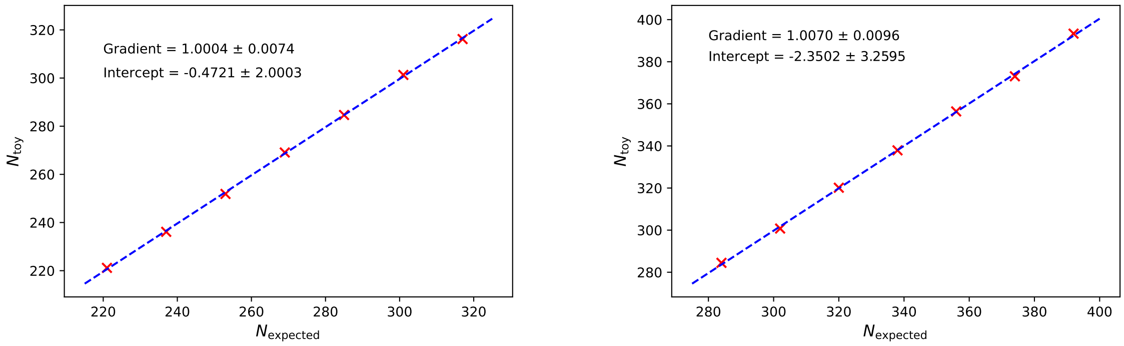


FIGURE 6.46. Results of the linear fit to the fitted yields from the toys (N_{toy}) versus the expected yield (N_{expected}). Tables 6.32 and 6.33 provide the values plotted in the left and right fits respectively. The linear fit is represented by the dashed blue line, with the gradient and intercept of the fitted line displayed on each plot. The fit on the left is for the $B^0 \to e^{\pm}\tau^{\mp}$ mode while the fit on the right is for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode.

A second linearity test is performed, this time in order to test the robustness of the fit in the presence of signal for the 365 fb^{-1} integrated luminosity case. The values for the signal yield that were tested (n_{true}) range from [0, 5.0] events in increments of 0.5 events. For this test, the following steps are taken for each value of n_{true} to be tested:

(1) Using the shape parameters from the fit presented in Figure 6.42, a PDF for the *two true lepton* component is defined. These are used to generate the data for each toy.
(2) For each of the 1000 toys in the ensemble test, the following steps are performed:
   (a) A dataset is generated using the fixed PDF from *Step 1*. The size of the dataset is randomly drawn from a Poisson distribution with mean equal to the 365 fb^{-1} expected yields for the *two true lepton* component from MC.
   (b) A new PDF is defined and fit to the generated dataset from *Step 2a*. The PDF is integrated across the *signal extraction region* to obtain a background estimate.
   (c) The observed number of events is drawn from a Poisson distribution with mean equal to n_{true} plus the background estimate from *Step 2b* and the background estimates of the other two components. The latter two background estimates are themselves drawn from Poisson distributions centred around the expected estimates provided in Table 6.22.
   (d) The signal yield (n_{fitted}) is then extracted by subtracting n_{true} from the observed number of events drawn in *Step 2c*.
(3) The median of the extracted signal yields from the 1000 toys is calculated and plotted against the corresponding value of n_{true} tested.

Figure 6.47 shows a plot of the extracted signal yield vs the true signal yield alongside a linear fit for both modes. They demonstrate a near-perfect linear fit in terms of the gradient, with a slight offset in the intercept. Therefore, it can be observed that the fit is well-behaved for different signal yields.

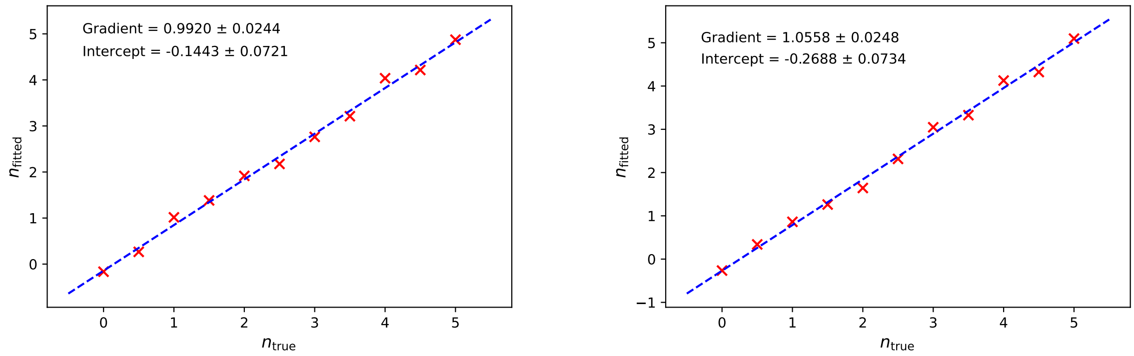


FIGURE 6.47. Results of the linear fit to the extracted signal yields (n_{fitted}) versus the injected signal yield (n_{true}). The linear fit is represented by the dashed blue line, with the gradient and intercept of the fitted line displayed on each plot. The fit on the left is for the $B^0 \to e^\pm \tau^\mp$ mode while the fit on the right is for the $B^0 \to \mu^\pm \tau^\mp$ mode.

6.7.2 Coverage Test

Prior to calculating the expected upper limits, a coverage test is performed in order to determine that the upper limit method provides a coverage consistent with a 90% confidence level. The test is performed for values of the signal yield (s) ranging from [0, 5] events in increments of 0.5 events. In order to make this test consistent with the box opening procedure, the background estimate in each case is derived from a fit to all background components within the *fitting region*. The PDF provided in Equation 6.15 is also used in this case, with projections of the fit provided in Figure 6.48. For each value of s being tested, the following method is performed:

(1) A PDF of fixed shape is defined using the parameters from the fit displayed in Figure 6.48. This is used to generate the data for each toy.

(2) For each of the 1000 toys in the ensemble test, the following steps are performed:
   - (a) A dataset is generated using the fixed PDF from *Step 1*. The size of the dataset equals the total background within the *fitting region* for 365 fb^{-1} which given in Table 6.21.
   - (b) A new PDF is defined and fit to the generated dataset from *Step 2a*. The PDF is integrated across the *signal extraction region* to obtain a background estimate and its corresponding error.
   - (c) The observed number of events is defined using the number of signal events, drawn from a Poisson of mean equal to s, plus the number of background events, drawn from a Poisson with mean equal to the MC estimate of the total background given in Table 6.22.
   - (d) The confidence interval is calculated, and it is checked to see if the value of s lies within the interval. If so, it is considered as covered.

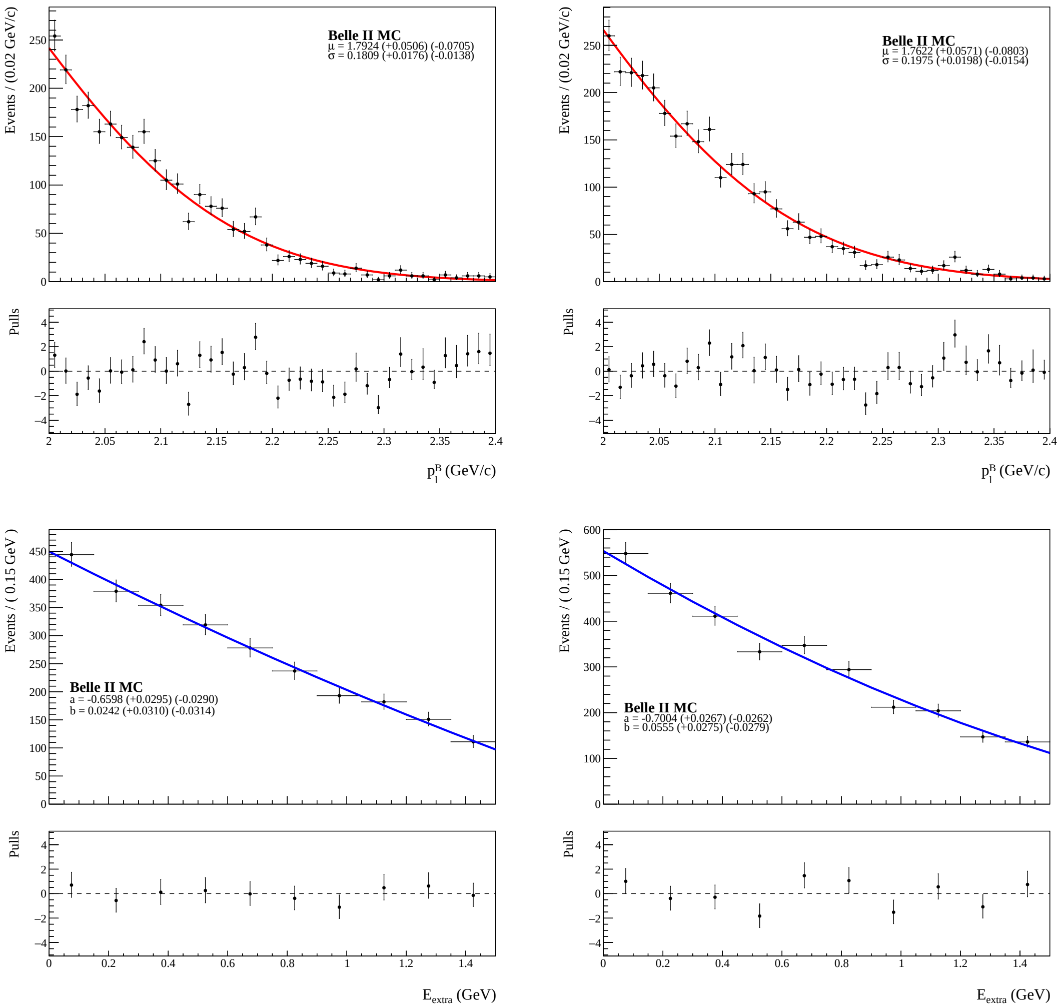


FIGURE 6.48. Fit projections to the all background components for the $B^0 \to e^{\pm}\tau^{\mp}$ mode (left) and $B^0 \to \mu^{\pm}\tau^{\mp}$ mode (right). The fitted values of the PDF shape parameters are provided on each sub-figure. This additional fit was conducted as part of the coverage test.

For all cases, the intention was to set the signal efficiency uncertainty to zero as the efficiency thus far is derived from MC samples. However, a value of zero caused issues within the `TRolke` class that is used to calculate the upper limit, so the signal efficiency uncertainty was set to 10^{-5} instead. Results for the coverage are presented in Table 6.34. For the $B^0 \to e^{\pm}\tau^{\mp}$ mode, there is slight under-coverage observed for most of the signal yields tested. This is a minor issue that can be accounted for by adjusting the confidence level from 90% to a value above 90% such that ultimately the coverage test shows the desired coverage. However, for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode, a clear issue with the coverage is observed. This stems from the unaccounted for peak in the *signal extraction region* which is now present when the *fake lepton* background component is included in the total background fit. This is demonstrated by Figure

6.49 which compares the shapes of the two main background components within the *fitting region*. From the fit in Figure 6.48, it is observed that the measured background from the fit is underestimated as the PDF does not account for the *fake lepton* peak. This resulted in shifted confidence intervals that no longer covered the true signal yield. Such an issue is not present for the $B^0 \to e^{\pm}\tau^{\mp}$ mode as the *fake lepton* background is a sub-leading component.

Further work is needed to address this peak, and this is detailed in Section 7.1. Ultimately, such a scenario is manageable by using the background estimation and partial unboxing strategies. As the *fake lepton* component is to be estimated from data during the partial unboxing stage, the underestimation of this component from the total background fit can be verified, especially as robust methods for the estimation of the other background components also exist.

TABLE 6.34. Percentage of upper limits that are covered for the $B^0 \to e^{\pm}\tau^{\mp}$ and $B^0 \to \mu^{\pm}\tau^{\mp}$ modes. The signal yield being tested in the coverage test is denoted by $\boldsymbol{s}$.

$\boldsymbol{s}$	**Coverage (%)**	
	$B^0 \to e^{\pm}\tau^{\mp}$ Mode	$B^0 \to \mu^{\pm}\tau^{\mp}$ Mode
0.0	91.2	15.5
0.5	88.6	20.9
1.0	89.4	24.2
1.5	87.5	20.7
2.0	86.7	29.9
2.5	87.4	29.3
3.0	89.5	31.7
3.5	90.0	32.3
4.0	87.6	35.6
4.5	88.8	37.4
5.0	88.4	40.2

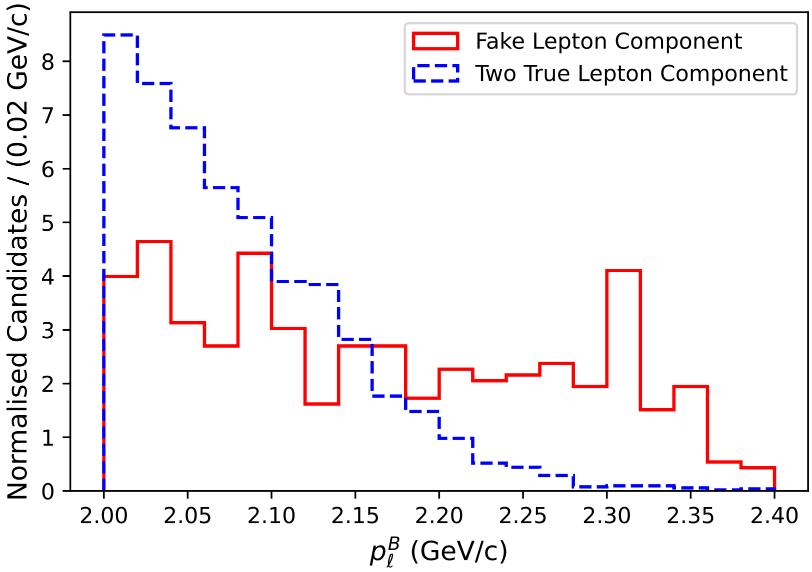


FIGURE 6.49. Comparison between the *two true lepton* and *fake lepton* p_ℓ^B shapes for background within the *fitting region*.

6.7.3 Expected Upper Limits

The expected upper limits for this analysis in its current version are provided in Table 6.35 for both modes. The table also includes the background estimates, their uncertainties, and the signal efficiencies as all of this information is used to calculate the upper limit. Until the control channel study is finalised, the uncertainty on the signal efficiency is assumed to be 5% which was the same as the Belle study [16].

TABLE 6.35. Expected branching fraction upper limits for the current version of the analysis. The background counts are scaled to 365 fb^{-1} and, along with the signal efficiency, are calculated within the *signal extraction region*.

	$B^0 \to e^\pm\tau^\mp$ **Mode**	$B^0 \to \mu^\pm\tau^\mp$ **Mode**
Total Background	2.9 ± 1.1	9.9 ± 2.8
Two True Lepton	1.6 ± 0.1	2.1 ± 0.1
Fake Lepton	0.4 ± 0.6	7.7 ± 2.8
J/ψ-Like	0.9 ± 1.0	0.1 ± 0.3
Efficiency (%)	0.0576	0.0591
Upper Limit ($\times 10^{-5}$)	1.8	2.5

The expected upper limit is close to the world best limit for the $B^0 \to e^\pm\tau^\mp$ mode, and is within a factor of 1.6 for the $B^0 \to \mu^\pm\tau^\mp$ mode. The reason for the larger upper limits is more so due to the lack of data currently available rather than the analysis setup itself. To demonstrate this point, a comprehensive numerical comparison between the past analyses and this analysis is provided in Table 6.36. An explanation of the terms is provided below:

- In order to compare data sizes, the number of B meson pairs $N_{\mathrm{B\overline{B}}}$ is quoted for all the B-factories. The number of B mesons N_B is the equivalent number for LHCb, however this value is not quoted in their paper. As the product of their signal efficiency ϵ and N_B can be derived from their normalisation constant α, it is the product $N_B \times \epsilon$ that is quantifiable for the LHCb analysis. Nevertheless, the N_B quantity is still calculated for the B-factories in order to understand how the value of $N_B \times \epsilon$ is derived for them.
- The effective background count $N_{\mathrm{bkg}}^{\mathrm{eff}}$ is quoted directly in the CLEO paper, while for the LHCb and Belle analyses, this number is derived from the uncertainties on their fitted signal yields. For the BaBar analysis, although the direct number is quoted in their paper as well, the background is instead calculated according to the number of data points within the region of the signal PDF following their fit as it is these background events that contribute to their sensitivity.
- The upper limit on the number of signal events N^{UL} is defined according to the effective background count. For high background analyses that operate in the Gaussian regime such as CLEO, LHCb and Belle, $N^{\mathrm{UL}} = 1.64 \times \sqrt{N_{\mathrm{bkg}}^{\mathrm{eff}}}$ where 1.64 corresponds to the 90% confidence interval for a Gaussian distribution. In the cases of BaBar and our analyses where the measurement is Poisson due to the low background count, $N^{\mathrm{UL}} = 1.64 \times f \times \sqrt{N_{\mathrm{bkg}}^{\mathrm{eff}}}$ where f is a factor that accounts for the difference between the 90% confidence intervals for a Gaussian and Poisson distribution.
- The estimated upper limit is then derived according to Equation 6.16.

$$\mathcal{B}^{\mathrm{UL}} = \frac{N^{\mathrm{UL}}}{N_B \times \epsilon} \tag{6.16}$$

Starting with the $B^0 \to e^\pm\tau^\mp$ mode, where the Belle study provides the world-leading upper limit, it is clear that our analysis approach yields a competitive upper limit with only half the data. Should this analysis also have access to 772×10^6 $B\overline{B}$ pairs, then the estimated upper limit for $B^0 \to e^\pm\tau^\mp$ would be $\approx 1.2 \times 10^{-5}$. Therefore, although the Belle analysis has the advantage of a higher efficiency due to its inclusive τ reconstruction, its sensitivity is impacted by the significant increase in background. Furthermore, its high background count prevents the Belle analysis approach from being scalable with higher integrated luminosities. A similar conclusion can be made for the $B^0 \to \mu^\pm\tau^\mp$ mode wherein our analysis approach yields an upper limit in the same order of magnitude as the world-leading limit from the LHCb study but with significantly less data. As with the Belle study, this is due to the large background included in their final fit. To further demonstrate the effectiveness of our approach with higher integrated luminosities, estimated upper limits assuming 3 ab^{-1} of data can be calculated by naively scaling the background estimates accordingly. For the $B^0 \to e^\pm\tau^\mp$ mode, the total background becomes 24.0 ± 9.1 which results in an estimated upper limit of 9.3×10^{-6} while for the $B^0 \to \mu^\pm\tau^\mp$

mode, the total background becomes 82.0 ± 23.2 leading to an estimated upper limit of 1.6×10^{-5}. The $B^0 \to e^{\pm}\tau^{\mp}$ upper limit would be world-leading, and although the background has increased to the same order of magnitude as the Belle approach, this is after considering four times as much data. While the $B^0 \to \mu^{\pm}\tau^{\mp}$ is still behind the current LHCb limit, it is important to note that with such high backgrounds, using the hybrid cut-and-count approach is sub-optimal, and a more appropriate estimate could be built by extracting the signal yield from a fit. This is discussed in Section 7.1.

TABLE 6.36. Numerical comparison of the sensitivities of all $B^0 \to \ell^{\pm}\tau^{\mp}$ searches. The number of $B\overline{B}$ and B mesons are given by $\boldsymbol{N_{B\overline{B}}}$ and $\boldsymbol{N_B}$ respectively. The signal efficiency is denoted by $\boldsymbol{\epsilon}$ while $\boldsymbol{N^{\text{eff}}_{\text{bkg}}}$ is the effective background estimate. The upper limit on the signal yield using a 90% confidence interval is given by $\boldsymbol{N^{\text{UL}}}$. For the BaBar and our analyses, an additional factor $\boldsymbol{f}$ is used in the calculation of $\boldsymbol{N^{\text{UL}}}$ in order to account for difference between the 90% confidence interval for the Gaussian and Poisson measurements. The estimated upper limit using $\boldsymbol{N_B} \times \boldsymbol{\epsilon}$ and $\boldsymbol{N^{\text{UL}}}$ is given by **Estimated UL** whereas **UL** is the published upper limits for each study. For LHCb, **UL** corresponds to their expected upper limits, while for CLEO, LHCb and Belle, the **UL** is their measured upper limit. The LHCb analysis does not state their $\boldsymbol{N_B}$ or $\boldsymbol{\epsilon}$ values which is why only their product is provided.

	CLEO [60]	**BaBar [61]**	**LHCb [15]**	**Belle [16]**	**Ours**
$\boldsymbol{N_{B\overline{B}}}$ $(\times\mathbf{10^6})$	9.6	378	–	772	387
$\boldsymbol{N_B}$ $(\times\mathbf{10^6})$	9	367	–	750	376
			$B^0 \to e^{\pm}\tau^{\mp}$		
$\boldsymbol{\epsilon}$ $(\times\mathbf{10^{-4}})$	154	3.2	–	9.8	5.76
$\boldsymbol{N_B} \times \boldsymbol{\epsilon}$ $(\times\mathbf{10^6})$	0.139	0.117	–	0.735	0.217
$\boldsymbol{N^{\text{eff}}_{\text{bkg}}}$	20	4	–	73	3
$\boldsymbol{N^{\text{UL}}}$ $(\boldsymbol{f})$	7.334	4.592 (1.40)	–	14.012	4.347 (1.56)
Estimated UL $(\times\mathbf{10^{-5}})$	5.276	3.925	–	1.906	2.003
UL $(\times\mathbf{10^{-5}})$	11	2.84	–	1.6	1.8
			$B^0 \to \mu^{\pm}\tau^{\mp}$		
$\boldsymbol{\epsilon}$ $(\times\mathbf{10^{-4}})$	220	2.7	–	11.0	5.91
$\boldsymbol{N_B} \times \boldsymbol{\epsilon}$ $(\times\mathbf{10^6})$	0.198	0.099	8.000	0.825	0.222
$\boldsymbol{N^{\text{eff}}_{\text{bkg}}}$	19	5	3364	62	10
$\boldsymbol{N^{\text{UL}}}$ $(\boldsymbol{f})$	7.149	4.987 (1.36)	95	12.913	6.481 (1.22)
Estimated UL $(\times\mathbf{10^{-5}})$	3.611	5.037	1.2	1.565	2.919
UL $(\times\mathbf{10^{-5}})$	3.8	2.20	1.6	1.5	2.5

Chapter 7

Conclusion

This thesis has presented the current status of the first Belle II search for the lepton flavour violating decay $B^0 \to \ell^\pm \tau^\mp$ ($\ell = e, \mu$).

Chapter 2 introduced lepton flavour violation within the Standard Model and explained how lepton flavour violating decays such as $B^0 \to \ell^\pm \tau^\mp$ are allowed by the Standard Model theory only in the presence of neutrino oscillations. The Standard Model prediction for the $B^0 \to \ell^\pm \tau^\mp$ branching fraction is far below any current or future experimental sensitivity. Therefore, should the $B^0 \to \ell^\pm \tau^\mp$ decay be observed, it would be clear evidence of new physics contributions leading to an enhanced branching fraction. Searches for $B^0 \to \ell^\pm \tau^\mp$ have been conducted before by the CLEO, BaBar, LHCb and Belle experiments, with all searches described in the chapter in order to set the experimental context for the study presented here.

Chapter 3 introduced the asymmetric e^+e^- SuperKEKB collider and how it was designed to mass produce pairs of B mesons. A significant portion of the chapter is dedicated to describing the Belle II experiment and the role that each of its sub-detectors play in the detection of particles produced from a collision at the interaction point. The chapter concludes by providing an overview of the methods by which decays such as $B^0 \to \ell^\pm \tau^\mp$ can be reconstructed using the sub-detector information.

The role of Chapter 4 was to discuss the various concepts that are pertinent to the work presented in subsequent chapters. Particular emphasis is given to multivariate analysis and machine learning concepts in the first half of the chapter in order to support the E_{extra} cleanup classifier study presented in Chapter 5. Within Chapter 5, the origin of the E_{extra} quantity is discussed in depth. With the performance degradation of E_{extra} as a key signal-background separation tool arising from beam background and fake photon clusters in the calorimeter, the need to develop classifiers that are trained to identify these contributions E_{extra} becomes strongly motivated. The training and testing of such classifiers is shown in depth within the chapter.

The latter half of Chapter 4 is dedicated to explaining the maximum likelihood fits and upper limit calculations used for the $B^0 \to \ell^\pm \tau^\mp$ search presented in Chapter 6. This chapter begins by comparing the analysis strategy used in this study with past searches. Like previous B-factory searches, this study utilises a hadronic tag approach in order to completely recover the B_{sig} kinematics. However, unlike past searches, this study partitions the leptonic τ decays into leptonic and hadronic modes, and designs specific signal extraction methods for each in order to maximise the sensitivity. For the current version of the analysis presented here, only the *leptonic partition* is considered. For these decays, three primary background components were identified and investigated. The first component is the *two true lepton* component which was estimated using specially-produced 24 ab^{-1} MC samples. Such high statistics MC generation was possible due to development of an `EvtGen` skim. The other background components consisted of the *fake lepton* component which is estimated from data. A closure test was performed which demonstrated good agreement between observed and expected background rates. The remaining background component is the J/ψ*-like decays* which were targeted using vetoes for the J/ψ and ψ(2S). An ongoing verification of the analysis is performed in the form of data-MC agreement checks, a control channel study using $B^0 \to D^-\pi^+$, and toy ensemble and linearity checks for all fits.

Estimations from MC of the branching fraction upper limit were obtained using a hybrid cut-and-count approach based on the Rolke-Lopez confidence interval method. The estimated limits are $\mathcal{B}(B^0 \to e^\pm\tau^\mp) < 1.8 \times 10^{-5}$ and $\mathcal{B}(B^0 \to \mu^\pm\tau^\mp) < 2.5 \times 10^{-5}$. Although they do not beat the world-leading upper limits, the results still demonstrate the improved sensitivity of the method. A detailed numerical comparison between all searches presented at the end of the chapter shows that for the $B^0 \to e^\pm\tau^\mp$ mode, our sensitivity is close to the Belle sensitivity despite using half the amount of data acquired by the Belle experiment. As for the $B^0 \to \mu^\pm\tau^\mp$ mode, our sensitivity is within a factor of 1.6 of the sensitivity for the LHCb study even though they have a significantly higher B meson production cross-section. Most of this improvement is due to the partitioning between the leptonic and hadronic τ decays. Although reconstructing the τ results in a decreased signal efficiency, the information gained on the τ decay through reconstruction allows for low background leptonic τ decay to be separated from the high background hadronic τ decay sample. The low background regime is advantageous in that it allows for loose selection cuts to be applied, thereby mitigating any further decreases in the signal efficiency, and it also permits the complete categorisation of all background contributions, which then allows for accurate background estimations. Furthermore, it is the presence of large backgrounds that prevent the LHCb and Belle analyses from having sensitivities that scale well with increased integrated luminosities as compared to the approach presented here. Due to this, the study presented here provides a suitable approach for future $B^0 \to \ell^\pm\tau^\mp$ searches at Belle II once more data has been accumulated.

7.1 Future Work

Currently, only expected upper limits based on MC are presented as the analysis requires further work for it to be ready use data to calculate the upper limits. The points of further work that are needed are outlined below in no particular order:

(1) Ultimately, the *fake lepton* background component will be estimated from data, and so far only a closure test has been performed which confirms the validity of the estimation method. Performing an estimation check on a data sideband would be preferred prior to unboxing to act

as an additional check. This is particularly important for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode for which the leading background is the *fake lepton* component.

(2) The systematic uncertainties for both the control channel and the $B^0 \to \ell^{\pm}\tau^{\mp}$ analysis remain to be calculated. The intention of the control channel is to have the corresponding systematic uncertainties largely cover the signal efficiency uncertainty for the $B^0 \to \ell^{\pm}\tau^{\mp}$ decay. The uncertainties which are covered by the control channel include the hadronic FEI calibration factors, lepton ID corrections and mid-to-high momentum tracking efficiency. Separate cross-checks are planned to verify the systematic uncertainty measured using the control channel. In addition to this, uncertainties associated with the shape of the *two true lepton* component PDF, the estimation of the *fake lepton* component, the number of $B\overline{B}$ pairs and the f_{00} quantity will also need to be included for the $B^0 \to \ell^{\pm}\tau^{\mp}$ decay.

(3) The pull plots for the fitted yields in the control channel study displayed smaller than expected widths for the *Signal* and D^* *Background* components. The pull plot for the combined yield was well-behaved indicating that the irregular pull plots stem from the correlations that exist between the *Signal* and D^* *Background* components. This requires further investigation, either in the way of modifying the PDF to account for such correlations, or by taking the correlations into account within the fitted yield uncertainties.

(4) The 24 ab^{-1} MC sample produced using the `EvtGen` skim looks well-behaved in terms of its shape within the *fitting region*. Whilst using the `EvtGen` skim does not lead to any sculpting of the background in the relevant regions, its normalisation exceeds what is expected from a naive scaling of the centrally-produced 2.8 ab^{-1} sample. The impact of this would be an overestimation in the *two true lepton* component. The causes of this are unknown and are currently being investigated. If the cause cannot be ascertained, some correction factor to this background component may be implemented based on checks in various data sidebands.

(5) Currently, all MC samples used are run-independent. Ideally, run-dependent MC samples would be used instead. Therefore, the next iteration of this analysis would aim to solely use run-dependent MC. This would include the `EvtGen` skimmed MC samples utilised for the *two true lepton* background estimation. On this point, it was mentioned that the upper cut on p_{CM} was, in hindsight, an incorrect cut to add to the skim as it shortened the tail around the *signal enhanced region* and could lead to an underestimation in the *two true lepton* background estimation. Therefore, when the relevant MC samples are regenerated to become run-dependent, this upper cut on p_{CM} will be removed.

(6) As this analysis approach was designed to be used for future studies when Belle II has collected multiple ab^{-1} of data, ideally an alternate signal extraction strategy for that regime would be explored within the setting of this current study. The hybrid cut-and-count approach employed here would not be suitable given the increased background levels, and instead an extraction

method based on fits using either a signal template from MC or analytical functions to describe the signal shape would be employed.

(7) It was shown that there are issues with the upper limit coverage for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode. This is attributed to the peak from the *fake lepton* component that is present when a total background fit is performed to the *fitting region*. To fix this issue, a modification to the PDF used for the total background fit is required, one that includes a component to describe the *fake lepton* peak. This should be conducted alongside a detailed look into the specific *fake lepton* background events that are causing this peak, to see if a veto can be implemented as opposed to changing the PDF. Additionally, as some under-coverage for the $B^0 \to e^{\pm}\tau^{\mp}$ mode was also seen, the coverage test will need to be performed again to determine what adjusted confidence level would yield a $\approx 90\%$ coverage.

Following completion of the above points, and any further lines of work that stem from such tasks, the analysis would then be ready for unboxing following an internal review of the analysis by our physics working group. An initial plan for the unboxing procedure involves unblinding the *fitting region* in stages so as to perform an early check on the background estimations before the *signal enhanced region* is unblinded. Once unboxing has been performed, the goal is to produce a publication based on this work.

A larger extension to the analysis involves including the *hadronic partition* in the upper limit calculation. So far, limited work using the *hadronic partition* has been performed and consists mainly of selection cut, best candidate selection studies and data-MC agreement checks. Therefore, the categorisation of the background components and their estimation methods, and a definition of the *fitting* and *signal enhanced regions* remains to be done. Furthermore, as the *hadronic partition* would experience more background, an entirely fitting-based strategy for signal extraction should be explored even with the current amount of data available. Although most of the sensitivity from the analysis approach presented here stems from the low-background *leptonic partition*, the *hadronic partition* would nonetheless add a small but non-zero amount of sensitivity making it an important long term plan for this study.

Bibliography

[1] N. Braun et al. 'The Belle II Core Software'. In: *Computing and Software for Big Science* 3 (Nov. 2018). DOI: 10.1007/s41781-018-0017-9.

[2] W. Sutcliffe et al. *WG1Template*. Privately developed python package. URL: https://github.com/MarkusPrim/WG1Template.

[3] W. Sutcliffe et al. *PIDvar*. Privately developed python package. URL: https://gitlab.desy.de/meliache/pidvar.

[4] D. Griffiths. *Introduction to Elementary Particles*. New York, USA: John Wiley Sons, 1987.

[5] T. Kobayashi. 'Experimental verification of the standard model of particle physics'. In: *Proceedings of the Japan Academy, Series B* 97.5 (2021), pp. 211–235. DOI: 10.2183/pjab.97.013.

[6] W. Altmannshofer et al. 'The Belle II Physics Book'. In: *PTEP* 2019.12 (2019). Ed. by E. Kou and P. Urquijo. [Erratum: PTEP 2020, 029201 (2020)], p. 123C01. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex].

[7] G. Aad et al. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[8] S. Chatrchyan et al. 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC'. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[9] M. López-Corredoira. 'Tests and Problems of the Standard Model in Cosmology'. In: *Foundations of Physics* 47.6 (Feb. 2017), pp. 711–768. DOI: 10.1007/s10701-017-0073-8.

[10] J. Ellis. *Limits of the Standard Model*. 2002. arXiv: hep-ph/0211168 [hep-ph].

[11] B. Abi et al. 'Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm'. In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 141801. DOI: 10.1103/PhysRevLett.126.141801.

[12] T. Aoyama et al. 'The anomalous magnetic moment of the muon in the Standard Model'. In: *Physics Reports* 887 (Dec. 2020), pp. 1–166. DOI: 10.1016/j.physrep.2020.07.006.

[13] A. D'Alise et al. 'Standard model anomalies: lepton flavour non-universality, g - 2 and W-mass'. In: *Journal of High Energy Physics* 2022.8 (Aug. 2022). DOI: 10.1007/jhep08(2022)125.

[14] Ed. A.J. Bevan, B. Golob, Th. Mannel, S. Prell, and B.D. Yabsley. 'The Physics of the B Factories'. In: *Eur. Phys. J.* (2014). C74 3026, SLAC-PUB-15968, KEK Preprint 2014-3.

[15] R. Aaij et al. 'Search for the lepton-flavour-violating decays $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$'. In: *Phys. Rev. Lett.* 123 (2019), p. 211801. DOI: 10.1103/PhysRevLett.123.211801.

[16] H. Atmacan et al. 'Search for $B^0 \to \tau^\pm\ell^\mp$ ($\ell = e, \mu$) with a hadronic tagging method at Belle'. In: *Phys. Rev. D* 104.9 (2021), p. L091105. DOI: 10.1103/PhysRevD.104.L091105. arXiv: 2108.11649 [hep-ex].

[17] G. Zweig. 'Memories Of Murray And The Quark Model'. In: *International Journal of Modern Physics A* 25.20 (Aug. 2010), pp. 3863–3877. DOI: 10.1142/s0217751x10050494.

[18] M. Gell-Mann. 'A schematic model of baryons and mesons'. In: *Physics Letters* 8.3 (1964), pp. 214–215. ISSN: 0031-9163. DOI: https://doi.org/10.1016/S0031-9163(64)92001-3.

[19] Y. Ne'eman. 'Derivation of strong interactions from a gauge invariance'. In: *Nuclear Physics* 26.2 (1961), pp. 222–229. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90134-1.

[20] E.D. Bloom et al. 'High-Energy Inelastic $e - p$ Scattering at 6° and 10°'. In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 930–934. DOI: 10.1103/PhysRevLett.23.930.

[21] M. Breidenbach et al. 'Observed Behavior of Highly Inelastic Electron-Proton Scattering'. In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 935–939. DOI: 10.1103/PhysRevLett.23.935.

[22] S. L. Glashow, J. Iliopoulos and L. Maiani. 'Weak Interactions with Lepton-Hadron Symmetry'. In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[23] J.J. Aubert et al. 'Experimental Observation of a Heavy Particle J'. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.

[24] J.-E. Augustin et al. 'Discovery of a Narrow Resonance in e^+e^- Annihilation'. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.

[25] M. Kobayashi and T. Maskawa. 'CP-Violation in the Renormalizable Theory of Weak Interaction'. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652.

[26] J.H. Christenson et al. 'Evidence for the 2π Decay of the K_2^0 Meson'. In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138.

[27] N. Cabibbo. 'Unitary Symmetry and Leptonic Decays'. In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[28] M.L. Perl et al. 'Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation'. In: *Phys. Rev. Lett.* 35 (22 Dec. 1975), pp. 1489–1492. DOI: 10.1103/PhysRevLett.35.1489.

[29] S. Abachi et al. 'Observation of the Top Quark'. In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2632–2637. DOI: 10.1103/physrevlett.74.2632.

[30] K. Abe et al. ‘"Observation of Large CP Violation in the Neutral B Meson System"’. In: *Physical Review Letters* 87.9 (Aug. 2001). DOI: 10.1103/physrevlett.87.091802.

[31] B. Aubert et al. ‘"Observation of CP violation in the B^0 meson system"’. In: *Physical Review Letters* 87.9 (Aug. 2001). DOI: 10.1103/physrevlett.87.091801.

[32] M.E. Peskin and D.V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.

[33] S.L. Glashow. ‘Partial-symmetries of weak interactions’. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[34] S. Weinberg. ‘A Model of Leptons’. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[35] A. Salam. ‘Weak and electromagnetic interactions’. In: *Selected Papers of Abdus Salam*, pp. 244–254. DOI: 10.1142/9789812795915_0034.

[36] F. Englert and R. Brout. ‘Broken Symmetry and the Mass of Gauge Vector Mesons’. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[37] P.W. Higgs. ‘Broken Symmetries and the Masses of Gauge Bosons’. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[38] M. Bentancourt. *Neutrino Cross Sections*. Fermilab NPC Neutrino University Lecture Series. June 2016.

[39] T. Dado. ‘Top-quark properties and mass measurements with the ATLAS detector’. In: *EPJ Web Conf.* 182 (2018). Ed. by Y. Aharonov, L. Bravina and S. Kabana, p. 02033. DOI: 10.1051/epjconf/201818202033.

[40] Wikimedia Commons. *Standard Model of Elementary Particles*. 2023. URL: https://en.wikipedia.org/wiki/File:Standard%5C_Model%5C_of%5C_Elementary%5C_Particles.svg.

[41] R.L. Workman et al. ‘Review of Particle Physics’. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[42] A. Mathad. ‘Lepton flavour universality tests in $b \to c\ell\nu$ decays at LHCb’. 57th Recontres de Moriond EW. 2023.

[43] R. Aaij et al. ‘Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays’. In: *Phys. Rev. D* 108 (3 Aug. 2023), p. 032002. DOI: 10.1103/PhysRevD.108.032002.

[44] R. Davis, D.S. Harmer and K.C. Hoffman. ‘Search for Neutrinos from the Sun’. In: *Phys. Rev. Lett.* 20 (21 May 1968), pp. 1205–1209. DOI: 10.1103/PhysRevLett.20.1205.

[45] Y. Fukuda et al. ‘Evidence for Oscillation of Atmospheric Neutrinos’. In: *Physical Review Letters* 81.8 (Aug. 1998), pp. 1562–1567. DOI: 10.1103/physrevlett.81.1562.

[46] Q.R. Ahmad et al. 'Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory'. In: *Physical Review Letters* 87.7 (July 2001). DOI: 10.1103/physrevlett.87.071301.

[47] B. Pontecorvo. 'Inverse β Processes and Nonconservation of Lepton Charge'. In: *Zhur. Eksptl'. i Teoret. Fiz.* 34 (Jan. 1958). URL: https://www.osti.gov/biblio/4349231.

[48] Z. Maki, M. Nakagawa and S. Sakata. 'Remarks on the Unified Model of Elementary Particles'. In: *Progress of Theoretical Physics* 28.5 (Nov. 1962), pp. 870–880. DOI: 10.1143/PTP.28.870.

[49] M.P. Bento et al. 'Multi-Higgs doublet models: physical parametrization, sum rules and unitarity bounds'. In: *Journal of High Energy Physics* 2017.11 (Nov. 2017). DOI: 10.1007/jhep11(2017)095.

[50] H.E. Haber and L.S. Haskins. 'Supersymmetric Theory and Models'. In: *Anticipating the Next Discoveries in Particle Physics*. WORLD SCIENTIFIC, May 2018. DOI: 10.1142/9789813233348_0006.

[51] C. Cornella, J. Fuentes-Martın and G. Isidori. 'Revisiting the vector leptoquark explanation of the B-physics anomalies'. In: *Journal of High Energy Physics* 2019.7 (July 2019). DOI: 10.1007/jhep07(2019)168.

[52] A.D. Smirnov. 'Vector leptoquark mass limits and branching ratios of $K_L^0, B^0, B_s \rightarrow l_i^+ l_j^-$ decays with account of fermion mixing in leptoquark currents'. In: *Mod. Phys. Lett. A* 33 (2018), p. 1850019. DOI: 10.1142/S0217732318500190.

[53] W. Buchmüller, R. Rückl and D. Wyler. 'Leptoquarks in lepton-quark collisions'. In: *Physics Letters B* 191.4 (1987), pp. 442–448. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(87)90637-X.

[54] J.C. Pati and A. Salam. 'Unified Lepton-Hadron Symmetry and a Gauge Theory of the Basic Interactions'. In: *Phys. Rev. D* 8 (4 Aug. 1973), pp. 1240–1251. DOI: 10.1103/PhysRevD.8.1240.

[55] J.C. Pati and A. Salam. 'Lepton number as the fourth "color"'. In: *Phys. Rev. D* 10 (1 July 1974), pp. 275–289. DOI: 10.1103/PhysRevD.10.275.

[56] The CMS collaboration, A. Tumasyan, W. Adam et al. 'Searches for additional Higgs bosons and for vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s}$= 13 TeV'. In: *Journal of High Energy Physics* 2023.7 (July 2023), p. 73. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2023)073.

[57] G. Aad, B. Abbott, K. Abeling et al. 'Search for pair production of third-generation leptoquarks decaying into a bottom quark and a τ-lepton with the ATLAS detector'. In: *The European Physical Journal C* 83.11 (Nov. 2023), p. 1075. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-12104-7.

[58] A. Dedes, J. Ellis and M. Raidal. ‘Higgs-mediated $B^0_{s,d} \to \mu\tau, e\tau$ and $\tau \to 3\mu, e\mu\mu$ decays in supersymmetric seesaw models’. In: *Physics Letters B* 549.1-2 (Nov. 2002), pp. 159–169. DOI: 10.1016/s0370-2693(02)02900-3.

[59] S. Dimopoulos and H. Georgi. ‘Softly broken supersymmetry and SU(5)’. In: *Nuclear Physics B* 193.1 (1981), pp. 150–162. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(81)90522-8.

[60] A. Bornheim et al. ‘Search for the lepton-flavor-violating leptonic B decays $B^0 \to \mu^\pm\tau^\mp$ and $B^0 \to e^\pm\tau^\mp$’. In: *Phys. Rev. Lett.* 93 (2004), p. 241802. DOI: 10.1103/PhysRevLett.93.241802.

[61] B. Aubert et al. ‘Searches for the decays $B^0 \to \ell^\pm\tau^\mp$ and $B^+ \to \ell^+\nu$ ($\ell = e, \mu$) using hadronic tag reconstruction’. In: *Phys. Rev. D* 77 (2008), p. 091104. DOI: 10.1103/PhysRevD.77.091104.

[62] M. Feindt et al. ‘A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories’. In: *NIMA* 654.1 (2011), pp. 432–440. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.06.008.

[63] T. Abe et al. ‘Belle II Technical Design Report’. In: (Nov. 2010). arXiv: 1011.0352.

[64] L. Aggarwal et al. ‘Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond’. In: (July 2022). arXiv: 2207.06307 [hep-ex].

[65] The Belle II Collaboration et al. *Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*. 2024. arXiv: 2407.00965 [hep-ex].

[66] P. Ahlburg et al. ‘The new and complete Belle II DEPFET pixel detector: Commissioning and previous operational experience’. In: *Nucl. Instrum. Meth. A* 1068 (2024), p. 169763. DOI: 10.1016/j.nima.2024.169763.

[67] K. Akai, K. Furukawa and H. Koiso. ‘SuperKEKB Collider’. In: *Nucl. Instrum. Meth. A* 907 (2018), pp. 188–199. DOI: 10.1016/j.nima.2018.08.017.

[68] W. Herr and B Muratori. ‘Concept of luminosity’. In: (2006). DOI: 10.5170/CERN-2006-002.361.

[69] SuperB Collaboration. *SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report*. 2007. DOI: 10.48550/ARXIV.0709.0451.

[70] F. Abudinén et al. ‘Precise measurement of the D^0 and D^+ lifetimes at Belle II’. In: *Phys. Rev. Lett.* 127.21 (2021), p. 211801. DOI: 10.1103/PhysRevLett.127.211801.

[71] H. Ye et al. ‘Commissioning and performance of the Belle II pixel detector’. In: *NIMA* 987 (2021), p. 164875. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164875.

[72] J. Kemmer and G. Lutz. ‘New Detector Concepts’. In: *Nucl. Instrum. Meth. A* 253 (1987), pp. 365–377. DOI: 10.1016/0168-9002(87)90518-3.

[73] B. Spruck et al. ‘Belle II Pixel Detector Commissioning and Operational Experience’. In: *PoS* Vertex2019 (2020), p. 015. DOI: 10.22323/1.373.0015.

[74] K. Adamczyk et al. ‘The design, construction, operation and performance of the Belle II silicon vertex detector’. In: *JINST* 17.11 (2022), P11042. DOI: 10.1088/1748-0221/17/11/P11042.

[75] N. Taniguchi. ‘Central Drift Chamber for Belle-II’. In: *Journal of Instrumentation* 12.06 (June 2017), p. C06014. DOI: 10.1088/1748-0221/12/06/C06014.

[76] V. Bertacchi et al. ‘Track finding at Belle II’. In: *Comput. Phys. Commun.* 259 (2021), p. 107610. DOI: 10.1016/j.cpc.2020.107610.

[77] K. Kojima. ‘The operation and performance of the TOP detector at the Belle II experiment’. In: *PoS* EPS-HEP2021 (2022), p. 803. DOI: 10.22323/1.398.0803.

[78] M. Yonenaga et al. ‘Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data’. In: *Progress of Theoretical and Experimental Physics* 2020.9 (Aug. 2020). 093H01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa090.

[79] J. Fast. ‘The Belle II imaging Time-of-Propagation (iTOP) detector’. In: *NIMA* 876 (2017). The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016), pp. 145–148. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2017.02.045.

[80] D. Kotchetkov et al. ‘Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector’. In: *NIMA* 941 (2019), p. 162342. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2019.162342.

[81] R. Pestotnik et al. ‘Aerogel RICH for forward PID at Belle II’. In: *NIMA* 732 (2013). Vienna Conference on Instrumentation 2013, pp. 371–374. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2013.06.080.

[82] T. Aushev et al. ‘A scintillator based endcap KL and muon detector for the Belle II experiment’. In: *NIMA* 789 (2015), pp. 134–142. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.03.060.

[83] D.J. Lange. ‘The EvtGen particle decay simulation package’. In: *NIMA* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4.

[84] T. Sjöstrand et al. ‘An introduction to PYTHIA 8.2’. In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024.

[85] S. Jadach, B.F.L. Ward and Z. Was. ‘The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions’. In: *Computer Physics Communications* 130.3 (2000), pp. 260–325. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(00)00048-5. URL: https://www.sciencedirect.com/science/article/pii/S0010465500000485.

[86] N. Davidson et al. ‘Universal interface of TAUOLA: Technical and physics documentation’. In: *Computer Physics Communications* 183.3 (2012), pp. 821–843. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2011.12.009.

[87] C.M. Carloni Calame et al. ‘The BABAYAGA event generator’. In: *Nuclear Physics B - Proceedings Supplements* 131 (2004). SIGHADO3, pp. 48–55. ISSN: 0920-5632. DOI: https://doi.org/10.1016/j.nuclphysbps.2004.02.008.

[88] F. A. Berends, P. H. Daverveldt and R. Kleiss. ‘Monte Carlo simulation of two-photon processes: I: Radiative corrections to multiperipheral $e^+e^-\mu^+\mu^-$ production’. In: *Computer Physics Communications* 40.2 (1986), pp. 271–284. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(86)90114-1.

[89] S. Agostinelli et al. ‘Geant4—a simulation toolkit’. In: *NIMA* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[90] K. Hirata. ‘An Introduction to SAD’. In: *2nd ICFA Advanced Beam Dynamics Workshop*. 1988, pp. 62–65. DOI: 10.5170/CERN-1988-004.62.

[91] T. Alexopoulos et al. ‘Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers’. In: *NIMA* 592.3 (2008), pp. 456–462. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.04.038.

[92] T. Bilka et al. *Implementation of GENFIT2 as an experiment independent track-fitting framework*. 2019. DOI: 10.48550/ARXIV.1902.04405.

[93] N. Braun. *Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment*. 2019. DOI: https://doi.org/10.1007/978-3-030-24997-7.

[94] W. Waltenberger, W. Mitaroff and F. Moser. ‘RAVE—a Detector-independent vertex reconstruction toolkit’. In: *NIMA* 581.1 (2007). VCI 2007, pp. 549–552. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2007.08.048.

[95] T. Keck. ‘FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification’. In: (2016). arXiv: 1609.06119.

[96] I.H. Witten, E. Frank and M.A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN: 0123748569.

[97] I.J. Good. ‘Rational Decisions’. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 14.1 (1952), pp. 107–114. ISSN: 00359246. URL: http://www.jstor.org/stable/2984087.

[98] W. Verkerke and D. P. Kirkby. ‘The RooFit toolkit for data modeling’. In: *eConf* C0303241 (2003). Ed. by L. Lyons and Muge Karagoz, MOLT007. arXiv: physics/0306116.

[99] W.A. Rolke and A.M. López. ‘Confidence intervals and upper bounds for small signals in the presence of background noise’. In: *NIMA* 458.3 (2001), pp. 745–758. ISSN: 0168-9002. DOI: 10.1016/s0168-9002(00)00935-9.

[100] A. Ali, S.M.H. Shamsuddin and A.L. Ralescu. ‘Classification with class imbalance problem: A review’. In: *SOCO 2015*. 2015.

[101] D.N. Brown, J. Ilic and G.B. Mohanty. ‘Extracting longitudinal shower development information from crystal calorimetry plus tracking’. In: *NIMA* 592.3 (2008), pp. 254–260. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.04.029.

[102] S. Longo et al. ‘CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders’. In: *NIMA* 982 (2020), p. 164562. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164562.

[103] T. Keck, F. Abudinén, F.U. Bernlochner et al. ‘The Full Event Interpretation’. In: *Comput Softw Big Sci* 3 (2019), p. 6. DOI: 10.1007/s41781-019-0021-8.

[104] J. Conrad and J. Lundberg. *TRolke*. Version 6.26. 2009. URL: https://root.cern.ch/doc/master/classTRolke.html.

Appendix A

Signal Extraction Region Optimisation

The p_ℓ^B dimension of the *signal extraction region* is chosen to have the boundaries [2.24, 2.4] in order to cover nearly all of the signal peak. This is important as p_ℓ^B provides most of the discrimination between signal and background. The boundaries for E_{extra} are harder to define as there is overlap between the signal and background peaks, and the signal tail is substantial. Therefore, optimisation is done to decide the upper bound of E_{extra} for the *signal extraction region*. The most optimal upper bound is decided according to whichever yields the best upper limit. An explanation of how the upper limit is calculated is given in Section 4.3. For each E_{extra} value tested, the optimisation procedure is as follows:

(1) Calculate the signal efficiency ϵ within the tested *signal extraction region*. As the error on the signal efficiency is not yet estimated, a standard 5% is used which is roughly equivalent to the error from the Belle analysis [16].

(2) Obtain the total background estimate n_{b}, scaled for 365 fb^{-1}

  (a) The *two true lepton* component is estimated using the fitted yield from 100 toy background fits, with the fit setup described in Section 6.7.1. The normalisation of the generated sample used in each fit is sampled from a Poisson distribution with a mean equal to the expected background yield from 365 fb^{-1}. The error on this estimate is obtained from the fit.

  (b) The *fake lepton* component is obtained by counting how many remain in the *signal extraction region* from the officially produced MC samples. The error on this estimate is the Poisson error.

  (c) The J/ψ*-like decay* component is obtained counting how many remain in the *signal extraction region* from the officially produced MC samples. The error on this estimate is the Poisson error.

  The combined error on the total background count is obtained by summing the error of each component in quadrature.

(3) To obtain a realistic set of values for the number of observed events n_{obs} for the background estimate in *Step 2*, a Poisson distribution with mean equal to the background estimate is drawn. The set of n_{obs} selected correspond to those with a reasonable non-zero Poisson probability.

For example, if the total background estimate was 1.10 events, the corresponding n_{obs} selected are those shown in Figure A.1 below.

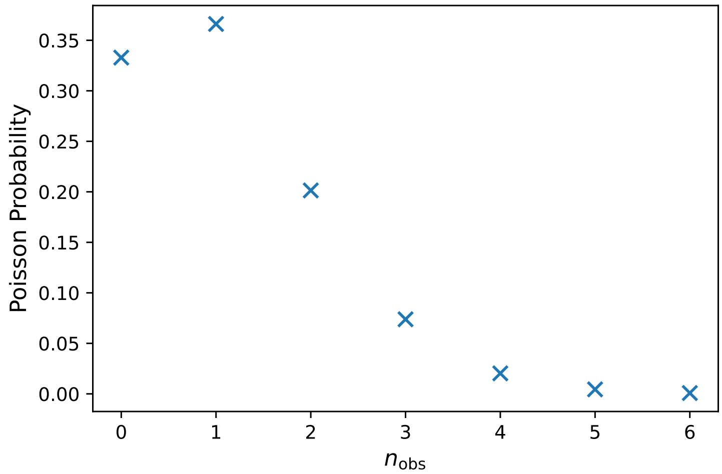


FIGURE A.1. Poisson probability distribution used to choose the range of n_{obs} values for an example background estimate of 1.10 events.

(4) The confidence interval (CI) is calculated for each n_{obs} selected from *Step 3* using `TRolke`. Except for n_{obs} values that lie in the tails of their respective Poisson distributions, the CI is an upper limit (UL).

(5) The overall UL is calculated by weighting each UL obtained in *S*tep 4 by the Poisson probability for the corresponding n_{obs}. Following on from the example in *Step 3*, supposing that below are example UL values from `TRolke` [104] and the Poisson probabilities for the corresponding n_{obs} values respectively

$$\mathrm{UL} = [1866, 5184, 8501, 11534, 14416, 17200, 19915]$$

$$p(n_{\mathrm{obs}}) = [0.3329, 0.3662, 0.2014, 0.0738, 0.0203, 0.0045, 0.0008]$$

then the weighted average and corresponding branching fraction upper limit using the Belle II dataset would be

$$\sum_i \mathrm{UL}_i \times p(n_{\mathrm{obs}})_i \approx 5469$$

$$\mathcal{B}_{\mathrm{UL}} < 1.41 \times 10^{-5}$$

A summary of the numbers obtained for each step along with the results of the optimisation are given in Table A.1 and Table A.2 for the $B^0 \to e^{\pm}\tau^{\mp}$ and $B^0 \to \mu^{\pm}\tau^{\mp}$ modes respectively. Plots showing the signal efficiency, total background and calculated upper limit for each mode is provided in Figure A.2. It can be seen from the results that the optimal values for the upper bound are either 0.20, 0.25 or 0.30, with 0.30 being chosen due to the following reasons: (1) the difference in upper limit values between these are down to statistical fluctuations as opposed to any meaningful difference in the efficiency or background yields; (2) choosing a wider E_{extra} region increases robustness against potential data-MC disagreements which are known to be most prominent at low E_{extra} values.

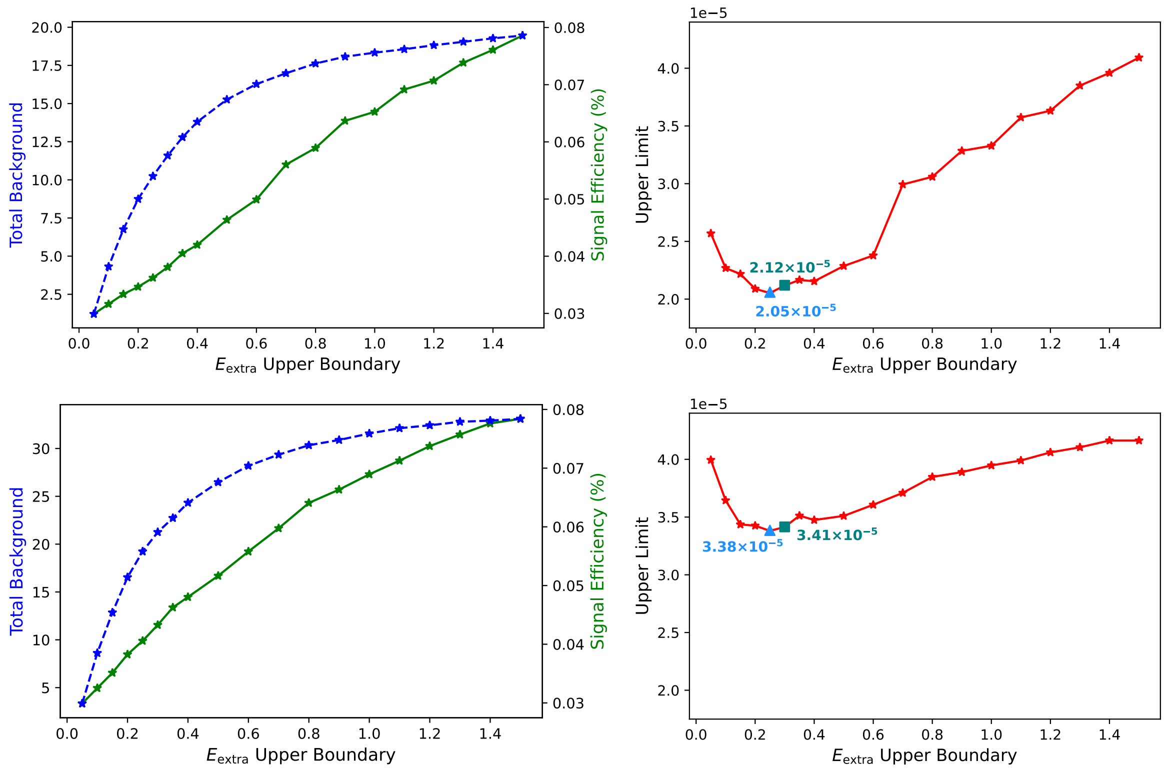


FIGURE A.2. Results of the *signal extraction region* optimisation for the $B^0 \to e^{\pm}\tau^{\mp}$ mode (top) and $B^0 \to \mu^{\pm}\tau^{\mp}$ mode (bottom). The numbers plotted here are found in Tables A.1 and A.2.

TABLE A.1. Summary of the results for the *signal extraction* region optimisation for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. The estimations for the *two true lepton*, *fake lepton* and J/ψ*-like decay* components are given by the $\boldsymbol{n_{\mathrm{TTL}}}$, $\boldsymbol{n_{\mathrm{FL}}}$ and $\boldsymbol{n_{J/\psi}}$ columns respectively. The total background is given by $\boldsymbol{n_b}$. All background counts are scaled according to 365 fb^{-1}. The range of observed events used to calculate the upper limit is given by $\boldsymbol{n_{obs}}$, while the resulting upper limit value is given by the **UL** column.

Cut	ϵ **(%)**	n_{TTL}	n_{FL}	$n_{J/\psi}$	n_b	n_{obs}	**UL** ($\times 10^{-5}$)
0.05	0.0299	0.56 ± 0.01	0.26 ± 0.51	0.39 ± 0.62	1.21 ± 0.80	[0, 6]	2.57
0.10	0.0382	1.08 ± 0.03	0.26 ± 0.51	0.52 ± 0.72	1.86 ± 0.88	[0, 7]	2.27
0.15	0.0447	1.60 ± 0.04	0.26 ± 0.51	0.65 ± 0.81	2.51 ± 0.96	[0, 8]	2.22
0.20	0.0500	2.08 ± 0.05	0.26 ± 0.51	0.65 ± 0.81	2.99 ± 0.96	[0, 10]	2.09
0.25	0.0540	2.54 ± 0.07	0.39 ± 0.62	0.65 ± 0.81	3.58 ± 1.02	[0, 11]	2.05
0.30	0.0576	2.98 ± 0.08	0.39 ± 0.62	0.91 ± 0.95	4.28 ± 1.14	[0, 12]	2.12
0.35	0.0608	3.37 ± 0.09	0.78 ± 0.88	1.03 ± 1.01	5.18 ± 1.34	[0, 14]	2.17
0.40	0.0635	3.80 ± 0.11	0.78 ± 0.88	1.16 ± 1.08	5.74 ± 1.40	[0, 15]	2.15
0.50	0.0674	4.54 ± 0.12	1.16 ± 1.08	1.68 ± 1.30	7.38 ± 1.69	[0, 17]	2.29
0.60	0.0701	5.22 ± 0.14	1.42 ± 1.19	2.07 ± 1.44	8.71 ± 1.87	[0, 19]	2.38
0.70	0.0720	5.83 ± 0.16	2.07 ± 1.44	3.10 ± 1.76	11.00 ± 3.42	[0, 23]	2.99
0.80	0.0737	6.40 ± 0.17	2.46 ± 1.57	3.23 ± 1.80	12.09 ± 3.60	[0, 25]	3.06
0.90	0.0749	6.91 ± 0.18	3.34 ± 1.83	3.62 ± 1.90	13.87 ± 4.06	[0, 27]	3.28
1.00	0.0756	7.37 ± 0.20	3.34 ± 1.83	3.75 ± 1.94	14.46 ± 4.18	[0, 27]	3.33
1.10	0.0762	7.80 ± 0.21	3.85 ± 1.96	4.27 ± 2.07	15.92 ± 4.70	[0, 29]	3.57
1.20	0.0769	8.18 ± 0.22	3.98 ± 1.99	4.40 ± 2.10	16.50 ± 4.83	[0, 31]	3.63
1.30	0.0775	8.53 ± 0.23	4.24 ± 2.06	4.91 ± 2.22	17.68 ± 5.33	[0, 32]	3.85
1.40	0.0781	8.85 ± 0.24	4.50 ± 2.12	5.17 ± 2.27	18.52 ± 5.59	[0, 33]	3.96
1.50	0.0786	9.15 ± 0.25	4.88 ± 2.21	5.43 ± 2.33	19.46 ± 5.87	[0, 35]	4.09

TABLE A.2. Summary of the results for the *signal extraction* region optimisation for the $B^0 \to \mu^\pm\tau^\mp$ mode. The estimations for the *two true lepton*, *fake lepton* and J/ψ*-like decay* components are given by the $\boldsymbol{n_{\mathrm{TTL}}}$, $\boldsymbol{n_{\mathrm{FL}}}$ and $\boldsymbol{n_{J/\psi}}$ columns respectively. The total background is given by $\boldsymbol{n_b}$. All background counts are scaled according to 365 fb^{-1}. The range of observed events used to calculate the upper limit is given by $\boldsymbol{n_{obs}}$, while the resulting upper limit value is given by the **UL** column.

Value	ϵ (%)	n_{TTL}	n_{FL}	$n_{J/\psi}$	n_b	n_{obs}	**UL** ($\times 10^{-5}$)
0.05	0.0299	0.75 ± 0.02	2.59 ± 1.61	0	3.34 ± 1.61	[0, 11]	3.99
0.10	0.0385	1.46 ± 0.03	3.49 ± 1.87	0	4.95 ± 1.87	[0, 14]	3.64
0.15	0.0454	2.15 ± 0.04	4.40 ± 2.10	0	6.55 ± 2.10	[0, 16]	3.44
0.20	0.0514	2.81 ± 0.06	5.66 ± 2.38	0	8.47 ± 2.38	[0, 19]	3.43
0.25	0.0558	3.45 ± 0.07	6.44 ± 2.54	0	9.89 ± 2.54	[0, 22]	3.38
0.30	0.0591	4.08 ± 0.08	7.34 ± 2.71	0.13 ± 0.36	11.55 ± 2.73	[0, 24]	3.41
0.35	0.0615	4.64 ± 0.09	8.61 ± 2.93	0.13 ± 0.36	13.38 ± 2.95	[0, 27]	3.51
0.40	0.0641	5.21 ± 0.10	9.13 ± 3.02	0.13 ± 0.36	14.47 ± 3.04	[0, 27]	3.47
0.50	0.0676	6.27 ± 0.12	9.90 ± 3.15	0.52 ± 0.72	16.69 ± 3.23	[0, 31]	3.51
0.60	0.0704	7.25 ± 0.14	11.46 ± 3.39	0.52 ± 0.72	19.23 ± 3.47	[0, 35]	3.61
0.70	0.0723	8.16 ± 0.16	12.85 ± 3.58	0.65 ± 0.81	21.66 ± 3.67	[0, 38]	3.71
0.80	0.0739	8.90 ± 0.17	14.76 ± 3.84	0.65 ± 0.81	24.31 ± 3.93	[0, 42]	3.85
0.90	0.0748	9.77 ± 0.19	15.28 ± 3.91	0.65 ± 0.81	25.70 ± 4.00	[0, 44]	3.89
1.00	0.0759	10.49 ± 0.20	16.16 ± 4.02	0.65 ± 0.81	27.30 ± 4.11	[0, 46]	3.95
1.10	0.0768	11.14 ± 0.21	16.94 ± 4.12	0.65 ± 0.81	28.73 ± 4.20	[0, 48]	3.99
1.20	0.0773	11.76 ± 0.23	17.84 ± 4.22	0.65 ± 0.81	30.25 ± 4.30	[0, 51]	4.06
1.30	0.0779	12.32 ± 0.24	18.49 ± 4.30	0.65 ± 0.81	31.46 ± 4.38	[0, 53]	4.10
1.40	0.0781	12.84 ± 0.25	19.13 ± 4.37	0.65 ± 0.81	32.62 ± 4.45	[0, 55]	4.16
1.50	0.0784	13.32 ± 0.25	19.13 ± 4.37	0.65 ± 0.81	33.10 ± 4.45	[0, 56]	4.16

Appendix B

Alternate Fitting Variables

In the context of the *leptonic partition*, two alternate fitting variables for E_{extra} were tested to determine if they provide the same or better discrimination between signal and background. The two variables tested were U_{CM} and $|\vec{p}_{\mathrm{miss}_{\mathrm{CM}}}|$. The signal and background shapes within the *fitting region* and a subset of the *signal extraction region* obtained using the cut $2.24 < p_{\ell}^{B} < 2.4$ GeV/c are compared between the two alternate fitting variables and E_{extra}. The distributions are provided in Figure B.1 and B.2 for the $B^0 \to e^{\pm}\tau^{\mp}$ and $B^0 \to \mu^{\pm}\tau^{\mp}$ modes respectively. In the case of U_{CM}, though there initially appears to be some separation between signal and background in the *fitting region*, looking at the subset of the *signal extraction region* demonstrates that all of this separation is already encapsulated by the $2.24 < p_{\ell}^{B} < 2.4$ dimension of the *signal extraction region*. Thus, within the *signal extraction region*, U_{CM} would add no extra useful information to separate signal from background. The situation is less favourable for $|\vec{p}_{\mathrm{miss}_{\mathrm{CM}}}|$ which shows no useful signal-background separation in either the *fitting region* or *signal extraction region*. In fact, some sculpting of the background appears to be introduced in the latter case. Comparing these with the E_{extra} distributions demonstrates that this variable still contains a valuable amount of signal-background separation in both regions. Specifically, it provides discrimination in addition to what is provided by p_{ℓ}^{B} within the *signal extraction region*, therefore justifying its use as a second fitting variable.

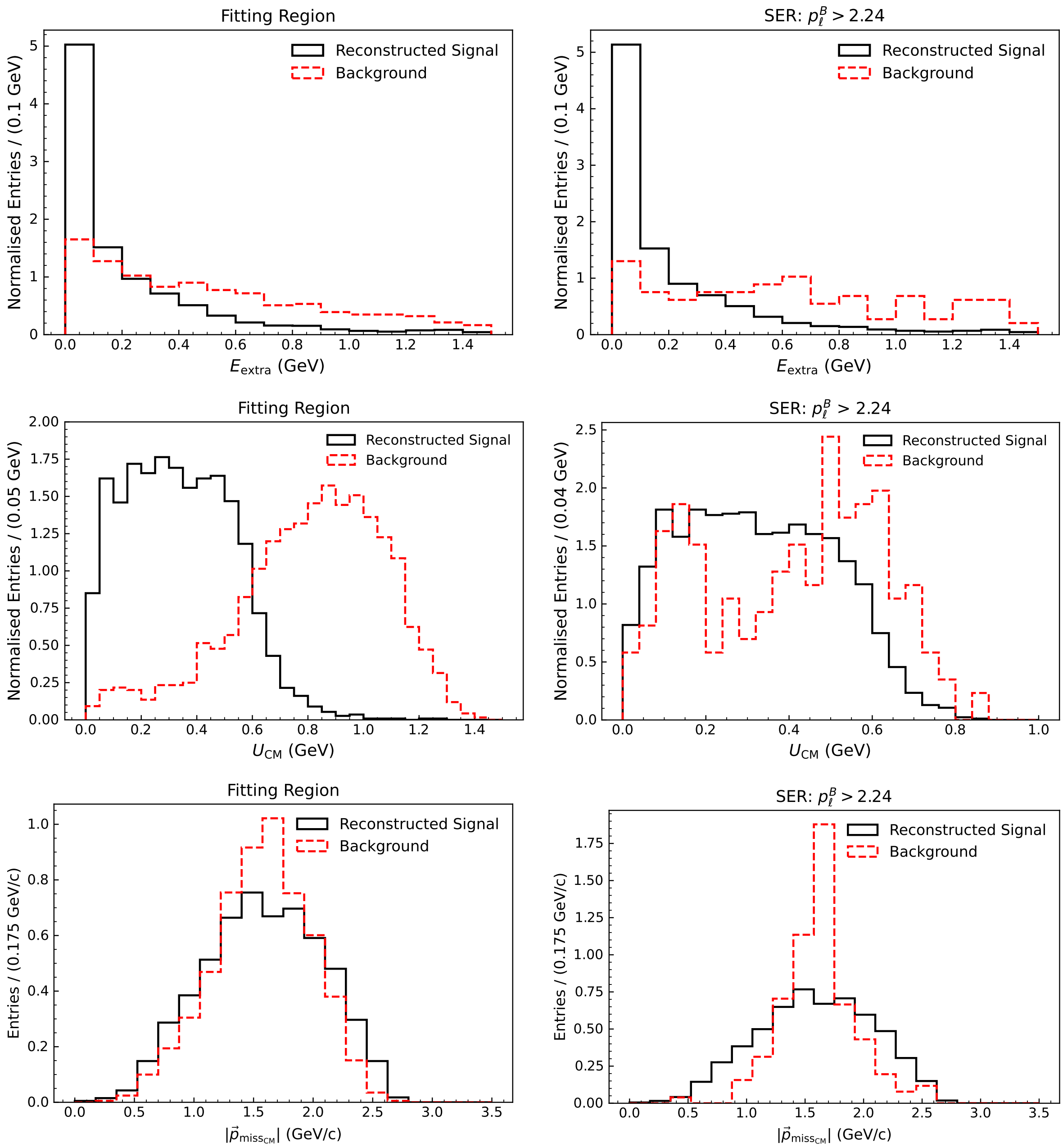


FIGURE B.1. Distributions of the current fitting variable E_{extra} compared with the alternate fitting variables U_{CM} and $|\vec{p}_{\text{miss}_{\text{CM}}}|$ for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. On the left are plots of the entire *fitting region* while the right-hand plots show the *signal extraction region* subset designated by $p_\ell^B > 2.24$ GeV/c.

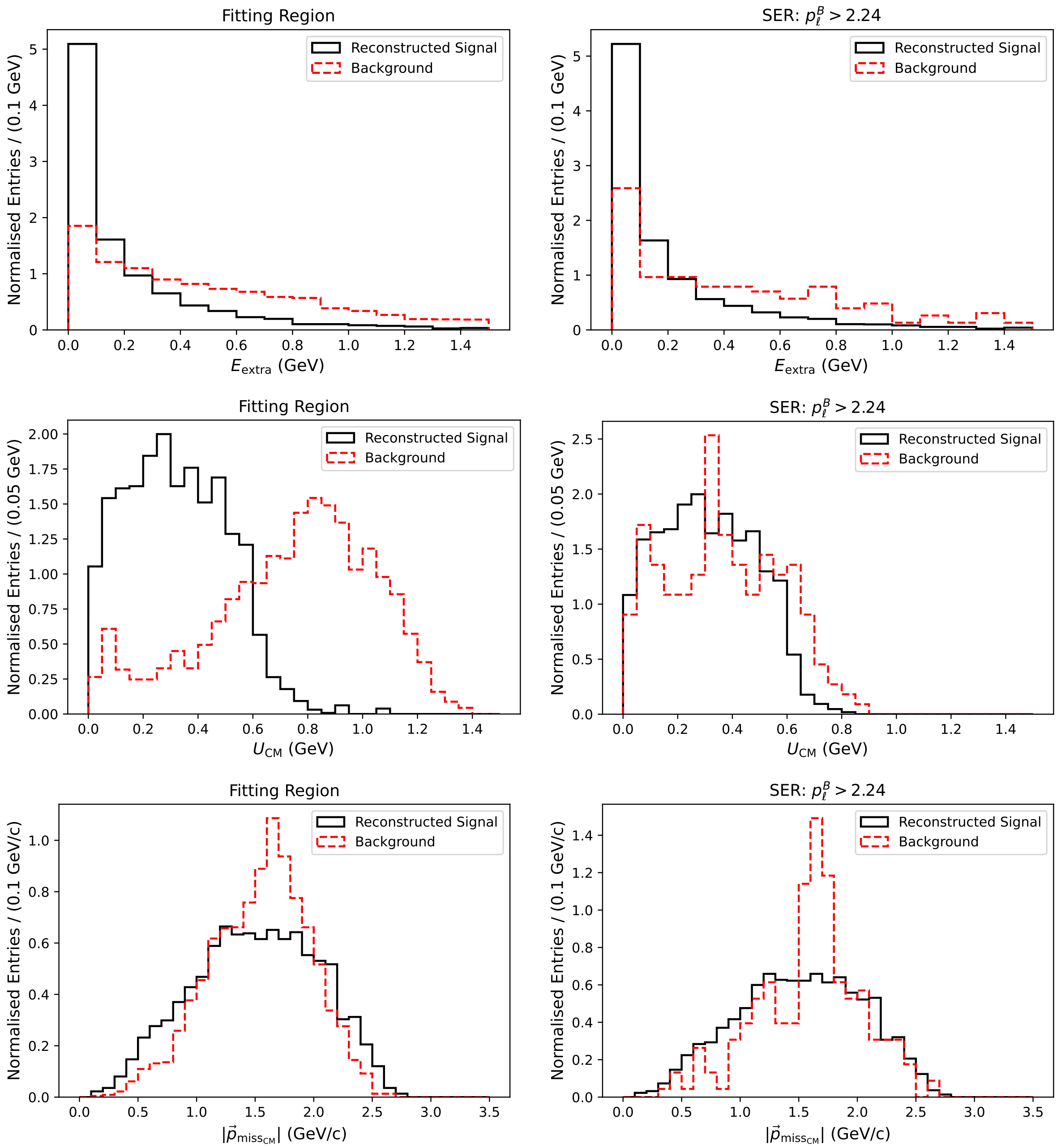


FIGURE B.2. Distributions of the current fitting variable E_{extra} compared with the alternate fitting variables U_{CM} and $|\vec{p}_{\text{miss}_{\text{CM}}}|$ for the $B^0 \to \mu^\pm\tau^\mp$ mode. On the left are plots of the entire *fitting region* while the right-hand plots show the *signal extraction region* subset designated by $p_\ell^B > 2.24$ GeV/c.

Appendix C

ROE Mask Study

Different sets of masks for ROE clusters were studied in order to determine which was most suitable. Table C.1 denotes each of the masks tried.

TABLE C.1. Cuts applied for ROE clusters along with the mask number. The current mask used for the analysis is written in magenta. The barrel, backward and forward endcap regions refer to the ECL sub-detector.

Mask	Mask Cuts
0	`inCDCAcceptance` == 1, $E > 0.055$
1	`inCDCAcceptance` == 1 $E > 0.055$ in barrel, $E > 0.1$ in backward endcap, $E > 0.08$ in forward endcap
2	**`inCDCAcceptance` == 1, $E > 0.055$** **`beamBackgroundSuppression` > 0.4, `fakePhotonSuppression` > 0.2**
3	`inCDCAcceptance` == 1 $E > 0.055$ in barrel, $E > 0.1$ in backward endcap, $E > 0.08$ in forward endcap `beamBackgroundSuppression` > 0.4, `fakePhotonSuppression` > 0.2
4	`inCDCAcceptance` == 1 $E > 0.055$ in barrel, $E > 0.1$ in backward endcap, $E > 0.08$ in forward endcap `minC2TDist` > 40, `|clusterTiming|` < 200

Figure C.1 includes comparisons between different masks based on the E_{extra} and N_{ROE} photon distributions for both modes within the *fitting region*. As the $B^0\overline{B}^0$ background is the most dominant, a MC sample of this background was used for the comparison. Multiple conclusions can be drawn from these distributions. Firstly, there is little difference between *Mask 0* and *Mask 1* for E_{extra}. This shows that region-specific energy cuts has little impact for this analysis. The minor difference between *Mask 2* and *Mask 3* further supports this idea. Going down the route of simplicity supports the use of the global energy cut $E > 0.055$ GeV for the ROE mask. There is a difference in the number of

ROE photons, however as the $E_{\rm extra}$ has marginal differences, it indicates that the photons affected by the region-specific energy cuts have small energies. Secondly, *Mask 4* provides the best $E_{\rm extra}$ distribution when considering the *signal extraction region*. The signal distribution is more peaked at lower $E_{\rm extra}$ values while for the background there is a smaller peak within the *signal extraction region*. The $N_{\rm ROE}$ photon distribution has fewer changes which indicates that higher energy photons are impacted more by *Mask 4*. From this last conclusion, it suggests that the most optimal ROE mask is the one that utilises the `minC2TDist` and `clusterTiming` variables. A likely reason for this is that the `beamBackgroundSuppression` and `fakePhotonSuppression` cuts are sub-optimal, and instead they should first be optimised before a comparison between these masks is made. As the current version of the analysis uses run-independent MC, it was advised by members within the collaboration to save this optimisation for when run-dependent MC is used.

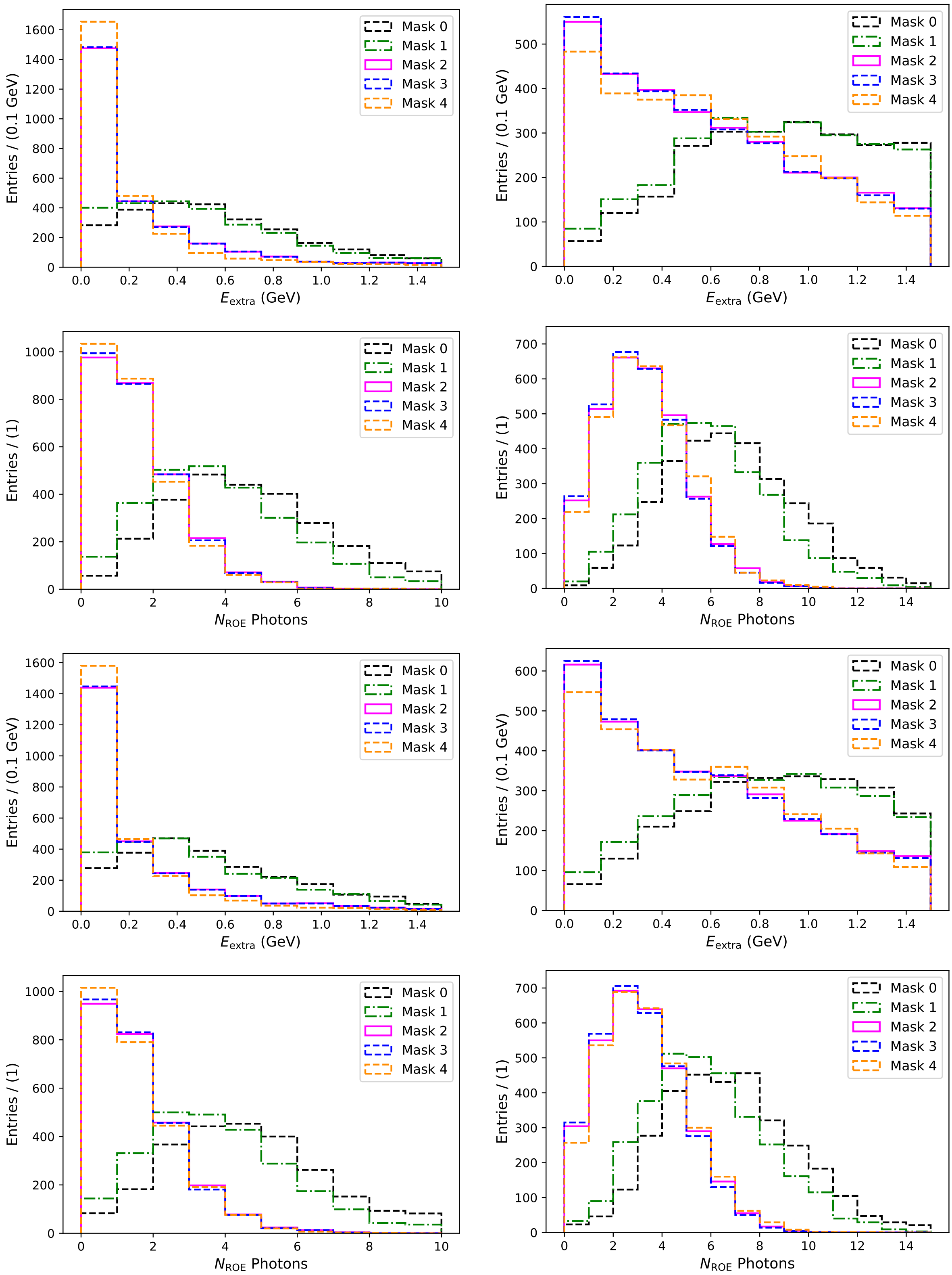


FIGURE C.1. Distributions of E_{extra} and the number of photons in the ROE for each of the different masks presented in Table C.1 for the $B^0 \to e^{\pm}\tau^{\mp}$ mode (top 4 plots) and $B^0 \to \mu^{\pm}\tau^{\mp}$ mode (bottom 4 plots). Reconstructed signal is on the left while $B^0\overline{B}^0$ background is on the right. This is within the *fitting region* for the leptonic τ modes.

Appendix D

BDT-Based PID Null Value Issue

As the BDT-based PID variable contains null values that lead to an efficiency loss for the leptonic τ modes, a study was conducted into what causes these null values. From an internal Belle II collaboration study on the BDT-based PID variable, it is mentioned that only tracks with at least one matched ECL cluster are provided with a PID score from the BDT. Further to this, clusters are required to have more than 55 MeV of energy, and tracks which are pointing to gaps within the ECL acceptance region are automatically vetoed. To check if these factors contributed to the null cases observed for this analysis, a version of the signal MC sample of $B^0 \to e^{\pm}\tau^{\mp}$ without any PID cuts on the τ daughter leptons was studied. As both $\tau \to e\nu\overline{\nu}$ and $\tau \to \mu\nu\overline{\nu}$ are included in this sample, the impact of the null cases on both electron and muon PID can still be studied without requiring the $B^0 \to \mu^{\pm}\tau^{\mp}$ sample. For the $B^0 \to e^{\pm}\tau^{\mp}$ sample, 3511 leptonic τ candidates had a null value for the BDT-based PID variable. Investigating these further, it is found that of these 3511 candidates:

- 3292 (93.7%) are lepton tracks that have no matched ECL cluster
- 23 (0.66%) are lepton tracks that lie in the ECL gaps between the endcaps and barrel
- 189 (5.38%) are lepton tracks have matched ECL clusters with $E < 55$ MeV

Given that track-cluster matching is responsible for nearly all the null cases, modifications on the analysis level in an attempt to fix this issue are unfeasible. Instead, the following two proposed solutions were discussed:

(1) Accept the efficiency loss
(2) Divide the dataset into four groups, depending on which combination of the signal or τ daughter lepton have a null value for the BDT-based PID variable. For the tracks with a null value, the likelihood-based PID variable is used instead.

It was decided to go with solution 1 and to accept the efficiency loss. The difficulty and tediousness of implementing solution 2 was one of the reasons why it was rejected. Furthermore, for the estimation of the *fake lepton* background which requires using the associated fake rate for the given lepton ID variable, using two different PID methods significantly complicates the estimation method.

With the decision to accept the efficiency loss, a study on why the affected τ daughter lepton tracks had no matching ECL cluster was performed. The track purities are categorised according the τ mode below, with transverse momentum and polar angle distributions provided in Figure D.1:

- *$\tau \to e\nu\overline{\nu}$ mode*
  66.89% of problematic tracks were electrons, 30.54% were pions, 1.53% were kaons
- *$\tau \to \mu\nu\overline{\nu}$ mode*
  56.51% of problematic tracks were muons, 37.14% were pions, 2.40% were kaons, 2.01% were electrons

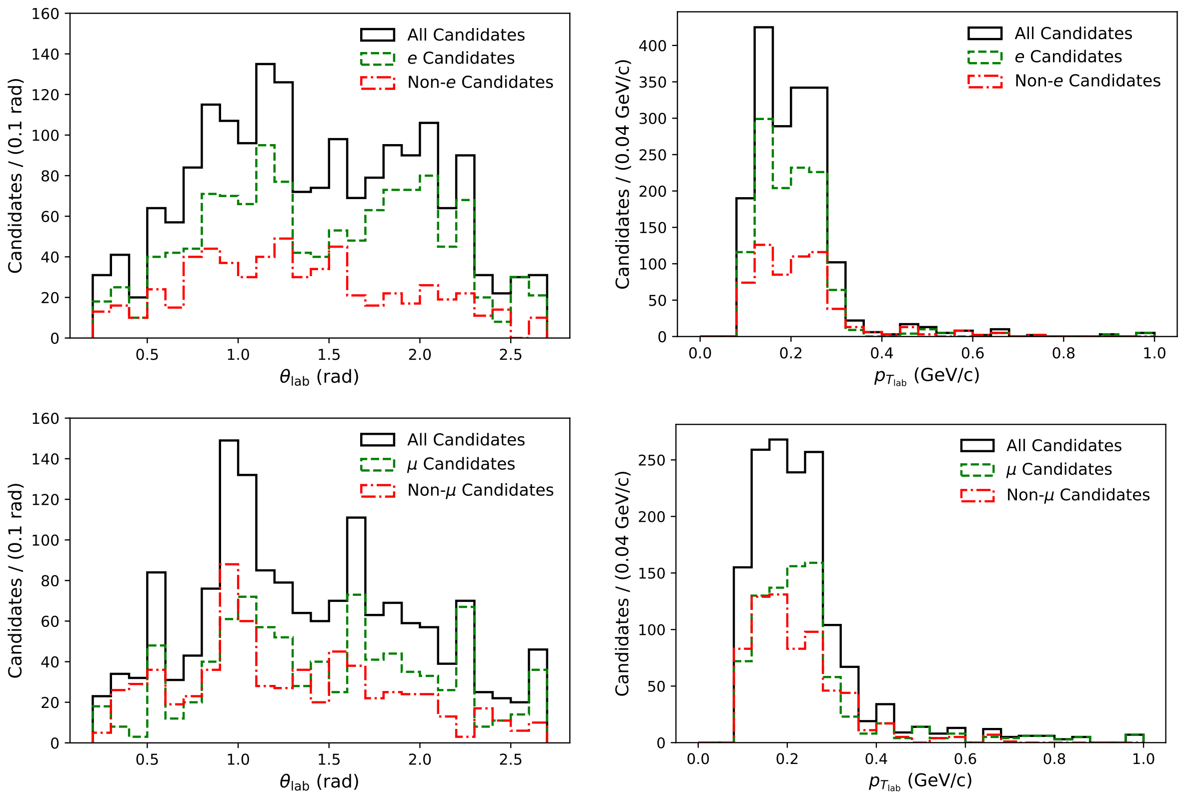


FIGURE D.1. Distributions of the transverse momentum p_T and polar angle θ for τ daughter lepton tracks with a null value for the BDT-based PID variable. Tracks for the $\tau \to e\nu\overline{\nu}$ is shown in the top two plots, with the bottom two plots showing tracks for the $\tau \to \mu\nu\overline{\nu}$ mode. The candidates in the legend refer to the tracks in the itemized list above.

From these results, it can be seen that the majority of signal events discarded had τ daughter lepton tracks that were of the correct particle identity. However, irrespective of the particle identity, most tracks suffered from low transverse momentum which explains why close to 94% had no matching ECL cluster. All had a polar angle distribution that resembles the typical ECL occupancy expected from such events.

Although solution 2 was rejected, it was recommended that a brief study be conducted on the impact on background retention when using the likelihood-based PID variable as a backup. For this, only the $B^0\overline{B}^0$ sample was used as it is by far the most dominant background contribution. Table D.1 outlines the results of using three different thresholds for the likelihood-based PID variable. Overall, as expected, more events are recovered as the likelihood-based PID threshold is lowered. Considering the most reasonable threshold value of 0.9, it can be seen that if solution 2 were implemented, the large increase in background retention compared to the small increase in signal efficiency shows that solution 2 would not yield any improvement to the sensitivity of the analysis but rather degrade sensitivity.

TABLE D.1. Results from different thresholds of the likelihood-based PID variable. The same threshold is applied for the electron and muon ID of τ daughter lepton tracks that have a null value for the BDT-based PID variable. The **Events Recovered** gives the percentage of events that are recovered by using the likelihood-based PID variable that would otherwise have been discarded due to the null issue. The **Events Surviving Fitting Region** gives the percentage of events that have recovered plus survive all the selection cuts up to and including the *fitting region* selection. The increase to overall efficiency/retention for signal/background within the *fitting region* is given in the **Efficiency/Retention Increase** row.

Quantity (%)	e, μ **ID** > 0.3		e, μ **ID** > 0.7		e, μ **ID** > 0.9	
	Signal	$B^0\overline{B}^0$	Signal	$B^0\overline{B}^0$	Signal	$B^0\overline{B}^0$
Events Recovered	72.56	59.11	22.76	9.46	13.82	3.99
Events Surviving Fitting Region	30.91	0.27	12.03	0.05	7.46	0.03
Efficiency/Retention Increase (%)	11.71	453.14	4.56	84.83	2.83	54.21

APPENDIX E

Lepton ID Correction Coverage Plots

Coverage plots for the lepton ID efficiency corrections are presented below for each mode. They are generated by the `PIDvar` package [3].

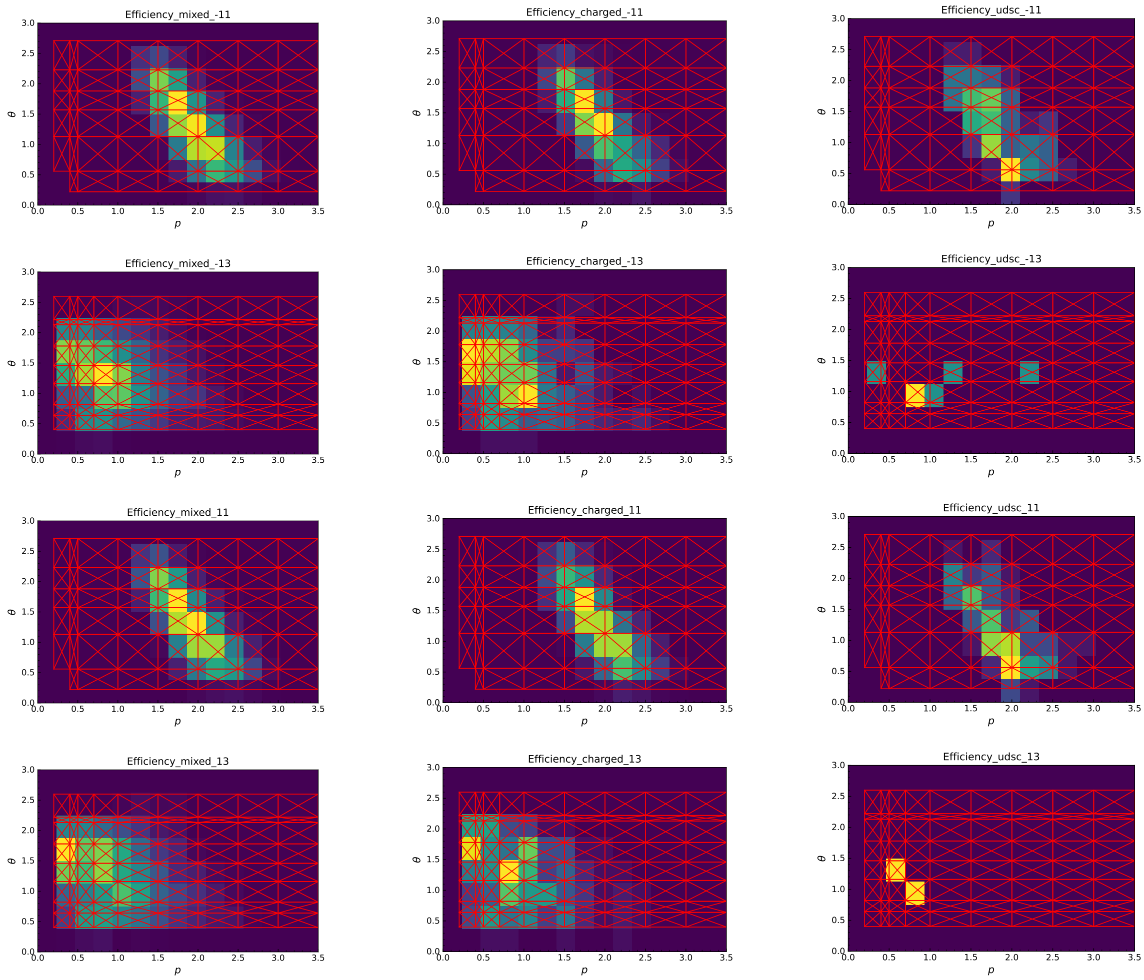


FIGURE E.1. Coverage distributions for the lepton ID efficiency corrections for the $B^0 \to e^{\pm}\tau^{\mp}$ mode. The red lines indicate the coverage provided, while the heatmap shows the distribution of the leptons that require correcting. The title of each sub-figure mentions the MC sample that is being corrected followed by the PDG code of the relevant lepton.

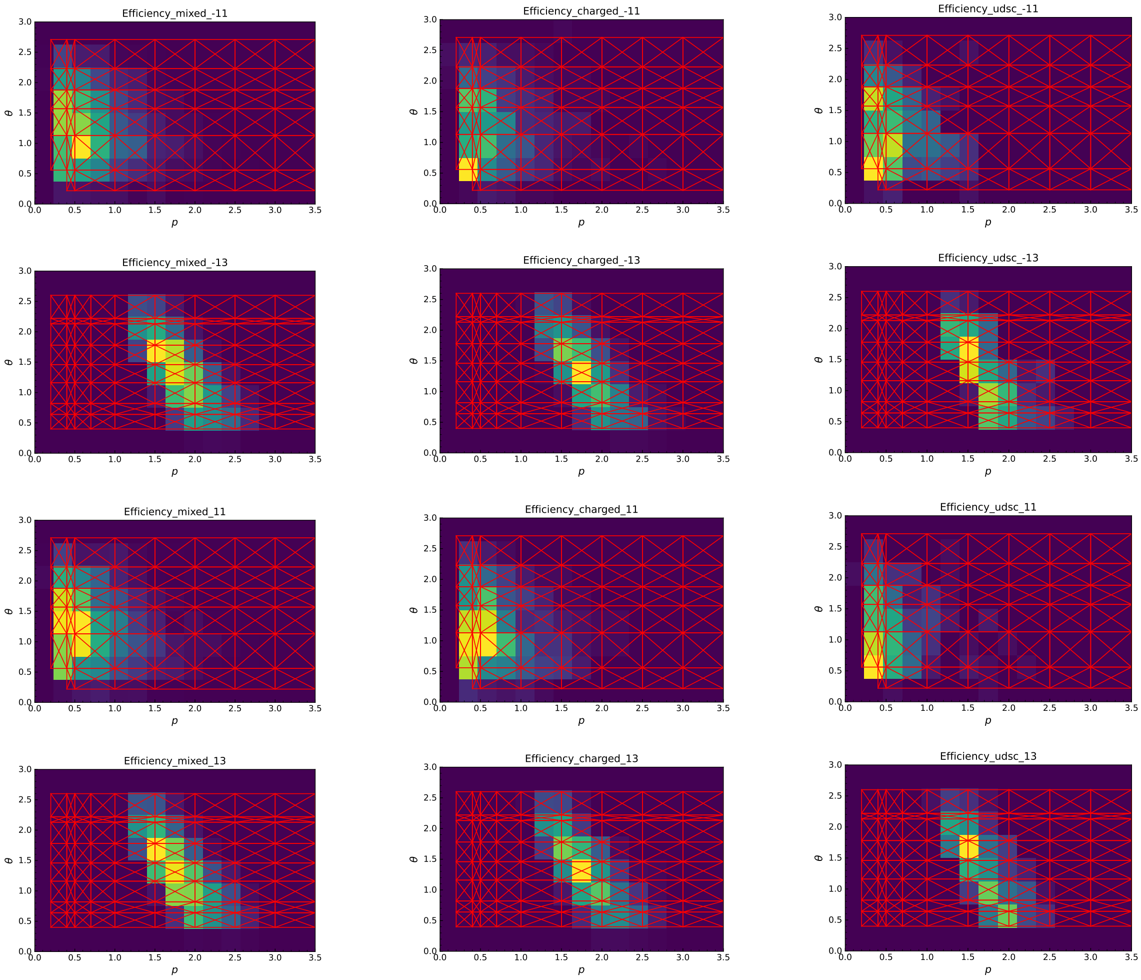


FIGURE E.2. Coverage distributions for the lepton ID efficiency corrections for the $B^0 \to \mu^{\pm}\tau^{\mp}$ mode. The red lines indicate the coverage provided, while the heatmap shows the distribution of the leptons that require correcting. The title of each sub-figure mentions the MC sample that is being corrected followed by the PDG code of the relevant lepton.

Appendix F

$D^{*-} \to D^-\pi^0$ Peak for the Control Channel Study

As observed in the E_{extra} fit provided in Figure 6.34, there is a peaking structure within the D^* *Background* component. Before including this in the modelling, it was imperative that the source of this peak be properly understood. Before any investigations took place, it was presumed that the peaking structure was due to the π^0 from $D^{*-}(\to D^-\pi^0)\,\pi^+$ background events. As only the D^- is explicitly reconstructed in the control channel study, the remaining π^0 is not reconstructed and it manifests as extra energy in the ROE. Given the kinematics of the $D^{*-}(\to D^-\pi^0)\,\pi^+$ decay, a peak around 0.3 GeV is expected for the remaining π^0 candidates. The most straightforward way to verify this hypothesis was to reproduce the E_{extra} distribution for the D^* *Background* component with a veto in place for the ROE that removes all contributions from π^0 candidates originating from a D^{*-} particle. Should the hypothesis be correct, this version of the E_{extra} distribution would have no peak structure. Figure F.1 provides the distribution for E_{extra} recalculated with the π^0 veto in place. It can be seen that the distribution is falling and that the peaking structure around 0.3 GeV is no longer present.

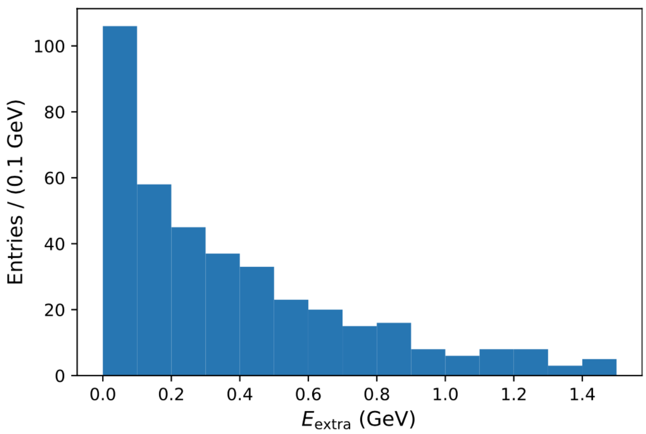


FIGURE F.1. Distributions of E_{extra} for the D^* *Background* component from the control channel study. This is calculated with the π^0 veto for the ROE which removes contributions from the π^0 candidates from $D^{*-}(\to D^-\pi^0)\,\pi^+$ background events.