Reconstruction Methods for Semi-leptonic Decays of B-mesons with the Belle II Experiment

Nadia Toutounji

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

Faculty of Science,
School of Physics,
University of Sydney

September 30$^{\text{th}}$, 2018

Abstract

The Belle II detector located in Tsukuba, Japan, building on the work of its predecessor Belle, is scheduled for long-term data collection from electron-positron e^+e^- collisions commencing in early 2019, for the purpose of studying rare B-meson decays in the search for new physics beyond the Standard Model. The Belle II Analysis Software Framework (BASF2) has been developed for physics analyses, with the Full Event Interpretation (FEI) being one such method designed for the reconstruction of B-meson decays from detector information. The FEI must be trained on simulated Monte Carlo (MC) data and introduces a signal-specific training process that can be tailored for a particular decay of interest in an attempt to increase the performance over signal-independent training processes such as those employed in Full Reconstruction (FR) methods at Belle. This study investigates the performance of the signal-specific and signal-independent FEI algorithms in the context of rare semi-leptonic B-meson decays, in comparison to leptonic decays, with the respective modes $B^+ \to \rho^0 \mu^+ \nu_\mu$ and $B^+ \to \tau^+ \nu_\tau$ chosen as working examples. The relative performance of the FEI methods implemented is evaluated via a number of key performance indicators including the reconstruction efficiency and purity of the reconstructed $\Upsilon(4S)$ event.

Acknowledgments

First and foremost, to my supervisor Professor Kevin Varvell, whose guidance and support was a constant presence throughout the course of this project. Thank you for the many hours spent reading, discussing and not to forget, debugging! To the Adelaide-Sydney Belle II gang - this project would not have been what it is without you. Thanks to Bruce, for the great discussions and even greater suggestions. Thank you to Frank and Chia-Ling for the fantastic advice, and to Sophie, who taught me a great many things for which I am extremely grateful. Thanks also to Lachlan, Bernie and Isabel, for the constant encouragement and moral support that always helped a great deal. And finally, thank you to my amazing family and friends, for being there for me in times of need providing a constant stream of love and support.

Statement of Student Contribution

All analysis work presented throughout this project was undertaken by the author using the Full Event Interpretation (FEI) multi-variate analysis technique developed by Dr. Thomas Keck, provided as part of the Belle II Analysis Software Framework (BASF2) and described in [1]. The various Python analysis scripts utilised within these studies took inspiration from or were adapted from example scripts available for use in BASF2, with extensive modifications implemented. As appropriately referenced within the text, the following contributions from external sources were incorporated into the research conducted:

- The generic FEI used for comparison with the specific FEI was trained by Dr. Thomas Keck, with the official training output obtained from the KEK computing system, KEKCC. The training of the various instances of the specific FEI described throughout this report were conducted by myself.

- The analysis recounted in [34] was used as inspiration for the selections chosen for the application of the FEI to $B^+_{sig} \to \tau^+\nu_\tau$, and the samples sizes used in the specific FEI training were decided upon in consideration with those utilised in [11].

- The description of the Belle II detector and associated figures provided in Chapter 2 were largely summarised from the detailed account presented by the Belle II Technical Design Report [22].

During the course of this project, I executed the generation and simulation of a large number of Monte Carlo samples using the third-party EvtGen [32] and Geant4 [33] packages as well as the training of multiple multi-variate algorithms. Additionally, all plots depicted in the following chapters were produced by me unless otherwise acknowledged. My supervisor, Prof. Kevin Varvell, read the draft chapters of this thesis in its entirety, with changes to content, style and language implemented based on the feedback received.

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

Table of Contents

Chapter 1

Introduction

The Standard Model of Particle Physics (SM), formulated through the collaborative efforts of many scientists and finalised during the 1970s, has achieved monumental success since its conception. Describing the existence of fundamental particles and the nature of their interactions, the model has been strengthened with numerous experimental observations and discoveries validating many of its features. Of these, some of the most notable achievements include the discovery of the top quark in 1995 [2], and the tau neutrino later in 2000 [3], the existence of both fundamental particles postulated years before their eventual discovery. The most recent success was the highly anticipated discovery of the Higgs boson in 2012 almost 50 years after it was first hypothesised [4], providing empirical evidence for the mechanism responsible for the origin of mass in fundamental particles.

The Standard Model, however, is unable to explain a number of known phenomena, and leaves several questions unanswered. It does not provide an explanation for the gravitational interaction, nor does it encompass dark matter and its physics. Additionally, cosmological models assume initial conditions of equal amounts of matter and anti-matter in the universe, but current observations reveal a considerable asymmetry, with an excess of matter forming the structures found in our universe today. By the conditions postulated in 1967 by Andrei Sakharov [5], such an asymmetry can only arise in the case of a violation of charge-parity (CP)-symmetry, which assumes that the laws of physics remain the same for a system that has undergone a conjugation of charge and inversion of spatial coordinates. The Standard Model does provide a mechanism for such a CP-violation [6], but this effect is far too small to explain the observed matter-antimatter asymmetry.

These mentioned points form only some of the multiple reasons why it is widely accepted that so-called physics Beyond the Standard Model (BSM) must exist, and an abundance of theories are under constant exploration and development. Many particle collider experiments are also actively involved in the global search for new physics, investigating particle decays for any significant deviations from SM predictions, guided by these theories. The Belle II Experiment located in Tsukuba, Japan, is one such collaboration, with the primary aim of using rare B-meson decays to search for new physics signatures [7]. It is the successor of the Belle Experiment, which was instrumental in discovering the existence of CP-violation in B-meson systems [8], together with the BaBar Experiment located in Stanford, California

[9]. Belle II aims to collect approximately 50 times more data than Belle with its long-term physics run scheduled for early 2019. It is expected that the substantial data-set will allow for rarer decay modes to be more thoroughly investigated than was possible before, with the hope that new physics effects will be able to be detected.

The focus of this study involves the testing and validation of one of the key methods used for B-meson reconstruction at Belle II, the Full Event Interpretation (FEI). At present, this technique is still relatively new and its performance has been investigated for very few B-meson decay modes of interest, including the rare leptonic decays $B^+ \to \tau^+\nu_\tau$ [10] and $B^+ \to \ell^+\nu_\ell\gamma$ [11]. The analysis presented throughout this report aims to validate the FEI when applied to the more physically complex semi-leptonic decays, with the rare decay $B^+ \to \rho^0\mu^+\nu_\mu$ taken as a working example.

The remainder of this chapter provides an overview of the Standard Model together with an introduction to B-meson decays. Chapters 2 and 3 describe the Belle II Experiment in detail, beginning with an in-depth summary of the particle accelerator and detector sub-components, and followed by an explanation of the software framework implemented in the experiment. The Full Event Interpretation software technique for B-meson reconstruction is then described in Chapter 4, with subsequent chapters reporting the results of its application. Chapter 5 first details the event selections found to be appropriate for the reduction of background in the reconstruction of $B^+ \to \rho^0\mu^+\nu_\mu$, which are then used within the application of the FEI given in Chapter 6. Finally, Chapter 7 reports the results obtained from the application of the FEI to $B^+ \to \tau^+\nu_\tau$, with a comparison of the FEI performance on the semi-leptonic and leptonic decay presented in the concluding Chapter 8.

A few points should first be noted before the commencement of the coming sections. All equations, variables and quantities shown throughout this report are expressed in natural units following the convention used in particle physics. That is, the speed of light c is set to unity, and all factors of c are henceforth not recorded. Additionally, as mentioned above, the Belle II Experiment is scheduled to begin long-term data collection in the early months of 2019. A preliminary five month long physics run was performed in mid-2018 with the omission of the vertex detector and corresponded to an integrated luminosity of 0.5 fb^{-1}, a sample too small for the adequate analysis of rare decays. As such, the validation studies reported in future chapters have been performed entirely on simulated data referred to as Monte Carlo (MC), and no real collision data has been employed within this analysis.

1.1 The Standard Model of Particle Physics

The Standard Model (SM) is a holistic theory that describes the fundamental particles existing in nature along with their interactions. These particles include both bosons, particles that have integer values of spin, and fermions, possessing half-integer spin and further classified into quarks and leptons due to different elementary properties. Figure 1.1 provides an overview of each particle incorporated within the SM.

In total, the Standard Model describes five bosons, four types of gauge bosons with spin 1, and the Higgs boson which possesses a spin of 0 [12]. The gauge bosons are the force-carrying particles responsible for the fundamental interactions between SM particles, specifically the electromagnetic, strong, and weak forces. The gravitational force is also considered to be fundamental in nature, though it is not incorporated into the SM and has no associated gauge boson within the model.

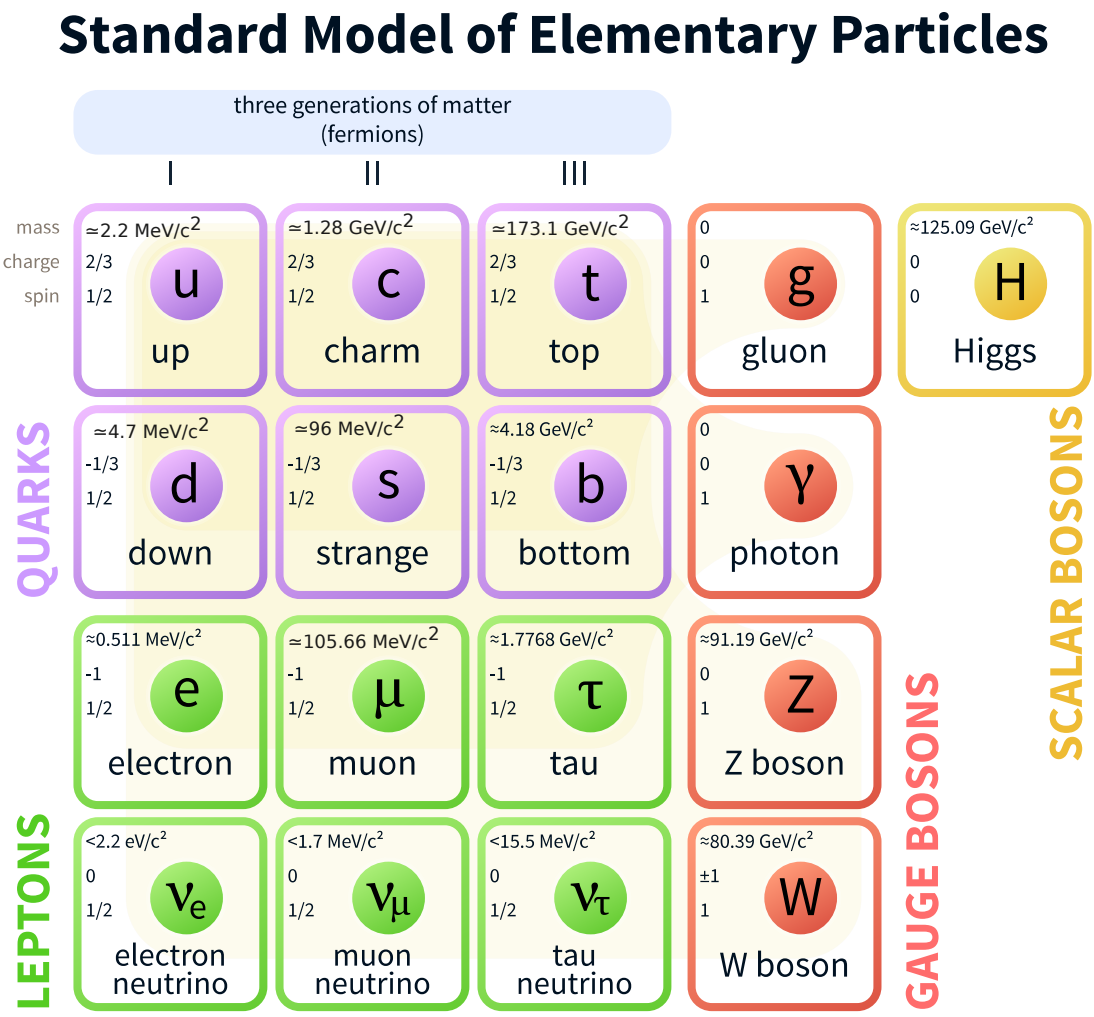


Figure 1.1: A depiction of the fundamental particles described by the Standard Model, with the mass, electric charge and spin of each particle also listed [13]. An anti-particle exists for all fermions shown (not pictured).

The photon γ is the mediator of the electromagnetic interaction which occurs between particles possessing non-zero electric charge. Gluons g are responsible for the strong interaction, experienced only by quarks and gluons themselves, due to their possession of an additional quantum property known as the colour charge. Analogous to the positive and negative representations of electric charge, three colours red, blue, and green are defined

such that zero net colour, or colourlessness, is achieved when all three colours are combined into a colour singlet. Each colour is also associated with an anti-colour, these being anti-red, anti-blue and anti-green, such that the combination of a corresponding colour and anti-colour also results in a net colour charge of zero. Eight effective gluon states exist, each containing a colour and anti-colour such that a net colour is achieved, and thus gluons can not only mediate the strong interaction, but participate in it themselves. In this way, the strong force differs significantly from the electromagnetic force, as photons are electrically neutral and do not participate in the electromagnetic interaction. The strong force was named for the fact that it is the strongest of the fundamental interactions, and it is the cause of the strong nuclear force which binds protons and neutrons together within the nuclei of atoms. The weak force, named similarly for its field strength, several orders of magnitude less than the strong and electromagnetic forces, is experienced by all fermions along with the Higgs boson. It is mediated by the electrically neutral Z-boson as well as the charged W-bosons, W^+ and W^-. The weak interaction is characterised by a number of unique properties, being the only interaction capable of changing the flavour of quarks as well as violating CP-symmetry, as expanded on in Section 1.1.1. The Z- and W-bosons are also the only massive gauge bosons, unlike the photon and gluon, which are massless. The last boson described by the SM is the Higgs boson H^0, represented as an excitation of a field termed the Higgs field which is responsible for the breaking of electroweak symmetry. In this way, the Higgs boson provides a mechanism by which all of the elementary particles (save for the photon and gluons) acquire mass.

The fermions of the SM are categorised into quarks and leptons, with each subsequently divided into three generations [12]. The six flavours of quarks are the up u and down d, charm c and strange s, and top t and bottom b quarks, each with a spin of $\frac{1}{2}$. Each quark also possesses both electric charge and colour, and they are the only fundamental particles that experience the electromagnetic, strong and weak forces. Additionally, for each quark there exists an equivalent anti-quark ($\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$, $\bar{b}$) with identical mass and spin, but opposite electric charge and anti-colour. Quarks and anti-quarks are arranged into composite particles called hadrons such that zero net colour is achieved, and these include both baryons and mesons. Baryons are formed by the combination of three quarks of different colours, and mesons from the combination of a quark and corresponding anti-quark.

Finally, the SM leptons similarly divided into three generations are the electron e^- and electron neutrino ν_e, muon μ^- and muon neutrino ν_μ, and the tau τ and tau neutrino ν_τ. The former lepton in each generation possesses electric charge, and experiences both the electromagnetic and weak forces. However, neutrinos are neutral particles, and therefore interact only via the weak force. Each lepton is also associated with a corresponding anti-particle, with these being the positron e^+, anti-muon or positive muon μ^+, anti-tau or positive tau τ^+, and the electron, muon and tau anti-neutrinos $\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$.

1.1.1 The Cabibbo-Kobayashi-Maskawa (CKM) Matrix

The Cabibbo-Kobayashi-Maskawa (CKM) Matrix is a 3x3 matrix, the elements of which are parameters put forward by the Standard Model and describe the strength of decays via the

weak interaction whereby a quark is changed from one flavour to another. When a quark emits or absorbs a charged $W^{\pm}$ boson, a flavour change must occur in order for electric charge to be conserved, as can be discerned from Figure 1.1 which includes the charges of each fundamental particle. It can additionally be deduced from this figure that the only quark flavour changes that do not violate charge conservation are those that involve the first quark of a generation transforming into the latter quark in either of the three generations, or vice versa. This ultimately gives rise to a 3x3 matrix, with elements as follows:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

Here, each element V corresponds to the strength of the weak decay with flavour transition given by the written subscript. For example, $|V_{ub}|$ characterises the $b \to u$ or $u \to b$ quark transition. The matrix must be unitary under normalisation in order to conserve probability, and verifying this unitarity has been the subject of substantial experimental efforts utilising various particle decays to determine the matrix elements with high precision. Only four parameters in total are required to specific the entirety of the CKM matrix. These include three real parameters known as the quark mixing angles as well as a single complex phase that provides the mechanism for CP-violation allowed for by the SM [14].

1.2 B-meson Decay

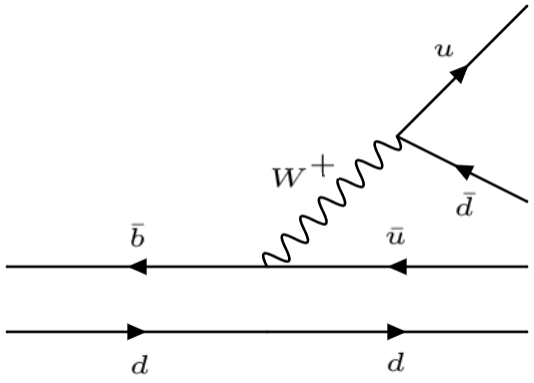
Of the existing mesons allowed for by the Standard Model, the B-meson is of particular importance, being the focus of many electron-positron collider experiments including Belle, Belle II and BaBar [15]. Both charged B-mesons, B^+ with anti-particle B^-, and neutral B-mesons, B^0 with associated anti-particle $\bar{B}^0$ exist, with the common feature between them being the presence of a bottom quark or anti-quark. Specifically, B^+ and B^0 mesons consist of a bottom anti-quark $\bar{b}$ paired with an up u or down d quark, respectively. Conversely, the anti-particle versions, B^- and $\bar{B}^0$ contain a bottom quark b together with an up $\bar{u}$ or down $\bar{d}$ anti-quark, respectively. Strange and charmed versions of these mesons also exist in which the up quark is replaced by either a strange s or charm c quark to form mesons with quark content B_s^0 ($s\bar{b}$), $\bar{B}_s^0$ ($\bar{s}b$), B_c^+ ($c\bar{b}$) and B_c^- ($\bar{c}b$).

B-mesons form the basis of electron-positron collider experiments, often referred to as B-factories, due to their quark content. The energies of these collisions can be tuned such that they result in the production of a meson known as the $\Upsilon(4S)$, a bound state of a bottom quark and anti-quark, $b\bar{b}$, which decays rapidly to a pair of charged B^+B^- or neutral $B^0\bar{B}^0$ mesons approximately 96% of the time [12]. The details of this process are left to Chapter 2, in which the Belle II experiment is described at length. Though each electron-positron collider experiment was unique in its goals and implementation, the key driving force behind both Belle and BaBar was the potential to discover the existence of CP-violation in the neutral B-meson system, a phenomenon already observed in neutral kaons in 1964 [16]. This aim was realised independently by both experiments in 2001 [8, 9], with empirical evidence

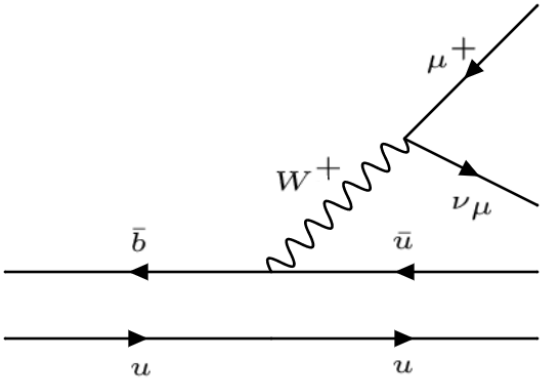
found for CP-violation in neutral B-mesons. As mentioned at the opening of the chapter, the Belle II Experiment has expanded this vision, with the larger expected data-set allowing for the more precise study of rare B-meson decays in which new physics signatures may be found. B-mesons decay in hundreds of allowed modes [12], each with an associated branching fraction, defined as the fraction of the total number of decays belonging to a particular decay mode. These decays can be divided into three categories, as described in the following sections.

1.2.1 Hadronic Modes

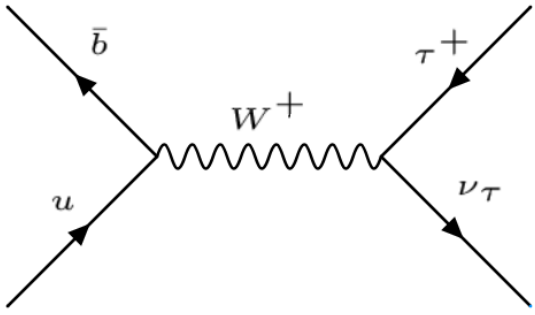
Hadronic decays of B-mesons are those in which the decay products are purely hadronic, or in other words, only bound states of quarks are formed, with no leptonic daughters present. An example of this includes the decay $B^0 \to \pi^+\pi^-$, a decay of a neutral B-meson into two charged π-meson daughters, π^+ and π^-, with quark content $u\bar{d}$ and $\bar{u}d$ respectively. A Feynman diagram depicting this decay is shown in Figure 1.2(a).



(a) $B^0 \to \pi^+\pi^-$



(b) $B^+ \to \rho^0\mu^+\nu_\mu$



(c) $B^+ \to \tau^+\nu_\tau$

Figure 1.2: Feynman diagrams illustrating example hadronic (a), semi-leptonic (b), and leptonic (c) B-meson decays.

1.2.2 Semi-leptonic Modes

In contrast, semi-leptonic B-meson decays include both hadronic and leptonic decay products. An example of a semi-leptonic mode includes the aforementioned rare decay $B^+ \to \rho^0 \mu^+ \nu_\mu$, which is studied extensively throughout the later chapters of this thesis. Here, a charged B-meson decays to a neutral ρ^0 meson with quark content $u\bar{u}$ or $d\bar{d}$ [1], along with a muon μ^+ and its associated muon neutrino ν_μ. This involves the $b \to u$ quark transition characterising the decay as rare, with any B-meson decay proceeding via other than the $b \to c$ transition given this classification [17]. The associated branching fraction obtained from the Particle Data Group is $1.58 \pm 0.11 \times 10^{-4}$ [12], and the decay is illustrated by means of the Feynman diagram in Figure 1.2(b). The semi-leptonic B-meson decay to a ρ^0 meson can similarly occur via the electron e^+ and τ^+ channels, represented by the general denotation $B^+ \to \rho^0 \ell^+ \nu_\ell$, where $\ell = e$, μ, or τ.

The ρ^0 meson resulting from this decay is a similarly short-lived particle, decaying to a pair of charged pions via the strong interaction $\approx 100\%$ of the time [12]. Due to this property, the ρ^0 is referred to as a resonance, with an invariant mass distribution calculated from the daughter pion 4-momentum vectors that is broadened around the ρ^0 rest mass with a width determined from the inverse of its lifetime [18]. This complexity buried within the ρ^0 decay corresponds to heightened backgrounds associated with its reconstruction, whereby particles can be misidentified as ρ^0 mesons due to the broadened mass distribution. The $B^+ \to \rho^0 \ell^+ \nu_\ell$ decay, together with other semi-leptonic B-meson decays involving the $b \to u$ quark transition including $B^+ \to \pi^0 \ell^+ \nu_\ell$, is of particular relevance to the testing of the unitarity of the CKM matrix via precision calculations of the element $|V_{ub}|$.

1.2.3 Leptonic Modes

Finally, B-mesons can also decay purely leptonically, with no quark content present at all amongst the B-meson daughters. Another pertinent example revisited in later chapters is the rare decay of a charged B-meson to a tau lepton τ^+ and corresponding tau neutrino ν_τ, $B^+ \to \tau^+ \nu_\tau$, depicted in Figure 1.2(c). The resultant τ meson also undergoes subsequent decay, with $\approx 85\%$ of these cases involving decay products consisting of no more than one charged particle, referred to collectively as 1-prong τ decays [12]. The $B^+ \to \tau^+ \nu_\tau$ studies presented in Chapter 7 restricted the decay of the τ to one such 1-prong decay, $\tau^+ \to \mu^+ \nu_\mu \bar{\nu_\tau}$, with a respective branching fraction of 17.39 ± 0.04 % [12]. Alternate τ decay modes including 3- and 5-prong processes also exist, with the relative branching fractions decreasing with decays to a greater number of charged daughters.

Purely leptonic B-meson decays, most notably $B^+ \to \tau^+ \nu_\tau$ due to its larger branching fraction compared with the electron and muon channels [12], are considered particular golden modes for the potential discovery of new physics. The Standard Model describes a single Higgs doublet, whereby the Higgs field corresponds to two electrically charged and two electrically neutral components. With the breaking of electroweak symmetry, the two charged

[1] The quark content of a ρ^0 meson is more accurately represented as $\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$, which portrays it as a superposition of two allowed quantum states.

components and a single neutral component are absorbed by the weak W- and Z-bosons respectively through the Higgs mechanism, making them massive, and the remaining neutral component is manifested as the single neutral Higgs boson. One proposed extension to the Standard Model is the Two Higgs Doublet Model (2HDM), which introduces an additional Higgs doublet such that symmetry breaking results in the predicted existence of three neutral (h, H, and A) and two charged (H^+ and H^-) Higgs bosons [19]. In particular, the Type II 2HDM restricts the coupling of one Higgs doublet to up-type (electric charge $q = \frac{2}{3}$) quarks, and the other to down-type ($q = -\frac{1}{3}$) quarks and charged leptons. In such a model, a charged Higgs $H^{\pm}$ could replace the $W^{\pm}$ boson mediating the $B^+ \to \tau^+\nu_\tau$ decay in Figure 1.2(c), with the amplitudes of both processess interfering in such a way as to alter the measured branching fraction from SM predictions [10]. Precision measurements of the $B^+ \to \tau^+\nu_\tau$ branching fraction with the substantial expected data-set of Belle II will thus provide direct testing of the Type II 2HDM extension to the SM.

Chapter 2

The Belle II Experiment

The High Energy Accelerator Research Organisation, abbreviated as KEK, was established in 1997 on the site of a particle physics lab located in Tsukuba, part of the Ibaraki prefecture in Japan, with the purpose of providing particle accelerators for study in fields such as high energy physics. Since its establishment, the highly successful Belle Experiment was conducted, in which an asymmetric electron-positron collider KEKB was operated during the years 1998 - 2010, and the Belle detector constructed to detect the products of these collisions. The experiment ceased data-taking in 2010 to make way for major upgrades to both the accelerator, now known as SuperKEKB, and the detector, now referred to as Belle II. The preliminary physics run of the Belle II experiment without the inner detector was completed mid-2018 after approximately five months of operation, with preparations ongoing for the long-term physics run scheduled for early 2019.

2.1 The Belle Experiment

The KEKB accelerator used in the Belle experiment consisted of two nearly circularly rings of circumference 3 km, one used to accelerate electrons to energies of 8.0 GeV, and the other responsible for accelerating positrons to energies of 3.5 GeV [20]. The beams would then collide in the center of the Belle detector, with decay products propagating outwards throughout the various detector layers. The beam energies were designed such that at the time of collision, they would have a combined energy equivalent to the rest mass of an $\Upsilon(4S)$ meson, 10.58 GeV, when calculated in its centre-of-mass frame. That is, these electron-positron collisions would result in the production of an $\Upsilon(4S)$, a meson consisting of a bottom b quark and $\bar{b}$ antiquark, which would be produced approximately at rest in its centre of mass frame, but boosted along the electron flight direction as seen in the laboratory frame of reference. The $\Upsilon(4S)$ decays almost instantaneously to a pair of B-mesons 96% of the time, with $\sim 51\%$ of these being decays to charged B-mesons, B^+B^-, and $\sim 49\%$ to neutral B-mesons, $B^0\bar{B}^0$ [12]. As the B-mesons have a very short lifetime of the order of 10^{-12} s [12], the observed boost in the laboratory frame was extremely useful in separating the decay vertices of the two decaying B-mesons, which recoiled against each other [15]. Despite the KEKB design energies of the electron and positron beams, collisions did not always result in the produc-

tion of the $\Upsilon(4S)$ $b\bar{b}$ bound state, and events in which $e^+e^- \to c\bar{c}$, $u\bar{u}$, $d\bar{d}$ or $s\bar{s}$ were also frequently occurring. These events, termed the continuum background, constituted a significant proportion of the backgrounds associated with the B-meson studies performed at Belle.

The primary purpose of the Belle experiment was to search for CP-violation in neutral B-meson systems, after it was first experimentally observed in neutral kaons in 1964 [16]. This goal was realised in 2001 when both the Belle and BaBar collaborations independently reported results indicating the existence of CP-violation in neutral B-mesons [8, 9]. Alongside the success of this discovery, the data collected by Belle during its operation has been extremely valuable in updating various Standard Model parameters including CKM-matrix elements, together with the branching fractions of a large number of B-meson decay modes. Many B-meson decays were also observed for the first time in Belle data, including the leptonic decay $B^+ \to \tau^+\nu_\tau$ introduced in Section 1.2.3 [21].

2.2 SuperKEKB

In a particle accelerator experiment, the luminosity L can be defined as the number of collision events N per unit time within a certain interaction cross-section σ:

$$L = \frac{1}{\sigma}\frac{dN}{dt}$$

In order to achieve the desired 40-fold increase in the luminosity for the Belle II Experiment, the KEKB accelerator has had to undergo necessary upgrades to form the accelerator now referred to as SuperKEKB. One of the most significant changes to the technical design is the implementation of the "Nano-Beam" scheme, which aims to reduce the overlap of the electron and positron beams at their interaction point [22]. Figure 2.1 is a representation of the beam collisions, with d representing the size of the region of overlap. Minimising the value of d results in the squeezing of the vertical beta functions (β^*_y) of each beam at the interaction point, which has a significant effect on the resultant luminosity L given the following relationship,

$$L = \frac{\gamma_\pm}{2er_e}\left(\frac{I_\pm \xi_{y\pm}}{\beta^*_{y\pm}}\right)\left(\frac{R_L}{R_{\xi y}}\right)$$

Here, the $\pm$ suffix refers to each beam, electron (-) and positron (+), and $\gamma_\pm$ is the Lorentz factor, e the elementary electric charge, and r_e the classical electron radius. The luminosity depends also on the beam currents $I_\pm$ and the vertical beam-beam parameters $\xi_{y\pm}$. Finally, R_L and $R_{\xi y}$ refer to reduction factors for the luminosity and vertical beam-beam parameters respectively, and these are determined using quantities including the crossing angle ϕ, with their ratio varying only slightly from unity. The Nano-Beam scheme boasts a reduction in the vertical beta functions (β^*_y) by a factor of 20 when compared to those present at KEKB. Combined with a doubling of the beam currents $I_\pm$, and assuming a similar vertical beam-beam parameter $\xi_{y\pm}$, the design luminosity of the SuperKEKB accelerator is $8\times10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, 40 times larger than the luminosity achieved by KEKB [22]. A comparison of the machine

parameters achieved by Belle and the design values for Belle II is presented in Figure 2.2.

However, such a substantial gain in the luminosity of SuperKEKB comes at the expense of a significant increase in the number of background processes arising from interactions within the individual beams and between the beams and foreign particles such as the surrounding residual gas in the evacuated beam pipe. These effects are collectively referred to as beam background, and encompass the following processes [22]:

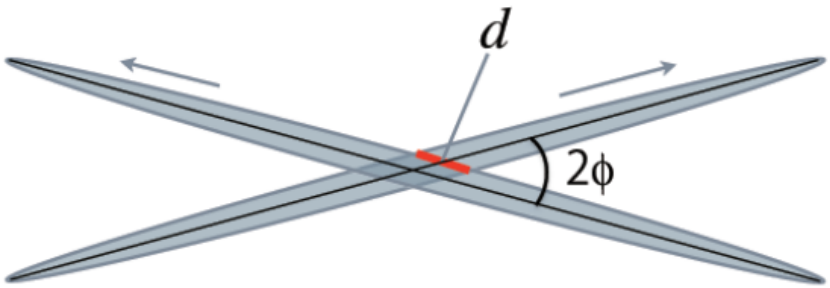


Figure 2.1: A depiction of the interaction between an electron and positron beam at SuperKEKB. The grey areas represent the bunches of colliding particles, with the size of the beam overlap d and the crossing angle ϕ minimised with the Nano-Beam design [22].

- Synchrotron radiation: The electrons and positrons moving at relativistic speeds within the beams emit photons as they are accelerated radially. These photons can produce false hits in the detector and contribute quite extensively to the observed backgrounds.
- Beam-gas scattering: Whilst the beams are contained within vacuum chambers, inevitably a small amount of gas particles remain with which the electrons and positrons can interact. These interactions include both Coulomb scattering, referring to elastic collisions that occur between charged particles, and bremsstrahlung scattering, whereby one of the electrons or positrons loses kinetic energy via the emission of a photon after hitting a gas particle. In each case, when a particle within a beam is scattered, its momentum is altered leading it to collide with the chamber walls or surrounding magnets, producing a shower of particles that contribute to the beam background.
- Touschek scattering: Particles within the beams can also scatter off one another, modifying their momentum and leading to shower particles when the electrons or positrons hit the surrounding chamber walls. The rate of Touschek scattering is inversely proportional to the beam size, and thus the background effects from Touschek scattering are expected to be quite severe with the greatly reduced beam sizes for Belle II.
- Radiative Bhabha scattering: The collision of the electron and positron beams does not always result in their annihilation, but rather the electrons and positrons can also

scatter off of one another in a process known as Bhabha scattering. Bremsstrahlung emission can also occur whereby one of the scattered particles releases a photon, a phenomenon known as radiative Bhabha, with $e^+e^- \rightarrow e^+e^-\gamma$. The resultant photons can travel along the beam-pipe and interact with the iron detector magnets, producing significant amounts of background neutrons. Being neutral hadrons, these neutrons form a major background for the detection of K_L^0 mesons in the K-Long and Muon (KLM) Belle II sub-detector, introduced in Section 2.3.6.

- Pair production: Via the electromagnetic interaction, an electron-positron pair can be created from a single photon in a process known as pair production. Similarly, the collision of an electron and positron can result in their annihilation to a pair of photons, as per the same fundamental interaction. The combination of these two effects forms an additional beam background termed pair production via the two-photon process, whereby colliding electrons and positrons annihilate to form two photons, both of which produce an electron and positron pair, $e^+e^- \rightarrow e^+e^-e^+e^-$. These background electrons and positrons enter the detector and contribute particularly to the background levels of the first sub-detector component, the Pixel Detector (PXD).

An additional change to the accelerator design for the Belle II Experiment that should be noted is a shift in the energies of the electron and positron beams. These have been changed from the Belle values, 8.0 GeV and 3.5 GeV, to 7.0 GeV and 4.0 GeV, for the electron and positron beams respectively. The new energy scheme still results in a centre-of-mass energy at the $\Upsilon(4S)$ resonance, but with a decreased boost in the laboratory frame along the electron flight direction. The separation between the decay vertices of both B-mesons is thus also reduced, but this effect is largely counteracted by the smaller beam sizes and consequently the narrower beam-pipe [22].

	KEKB Achieved	SuperKEKB
Energy (GeV) (LER/HER)	3.5/8.0	4.0/7.0
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
Luminosity ($10^{34}\text{cm}^{-2}\text{s}^{-1}$)	2.11	80

Figure 2.2: A comparison of the key machine parameters achieved by the KEKB accelerator and predicted for the SuperKEKB accelerator. Where relevant, values for the positron and electron beams are quoted as e^+/e^- [22].

2.3 The Belle II Detector

The Belle II detector consists of a number of individual sub-detector layers, each with defined roles relating to the measurement of kinematic quantities and particle detection, identification and discrimination. The following sections present an overview of the structure

and functionality of each detector component, beginning with those closest to the electron-positron beam line and continuing through to the outer regions of the detector. A schematic of the Belle II detector illustrating the individual layers is displayed in Figure 2.3.

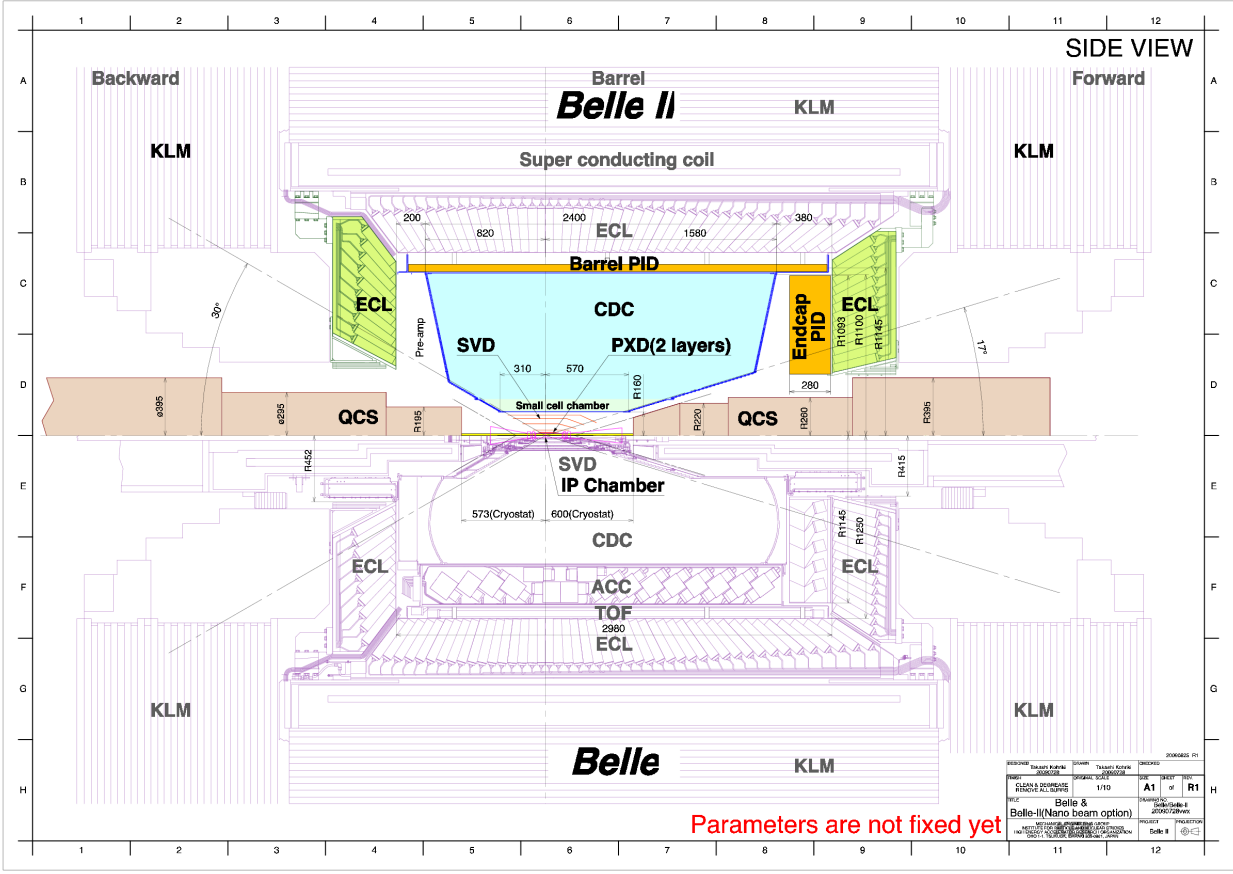


Figure 2.3: A side view of the Belle II detector (top), compared with its predecessor Belle (bottom). The various sub-detectors are labeled [23].

2.3.1 Pixel Detector (PXD)

The Pixel Detector is the innermost component situated close to the beam-pipe housing the incoming electron and positron beams. Due to the Nano-Beam scheme chosen for Belle II, the radius of the beam-pipe in the interaction region of the beams is only approximately 10 mm, with the pixel detector consisting of two layers of sensors positioned at radii of 14 mm and 22 mm [22]. The PXD together with the Silicon Vertex Detector (SVD) and Central Drift Chamber (CDC) introduced in the following sections form the region of the detector responsible for tracking the trajectories of charged particles as they curve within the supplied solenoidal 1.5 T magnetic field, allowing for the determination of their momentum. Another main goal of these detector regions is to use this information for the precise determination of the decay vertices of particles, including those of the parent B-mesons.

The sensors used within the PXD are comprised of DEPleted Field Effect Transistor (DEPFET) pixels, with 8 million pixels in total divided between the two layers. The DEPFET technology is semi-conductor based, with hits recorded by collecting electrons that have been freed by the incidence of charged particles colliding with the transistors. The nature of the DEPFET structure is such that the sensors consume very little power, requiring only air cooling, and can be built to be quite thick, with a thickness of only 50 microns.

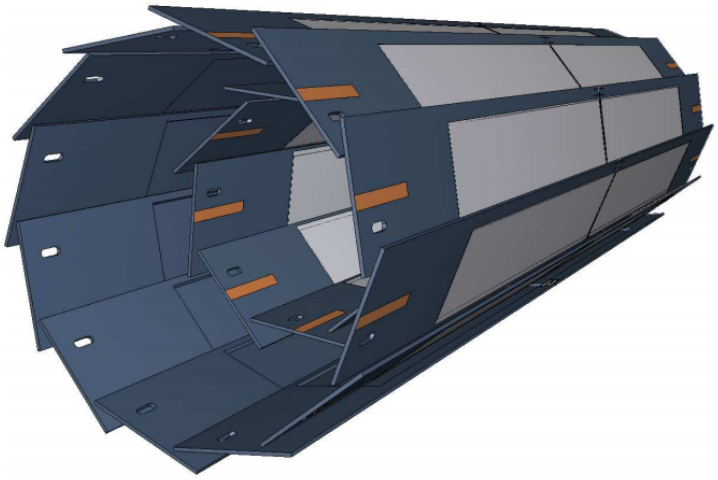


Figure 2.4: A schematic of the PXD as installed within Belle II, corresponding to a total length of 174 mm and diameters of 22 mm and 44 mm for the inner and outer pixel layers, respectively [22].

The PXD is a new addition to the Belle II detector, with Belle containing only the SVD for particle decay vertexing. Its introduction was necessary to contend with the substantial levels of beam background expected with the higher luminosity SuperKEKB, as the pixels provide significantly more channels to accept hits from beam background particles produced within the beam-pipe. The introduction of the PXD also allowed for the full vertex detector to be positioned closer to the beam-pipe than was possible in Belle, resulting in better vertex resolution.

2.3.2 Silicon Vertex Detector (SVD)

The Silicon Vertex Detector, as mentioned above, is an additional detector component designed for charged particle tracking and decay vertexing. It consists of four layers of strip sensors made of silicon, with the innermost layer positioned 16 mm from the outer layer of the PXD [22]. Given the high luminosity and expected levels of beam background, this is the safest distance to introduce strip sensors. It is not feasible to use pixel layers throughout the entire inner region as both the readout from the large number of channels and the estimated cost would be too large. Complementing the PXD, the SVD also allows for the measurement

of decay vertices displaced further from the interaction point.

The sensors used in each layer are double-sided silicon strips (DSSDs), with n-doped silicon at one side and p-doped silicon at the other, with electrons liberated by charged particle hits traversing the strips towards the n-doped side.

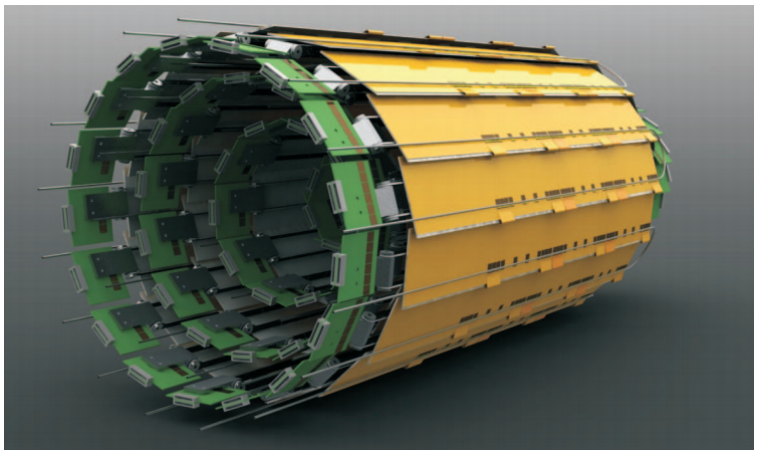


Figure 2.5: The Belle II SVD, containing four layers of silicon strip sensors with a distance of 20.4 cm separating the inner and outermost layers [22].

2.3.3 Central Drift Chamber (CDC)

The Central Drift Chamber, located at a radial distance of 160 mm, 20 mm out from the fourth layer of the SVD, is a large chamber designed for a number of key purposes. It reinforces the tracking capabilities of the PXD and SVD and forms a crucial component in the reconstruction of charged particle tracks and the precise determination of track momenta. However, it holds the additional benefit of also being able to provide information on the identity of the charged tracks passing through the chamber. The CDC extends outwards to a radial distance of 1130 mm around the beam-pipe, making it the largest of the inner detector layers responsible for charged particle tracking [22].

The Belle II CDC is modeled closely on the CDC incorporated within Belle, which was highly successful in its performance. It consists of a total of 14336 drift cells arranged in layers, an improvement on the 8400 employed at Belle [20], with each drift cell filled with a mixture of helium and ethane gases in equal proportions. Charged tracks ionise the gases as they pass through the drift cells, releasing electrons that then drift towards sensor wires that register and amplify the signals. The measurement of the drift time of the electrons together with the location of the sensor wire hits can be used for the reconstruction of charged particle tracks. Additionally, the level of kinetic energy lost by each track as it ionises the gas within a cell can be measured. It is here that the CDC is able to contribute to particle identification, with different charged particles possessing characteristic signatures in their distributions of $\frac{dE}{dx}$, the energy loss as a function of distance. In this way, tracks

with particularly low momentum that are unable to reach the additional detector components designed for particle identification (described in the upcoming section) can often be identified using the CDC alone. The arrangement of wires within the CDC is illustrated in Figure 2.6, with a higher density observed in the inner layers to accommodate the increased levels of beam background.

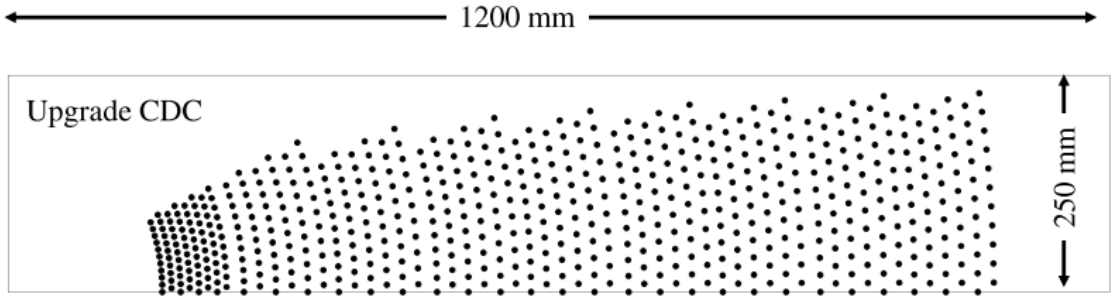


Figure 2.6: The sensor wire configuration of the Belle II CDC [22].

2.3.4 Particle Identification Components

Whilst the CDC proves useful in determining the identity of charged tracks, two independent particle identification systems are also incorporated into the Belle II detector just beyond the CDC, and boast a superior performance. These are the Time-Of-Propagation (TOP) counter and the Aerogel Ring-Imaging Cherenkov detector (ARICH), each located at designated regions for sufficient coverage of the entirety of the detector.

The beam-pipe housing the two incoming electron and positron beams must pass through the detector, and thus when referring to Belle II, it is common to divide the detector into two sections, namely the barrel region and the endcaps. The barrel refers to the region of the detector located at radial distances from the beam-pipe, with the two endcaps referring to the left- and right-most detector walls surrounding the beam-pipe at each end. Additionally, due to the beam asymmetry and the consequent lab-frame boost, the two endcaps are termed forward and backward, with the forward endcap defined as the one facing the incoming electron beam and resultant boost. The structure of the barrel and endcap regions can be clearly distinguished in Figure 2.3.

The TOP counter lies in the barrel region of the detector, and consists of 16 2.7 m long modules distributed azimuthally around the beam-line [24]. Each module is comprised of a length of quartz acting as a radiator, with a spherical mirror attached to one end, and a prism wedge at the other. When a charged particle passes through the quartz bars with a velocity higher than the associated speed of light for the medium, Cherenkov radiation is produced, with the resulting photons traversing the length of the bar via total internal reflection, arriving at a series of photo-multiplier tubes attached to the prism. A 3-dimensional Cherenkov image can then be simulated using the spatial coordinates and time of arrival of the Cherenkov photons, and charged particles can be discriminated based on these generated

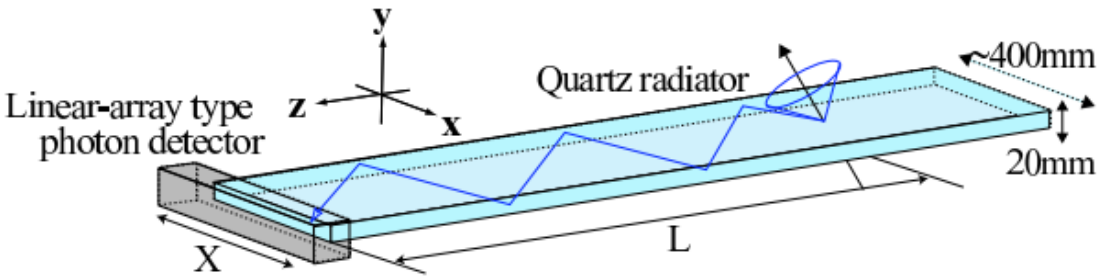


Figure 2.7: An illustration of the Belle II TOP counter depicting the total internal reflection of a Cherenkov photon through the quartz radiator [22].

images. For Belle II, the TOP counter has replaced the equivalent Belle detector component, the Time-Of-Flight (TOF) detector, with the TOP designed particularly to improve the kaon-pion (K/π) separation capabilities.

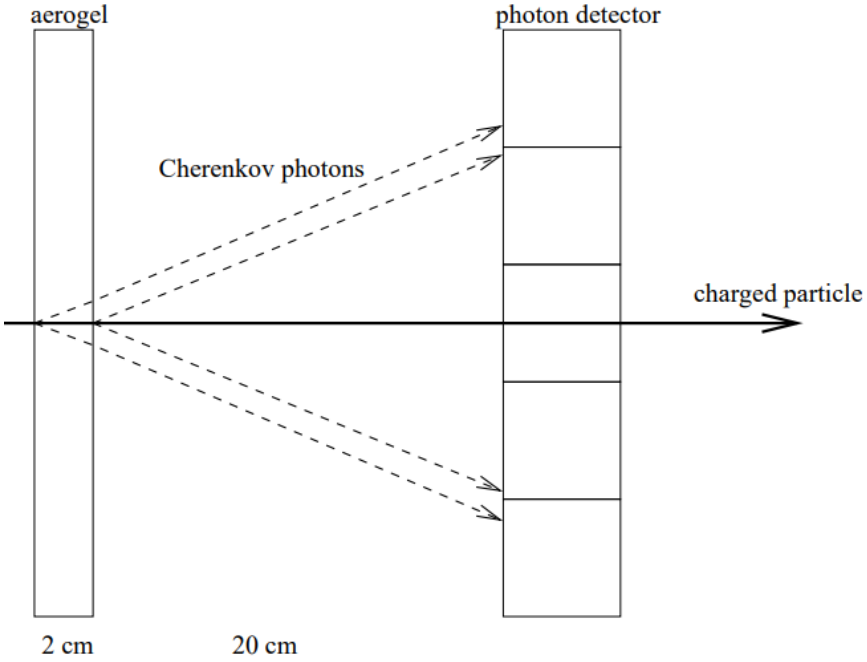


Figure 2.8: A representation of the Cherenkov ring-imaging technique employed by the ARICH detector [22].

Conversely, the ARICH detector is located at the forward endcap, and whilst differing in structure to the TOP counter, it also relies on Cherenkov radiation imaging for charged particle identification. It spans a 3.5 m^2 area of the endcap in two layers, with a radiator layer of aerogel tiles separated from a plane of photon detectors by a distance of 200 mm [25]. Planar mirrors surround the edges of the entire system to redirect the Cherenkov photons produced in the aerogel to the photon detector region. The large volume between the two layers allows the trajectories of the emerging photons to diverge enough that the incident

photons form a ring on the detector plane, resulting in the Cherenkov image used for particle discrimination. The ARICH has similarly replaced the Aerogel Cherenkov Counter (ACC) configuration used at Belle, with improved kaon-pion separation over a large momentum range, and reasonable separation between pions, muons and electrons at momenta under 1 GeV [22].

2.3.5 Electromagnetic Calorimeter (ECL)

The Electromagnetic Calorimeter, extending over both the barrel and endcap regions of the detector, encases all of the afore-mentioned components and forms the second-to-last detector layer of Belle II. It is responsible for a number of key tasks including the detection of photons which, being neutral particles, are not detected by any of the former components. The ECL not only detects these photons, but is able to determine their energies and angular coordinates with good precision. Another primary role of the ECL is in determining the energy of charged tracks, a property ultimately used for the identification of electrons. The ECL also contributes to the detection of K_L^0 mesons alongside the KLM introduced in the following section.

As for the structure of the ECL, a total of 8736 scintillator crystals cover a polar angle range of $\sim 12^\circ < \theta < 155^\circ$, with 6624 and 2112 divided between the barrel and endcaps, respectively [22]. When ionising particles enter the scintillator crystals, their energy is absorbed and re-emitted in the form of photons, which are then detected via photo-diodes glued to each crystal. The same crystals have been inherited for the Belle II ECL, due to only minimal damage received during the run-time of KEKB. At present, the crystals are made of Thallium-activated Cesium Iodide, CsI(Tl), a scintillating material with a considerably large light output. However, discussion and research are currently ongoing to upgrade the crystals to pure Cesium Iodide (CsI), at least in the end-cap region [22]. Although these produce a considerably lower light output, the decay time between the energy deposits and the emittance of photons is substantially shorter. The ECL electronics have also been significantly upgraded since Belle, with major improvements to the readout time achieved.

2.3.6 K-Long (K_L^0) and Muon (μ) Detector (KLM)

Finally, the K_L^0 and μ detector, the outermost layer of the detector similarly sectioned into the barrel and endcap regions, is designed for the detection and identification of muons and neutral K_L^0 mesons. It is located outside of the super-conducting solenoid used to provide the 1.5 T magnetic field to the detector, which occupies the space between the ECL and the KLM, as can be seen in Figure 2.3. The angular coverage of the KLM is comparable to that of the ECL with a $20^\circ < \theta < 155^\circ$ range in polar angle, and polyethylene sheets shield the KLM from multiple backgrounds including neutrons produced from beam background processes. Resembling closely the successful Belle KLM, detector elements known as resistive plate chambers (RPCs) are wedged between 4.7 cm thick iron plates in the barrel. When hadrons deposit energy into the ECL or collide with the iron plates, they interact

strongly such that a characteristic shower of secondary hadrons is produced. These traverse the iron plates and are detected by the RPCs via the charged hadrons present in the shower. Each RPC consists of two parallel electrode sheets made from an insulating substance known as float glass across which a high voltage is applied. Charged particles can ionise the gas between the electrodes, with electrons and ions accelerated to the anode and cathode, respectively. At the end-caps, however, RPCs are no longer optimal when contending with the higher luminosity beam background effects, and polystyrene scintillator strips are used in lieu of the RPCs within both the forward and backward endcaps [26].

Much like photons, neutral hadrons including K_L^0 mesons cannot be detected by the charged particle tracking sub-detectors, and are only able to deposit energy within the ECL and/or the KLM. Given this, clusters registered only in the KLM, or aligned clusters in both the ECL and KLM that cannot be matched to any charged particle track, strongly indicate the presence of a neutral hadron. Due to the neutron shielding, such cases are interpreted as arising from K_L^0 mesons, with neutrons contributing to the observed background. In general, the energies characteristic of K_L^0 mesons are quite low, resulting in a poor energy resolution for K_L^0 mesons detected in the KLM due to statistical fluctuations.

The nature of the momentum of the muons produced at Belle and Belle II is such that the muons are most often low ionising, and can traverse the ECL and even the iron plates with minimal energy loss. As such, muons produce non-shower hits in the KLM that can be matched to tracks in the CDC [27]. This provides a useful feature for the discrimination between muons and other charged hadrons, particularly pions, that do generate hadronic showers within the KLM.

2.3.7 Triggering and Data Acquisition

As explained, the sub-detectors described above are designed for particle tracking and identification based on certain signatures or hits occurring within the component geometry. This functionality is achieved through an efficient data acquisition system including a number of triggers associated with certain criteria by which the various physics processes occurring within the Belle II environment, including beam background, continuum and $\Upsilon(4S)$ events can be detected. The triggering system developed for Belle II was similarly modeled after the Belle system with advancements in the technology used, and consists of five sub-triggers activated by hits in the various detector components. These sub-trigger signals are then passed to the Global Decision Logic (GDL), which combines the information received from each sub-trigger system to make a final decision regarding the nature of the detected event, as per a number of secondary criteria [28]. This triggering system is depicted in Figure 2.9, with the sub-triggers shown associated with the CDC tracking, ECL and KLM clustering and the particle identification capabilities of the ECL (EPID) and TOP and ARICH detectors located in the barrel (BPID).

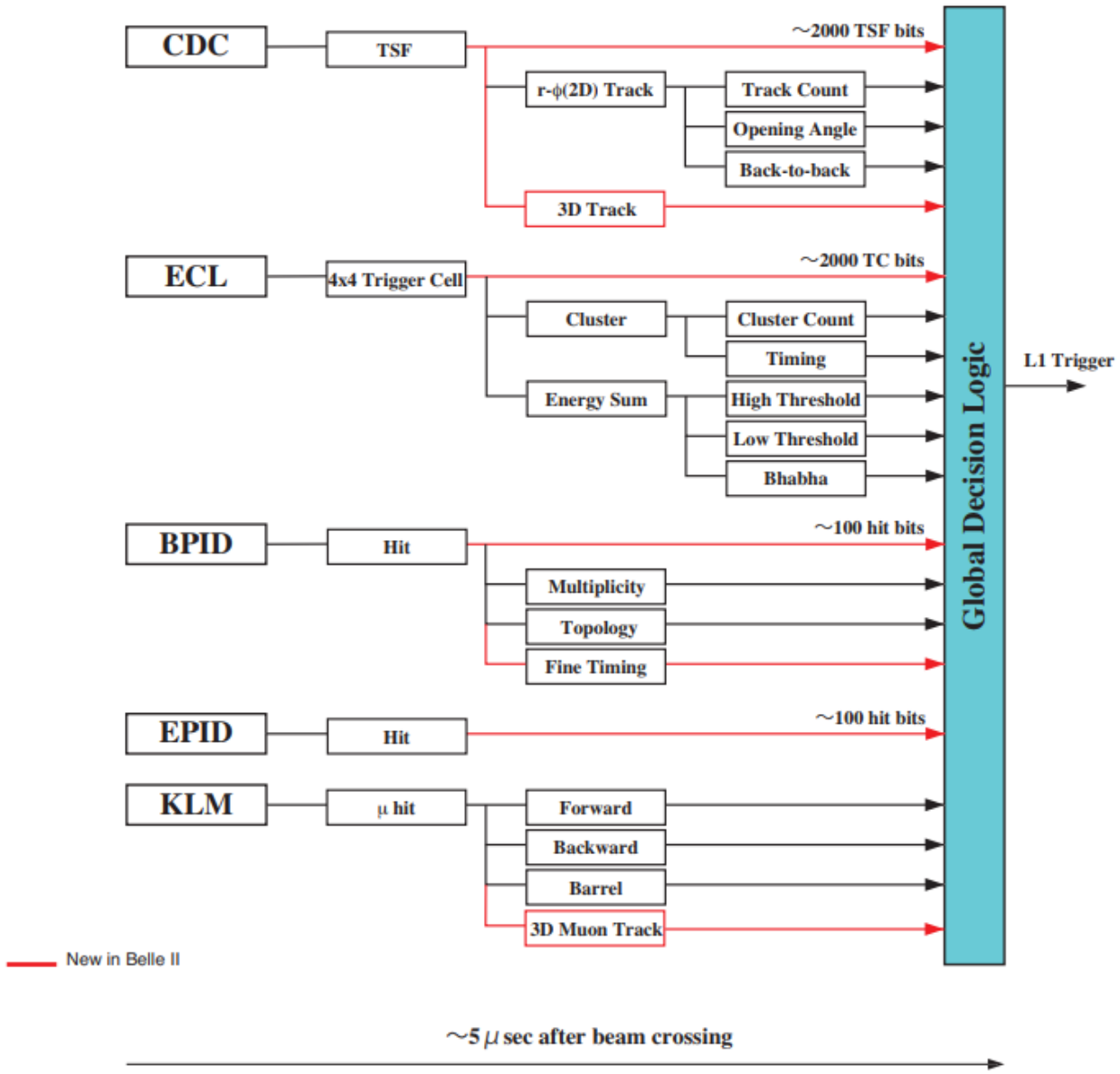


Figure 2.9: The structure of the Belle II triggering system, with the quoted information from each sub-trigger passed to the Global Decision Logic (GDL) [22].

Chapter 3

The Belle II Analysis Software Framework (BASF2)

The Belle II Analysis Software Framework, abbreviated to BASF2, is the term used to refer to the entire suite of software utilised throughout the Belle II experiment. This extends not only to software used for physics analysis, but software related to the simulation and monitoring of the detector. The framework is primarily programmed using C++, but Python is the main language chosen for writing scripts that get executed by the framework [29]. This is a significant shift from the Belle Analysis Software Framework (BASF) used in the Belle Experiment, which was programmed using a mixture of C, C++ and Fortran, with C++ being the primary language used for writing analysis scripts [30]. BASF2 also incorporates multiple independent libraries into the framework, with the most notable one being ROOT [31], a scientific software framework used for physics analysis developed at the European Organization for Nuclear Research, CERN, in Geneva. Other relevant external packages include EvtGen [32], used for the simulation of particle decays, particularly those of B and D-mesons, and Geant4 [33], responsible for simulating the interaction and consequent detection of particles by the various sub-detectors.

3.1 Framework Structure

3.1.1 Overview

The processing of data within BASF2 relies on the use of modules, which are linearly arranged in what is termed a path [29]. The modules are self-contained processing blocks responsible for particular tasks which span a vast range of applications, such as reading information from and writing information to files, simulating particle decay and detection, and multivariate classification, which is discussed in detail in Section 3.3. When processing a path, the individual modules are read and implemented in order.

Upon executing a path with multiple modules, data often needs to be shared between them. A primary example would be an analysis script that generates a number of $\Upsilon(4S)$ particles, and simulates their decay into subsequent particles which propagate through and

interact with the detector. Such a script is known as a steering file, written in Python, and first creates and then executes a path containing the relevant modules. In this example, data including the list of particles involved in the $\Upsilon(4S)$ decays, ECL and KLM cluster information, and particle track information needs to be accessed by multiple modules. This is implemented within BASF2 using the DataStore, which is a common storage available within the framework. Relevant data is stored in the DataStore as data-objects, with the aforementioned examples being `ParticleLists`, `ECLClusters`, `KLMClusters` and `Tracks`. All modules are able to read from and write to the DataStore, forming an efficient method of data processing. A schematic of the BASF2 data processing chain is depicted in Figure 3.1.

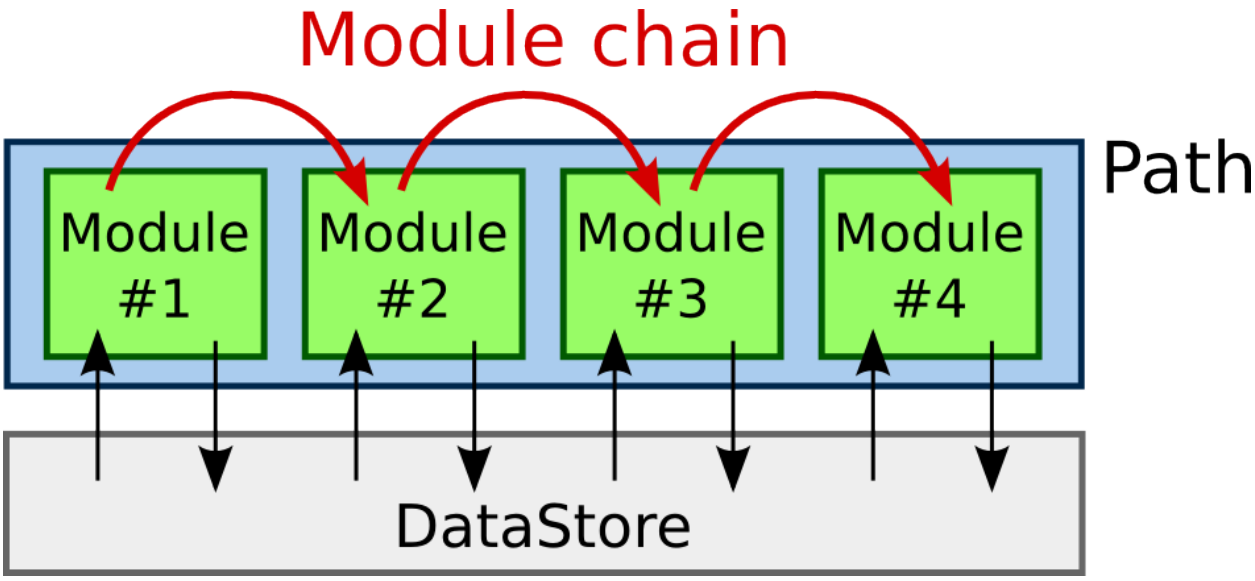


Figure 3.1: The BASF2 data processing chain. Modules are arranged and executed linearly in a path, exchanging information using the DataStore [29].

3.1.2 Packages

Additionally, some modules are categorised and grouped together into packages for convenience and ease of access. These packages represent various applications within the Belle II Experiment, and those available in BASF2 include the following:

- Framework: Contains modules relating to file input and output, Python steering and parallel processing.
- Geometry: Contains modules relating to building the detector geometry for simulation.
- Simulation: Contains modules relating to the generation of particle decay events and the simulation of their detection.
- Subdetectors: contains libraries with geometry information relating to the various detector layers such as the SVD, as well as modules that simulate the detector response to incident particles.

- Tracking: Contains modules used for identifying particle tracks, and fitting tracks to decay vertices.
- Analysis: Contains modules and libraries relevant for physics analyses, such as decay event reconstruction and event selections.
- Display: Contains the `EventDisplay` module which provides a 3-D visualisation of the detector along with individual decay events.
- MVA: Contains modules used to perform multi-variate analysis, a machine learning technique that is described in Section 3.3.

3.2 Monte Carlo (MC)

Simulated data, or Monte Carlo, is a valuable tool in many high energy physics applications. Within the scope of Belle II, Monte Carlo refers to the generation of particle decay events, such as those of an $\Upsilon(4S)$, and the simulation of how these decay products interact and are subsequently detected by the relevant Belle II sub-detectors. Through studying the simulation of the detector response to various generated collisions, analysis techniques can be optimised for use on real data. The Belle II collaboration conducts official MC productions for many different decay processes that are available for use, and are referenced throughout this report. Alternatively, BASF2 users are able to generate individual MC samples within the analysis framework. The production of Monte Carlo at Belle II consists of two stages:

1. Event Generation: Using the EvtGen package [32], full particle decay events are simulated. Users can select the required number of decay events, and can choose to specify any desired decay modes and their amplitudes, or default to current theoretical and experimental values. The particle mother may decay directly to the final particles in the chain, or to intermediate particles that further decay to produce these final-state particles. This step is entirely detector-independent, and is referred to as generator-level MC.
2. Event Simulation: The generated particles are then propagated throughout the detector. Final-state particles interact with the various sub-detectors including the ECL and KLM, the simulation of which is performed using the Geant4 package [33]. This results in simulated clusters in the ECL and KLM, as well as simulated particle tracks that can be matched to these clusters.

The result of this two-step process is a set of simulated ECL and KLM clusters along with particle tracks. This information can be used to reconstruct the final-state particles and subsequently the entire decay chain.

3.2.1 MC Truth Matching

Unlike real data, MC provides the unique opportunity to access not only the result of the reconstruction from detector information, but also the input generated particles. As such,

the generator-level MC is often referred to as the MC truth. BASF2 implements a process by which reconstructed-level and generator-level information can be compared, or matched, in order to determine whether a certain particle has been reconstructed correctly. This process is invaluable in providing estimates of the quality of reconstruction that can be extrapolated to real data. The BASF2 module responsible for MC truth matching is the `MCMatcherParticles` module, which offers two ways in which a match can be executed. The standard method, the function `matchMCTruth()`, performs an exact match to truth information. That is, for a certain particle decay, all daughter particles must have a common mother that is found in the truth. Alternatively, the `looseMCTruth()` function performs a more flexible match that does not require that all daughter particles arise from a common mother, but rather only the majority of them.

After applying these functions, the output of the matching is presented as a set of error bit flags that describe the nature of the matching errors, if any [40]. The complete list of errors is not detailed here, but some examples include particles that are missing in the reconstructed case which were present in the truth, or particles that have been misidentified after reconstruction. An error flag also exists for the presence of a neutrino in the truth that is not present in the reconstruction, which is of particular relevance to the study of semi-leptonic and leptonic decays. As neutrinos cannot be detected by Belle II, this error flag is unavoidable.

Another closely related set of variables that can be accessed after truth matching is the `isSignal` class of variables. These provide a more compact way of using truth information to assess whether a particle was correctly reconstructed, with `isSignal` given an integer value of 0 or 1 for incorrect and correct reconstruction, respectively. The `isSignal` variable is defined such that a value of 1 is returned for a particle match that is exact or contains either of two error flags that are deemed somewhat acceptable - namely the existence of an intermediate particle that was not reconstructed between the decay of its mother and daughters, and the absence of reconstructed photons radiated from final-state particles. Extended versions of `isSignal` exist that impose softer restrictions on these truth matched error flags. The `isSignalAcceptMissingNeutrino` variable is one such case designed for leptonic and semi-leptonic decays in order to bypass the afore-mentioned error associated with neutrinos that are generated but not reconstructed. It is identical to `isSignal` but also returns an integer value of 1 if the missing neutrino error flag is present. Similarly, for hadronic decays, an `isExtendedSignal` variable exists that additionally accepts the misidentification of a charged final-state particle.

3.3 Multivariate Classification

Multivariate classification is an extremely valuable tool used in a variety of different applications, with language recognition technology, internet search engines and email spam filters being some of a multitude of examples. It is a type of machine learning technique in which, using a number of input variables, a data element can be classified into one of several given categories. In order for this classification to be meaningful, the technique must be trained

on large samples of data that have been categorised correctly.

Multivariate classification can be performed in several ways, including categorical methods (sorting data into independent categories such as languages) and ordinal methods (sorting data into ordered categories such as whether a duration is 'short', 'medium', or 'long'). However, most relevant for physics analyses, and the classification method that is described in detail below, is a numerical-based method. A typical use for multivariate classification in high energy physics analyses such as those that will be performed at Belle II, would be the classification of particles or decay events as being either signal or background. When studying a particle decay of interest in particle collider experiments, these decay events, termed signal events, are produced alongside a multitude of other particle decays forming various types of backgrounds (described in Chapter 5). Multivariate classification methods, after being trained on large samples of Monte Carlo containing a mixture of known signal and background events, aim to classify unseen data into one of these two categories.

In practice, the numerical method of multivariate classification employed in the techniques discussed throughout this report involves defining a multi-dimensional vector $\mathbf{x}$ for each reconstructed particle candidate involved in a decay [34]. The elements of the vector, termed features, can include variables such as particle momentum and decay vertex information among many other quantities. This multi-dimensional feature vector $\mathbf{x}$ can then be mapped onto a one-dimensional number known as a test-statistic T. Upon training the classification on large numbers of simulated Monte Carlo containing both signal and background decay events, the distribution of the test-statistic provides a meaningful way in which to discriminate between the two categories.

Figure 3.2 shows an example of a test-statistic distribution for a particular feature vector $\mathbf{x}$, in which two distinct peaks are observed. Given a particular null hypothesis H_0, such as that corresponding to the default assumption that $\mathbf{x}$ is signal, an alternate hypothesis H_1 exists capable of rejecting the null hypothesis, in this case referring to the determination of $\mathbf{x}$ as background. These conditions can be associated with probability density functions (PDFs) $T(\mathbf{x}|H_0)$ and $T(\mathbf{x}|H_1)$ for signal and background respectively, where

$$\int_{T_a}^{T_b} T(\mathbf{x}|H_0), \qquad \int_{T_a}^{T_b} T(\mathbf{x}|H_1)$$

represent the probability that $\mathbf{x}$ will be considered signal or background within a certain range of the test-statistic ($T_a < T < T_b$). These PDFs are normalised to satisfy the conditions of probability:

$$\int_{-\infty}^{\infty} T(\mathbf{x}|H) = 1$$

Upon training, a critical value of the test-statistic T_c can be chosen above which particle candidates can be considered signal. This naturally results in errors whereby candidates are misidentified. As seen in Figure 3.2,

$$\alpha = \int_{-\infty}^{T_c} T(\mathbf{x}|H_0)$$

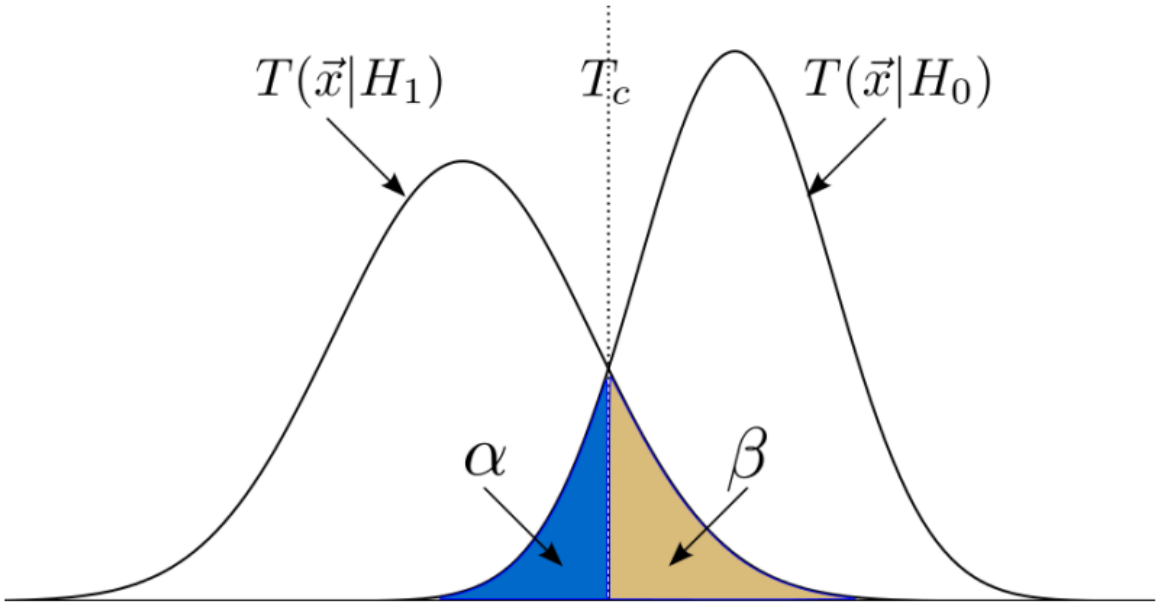


Figure 3.2: An example distribution of the test-statistic T for signal $T(\mathbf{x}|H_0)$ and background $T(\mathbf{x}|H_1)$ PDFs. Above a critical value T_c, all candidates are considered signal. α and β represent the errors associated with candidates misidentified as background and signal, respectively [34].

represents the probability of errors in which signal candidates are misidentified as background, and

$$\beta = \int_{T_c}^{\infty} T(\mathbf{x}|H_1)$$

represents the probability of errors in which background candidates are misidentified as signal.

3.3.1 Multivariate Classification in BASF2

3.3.1.1 Artificial Neural Networks

Artificial neural networks are computing applications that consist of multiple different interconnected nodes, or neurons, present in different layers. Each neuron in a layer is connected to each neuron in the adjacent layer, forming a complex network resembling the structure of the biological brain, after which the technique is named. A signal can be transferred along the connections between neurons, with each connection given an associated weight, w, that represents the strength of the connection. These weights are modified throughout the training process in what is known as machine learning. An example of the structure of an artificial neural network is depicted in Figure 3.3. In practice, the neurons of the first layer correspond to the elements of the input feature vector described above, with the final layer being synonymous with the output test-statistic. The NeuroBayes package [36] was the chosen software for artificial neural network analysis in the Belle Experiment. BASF2, however, now employs the use of the external ROOT-based Toolkit for Multivariate Analysis, or TMVA package [37], which is tailored for high energy physics applications and provides multiple different techniques for multivariate analysis.

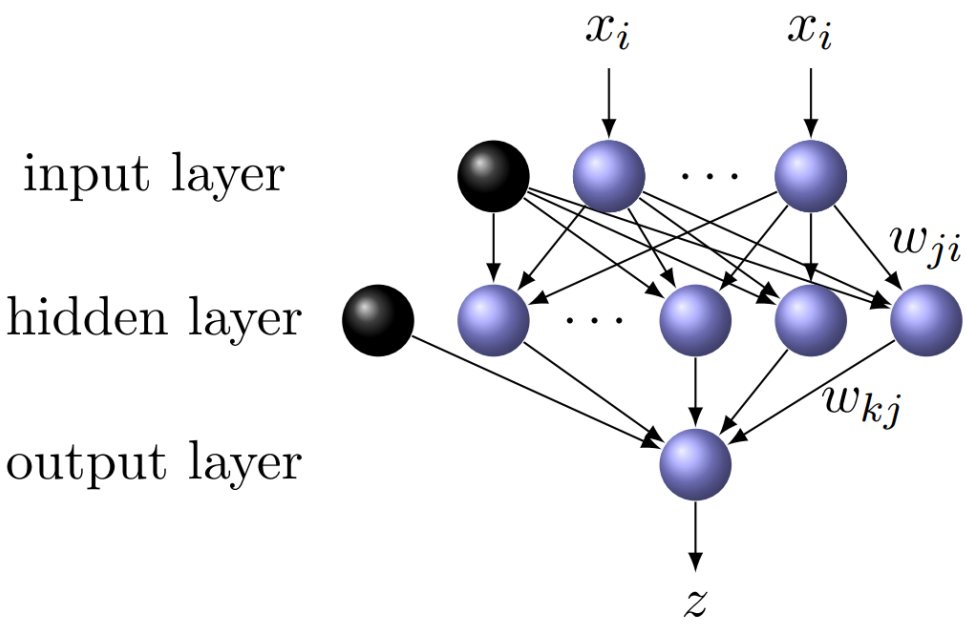


Figure 3.3: A schematic of an artificial neural network, with neurons interconnected between layers [35].

3.3.1.2 Boosted Decision Trees (BDTs)

An alternative method for the execution of multi-variate analysis available in BASF2 involves the use of decision tree classification. In this method, input features are evaluated

against particular criteria by which they are then split into categories. This classification then propagates down a structure referred to as a decision tree, whereby categories at each node are further split based on relevant variables, and a multi-variate classifier is trained. The typical structure of a decision tree is illustrated in Figure 3.4. Individual decision trees, however, are particularly sensitive to over-training [34], and an extended method known as boosted decision tree learning exists with the purpose of counteracting this effect. In this technique, multiple decision trees are used, each with a minimal number of nodes to prevent over-training. Each tree is assigned a weight, w_i, and the final classifier output is obtained through the weighted sum of the classifiers resulting from these weaker trees in order to produce a stronger classifier output.

The `FastBDT` software [38] is the primary BDT-learning technique incorporated within BASF2, and is more specifically an example of a stochastic gradient-boosted decision tree method. The gradient boost refers to one of multiple potential algorithms existing for the determination of the weights described in detail here [39], with the stochastic descriptor referring to the fact that only a random subset of the input statistics are chosen for the classifier training. This in turn allows the resultant classifiers to better discriminate between signal and background when applied to unseen data [34]. The Full Event Interpretation, being the focus of this study and described in detail in the following chapter, utilises `FastBDT` as its machine learning algorithm of choice.

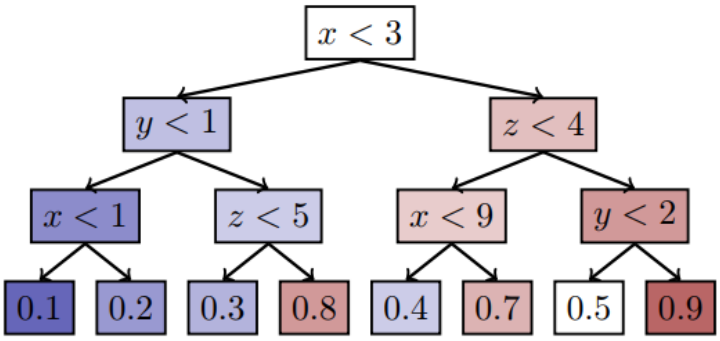


Figure 3.4: The structure of a single decision tree, depicting the categorisation of data at each node and the resultant trained classifiers [38].

Chapter 4

Full Event Interpretation

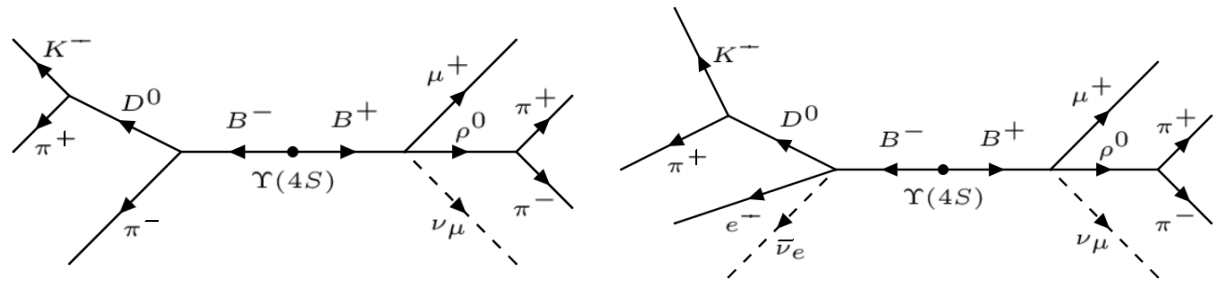
In the context of electron-positron collider experiments such as Belle and Belle II, the reconstruction of full $\Upsilon(4S)$ events containing B-meson decay modes of interest is achieved through two main reconstruction methods, namely tagged and untagged analyses. In such events, the decays of the daughter B-mesons can be precisely established, inferring the existence of the parent $\Upsilon(4S)$ meson with an initial state that can be deduced from the finely tuned energies of the incident electron and positron beams. Both reconstruction methods benefit from this knowledge, with kinematic information corresponding to the desired signal decay, or B_{sig}, able to be determined from 4-momentum conservation of the entire event.

4.0.1 Tagged Analysis

In tagged analyses involving the decay of an $\Upsilon(4S)$ to two B-meson daughters, in addition to the reconstruction of the desired B_{sig} decay mode, the second B-meson associated with the event, the B_{tag}, is also reconstructed using a number of exclusive decay channels. In this way, the reconstruction of the entire $\Upsilon(4S)$ decay chain is achieved. The B_{tag} decay modes can be either hadronic or semi-leptonic in nature, with examples of these presented in Figures 4.1(a) and 4.1(b), respectively, for the B_{sig} decay mode of particular relevance to this study, $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$.

Hadronic and semi-leptonic tags differ significantly in their treatment of leptonic and semi-leptonic B_{sig} decay modes, due to the presence of neutrinos. As described in Chapter 1, neutrinos are the lone fundamental particles known to interact only via the weak force. As such, they possess a significantly low interaction cross-section, and are unable to be detected in collider experiments such as Belle II, subsequently carrying energy and momentum away from systems in which they are present in the decay products. For (semi-)leptonic B_{sig} decay modes, $\Upsilon(4S)$ events consisting of hadronic B_{tag} daughters therefore possess a single source of missing 4-momentum assigned to the signal side. Thus, taking into account the initial state of the $\Upsilon(4S)$, the 4-momentum vector of the signal-side neutrino can be determined via 4-momentum conservation. This is not the case, however, for $\Upsilon(4S)$ decays to semi-leptonic B_{tag} daughters, in which neutrinos can be present on both the tag and signal sides of the event. Under such circumstances, the 4-momentum of the signal-side neutrino can no longer be exactly determined via 4-momentum conservation, but rather only estimated with the

help of appropriate signal-side selections.



(a) Hadronic tag

(b) Semi-leptonic tag

Figure 4.1: Examples of Υ(4S) events containing a B_{tag} and B_{sig} decaying as $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$, with $\rho^0 \to \pi^+\pi^-$. Dashed lines represent undetected neutrinos.

4.0.2 Untagged Analysis

Conversely, untagged analyses require only the reconstruction of the B_{sig} daughter from an Υ(4S) decay, with all remaining particles in the event built from detector information assumed to have arisen from the second B-meson. Since no explicit B_{tag} is reconstructed, in the combination of the 4-momentum of these particles suitable selections must be implemented to retain only those events consistent with the nature of B-meson decays. Given that no neutrinos interact with the detector, it is likewise assumed that the only neutrinos present in an Υ(4S) event are associated with the B_{sig} decay. In this way, a value for the 4-momentum of the signal-side neutrino can be determined, though the precision of such a measurement is secondary to the hadronic tagged analysis able to reconstruct the entirety of the Υ(4S) decay chain excepting the signal neutrino. A representation of a typical untagged analysis in depicted in Figure 4.2, for the semi-leptonic signal decay $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$.

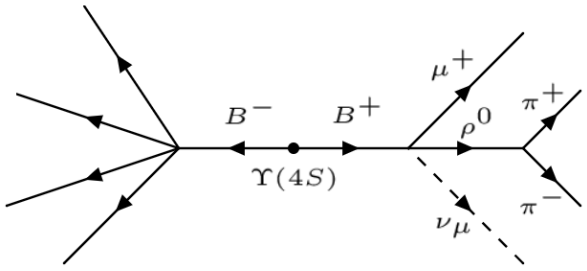


Figure 4.2: Untagged analysis of an Υ(4S) event with daughter B_{sig} decaying as $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$, with $\rho^0 \to \pi^+\pi^-$. No second B-meson is explicitly reconstructed.

The studies performed and detailed throughout the remainder of this report take the tagged approach, utilising both hadronic and semi-leptonic tagging methods in a compre-

hensive algorithm known as the Full Event Interpretation (FEI) [1]. The FEI is the signature analysis tool available in BASF2 for conducting tagged analyses and employs the use of multi-variate classification to effectively discriminate between signal and background decay events. It follows the same general approach as the equivalent technique that was used for the Belle Experiment, termed Full Reconstruction (FR), but with several improvements and added features that are discussed in detail in the coming sections.

In a collection of $\Upsilon(4S)$ decays, a typical analysis using a tagged approach will select those events in which both B_{sig} and B_{tag} daughters can be reconstructed. However, event reconstruction is very rarely perfect and often includes errors such as final state particles being misidentified, an example of this being a kaon incorrectly identified as a pion, or some daughter particles being assigned to the wrong B-meson. It is therefore extremely valuable to optimise the reconstruction method to produce as many correctly reconstructed candidates as possible. The FEI algorithm achieves this through first being trained on large amounts of Monte Carlo (MC) data in which MC matching can be used to access the truth information for each candidate. Using multi-variate analysis, the FEI utilises this truth information to learn to reconstruct full $\Upsilon(4S)$ decay events with the highest possible efficiency and purity.

4.1 Overview of Algorithm

The FEI tagging algorithm is trained in stages using a hierarchical approach that first uses detector information including tracks, ECL and KLM clusters to build final-state particle candidates [10]. A multi-variate classifier (MVC) is trained for each final-state particle using a number of relevant input variables detailed below, the result of which is summarised in a variable named the `SignalProbability`. This variable is related to the probability that a certain candidate is correctly reconstructed, and thus acts as a useful discriminant between signal and background events. The final-state particle candidates are then reconstructed into intermediate particle candidates within a variety of allowed decay channels, and a multi-variate classifier is trained for each individual channel that uses the final-state particle MVCs as input amongst other relevant variables. Finally, these intermediate candidates are used to build the mother B-mesons, with yet another MVC being trained for each reconstructed decay channel. In this way, the algorithm is trained stage by stage, with the `SignalProbability` of the candidates reconstructed at each level forming part of the input for the training of the MVCs performed in the following stages. This hierarchical approach is depicted in Figure 4.3.

4.2 Input Variables Used for Classifier Training

The following variables are used as input into the multi-variate classifiers to determine the resulting signal probability of each candidate [40]:

4.2.1 Final-state Particles (FSPs)

Charged FSPs: $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$

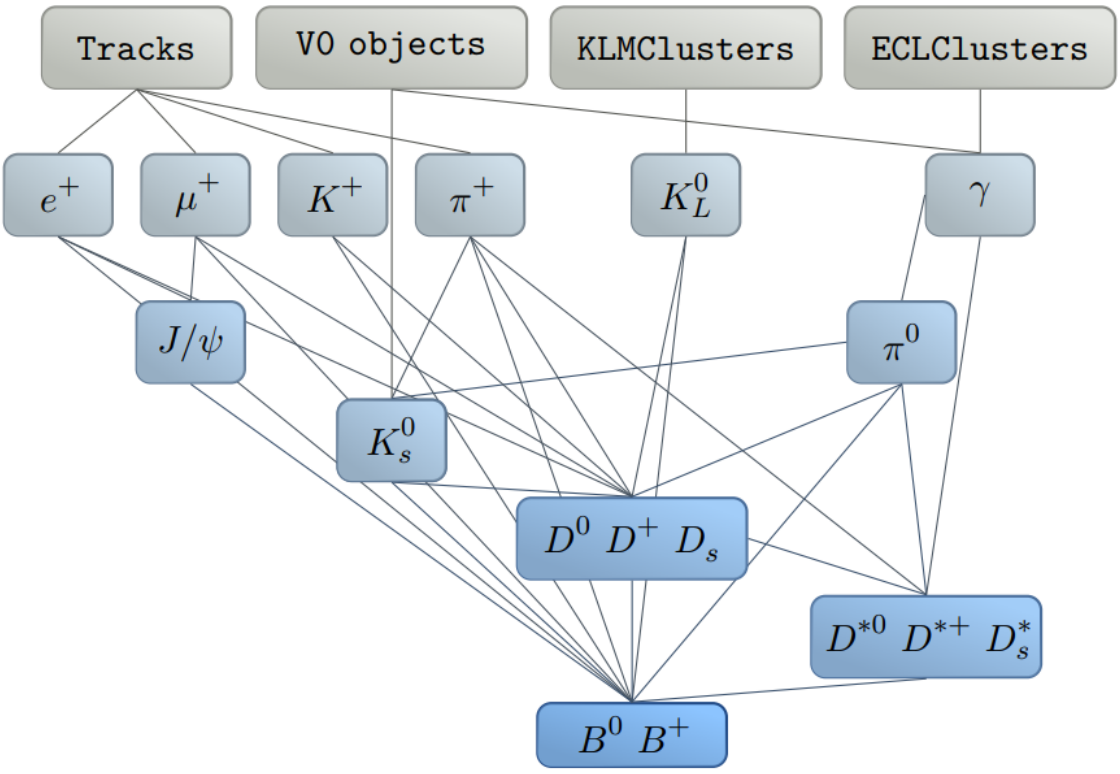


Figure 4.3: The hierarchical structure of the FEI tagging algorithm. Detector information including tracks and clusters is first used to reconstruct and train a classifier for each final-state particle. Intermediate particles are then reconstructed in stages, with the classifier output of each stage taken as input for the classifier training of the successive stage [1].

- **Particle identification (ID) variables:** In BASF2, a particle ID exists for each charged final-state particle, namely the `electronID`, `muonID`, `pionID`, and `kaonID`. For all but the `pionID`, these variables represent the likelihood ratio that a particular charged track is in fact the named particle as opposed to a charged pion. The `pionID` is the exception, with this variable representing the probability that a particular charged track belongs to a pion over a kaon. The proton and deuteron are two additional charged final-state particles with associated particleIDs (`protonID` and `deuteronID` respectively), but these are not used by the FEI.

- **Track momentum variables:** These include the magnitude of the total momentum p, as well as the individual transverse and z components of the momentum, p_t and p_z respectively. These variables are useful for discriminating between tracks originating from different final-state particles with characteristic momentum distributions.

- **Track impact parameters:** The interaction point is defined as the point on the beam line at which the electron-positron collision occurs. For any given track, the distance between the interaction point and the point closest to this on the track can be calculated. The radial, d_r, and z-components, d_z, of this distance are referred to

as impact parameters and they provide a way to discriminate well between particles produced at earlier stages in the decay chain, and those produced in subsequent decays with tracks originating further from the interaction point.

- **Track χ^2 probability:** All tracks are built from detector information and are fitted with an associated χ^2 probability. Higher values of χ^2 indicate a more precise track fit, and so this variable is also passed to the MVC training.

- **Particle rank:** When reconstructing particles at any point in a decay chain, there are often many candidates in the output that fit the reconstruction criteria, though only one of these candidates actually corresponds to the true particle. As such, in order to reduce combinatorics and unnecessary computing time, the FEI utilises a BASF2 feature known as the best candidate selection, whereby a list of candidates can be ranked in ascending or descending order based on the values they possess for a particular variable. The FEI employs this technique for most particles to select only a subset of candidates most likely to be correct.
  For the charged final-state particles reconstructed by the FEI, an initial pre-classifier cut is made on the impact parameters of all charged tracks. A discussion on the pre- and post-classifier cuts made during the FEI is left to Section 4.3, with this simply noted here. All candidates surviving this selection are then ranked from highest to lowest according to the relevant particle ID for each FSP (`electronID`, `muonID`, `pionID`, and `kaonID`), and only the best 20 candidates are kept for the pion and kaon cases, and only 10 for the electron and muon cases. The rank of each candidate in the list is stored and this rank is also used as an input variable for training. Illustrating this with an example, for the training of the classifier of an electron, all tracks surviving the above pre-cut are then ranked in order of `electronID`, with the most electron-like candidate given the rank of first. Only the 10 most electron-like candidates are kept, and their associated rank in the list is taken into account by the classifier training.

Photons: γ

- **Number of ECL cluster hits:** When a particle deposits energy into the ECL, multiple crystals may be hit producing an associated cluster. The number of crystals hit for any particular cluster, `clusterNHits`, is a useful discriminating variable between particles with characteristic energy deposits.

- **ECL cluster region:** The `clusterReg` variable simply returns which of the three regions of the ECL the cluster was found in, that is, the forward, backward and barrel regions. This is valuable information as the geometry of the ECL crystals differs between these three regions.

- **ECL cluster timing:** The time of detection of each cluster is recorded and the `clusterTiming` variable forms another input for the MVC training. The timing of cluster detection from known electron-positron conditions can be distinguished relatively well from that of clusters produced from beam background events.

- **ECL cluster shape:** Another significant aspect of cluster discrimination is the cluster shape, which is primarily described by the variable `E9E25`. This represents the ratio of the energy deposited in the inner 3x3 crystals of the cluster and the energy deposited in the outer 5x5 crystals. The value of `E9E25` can be used to distinguish between particles with different characteristic deposits in the calorimeter.
- **Energy and momentum:** The energy E and transverse and z-components of the momentum, p_t and p_z of the photon candidates are also passed to the classifier training, with the momentum of the detected photons determined by the the interaction point.
- **Particle rank:** The appropriate pre-cut is applied to all photon candidates and they are then ranked from highest to lowest according to their energy. The best 40 candidates are kept and the ranks of each candidate are used as input for the MVC training.

K_L^0 mesons

- **Energy:** The energy E of each K_L^0 candidate is passed to the classifier training.
- **KLM cluster timing:** The time of detection of each cluster in the KLM is also used, analogous to the use of ECL cluster timing for training photon classifiers.

4.2.2 Intermediate Particles

π^0 mesons

- **Invariant mass:** The invariant or rest mass of π^0 candidates is an important input variable due to the fact that the π^0 mass is well-known and experimentally established to be 135 MeV, as recorded by the Particle Data Group (PDG) [12]. Candidates with an invariant mass close to this value are more likely to correspond to true π^0 mesons. Two variables are passed to the classifier training, namely, the invariant mass M itself, and the variable dM, which is defined as the difference between the mass of the candidate and the nominal mass as given by the PDG.
- **Angle between photon daughters:** π^0 mesons decay to a pair of photons approximately 99% of the time [12]. The angle between the flight directions of both daughter photons is a useful quantity containing kinematic information about this decay.
- **Daughter signal probability:** As the π^0 mesons are intermediate particles built from final-state neutral photons, the output from the photon classifier training, that is, the photon signal probability, is used as part of the input to the π^0 classifier. This reflects how the hierarchical structure of the FEI is accomplished in practice.
- **Energy and momentum:** The energy E and transverse and z-components of the momentum, p_t and p_z are also included for π^0 mesons.
- **Particle rank:** After the relevant pre-cut for π^0 mesons is applied, the candidates are then ranked from those with the lowest values of $|dM|$ to those with higher values. The best 20 candidates are then kept and the ranks of each are also passed to the classifier training.

K_S^0 mesons

- **Distance variables:** For particles that produce charged tracks, the impact parameters are measured between the interaction point and the closest point of the track, as mentioned above. In the case of K_S^0 mesons which do not leave charged tracks but decay within the geometry of the detector, the equivalent distance is that between the interaction point and the decay vertex of the K_S^0 reconstructed from its daughters. The magnitude of the total displacement vector, given the variable name of `distance` in BASF2, as well as the radial and z-components of this distance, dr and dz, are included as input variables. The significance of the distance forms another variable passed to the classifier, and this is defined as the ratio between the measured distance and the uncertainty of this measurement. This indicates that the more precise our determination of the K_S^0 vertex position, the more meaningful distance variables will be when considered by the MVC training.

- **χ^2 probability:** The χ^2 probability for K_S^0 mesons and other intermediate particles is associated with the quality of the decay vertex position fit, as opposed to the track fit when used as input for charged FSPs.

- **Angle between vertex position and K_S^0 momentum:** This refers to the angle between the K_S^0 momentum vector and the distance vector formed between the interaction point and the K_S^0 decay vertex, and is given the name `cosAngleBetweenMomentumAndVertexVector` in BASF2. This variable contains relevant information related to the K_S^0 flight direction and the kinematics of its decay.

- **Invariant mass:** Much like for π^0 mesons, the invariant mass M of the reconstructed K_S^0 candidates as well as the difference dM between these values and the nominal mass of 498 MeV [12] as quoted by the PDG, form input variables to the classifier.

- **Energy:** The energy E of the K_S^0 candidates calculated in their rest frames is also included.

- **Angle between daughters:** The angle between the flight direction vectors of the K_S^0 daughters provides useful kinematic information in addition to the energy, and this forms yet another input variable.

- **Daughter momenta and distance variables:** The momentum p of each K_S^0 daughter candidate calculated in their rest frames is also passed as kinematic information to the classifier. Additionally, the radial and z- distance variables d_r and d_z of each daughter are used, and these are crucial to the determination of the K_S^0 decay vertex.

- **Daughter signal probability:** As previously discussed, the signal probability determined from the classifier training of each daughter is used as input to the classifier training of the parent particle.

- **Particle rank:** Finally, the particle rank is included in a similar way to the treatment of π^0 mesons. After the relevant pre-classifier cut, K_S^0 candidates are ranked from

lowest to highest according to their values of $|dM|$. The best 20 candidates are kept, and their ranks are passed to the classifier training.

J/Ψ mesons, D-mesons: D^0, $D^{\pm}$, D_s, D^{*0}, $D^{*\pm}$, D_s^*

- **Daughter signal probability:** The signal probability of each daughter is used as input, as well as the product of these probabilities for each daughter.

- **χ^2 probability:** The χ^2 probability associated with the decay vertex fit is used, as well as the relevant χ^2 probability for each daughter. For charged final-state daughters, this is the χ^2 probability of the track fit, or alternatively for daughters that undergo subsequent decay, this refers to the χ^2 probability of the fit of the secondary decay vertex.

- **Invariant mass:** The difference between the invariant mass of the candidates and the established nominal mass, $|dM|$, is also used as classifier input for these intermediate particles, together with the invariant mass M of each daughter candidate.

- **Angle between vertex position and particle momentum:** This is defined in the same way as described above for K_S^0 mesons, and contains kinematic information.

- **Daughter momenta and distance variables:** In a similar manner to the treatment of K_S^0 mesons, the momentum p of each particle daughter candidate calculated in its rest frame forms an input variable, together with the magnitude of the distance vector from the interaction point to each daughter track or secondary vertex.

- **Decay angles:** A number of relevant decay angles associated with these intermediate particles are also passed to the classifier training. These include the angles between the mother momentum vector and the momentum vector of each daughter, as well as the angles between the momentum vectors of each daughter particle with one another.

- **Released energy in decay:** This corresponds to a quantity denoted Q and is calculated by subtracting the sum of the invariant mass of all daughter candidates from the invariant mass of the parent intermediate particle. A related variable dQ is also defined and corresponds to the difference between the released energy for each candidate and the established nominal value. Both Q and dQ form input variables to the classifier.

- **Particle rank:** After the relevant pre-cuts, D^{*0}, $D^{*\pm}$ and D_s^* candidates are ranked from lowest to highest according to their values of dQ, with 20 candidates retained. J/Ψ, D^0, $D^{\pm}$ and D_s mesons are ranked according to dM, likewise with 20 candidates kept. These particle ranks are similarly passed to the classifier training.

4.2.3 B-mesons: B^0, $B^{\pm}$

Many of the input variables passed to the B-meson classifiers have been described in detail in the previous sections. These are simply listed below, with unique variables accompanied by descriptions.

- **Daughter signal probability:** the signal probability of each daughter, and the product of the signal probability of all daughters.

- χ^2 **probability:** of B-meson decay vertex fit, and daughter track or secondary decay vertex fit.

- **Daughter momenta and distance variables:** Momentum p and magnitude of the distance vector for each daughter.

- **Distance variables:** magnitude of the distance vector, along with individual radial, x-, y-, and z- components, d_r, d_x, d_y, d_z. The significance of the distance is also included.

- **Angle between vertex position and particle momentum.**

- **Decay angles:** between the B-meson and each daughter, and between each daughter with all other daughters.

- **Energy difference:** The energy difference ΔE is defined as the difference between the energy E_B of the B-meson candidate in the $\Upsilon(4S)$ centre-of-mass (CMS) frame and the CMS beam energy E_{beam}, or $\Delta E = E_B - E_{beam}$. The beam energy corresponds to approximately half of the available energy in the centre-of-mass system. Since the $\Upsilon(4S)$ decays to two B-mesons of equal mass, true B-meson candidates should possess a value of ΔE close to 0. For this reason, ΔE is a crucial variable used for discriminating between true and fake B-meson candidates, and is therefore used as an input variable for the training of the classifiers.

- **Particle rank:** The B-meson candidates are ranked from highest to lowest according the value of the product of the signal probability of all daughters. The best 20 candidates are kept, and their ranks are passed to the classifier training.

4.3 Cuts Made Within FEI Algorithm

The FEI algorithm is nominally designed to be trained on MC samples of the order of 100 million $\Upsilon(4S)$ decay events. Multiple candidates are reconstructed for each particle, with each particle mother being built with all possible combinations of daughter candidates. The hierarchical nature of the FEI therefore leads to a significant increase in combinatorics at each reconstruction stage, with the final reconstructed B-mesons being the result of millions of possible combinations. As such, the computing time and CPU usage that the FEI would require without making any selections to reduce the number of candidates at each stage would be unreasonably large. Three categories of selection cuts are thus made throughout the training of the FEI algorithm, and include cuts made before and after the training of MVCs for each particle type, as well as user-defined cuts that can be applied to the parent B-mesons [40]. With all selections in place, the FEI training takes between 2-5 days to run on the KEK computing system, KEKCC, and requires 10-20 TB of disk space during the training process.

4.3.1 Pre-classifier Cuts

One of the most useful ways of reducing combinatorics is to apply appropriate selections to the particle candidates before any classifiers are trained. These cuts should be reasonably loose and aim to remove only candidates that are very unlikely to correspond to true particles, such that computing time is not wasted passing variables from these candidates to the classifier training. These pre-classifier cuts were alluded to in Section 4.2, and are quoted below for each particle type as defined by the FEI algorithm.

Charged final-state particles ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$): A simple cut on the track impact parameters is made to ensure the tracks are close enough to the interaction point:

$$d_r < 2 \text{ cm, and } |dz| < 4 \text{ cm}$$

Only the 20 $\pi^{\pm}$ and $K^{\pm}$ candidates and 10 $e^{\pm}$ and $\mu^{\pm}$ candidates with the highest particle ID values are retained.

Photons (γ): Depending on whether the ECL cluster associated with a photon candidate is located in the forward region, barrel, or backward region, an appropriate cut on the photon energy is made to distinguish candidates that have energies characteristic of true photons. The values 1, 2, and 3 of the `clusterReg` variable correspond to the forward, barrel and backward regions, respectively:

$$\begin{aligned}&(\text{clusterReg} == 1 \text{ and } E > 0.10 \text{ GeV}) \text{ or } (\text{clusterReg} == 2 \text{ and } E > 0.09 \text{ GeV}) \text{ or} \\ &(\text{clusterReg} == 3 \text{ and } E > 0.16 \text{ GeV})\end{aligned}$$

Only the 40 photon candidates with the highest energies are retained.

K_L^0 mesons: No pre-classifier cut is applied to the K_L^0 mesons. Only the 20 K_L^0 candidates with the lowest values of $|dM|$ are retained.

D^{*0}, $D^{*\pm}$, D_s^* mesons: Selections are made on the released energy Q of the D-meson decays to ensure these are close to the nominal values:

$$0 \text{ GeV} < Q < 0.3 \text{ GeV}$$

Only the 20 D-meson candidates with the lowest values of $|dQ|$ are retained.

The pre-classifier cuts applied to the remainder of the intermediate particle lists correspond to restricting the invariant mass to a suitable mass window, and are summarised in Table 4.1.

$B^{\pm}$, B^0 mesons: For hadronic channels, cuts are applied to two meaningful kinematic variables, the first of which being the energy difference ΔE introduced in Section 4.2.3. As mentioned, true B-meson candidates are expected to have a value of ΔE close to 0, and this is reflected in an appropriate pre-classifier cut:

$$|\Delta E| < 0.5 \text{ GeV}$$

Intermediate particle	Invariant mass cut (GeV)	Number of candidates retained with lowest values of $\|dM\|$
π^0	$0.08 < M < 0.18$	20
K_S^0	$0.4 < M < 0.6$	20
D^0, $D^\pm$	$1.7 < M < 1.95$	20
D_s^+	$1.68 < M < 2.1$	20
J/Ψ	$2.8 < M < 3.5$	20

Table 4.1: The invariant mass pre-cuts applied to a subset of the intermediate particles reconstructed by the FEI.

The other variable used for selection is the beam-constrained mass M_{bc}, and this is defined by the following formula:

$$M_{bc} = \sqrt{E_{beam}^2 - \vec{p}^2},$$

where E_{beam} is the aforementioned beam energy defined in the $\Upsilon(4S)$ rest frame, and $\vec{p}$ corresponds to the reconstructed 3-momentum of the B-meson candidate. In essence, M_{bc} is a quantity that is analogous to the invariant mass, but with the beam energy replacing the B-meson energy such that the calculation does not rely on the B-meson mass determined from its daughters. True B-meson candidates should have M_{bc} values close to the B-meson invariant mass of 5.28 GeV [12], and a loose cut is therefore employed to reduce combinatorics before classifier training:

$$M_{bc} > 5.2 \text{ GeV}$$

Only the 20 B-meson candidates with the highest product of the signal probabilities of each daughter are retained. Whilst being an effective discriminator between B-mesons likely to be correctly or incorrectly reconstructed along the decay chain, this cut leaves the method prone to discarding good B-meson candidates in such a case that the `SignalProbability` of a single daughter is set to 0 due to a failing of the algorithm. No pre-classifier cuts are applied to the B-meson semi-leptonic decay modes.

It should be noted that in all of the above cases, the number of candidates retained corresponds to the maximum possible number, that is, a channel with a number of candidates less than or equal to the defined cut-off will retain all candidates after such a pre-cut is applied. It is therefore pertinent that the cut-off values are reasonably defined for each particle type to remain considerably loose whilst also avoiding such cases in which all candidates are able to pass the implemented cut. All values quoted represent the default configuration of the FEI given the suggested order of input training events, with the opportunity available for individual users to override these cut-off values with their desired alternatives.

4.3.2 Post-classifier Cuts

After selecting a number of candidates to proceed to the classifier training, additional cuts can then be made on the list of candidates once the classifiers have been trained. Each candidate

has a `SignalProbability` now associated with it as a result of the training, and this can be used as an additional variable for selection, given that it is expected that candidates with a relatively low `SignalProbability` are less likely to correspond to the true particles. Post-classifier cuts on this `SignalProbability` are enforced to further reduce combinatorics by decreasing the number of candidates proceeding to the subsequent reconstruction stage. The default cuts as implemented within the FEI algorithm are depicted in Table 4.2 for each particle type, and these can be overridden by the user as desired in the same way as for the pre-classifier cuts described above. No post-classifier cuts are applied to the B-mesons, as this is the final stage of the FEI reconstruction.

Particle	Post-classifier cut	Number of candidates retained with highest values of `SignalProbability`
$K^{\pm}$, $\pi^{\pm}$	`SignalProbability` > 0.01	10
$e^{\pm}$, $\mu^{\pm}$	`SignalProbability` > 0.01	5
γ	`SignalProbability` > 0.01	20
K^0_L, K^0_S, π^0	`SignalProbability` > 0.01	10
J/Ψ, D^0, $D^{\pm}$, $D^{\pm}_s$, D^{*0}, $D^{*\pm}$, D^*_s	`SignalProbability` > 0.001	10
$B^{\pm}$, B^0	-	all

Table 4.2: The post-classifier cuts applied to the particles reconstructed by the FEI.

4.3.3 User Cuts

Finally, the FEI algorithm provides the user with an opportunity to insert their own selections relevant to their individual physics analyses, though these can only be applied to the parent B-mesons. The pre- and post-classifier cuts for each particle are set by the algorithm and not subject to change by the user. Any user cuts made are applied to the B-meson candidate lists before any other cut is applied, and consequently, only candidates surviving these cuts will be passed through to the training of the B-meson classifiers.

4.4 Comparison with Full Reconstruction

As mentioned in the opening of the chapter, the Belle Full Reconstruction (FR) algorithm was similar in approach to the FEI, with the hierarchical structure and the training of the classifiers at each reconstruction stage being a common feature between the two methods. However, the FEI has improved upon various aspects of the FR, and benefits from a number of added features.

The FR algorithm involved the use of the external NeuroBayes package[36], with the multivariate analysis technique chosen for the FEI being that of the stochastic boosted decision tree method, FastBDT [38]. The training of the FR algorithm was performed centrally at KEK by the Belle collaboration, with individual users accessing the results of this training

to incorporate into their own analyses. The training process was also slow, taking of the order of weeks. In contrast, the FEI is much faster, taking only of the order of days for the training process to be completed, and is designed such that users can perform the training themselves. The FEI is significantly more user-friendly, with the entire algorithm automated such that users need only to specify the initial configuration they would like for the training, such as defining any desired user cuts [34]. The algorithm is then efficiently trained as per the specified configuration with no further intervention required by the user. The FEI thus effectively provides the user with a high degree of customisation at the training stage, a feature that could not be realised by the FR.

Another major difference between the two methods is that FR only employed the use of hadronic tags, with the algorithm designed to train only hadronic B-meson decay modes. The FEI incorporates both hadronic and semi-leptonic decay modes, allowing for tagged analyses using both types of tags [10]. The algorithm additionally affords users the option of turning either the hadronic or semi-leptonic modes off during the training in order to save computing time in the case that the user is interested in only one of the two tagging methods. One further advantage of the FEI is that it supports more B-meson hadronic decay channels than were supported by the FR. At the commencement of this study, the FEI incorporated a total of 29 hadronic charged and 26 hadronic neutral B-meson decay modes, a significant increase from the equivalent 17 and 15 respective modes characteristic of the FR [40]. With the inclusion of the 42 charged and 39 neutral semi-leptonic modes, the FEI is able to reconstruct a total of over 4000 exclusive decay channels compared with the rough 1000 achieved by FR. Only the charged B-meson modes were relevant to the studies detailed in later chapters, with lists of these modes presented in Tables 6.1 and 6.2. Furthermore, a direct comparison between the FEI and FR algorithms was performed in [41], which applied the FEI to the full Belle dataset of approximately 3 billion events. In this study, ten semi-leptonic signal channels with well-known branching fractions were used to determine calibration factors ϵ for each individual hadronic tag channel, defined as the ratio between the number of signal events ($\Upsilon(4S) \rightarrow B_{sig}B_{tag}$) found in data and the number expected using MC simulations, $\epsilon = N_{data}/N_{MC}$. For the hadronic channels incorporated by both algorithms, the calibration factors determined using the FEI were found to be consistent with the equivalent values produced by the FR study in [42].

Finally, the largest and arguably the most noteworthy of improvements made by the FEI is the inclusion of the means to incorporate a specific B_{sig} decay mode reconstruction within the tagging algorithm itself. The FR algorithm was trained on large numbers of generic $B^0\bar{B}^0$ and B^+B^- MC events where the B-mesons were allowed to decay in multiple supported channels, and hadronic tags were reconstructed accordingly. Individual users interested in studying a particular decay mode, B_{sig}, would then search for this mode out of a subset of $\Upsilon(4S)$ decay events in which a hadronic tag could be built. This method, however, introduced an element of bias, as the algorithm was centrally trained on a sample of MC with certain features, consisting of a number of B-meson decay modes present in quantities characteristic of their branching fractions. However, after applying a signal-side selection to a set of events, it could not be assumed that the underlying features of the MC sample used for training and the subset of events used for analysis were the same. In essence, any effects

of the signal-side selection on the distributions of the tag-side B-mesons were ignored, with users applying a trained algorithm that ran on certain assumptions that did not necessarily pertain to the analysis sample used.

The FEI developed for Belle II proposes a solution to this issue of bias by providing two possible methods of FEI application, the **Generic FEI** and the **Specific FEI** [10]. The generic FEI is analogous to the FR technique, whereby it is trained on large MC samples of $B^0\bar{B}^0$ and B^+B^- MC events with no signal-side selection implemented at the training stage. However, the specific FEI, a unique method developed for Belle II, is trained not only on a set of $B^0\bar{B}^0$ and B^+B^- MC events, but also on $\Upsilon(4S)$ events in which one B-meson decays as B_{sig}, the decay of interest for analysis. The B_{sig} mesons are first reconstructed with all relevant physics selections decided by the user, and the FEI algorithm is then trained only on the particles present in the rest of the $\Upsilon(4S)$ event. In this way, each specific FEI training performed is tailored to the decay mode of interest, effectively removing any potential bias. Due to the extended run-time of the FR, it was simply not feasible for individual users to perform their own MVA training. The FEI, however, given its efficient run-time and user-friendly nature has made it possible for users to not only perform the training themselves, but to tailor it towards their specific physics analysis. A detailed description of the generic and specific FEI is given below, and the comparison of the performance of each method forms the focus of the study detailed in the upcoming chapters.

4.5 Generic FEI

As introduced in the opening of this chapter, the FEI forms an effective tool for full $\Upsilon(4S)$ decay event reconstruction, though it must first be trained on MC data in which the truth information of all particles is known. The sections above offered a comprehensive explanation of how the FEI tagging algorithm is trained. Naturally, it follows that the trained algorithm can then be applied to various independent MC samples as well as real data, reconstructing B-meson and $\Upsilon(4S)$ candidates effectively. As such, any implementation of the FEI is usually described in terms of these two stages, namely the training and the application of the FEI. Both stages differ in practice between the generic and specific FEI, and these are delineated below.

4.5.1 Training

The training of the generic FEI is typically performed on the order of 100 million generic charged B^+B^- and mixed $B^0\bar{B}^0$ events,[1] and hadronic and semi-leptonic tags are built across a number of allowed channels. All pre- and post-classifier cuts are applied to each particle in the hierarchical structure, as well as any user cut specified for the parent B-mesons. The training is entirely independent of any signal-side selection.

[1] $\Upsilon(4S)$ decays to neutral pairs of B-mesons are often referred to as *mixed* events, the name arising from the oscillations predicted by the Standard Model between neutral matter and anti-matter particles.

A successful training yields a collection of weight files for each decay channel that forms part of the input to the application stage, as well as four `ParticleLists` that the user can access and manipulate during the application. These are the `B0:generic`, `B+:generic`, `B0:semileptonic` and `B+:semileptonic` lists, corresponding to the hadronic and semi-leptonic tag lists for the neutral and charged B-mesons, respectively.

4.5.2 Application

When applying the FEI to an independent sample of MC or real data, the weight files produced from the training stage must first be supplied. The most basic application of the generic FEI would then include using the given `ParticleLists` to create and store B-meson tag candidates. Certain kinematic or other variables of interest can be read out for each candidate.

Users wanting to perform a tagged analysis for a B_{sig} decay mode of choice can apply the generic FEI to a data or MC sample to produce a list of tags alongside building B_{sig} candidates, and can then combine these using the `reconstructDecay` function in order to reconstruct the entire $\Upsilon(4S)$ decay event. In the case of the generic FEI application, both the B_{sig} and B_{tag} candidates are built within the same path, with the `reconstructDecay` function ensuring that no B_{sig} and tag candidates associated with the same daughters are combined together into an $\Upsilon(4S)$.

4.6 Specific FEI

4.6.1 Training

The training of the specific FEI has a number of key differences in its implementation when compared with the generic FEI, the first of these being the nature of the MC samples used for the training. While the generic FEI is typically trained on the order of 100 million generic charged B^+B^- and mixed $B^0\bar{B}^0$ events, the specific FEI must be trained on these events in addition to a MC signal sample also recommended to be of the order of 100 million events. The signal sample here refers to a set of $\Upsilon(4S)$ decay events in which one resultant B-meson decays as B_{sig} in a decay channel of interest, and the other B-meson decays generically to any allowed mode.

Before any classifier training begins, B_{sig} candidates are reconstructed within each input MC sample, including both the signal events and the generic $B\bar{B}$ events that do not contain the B_{sig} decay, hence forming a significant background. A rest-of-event (ROE) object is then created for each B_{sig} candidate that contains all of the remaining reconstructed particles in the event that were not assigned to B_{sig}. MC truth matching is applied to each B_{sig} candidate, with the `isSignalAcceptMissingNeutrino` variable used to determine correctly reconstructed candidates for semi-leptonic B_{sig} decay modes, and either `isSignal` or `isExtendedSignal` for hadronic B_{sig} decay modes depending on user preference.

The FEI classifier training is then only executed on the rest-of-event of B_{sig} candidates in one of two categories:
1. B_{sig} candidates built from signal events, that pass truth matching and are correctly reconstructed, or;
2. B_{sig} candidates built from background events that fail truth matching and are incorrectly reconstructed.

Due to the fact that the the algorithm is only trained on a subset of the number of input events, large statistics of both signal and background are necessary for the efficient training of the specific FEI. It is worth noting that the quoted training sample sizes required for the specific and generic FEI are largely recommended and somewhat in need of validation, a notion expanded on within the FEI studies detailed in the upcoming chapters.

As with the generic FEI, the specific training produces a collection of weight files that are necessary input for the application stage, as well as the four `ParticleLists` associated with each B-meson tag.

4.6.2 Application

In the application of the trained specific FEI to an independent MC or real data sample, B_{sig} candidates are first reconstructed in much the same way as executed in the training, and a rest-of-event object is built for each candidate. However, these ROE objects are then copied to a separate analysis path, often termed the rest-of-event path, and it is here that the trained weight files are applied and the four B_{tag} `ParticleLists` are available for manipulation. In this way, the B_{tag} candidates are only built using particles not already assigned to the B_{sig}, reflecting the conditions under which the specific FEI algorithm was trained.

The B_{tag} candidates can then be copied back to the main analysis path and combined with the B_{sig} candidates using `reconstructDecay`, effectively reconstructing the entire $\Upsilon(4S)$ decay. For optimum results, the selection cuts employed in the reconstruction of B_{sig} candidates should be expected to be the same in both the training and reconstruction stages of the specific FEI.

Chapter 5

Signal Side Selection: $B^+ \to \rho^0 \mu^+ \nu_\mu$

Before investigating the performance of the Full Event Interpretation tagging algorithm on $\Upsilon(4S)$ decays with $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$ chosen as the signal decay of interest, appropriate signal-side selections must be determined. The reconstruction of B_{sig} candidates is executed independently of the FEI, with the final list of candidates produced forming part of the input to the FEI. As discussed in Section 4.6, in the case of the specific FEI the B_{sig} reconstruction is implemented at the training stage, and the algorithm is trained within the rest-of-event of each correctly reconstructed B_{sig} candidate. A similar approach is employed in the application of the specific FEI, with B_{tag} candidates reconstructed within these B_{sig} rest-of-event objects. It is therefore imperative that the selections decided on in reconstructing the B_{sig} decay chain are reasonable for use in the specific FEI. Tight selections that exclude a substantial number of B_{sig} candidates may leave too few statistics for the FEI algorithm to be trained on, whilst very loose selections may not provide sufficient discrimination between true B_{sig} candidates and background. These criteria can be summarised through the definition of two key quantities, namely the B_{sig} reconstruction efficiency and the purity.

The B_{sig} reconstruction efficiency is a measure of the proportion of total events in a particular sample in which B_{sig} candidates are able to be reconstructed, after a number of event selections. When considering Monte Carlo in which truth information is known and can be accessed, the reconstruction efficiency can be defined in either of two ways:

$$\textbf{Reconstruction efficiency} \equiv \frac{\text{Number of events with at least one reconstructed } B_{sig} \text{ candidate}}{\text{Number of total events}}$$

$$\textbf{Reconstruction efficiency (correct)} \equiv \frac{\text{Number of events with a } \textit{correctly} \text{ reconstructed } B_{sig} \text{ candidate}}{\text{Number of total events}}$$

Here, a *correctly* reconstructed candidate refers to one that can be matched to a particle in the MC truth using the appropriate `isSignal` variable introduced in Section 3.3.1, this being `isSignalAcceptMissingNeutrino` for the semi-leptonic decay $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$. Given that the following analysis is an entirely MC-based study, it was decided to adopt this second definition for all reconstruction efficiencies calculated throughout the remainder of this report. The training of the specific FEI itself relies on the reconstruction efficiency defined in

this way, with the training on signal MC samples performed only within the rest-of-event of those B_{sig} candidates that have been correctly reconstructed. Henceforth, all reconstruction efficiencies quoted follow this second definition, with the '(correct)' descriptor dropped from subsequent mentions.

The purity is another useful indicator of performance, and represents the proportion of those reconstructed B_{sig} candidates that can be matched to MC truth and considered to have been reconstructed correctly:

$$\textbf{Purity} \equiv \frac{\text{Number of events with } \textit{correctly} \text{ reconstructed } B_{sig} \text{ candidates}}{\text{Number of events with reconstructed } B_{sig} \text{ candidates}}$$

Due to the way in which they are defined, increases in purity are often met with decreases in efficiency and vice versa, with tighter selections rejecting many B_{sig} candidates likely to be incorrectly reconstructed at the expense of also losing a number of correct candidates. The remainder of this chapter details the selections decided on to achieve a reasonable balance between these two quantities, given the following decay chain:

$$B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$$
$$\hookrightarrow \pi^+ \pi^-$$

5.1 Description of Monte Carlo

In addition to B_{sig} candidates that can be incorrectly reconstructed from a sample of signal events where the B_{sig} decay is present in truth, fake B_{sig} candidates can be reconstructed from multiple background processes which do not contain the B_{sig} decay. The effect of various selections on the number of B_{sig} candidates reconstructed from both signal and background events was investigated by running the B_{sig} reconstruction on MC samples of the following sizes:

- 326 673 *signal* events, with $\Upsilon(4S) \to B^+_{sig} B^-$, $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$.
- 1 million *charged background* events, in which charged B-meson pairs (B^+B^-) from an $\Upsilon(4S)$ parent decay via any allowed modes.
- 1 million *mixed background* events, in which neutral B-meson pairs ($B^0\bar{B}^0$) from an $\Upsilon(4S)$ parent decay via any allowed modes.
- 1 million $c\bar{c}$ *continuum background* events, in which the electron-positron collisions do not produce an $\Upsilon(4S)$ meson (quark content $b\bar{b}$), but rather $e^+e^- \to c\bar{c}$.
- 1 million *uds continuum background* events, where $e^+e^- \to u\bar{u}$, $d\bar{d}$ or $s\bar{s}$.

All MC samples were taken from the seventh official Belle II MC campaign (MC7) located on the KEK computing system, KEKCC. Additionally, each sample used was produced alongside the standard estimate of beam background chosen for Belle II, referred to as

BGx1. Alternative samples with higher estimates of the beam background including twice (BGx2) or five times (BGx5) as large are also available, though none have been utilised in this study.

5.2 Final-State Particle Selections

The first step in reconstructing any decay chain using BASF2 is to build lists of the final-state particles using the `fillParticleList` function, within which relevant selections can be implemented. For the chosen B_{sig} decay chain, the final-state particles with their associated selections were as follows:

μ^+:

- `muonID` > 0.5. A particleID cut is necessary when building particle lists from charged tracks as without such a cut, tracks from all charged particles would be included in the list. With particleID variables defined between zero and unity, a moderate choice was decided on for the muon list in order to retain a large number of candidates at the expense of some contamination by other charged final-state particles.
- $d_r < 1$ cm and $|d_z| < 2$ cm. A suitable cut on the impact parameters of the tracks (see Section 4.2.1) used to build the list was also made to ensure the muon originated close to the interaction point.

π^+, π^-:

- `pionID` > 0.5. Much like for the muon case, a relatively loose cut on the `pionID` was chosen for the pion daughters produced from the ρ^0 meson.
- $d_r < 1$ cm, and $|d_z| < 2$ cm. An identical cut on the impact parameters of the pion tracks was also implemented.
- $0.5 < p_{CMS} < 2.8$ GeV. A substantially broad selection on the pion momentum calculated in the centre-of-mass frame was also made, excluding tracks with momenta highly unlikely to be characteristic of pions produced in pairs from a parent ρ^0 meson.

5.3 Further Selections

The neutral ρ^0 resonance [18] introduced in Section 1.2.1 is associated with a decay width established experimentally to be 147.8 ± 0.9 MeV for the given value of the ρ^0 rest mass, 775.26 ± 0.25 MeV [12]. This broadened distribution of the ρ^0 invariant mass can be seen in Figure 5.1(a), which depicts the MC truth distribution for the ρ^0 mesons originating from B_{sig} in the chosen signal sample. Figure 5.2(b) illustrates the same distribution for the reconstructed ρ^0 mesons in signal, the clean peak observed in truth now contaminated by fake entries whereby ρ^0 candidates have been built from incorrect pion combinations or daughters misidentified as pions. All possible combinations passing the pion list criteria introduced in the previous section are included in this plot, with multiple ρ^0 candidates present per event.

Finally, the ρ^0 invariant mass distributions produced from applying the reconstruction to each of the various background samples are also provided, for charged in 5.2(c), mixed in 5.2(d), $c\bar{c}$ in 5.2(e), and uds in 5.2(f). Given these distributions, an appropriate cut was chosen for the ρ^0 invariant mass to select the observed peak in reconstructed signal and reject fake combinations and background events.

ρ^0 invariant mass cut:

- $0.5 < \mathrm{M} < 1.4$ GeV

In order to use the B_{sig} reconstruction developed in the training of the specific FEI, a rest-of-event (ROE) object must be built for each reconstructed candidate. This is achieved through the `buildRestOfEvent` function, which uses all tracks and clusters present in a particular event that have not been assigned to a B_{sig} candidate to create its associated ROE. As the B_{tag} candidates are reconstructed inside of these objects in the specific FEI, it is useful to first clean them up, removing tracks or clusters likely to be the result of beam background events in order to improve the purity of the resultant B_{tag} candidates. BASF2 accommodates such a purpose through the inclusion of a feature known as an ROE mask, which is a set of user-defined selections that can be applied to the ROE objects of candidates reconstructed in an event. The selections included within the mask applied to the ROE of each B_{sig} candidate were as follows:

ROE ECL cluster selections:

- `clusterE9E25` > 0.9. Lower values of this ratio (see Section 4.2.1) correspond to ECL cluster shapes characteristic of beam background events.

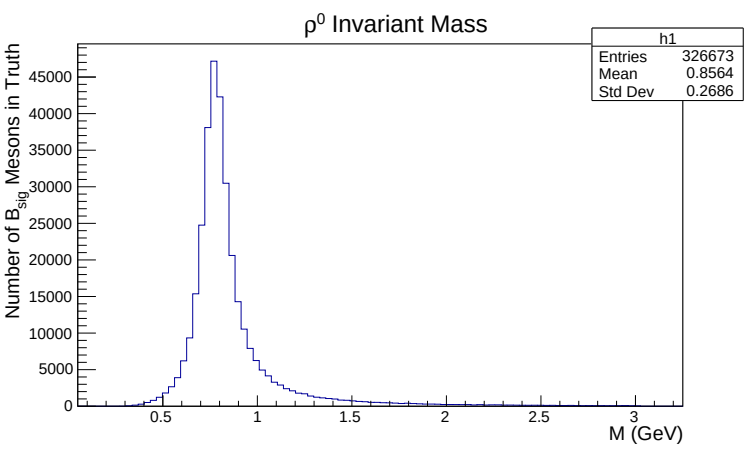
- `clusterTiming` < 50 ns. After an initial electron-positron collision, the time taken for the detection of ECL clusters produced from the decays of pairs of B-mesons is distinguishably shorter than the timing due to beam background.

- `goodGamma` $== 1$. This variable selects only photon candidates from the ECL with energies above thresholds likely to exclude beam background photons. These thresholds are defined as 100 MeV, 90 MeV and 160 MeV for the forward endcap, barrel, and backward endcap regions respectively.

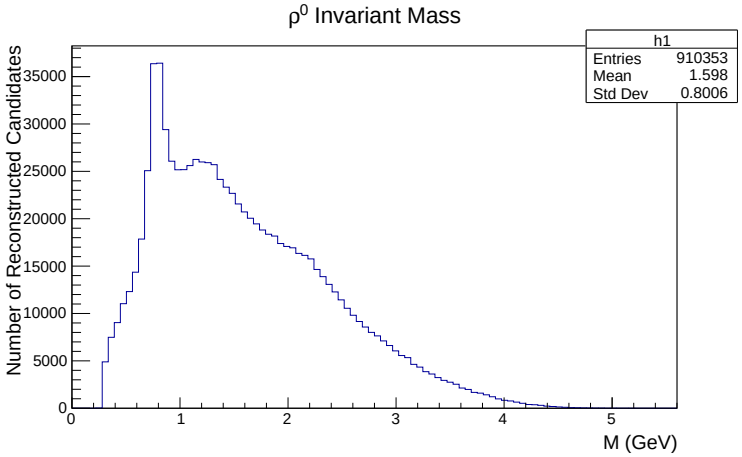
- `trackMatchType` $== 0$. This variable configuration is used in conjuction with the above `goodGamma` condition and selects ECL clusters with no matched tracks to build photon candidates.

ROE track selections:

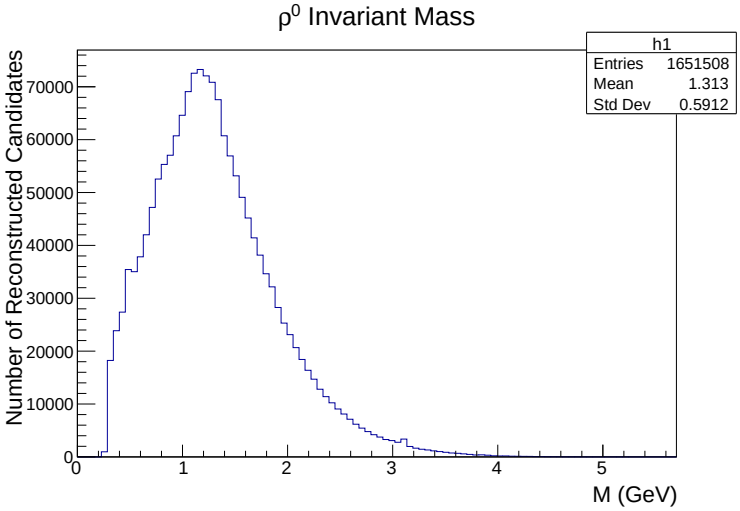
- $d_r < 2$ cm, and $|d_z| < 4$ cm. The impact parameters for charged tracks are restricted to regions close to the interaction point to exclude tracks likely arising from beam background processes.



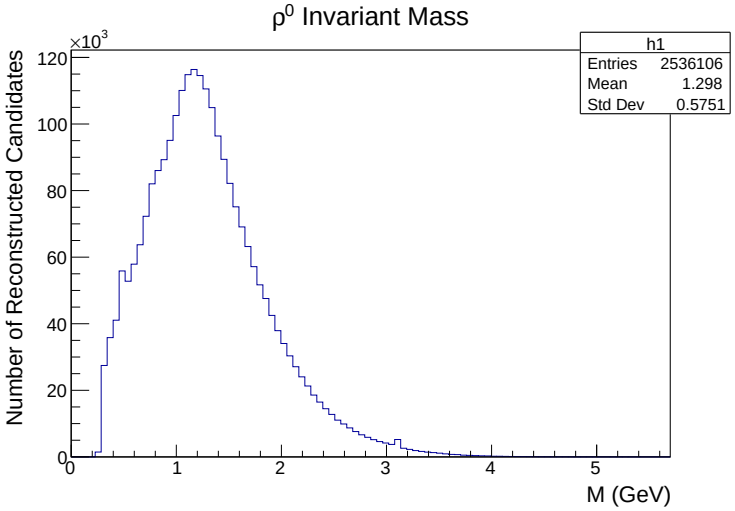
(a) MC truth



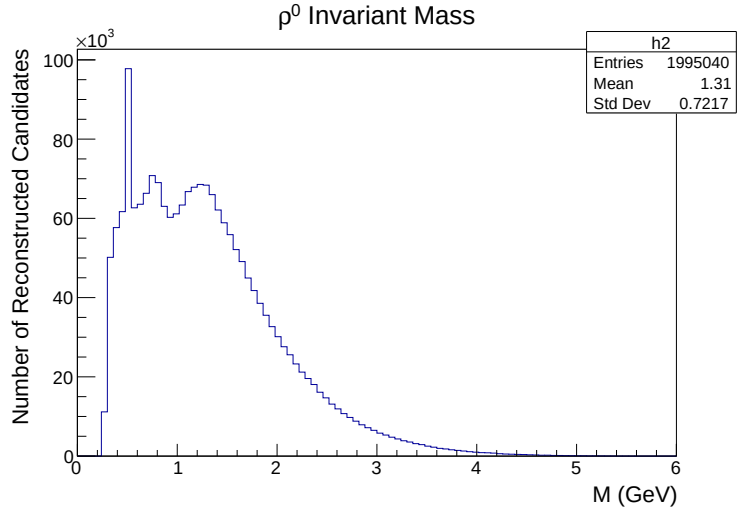
(b) Signal



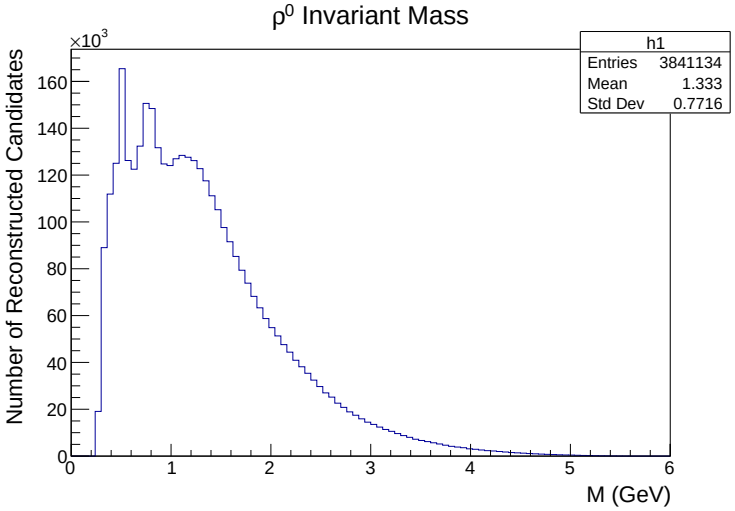
(c) Charged



(d) Mixed



(e) $c\bar{c}$



(f) uds

Figure 5.1: Distributions of the ρ^0 invariant mass in MC truth and reconstruction from signal and multiple background samples. All final-state pion cuts have been applied.

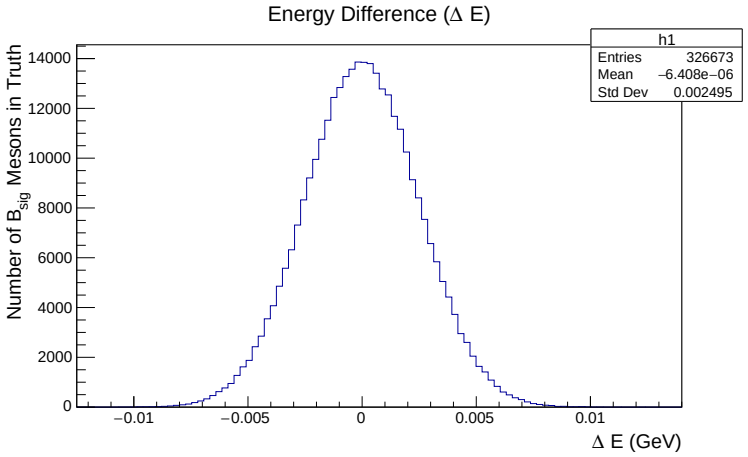
Simply applying the mask does not affect the efficiency or purity of the B_{sig} candidates themselves, but rather only modifies their corresponding ROE objects. However, selections can be made on B_{sig} candidates according to the contents of their ROE, either with or without the application of a mask. This is especially relevant to the determination of the final two selections made in the reconstruction of the B_{sig} decay chain. These are cuts on two key variables associated with Belle II B-meson analysis, the beam-constrained mass M_{bc} and the energy difference ΔE, introduced in Chapter 4.

As the chosen B_{sig} decay mode is semi-leptonic, $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$, the undetected neutrino carries energy and momentum away from the system. This results in some discrepancy when calculating M_{bc} and ΔE, as these variables depend on the B-meson energy and momentum which are in turn determined by the 4-momentum vectors of the B-meson daughters. In order to compensate for the unknown neutrino 4-momentum in reconstructed decays involving missing energy, alternative variables for M_{bc} and ΔE are included in BASF2, namely `correctedB_mbc` and `correctedB_deltaE`, which use ROE information to determine the neutrino 4-momentum. The configuration of the electron-positron beam is such that the 4-momentum of the initial Υ(4S) resonance is well-known, and thus, through 4-momentum conservation for a particular B_{sig} candidate:

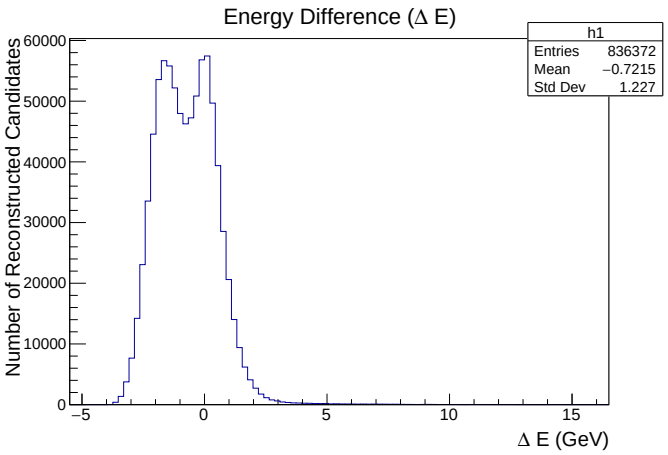
$$\mathbf{p}_{miss} = \mathbf{p}_{\Upsilon(4S)} - \mathbf{p}_{ROE} - \mathbf{p}_{reconstructed}$$

where $\mathbf{p}_{miss}$ is the missing 4-momentum in the event ascribed to the neutrino, $\mathbf{p}_{\Upsilon(4S)}$ is the 4-momentum of the initial Υ(4S), $\mathbf{p}_{ROE}$ is the sum of the 4-momentum vectors of all objects included in the B_{sig} ROE, and $\mathbf{p}_{reconstructed}$ is the sum of the 4-momentum of all reconstructed B_{sig} daughters. The corrected M_{bc} and ΔE variables are then calculated using the B-meson energy and momentum determined by the addition of the $\mathbf{p}_{reconstructed}$ and $\mathbf{p}_{miss}$ 4-vectors.

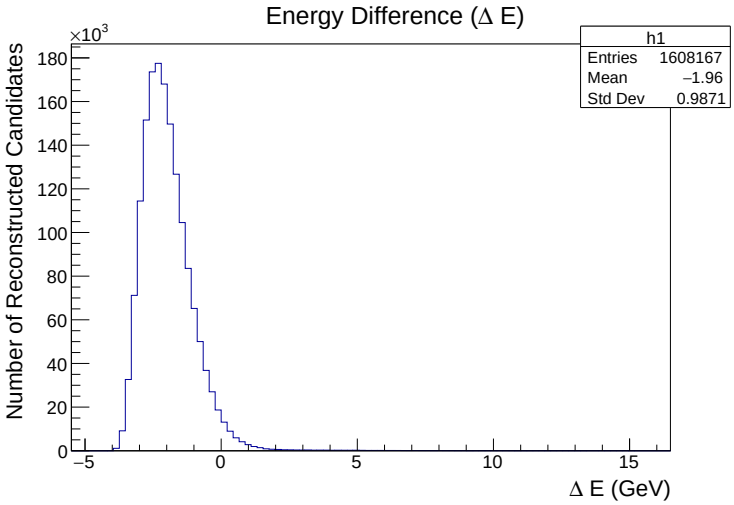
Figure 5.2 displays the ΔE distributions for the signal and background samples used, with all previous selections including the final-state particle and ρ^0 invariant mass cut applied. The standard ΔE variable is used for the MC truth distribution in 5.2(a) in which the neutrino is present, with the corrected version of ΔE plotted for the B_{sig} reconstruction in signal, 5.2(b), followed by the various backgrounds. Additionally, all reconstruction plots contain multiple B_{sig} candidates per event, with values of ΔE extending far beyond the plotted range as a result of beam background contamination in the B_{sig} ROE. Thus, a cut of $\Delta E_{corrected} < 15$ GeV was applied to each reconstructed plot in order to isolate the structure surrounding the signal region. As can be seen through the scales of the x-axes in the truth and reconstructed plots, the beam background contamination was quite substantial, with the ΔE distribution in truth defined over a very narrow range. In particular, for the B_{sig} candidates reconstructed from signal MC in Figure 5.2(b), this background contamination was manifested in the observed dual peak structure. Correctly reconstructed candidates satisfying the `isSignalAcceptMissingNeutrino == 1` condition populate the right-most peak, with the additional peak at energy differences of -2 GeV and below consisting almost entirely of fake candidates. The final ΔE selection thus decided on is listed below.



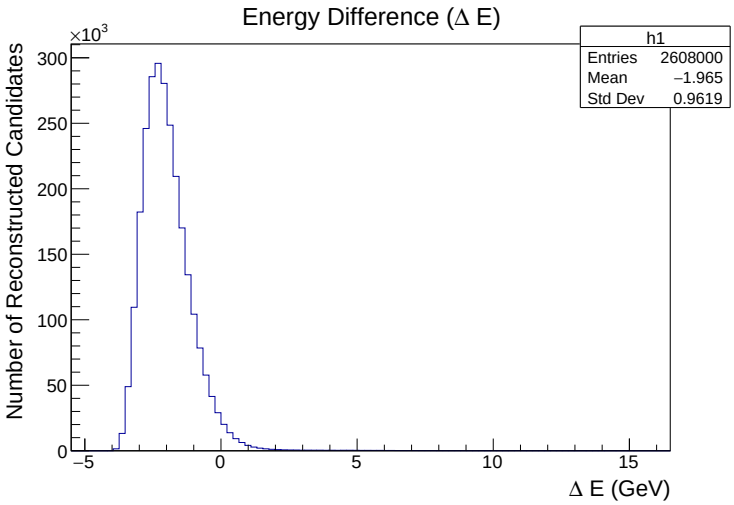
(a) MC truth



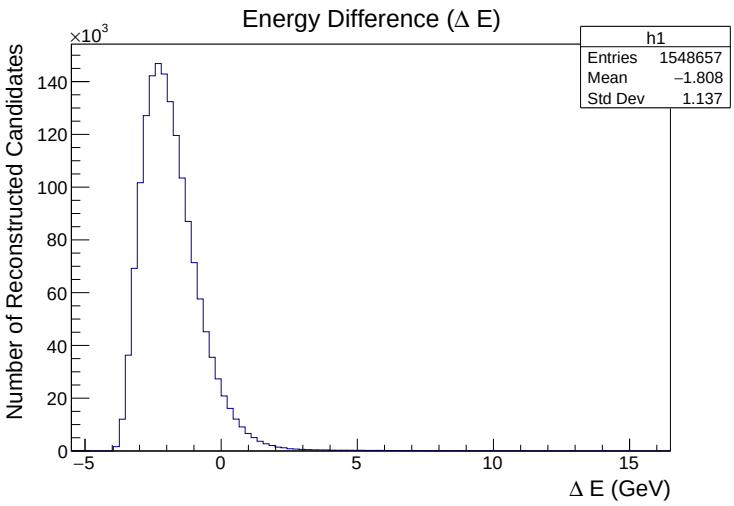
(b) Signal



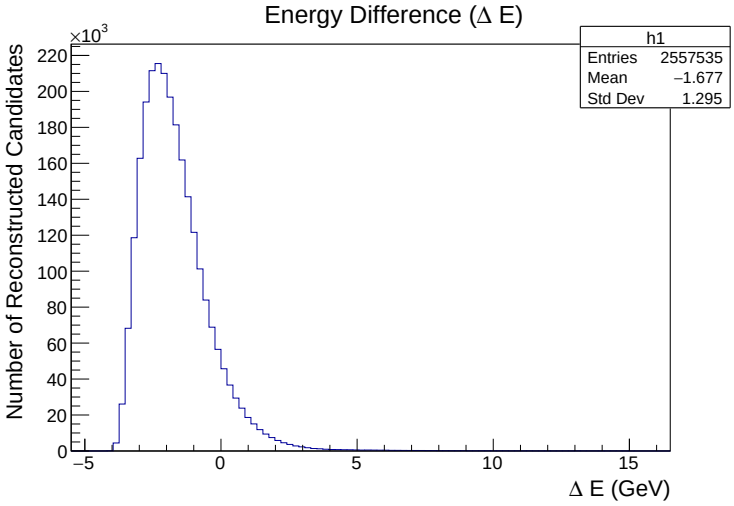
(c) Charged



(d) Mixed

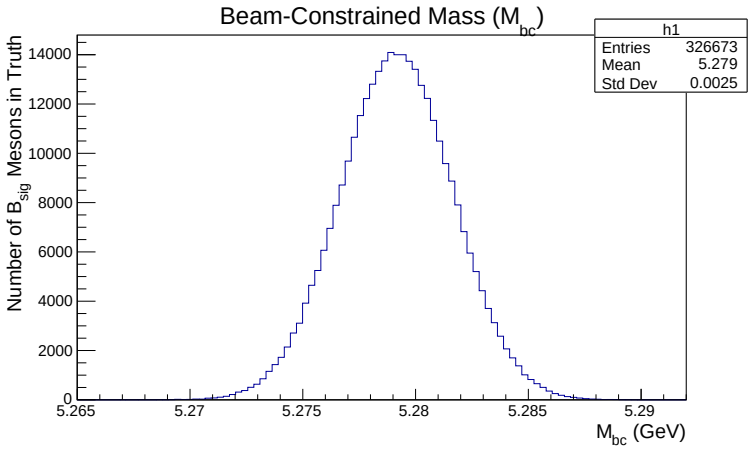


(e) $c\bar{c}$

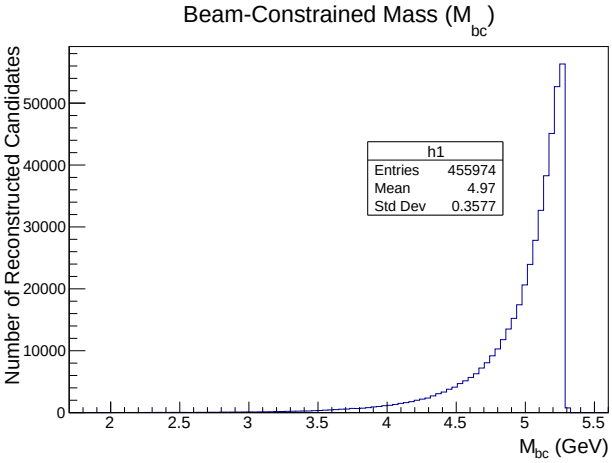


(f) uds

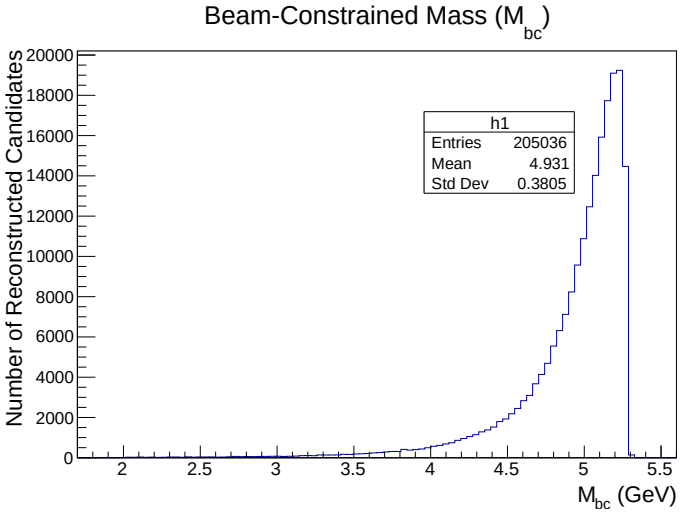
Figure 5.2: ΔE distributions for MC truth and reconstruction from signal and background. Reconstruction plots have been corrected for the missing neutrino 4-momentum, and only the $\Delta E < 15$ GeV range is shown.



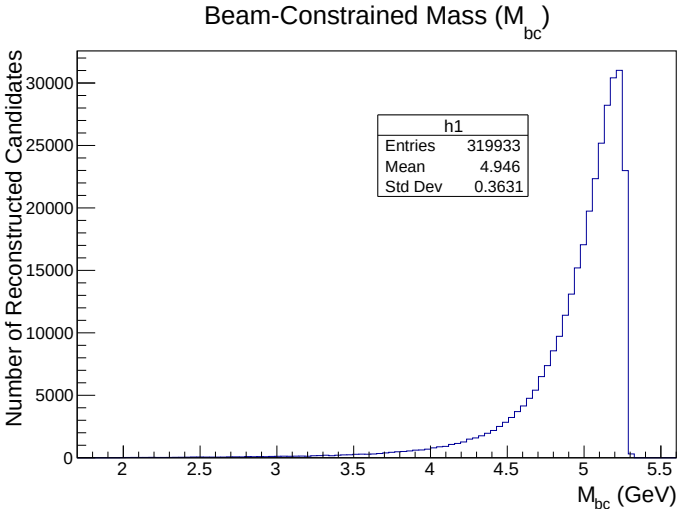
(a) MC truth



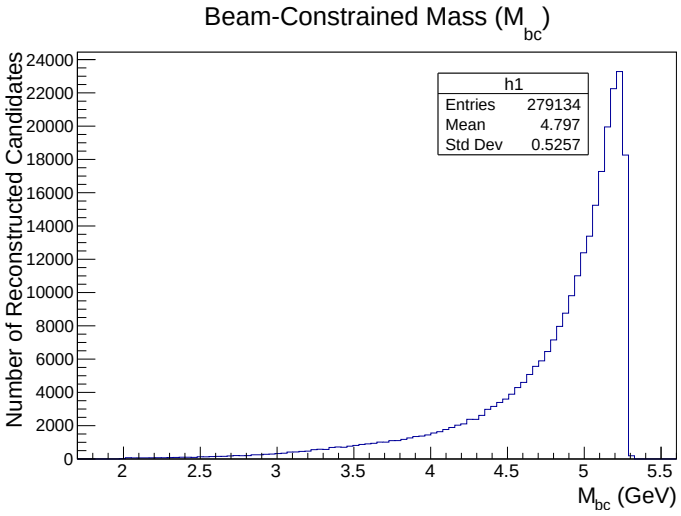
(b) Signal



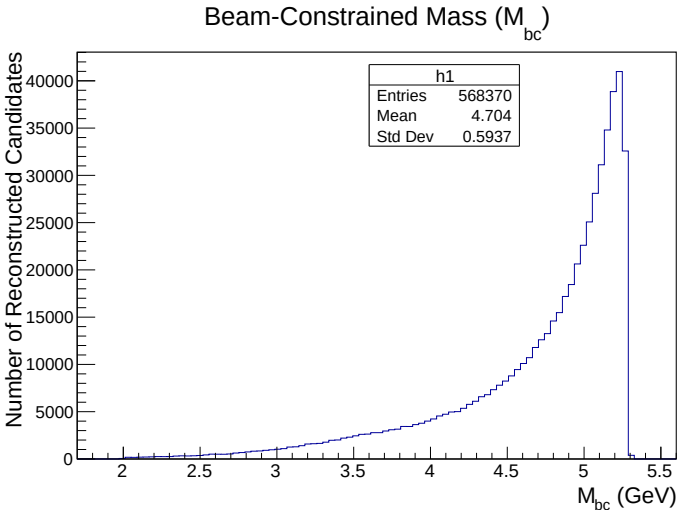
(c) Charged



(d) Mixed



(e) $c\bar{c}$



(f) uds

Figure 5.3: M_{bc} distributions for MC truth and reconstruction from signal and background. Reconstruction plots have been corrected for the missing neutrino 4-momentum, and only the $M_{bc} > 2$ GeV range is shown.

Corrected B_{sig} energy difference cut:

- -1.0 < ΔE (corrected) < 2.0 GeV

Finally, the distributions of the beam-constrained mass M_{bc} are shown in Figure 5.3, again with all preceding selections including the corrected ΔE cut implemented and the standard and corrected M_{bc} variables used for the truth and reconstructed distributions, respectively. Similarly, a cut of M_{bc} (corrected) > 2 GeV was applied to the reconstruction plots to better observe the distribution close to the signal region. A similar structure to the equivalent ΔE distributions was seen, whereby the narrow M_{bc} range characteristic of the truth was considerably broadened in the reconstruction, as evidenced by the different scales. An appropriate cut was subsequently chosen for the B_{sig} beam-constrained mass.

Corrected B_{sig} beam-constrained mass cut:

- 5.0 < M_{bc} (corrected) < 5.3 GeV

5.4 B_{sig} Reconstruction Efficiency and Purity

The effect of the successive cuts made throughout the reconstruction of the B_{sig} decay chain on both signal and background samples is summarised in Table 5.1, which quotes the total number of reconstructed B_{sig} candidates at each stage. Implementing each cut substantially reduces the number of fake B_{sig} candidates reconstructed from background as well as the incorrectly reconstructed B_{sig} candidates from signal events, whilst retaining a moderate amount of true signal candidates. After all cuts have been made, a total of 285 952 B_{sig} candidates remain in signal, corresponding to 141 251 individual events out of the total 326 673 events included in the sample. Of these candidates, 69 274 in 56 449 individual events can be matched to a true B_{sig} particle in the MC truth using the `isSignalAcceptMissingNeutrino == 1` matching condition. The final reconstruction efficiency and purity were thus calculated to be 17.28 ± 0.07 % and 39.96 ± 0.13 %, respectively.

It should be noted that these selections were chosen to provide a reasonable B_{sig} reconstruction for input into the training of the specific FEI, and it was never intended for these selections to be solely responsible for the rejection of the various backgrounds. In fact, the classifier output of the training of the FEI, the `SignalProbability`, forms one of the major discriminating variables between signal and background and the effect of `SignalProbability` cuts on the total $\Upsilon(4S)$ reconstruction efficiency, purity and background rejection is investigated extensively in the coming chapters.

Selection	Signal $B^+ \to \rho^0 \mu^+ \nu_\mu$	Signal $B^+ \to \rho^0 \mu^+ \nu_\mu$ (MC truth-matched)	Charged $B^+ B^-$	Mixed $B^0 \bar{B}^0$	Continuum $c\bar{c}$	Continuum uds
All final-state particle cuts	1701705	84518	2830166	4516078	3076124	5020701
ρ^0 invariant mass cut	836774	77220	1609226	2609985	1549863	2560085
Corrected B_{sig} energy difference cut	456575	75365	205734	320992	280841	573012
Corrected B_{sig} beam-constrained mass cut	285952	69274	117412	187020	134807	240778

Table 5.1: Cutflow table demonstrating the number of reconstructed B_{sig} candidates from signal and background MC samples after successive selections.

Chapter 6

FEI in Practice: Semi-Leptonic Decay $B^+ \to \rho^0 \mu^+ \nu_\mu$

Since its development, the generic Full Event Interpretation tagging algorithm has been trained multiple times on MC samples containing pairs of charged and neutral B-mesons, with the results of these trainings accessible to users on the KEK computing system, KEKCC. Consequently, the generic FEI is currently utilised for a number of ongoing analyses within the Belle II collaboration. The rationale behind the development of the specific FEI, however, was to provide the opportunity to train an unbiased tagging algorithm tailored towards individual analyses. Despite this, at present the performance of the specific FEI has only been investigated largely in the context of rare leptonic decays, namely $B^+ \to \tau^+ \nu_\tau$ [10, 34] and $B^+ \to \ell^+ \nu_\ell \gamma$ [11]. Evaluating the performance of the specific FEI when applied to semi-leptonic decays is hence both new and significant, with rare semi-leptonic decays such as $B^+ \to \rho^0 \ell^+ \nu_\ell$ also being sensitive to new physics. Furthermore, the previous leptonic analyses have been conducted primarily on Belle MC and data that have been converted to a format compatible with BASF2, where the beam background is substantially lower than the expected levels for the higher luminosity design of Belle II. Examining the effectiveness of the specific FEI on Belle II MC samples produced alongside this considerable beam background is therefore also quite a crucial exercise. The analysis presented throughout the remainder of this chapter aims to both validate and investigate the general performance of the FEI technique, particularly through the comparison of the specific and generic FEI for the signal decay $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$. This represents one of two modes for the $B^+ \to \rho^0 \ell^+ \nu_\ell$ decay, with $\ell = e$ or μ. The inclusion of both channels would have corresponded to double the required signal samples for the specific FEI training, and it was decided to use the muon channel as the working example.

6.1 FEI Training

6.1.1 Generic FEI

The generic FEI utilised in the following analysis was trained centrally as per the method described in Section 4.5.1, with the resulting weight files obtained from KEKCC. All MC samples used as input to the training were taken from the seventh official Belle II Monte Carlo campaign, MC7, and were subjected to the standard estimate of the beam background, BGx1. This amounted to 180 million events in total with the following composition:

- 90 million charged $\Upsilon(4S) \rightarrow B^+B^-$ events, with B^+/B^- decaying via any allowed modes.
- 90 million mixed $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ events, with $B^0/\bar{B}^0$ decaying via any allowed modes.

6.1.2 Specific FEI

The training of the specific FEI was conducted personally, following the technique detailed in Section 4.6.1. The B_{sig} decay chain reconstruction used for the training was identical to the method outlined in Chapter 5, with all selections in place. Two further event-level selections were applied for the purpose of reducing the computing time required by the training, and these were:

- An initial skim to remove any event with greater than 12 tracks satisfying the impact parameter conditions, $d_r < 2$ cm, and $|d_z| < 4$ cm. As discussed several times prior, these constraints on the track impact parameters correspond to charged particles originating close to the interaction point and thus likely resulting from B-meson decay as opposed to beam background processes. Given that only 3 tracks in total were anticipated from the B_{sig} decay (the muon and two pions), any $\Upsilon(4S)$ decay event with 12 or more such tracks was expected to have suffered from misidentification or otherwise incorrect reconstruction, and these events were hence discarded pre-training so as not to waste unnecessary computing time.
- A further related skim was enforced as a user cut to the B_{tag} lists generated by the FEI algorithm, a concept introduced in Section 4.3.3. For each B_{tag} candidate built, a cut of $3 \leqslant$ `nRemainingTracksInEvent` $\leqslant 7$ was applied, which simply retained only those events containing a B_{tag} and between 3 and 7 additional tracks, corresponding to those having originated from B_{sig}. As mentioned, only 3 B_{sig} tracks were expected, and this was therefore a substantially broad selection chosen primarily to reduce computing time and discard events in which less than the required three tracks assigned to B_{sig} were present.

All MC samples used for the training were similarly inclusive of the standard beam background estimate, BGx1, and were either taken from the official MC7 production or generated independently where required. The size of the samples for use in the specific FEI training was the subject of particular consideration throughout the course of this study. As very few

analyses utilising the specific FEI had thus far been performed, none of which using beam-background Belle II MC, the optimal sample size was not well established. Additionally, as mentioned, the specific FEI algorithm is trained on only a subset of the supplied events determined by the B_{sig} reconstruction efficiency. By this very nature of the specific FEI, defining a general recommended size for the signal and background training samples is non-trivial due to the differences in these reconstruction efficiencies for various B_{sig} decay modes or assumptions of the beam background. The samples used were ultimately decided on after considering the sizes chosen for the existing specific FEI analyses [10, 11] in comparison with the generic FEI, as well as the feasibility given the available computing resources and time constraints. In this way, investigating the effectiveness of the specific FEI given the training sample sizes was itself a meaningful comparative test against the generic FEI, as the total number of events used for the training exceeded the generic FEI by over 100 million, and comprised the following:

- 100 million signal $\Upsilon(4S) \to B^+_{sig}B^-$ events, with $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ and B^- decaying via any allowed mode. These events were generated personally as the desired sample size was considerably higher than the number of events available in MC7.
- 90 million charged $\Upsilon(4S) \to B^+B^-$ background events, with B^+/B^- decaying via any allowed modes, taken from MC7.
- 90 million mixed $\Upsilon(4S) \to B^0\bar{B}^0$ background events, with $B^0/\bar{B}^0$ decaying via any allowed modes, taken from MC7.
- 4 million charged $\Upsilon(4S) \to B^+B^-$ events, with $B^+ \to X^0_u\ell^+\nu_\ell$ and B^- decaying via any allowed mode. Here, X^0_u refers to any neutral meson containing an up quark u or antiquark $\bar{u}$, such as the ρ^0 meson or neutral pion π^0 (each with quark content $u\bar{u}$), with $\ell = e$, μ, or τ. An additional source of background for $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ is formed by other semi-leptonic decays of this type, as the similar nature of the decay and its products can result in frequent misidentification. A relatively small number of these background events was included in the training, as these were taken from MC7 where only a small quantity was available.
- 3 million mixed $\Upsilon(4S) \to B^0\bar{B}^0$ events, with $B^0 \to X^-_u\ell^+\nu_\ell$ and $\bar{B}^0$ decaying via any allowed mode. This is the cross-feed background analogous to the above for the neutral B-meson case, with X^-_u referring to charged mesons containing an up quark u or anti-quark $\bar{u}$, such as the charged ρ^- meson with quark content $\bar{u}d$. These events were similarly obtained from MC7.

6.2 Applying the FEI

The trained generic and specific FEI were subsequently applied to both signal and background MC samples inclusive of BGx1 beam background. These samples were previously introduced in Chapter 5, and consisted of the following:

- 326 673 signal $\Upsilon(4S) \to B^+_{sig}B^-$ events, with $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ and B^- decaying via any allowed mode.

- 1 million charged $\Upsilon(4S) \to B^+B^-$ background events, with B^+/B^- decaying via any allowed modes.
- 1 million mixed $\Upsilon(4S) \to B^0\bar{B}^0$ background events, with $B^0/\bar{B}^0$ decaying via any allowed modes.
- 1 million $c\bar{c}$ continuum background events, where $e^+e^- \to c\bar{c}$.
- 1 million uds continuum background events, where $e^+e^- \to u\bar{u}$, $d\bar{d}$ or $s\bar{s}$.

As detailed in Sections 4.5.2 and 4.6.2, the application stage of the FEI differs between the specific and generic configurations, with the specific FEI reconstructing B_{tag} mesons only within the rest-of-event objects of those B_{sig} candidates that have been correctly reconstructed. The B_{tag} and B_{sig} candidates can then be combined to create candidates for the parent $\Upsilon(4S)$, thereby reconstructing the entire event. Conversely, the generic FEI reconstructs both the B_{tag} and B_{sig} candidates independently within the same analysis path, with $\Upsilon(4S)$ candidates built from allowed combinations of these daughter candidates.

Despite the notable differences between the execution of the specific and generic FEI both at the training and application stages, the selections employed during the FEI application were consistent between the two methods and included all of the B_{sig} reconstruction selections outlined in Chapter 5 together with the initial 12-track skim and B_{tag} user cut described in the above section in order to minimise computing time. Both the hadronic and semi-leptonic tagging capabilities of the FEI were utilised in this study, with $\Upsilon(4S)$ candidates built from the combinations of B_{sig} with each B_{tag} type, henceforth referred to as B_{had} and B_{SL}. Furthermore, for each tagging method, after the reconstruction of the entire event, only the $\Upsilon(4S)$ candidate built from the B_{tag} daughter with the highest value of the FEI classifier output, the `SignalProbability`, was kept. In other words, only one $\Upsilon(4S)$ candidate remained per event, this being the one built from the B_{tag} daughter deemed by the FEI algorithm to be the most likely correctly reconstructed.

6.3 Comparison of Performance

When aiming to compare the effectiveness of the specific and generic FEI, it is necessary to define the key indicators of performance upon which such a comparison should be based. The analysis presented in the coming sections expresses the relative performance between these two methods in the context of the following primary features:

- The number of B_{tag} decay channels for which a multi-variate classifier was successfully trained by the FEI.
- The proportion of these channels with reconstructed candidates possessing non-zero values of the classifier output, the `SignalProbability`.
- The reconstruction efficiency and purity of the parent $\Upsilon(4S)$ meson determined in the application the FEI to signal MC events.

- The rejection of the various MC backgrounds achieved by selections on the `SignalProbability` enforced at the application stage.

6.3.1 Decay Channel Breakdown

Upon its completion, the FEI training produces a report that details, following the hierarchical structure of the algorithm, whether a multi-variate classifier was successfully trained for each final-state particle and subsequent decay channel, leading ultimately to the classifiers trained for the parent B-meson tagging channels. In order for a classifier to be trained for a certain decay, a minimum of 500 events in which the channel has been reconstructed and successfully truth-matched via the relevant `isSignal` variable is required, together with at least another 500 events in which the reconstruction failed the truth-matching condition. In this way, the classifier is trained with information pertaining to the circumstances by which a certain channel may either be correctly or incorrectly reconstructed. If the algorithm fails to train a classifier for a particular channel due to low statistics, it is excluded from the list of channels available for reconstruction at the application stage of the FEI.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0\pi^+$	$(4.81 \pm 0.15) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0\pi^+\pi^0$	(1.34 ± 0.18) %	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^0\pi^0$	$< 0.98\%$ *	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-$	$(5.7 \pm 2.2) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-\pi^0$	> 1.5 % *	✓	✓	✓	✓
$B^+ \to \bar{D}^0 D^+$	$(3.8 \pm 0.4) \times 10^{-4}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^+ K^0_S$	$(1.55 \pm 0.21) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^+ K^0_S$	$(2.1 \pm 0.5) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^{*+} K^0_S$	$(3.8 \pm 0.4) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^{*+} K^0_S$	$(9.2 \pm 1.2) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^0 K^+$	$(1.45 \pm 0.33) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^0 K^+$	$(2.26 \pm 0.23) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^{*0} K^+$	$(6.3 \pm 0.5) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^{*0} K^+$	(1.12 ± 0.13) %	✓	✓	✗	✗

cont.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to D_S^+ \bar{D}^0$	$(10.0 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} \pi^+$	$(5.18 \pm 0.26) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} \pi^+ \pi^0$	$(9.8 \pm 1.7) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} \pi^+ \pi^0 \pi^0$	5×10^{-4} *	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} \pi^+ \pi^+ \pi^-$	(1.03 ± 0.12) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} \pi^+ \pi^+ \pi^- \pi^0$	(1.8 ± 0.4) %	✓	✓	✓	✓
$B^+ \to D_S^{*+} \bar{D}^0$	$(7.6 \pm 1.6) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to D_S^+ \bar{D}^{*0}$	$(8.2 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 K^+$	$(3.65 \pm 0.33) \times 10^{-4}$	✓	✓	✗	✓
$B^+ \to D^- \pi^+ \pi^+$	$(1.07 \pm 0.05) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to D^- \pi^+ \pi^+ \pi^0$	0.20 % *	✓	✓	✓	✓
$B^+ \to J/\psi K^+$	$(1.016 \pm 0.033) \times 10^{-3}$	✗	✓	-	✗
$B^+ \to J/\psi K^+ \pi^+ \pi^-$	$(8.1 \pm 1.3) \times 10^{-4}$	✗	✓	-	✓
$B^+ \to J/\psi K^+ \pi^0$	1×10^{-4} *	✗	✓	-	✓
$B^+ \to J/\psi K_S^0 \pi^+$	0.11 % *	✗	✓	-	✓
	Total channels:	25/29	29/29	10/29	23/29

Table 6.1: The hadronic tagging channels incorporated by the FEI, with their associated branching fractions as recorded by the Particle Data Group (PDG) [12]. * indicates channels for which the branching fraction was not found in the PDG, with the quoted values taken or estimated to be those used in the generation of the standard Belle II MC7 production. The channels that passed the training and application stages of the specific and generic FEI on BGx1 MC are also indicated.

The charged B-meson tagging modes incorporated by the FEI along with their corresponding branching fractions are presented in Tables 6.1 and 6.2, for hadronic and semi-leptonic modes respectively. The results obtained from the FEI automatic reporting for the training of both the specific and generic FEI are also quoted, demonstrating whether a classifier was trained successfully for each individual channel. All 42 semi-leptonic tagging channels were successfully trained by both FEI methods. However, whilst the generic FEI was additionally capable of training a classifier for each of the 29 hadronic channels, the specific FEI managed only to successfully train a total of 25 hadronic channels. Interestingly, though one of these failed channels, $B^+ \to J/\psi K^+ \pi^0$, possesses the smallest branching fraction of the included hadronic modes, the remaining three failed modes each have branching fractions higher than at least one other hadronic mode that was successfully trained. As the

number of correctly reconstructed events for a particular channel depends not only on its branching fraction but also on its reconstruction efficiency, this observation is not exceedingly unusual.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0 e^+$	(2.26 ± 0.11) %	✓	✓	✓	✓
$B^+ \to \bar{D}^0 \mu^+$	(2.26 ± 0.11) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} e^+$	(5.70 ± 0.19) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} \mu^+$	(5.70 ± 0.19) %	✓	✓	✗	✓
$B^+ \to D^- \pi^+ e^+$	$(4.2 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^- \pi^+ \mu^+$	$(4.2 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^{*-} \pi^+ e^+$	$(6.1 \pm 0.6) \times 10^{-3}$	✓	✓	✓	✗
$B^+ \to D^{*-} \pi^+ \mu^+$	$(6.1 \pm 0.6) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} \pi^+$	0.8×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} \pi^+ \pi^0$	0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} \pi^+ \pi^0 \pi^0$	$< 0.16\%$	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} \pi^+ \pi^+ \pi^-$	0.9×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} \pi^+ \pi^+ \pi^- \pi^0$	> 0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} D^+$	0.6×10^{-4}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^+ K^0_S$	0.2×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0}_{SL} D^+ K^0_S$	0.3×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^{*+} K^0_S$	0.6×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0}_{SL} D^{*+} K^0_S$	1.5×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^0 K^+$	0.2×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} D^0 K^+$	0.4×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^{*0} K^+$	1.0×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} D^{*0} K^+$	0.2 %	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^+_{SL}$	1.0×10^{-4}	✓	✓	✗	✓

cont.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0 D^+_{SL} K^0_S$	0.4×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^+_{SL} K^0_S$	0.5×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^{*+}_{SL} K^0_S$	0.7×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^{*+}_{SL} K^0_S$	1.7×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^0_{SL} K^+$	0.2×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^0_{SL} K^+$	0.4×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^{*0}_{SL} K^+$	1.0×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^{*0}_{SL} K^+$	0.2 %	✓	✓	✗	✗
$B^+ \to D^+_S \bar{D}^0_{SL}$	1.6×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+$	0.8×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^0$	1.6×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^0 \pi^0$	0.8×10^{-4}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^+ \pi^-$	0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^+ \pi^- \pi^0$	0.3 %	✓	✓	✗	✓
$B^+ \to D^{*+}_S \bar{D}^0_{SL}$	1.2×10^{-3}	✓	✓	✗	✓
$B^+ \to D^+_S \bar{D}^{*0}_{SL}$	1.3×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^0_{SL} K^+$	0.6×10^{-4}	✓	✓	✗	✓
$B^+ \to D^-_{SL} \pi^+ \pi^+$	0.3×10^{-3}	✓	✓	✓	✓
$B^+ \to D^-_{SL} \pi^+ \pi^+ \pi^0$	0.1 %	✓	✓	✓	✓
	Total channels:	42/42	42/42	16/42	28/42

Table 6.2: The semi-leptonic tagging channels reconstructed by the FEI. Undetected neutrinos are omitted from the above descriptions of the former 8 channels, with the associated branching fractions shown obtained from the PDG [12]. The remaining 34 channels involving intermediate semi-leptonic D-meson decays are paired with approximate branching fractions calculated from the values taken from the PDG or estimated to be those used in the generation of the standard Belle II MC7 production. The channels that passed the training and application stages of the specific and generic FEI on BGx1 MC are also indicated.

Given the knowledge of the decay channels for which a classifier was successfully trained by each of the two FEI methods, the FEI was first applied to the designated signal sample and the distributions of the classifier output, the `SignalProbability`, were examined for each individual B_{tag} channel. An unexpected outcome was subsequently observed whereby a number of B_{tag} channels, despite being successfully trained, possessed values of `SignalProbability` that were exactly zero for every reconstructed B_{tag} candidate within the channel. Conversely, channels that did not exhibit this behaviour contained B_{tag} candidates each with non-zero

values for the `SignalProbability`. In other words, each particular B_{tag} decay channel corresponded to either exclusively a zero or non-zero `SignalProbability` for all reconstructed candidates, with no overlap observed between these two cases within any one channel. This ultimately led to the definition of an additional failure condition for each tagging mode manifested at the application stage of the FEI. That is, only B_{tag} channels consisting of reconstructed candidates with non-zero values of the `SignalProbability` were deemed to be a successful application of the FEI, as channels with exclusively zero values were unable to provide any insight into the likelihood of correct reconstruction for any one B_{tag} candidate. Whilst the exact cause behind this observation is unknown, a likely hypothesis is that these failed channels were able to meet the threshold above which a classifier was trained for the channel, but contained too few statistics for the classifier to be trained adequately.

Tables 6.1 and 6.2 also report the evaluation of this secondary failure condition for each individual hadronic and semi-leptonic B_{tag} mode, indicating whether a channel passed or failed the application of the FEI to the aforementioned signal sample. It is here that a significant discrepancy between the performance of the specific and generic FEI arises. In first considering the B_{had} tagging modes, despite 25 out of the total 29 channels having been successfully trained in the specific FEI, only 10 of these channels subsequently passed the application stage. This is in stark contrast to the generic FEI, with only 6 trained channels failing the FEI application, leaving a total of 23 hadronic channels passing both the training and application. This trend is reflected in the semi-leptonic channels, with the application of the specific and generic FEI resulting in 16 and 28 successful channels respectively, despite each method being capable of training a classifier for each of the 42 B_{SL} channels. Although considerably more events were used as input into the training of the specific FEI, the generic FEI was consistently able to take advantage of almost double the tagging channels incorporated within the FEI framework.

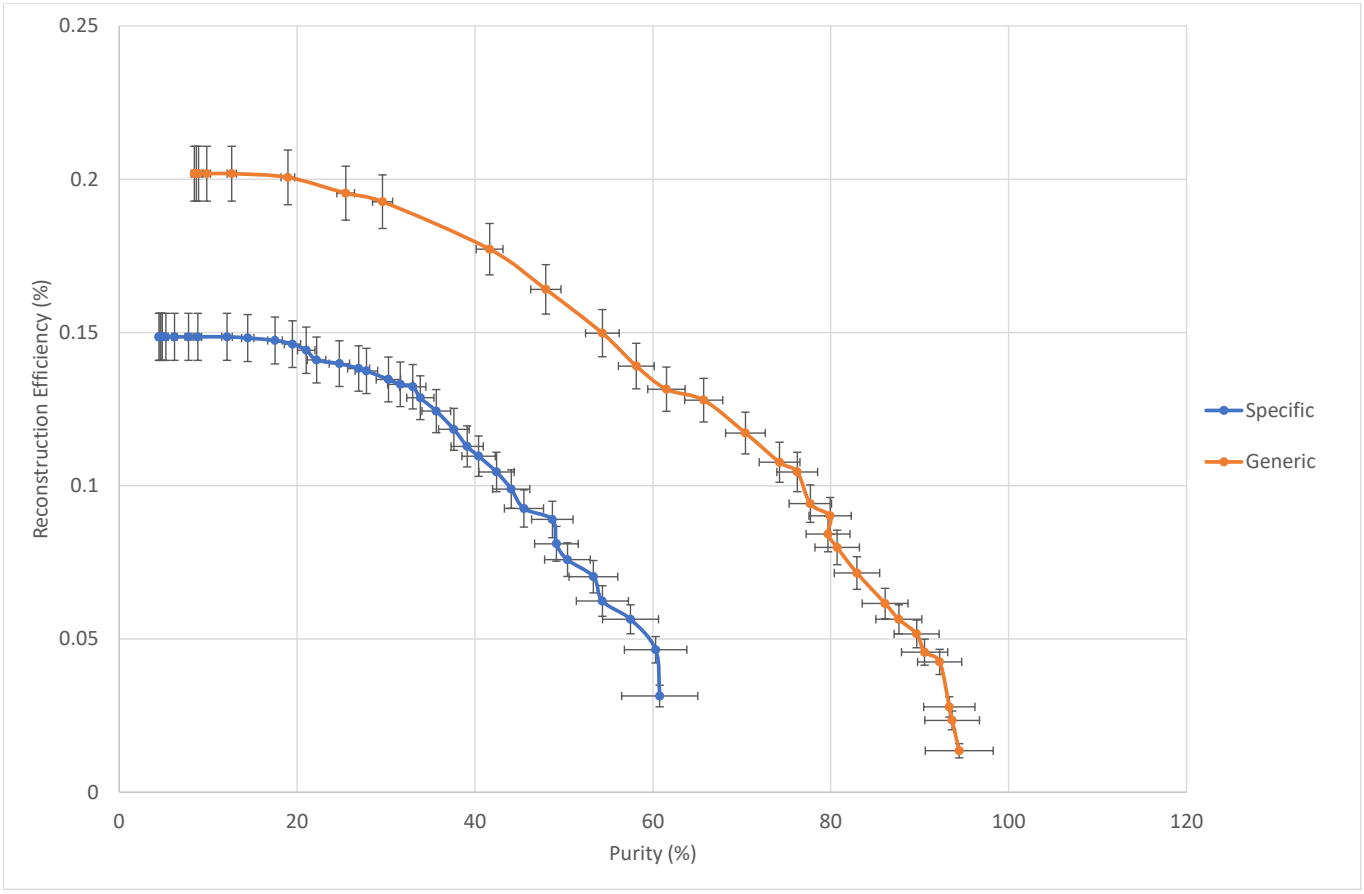
6.3.2 Performance on Signal $B^+ \to \rho^0 \mu^+ \nu_\mu$

In the application of the FEI to a signal MC sample, a reconstruction efficiency and purity for the entire $\Upsilon(4S)$ event can be determined, analogous to those defined for the B_{sig} in Chapter 5:

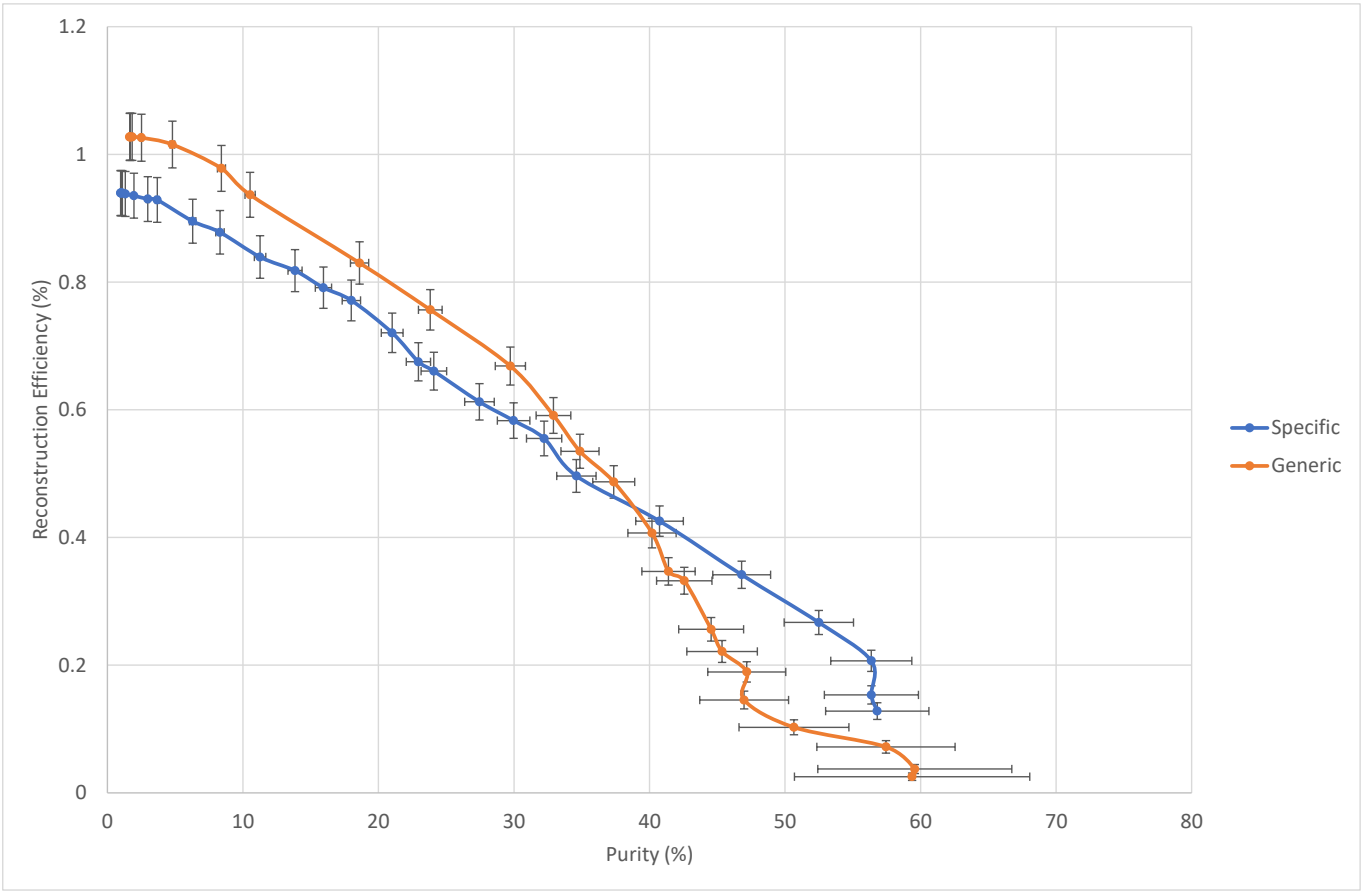
$$\textbf{Reconstruction efficiency} \equiv \frac{\text{Number of events with a } \textit{correctly} \text{ reconstructed } \Upsilon(4S) \text{ candidate}}{\text{Number of total events}}$$

$$\textbf{Purity} \equiv \frac{\text{Number of events with a } \textit{correctly} \text{ reconstructed } \Upsilon(4S) \text{ candidate}}{\text{Number of events with reconstructed } \Upsilon(4S) \text{ candidates}}$$

A wide range of B_{tag} `SignalProbability` cuts were applied to the output of the FEI application, with the reconstruction efficiency and purity of the parent $\Upsilon(4S)$ meson calculated after each successive cut. Scanning over the range of applied cuts between `SignalProbability` > 0 and `SignalProbability` > 0.95, a total of 36 individual points were sampled.



(a) Hadronic tagging



(b) Semi-leptonic tagging

Figure 6.1: Efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed from B_{sig} and B_{tag} daughters in BGx1 signal MC.

Table 6.3 demonstrates an example of this procedure evaluated at the initial `SignalProbability` cut of > 0, listing the number of corresponding reconstructed and correct candidates and the resultant efficiencies and purities calculated. The reconstruction efficiencies were then plotted against the purities for both the specific and generic FEI, and these distributions are displayed in Figures 6.1(a) and 6.1(b) for $\Upsilon(4S)$ mesons built from hadronic and semi-leptonic tags, respectively.

Examining these distributions in the case of hadronic tagging reveals a sizable disparity between the specific and generic FEI. At common values of the purity, the generic reconstruction efficiency exceeds that of the specific FEI, with equivalent efficiencies between the two methods similarly corresponding to higher purities in the generic FEI. Moreover, this trend is observed consistently over the entire sampled range, with the specific FEI unable to reach the level of performance achieved by the generic FEI at any point.

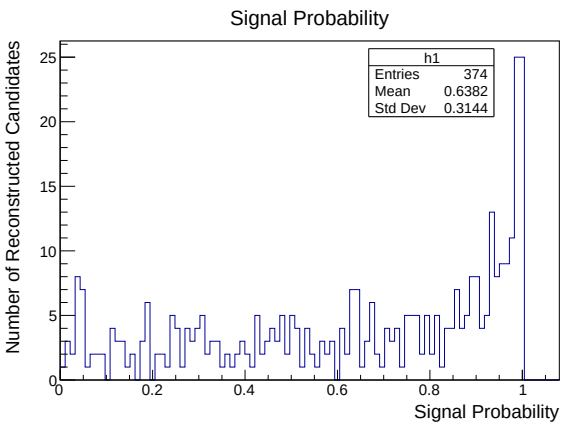
`SignalProbability` > 0	B_{sig}	B_{had}	$\Upsilon(4S)_{had}$	B_{SL}	$\Upsilon(4S)_{SL}$
Number of events in MC truth	326673	251729		74944	
Specific FEI					
Number of reconstructed candidates	285952	145910	8382	3240438	72133
Number of correctly reconstructed candidates	69274	1470	374	1106	704
$\Upsilon(4S)$ reconstruction efficiency			**(0.15± 0.01) %**		**(0.94 ± 0.04) %**
Purity			**(4.46 ± 0.23) %**		**(0.98 ± 0.04) %**
Generic FEI					
Number of reconstructed candidates	285952	574348	6014	3743973	46776
Number of correctly reconstructed candidates	69274	4431	508	4861	770
$\Upsilon(4S)$ reconstruction efficiency			**(0.20 ± 0.01) %**		**(1.03 ± 0.04) %**
Purity			**(0.98 ± 0.04) %**		**(1.65 ± 0.06) %**

Table 6.3: The number of B_{sig}, B_{tag} and $\Upsilon(4S)$ candidates reconstructed and truth-matched with the application of the FEI to signal BGx1 MC for a single chosen `SignalProbability` cut. The subsequent values for the $\Upsilon(4S)$ reconstruction efficiency and purity are also included.

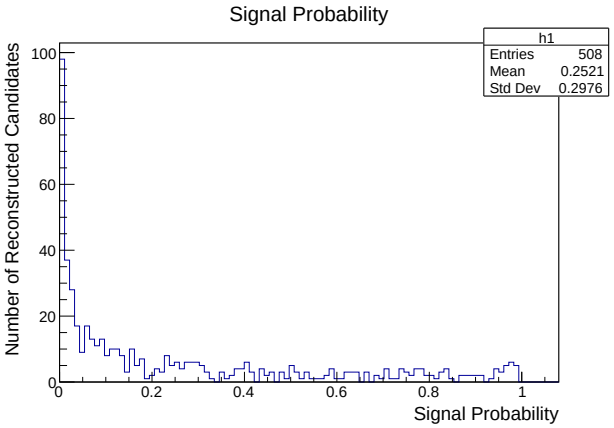
For the signal events built from semi-leptonic tags, the relative under-performance of the specific FEI described above is observed at lower purity values, where the reconstruction efficiencies of the generic FEI surpass those of the specific FEI. However, at higher purities corresponding to tighter cuts on the B_{SL} `SignalProbability`, the specific FEI efficiency overtakes that of the generic FEI, with more correctly reconstructed $\Upsilon(4S)$ candidates built from B_{SL} daughters surviving these harsher selections. This can be seen explicitly in Figure 6.2, which portrays the `SignalProbability` distributions of the B_{tag} daughter candidates from correctly reconstructed $\Upsilon(4S)$ signal events, for both the hadronically and semi-leptonically tagged specific and generic FEI. Only $\Upsilon(4S)$ candidates truth-matched via

the `isSignalAcceptMissingNeutrino` condition with B_{tag} daughters possessing non-zero `SignalProbability` values, thus arising only from channels passing both the FEI training and application stages, are included in each plot.

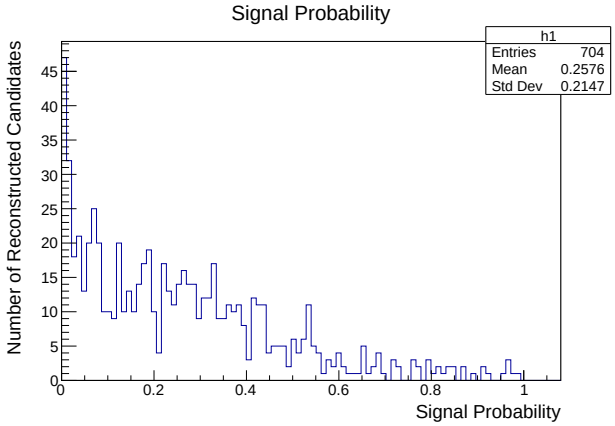
The specific FEI corresponds to `SignalProbability` distributions skewed further towards 1.0 than the generic FEI for both hadronic and semi-leptonic tagging modes, and in both cases, more correctly reconstructed $\Upsilon(4S)$ candidates survive harsher `SignalProbability` cuts in the specific FEI when compared with the generic FEI. However, a significant number of incorrectly reconstructed $\Upsilon(4S)$ candidates also exist with B_{had} daughters possessing these higher `SignalProbability` values, resulting in lower purities that bring down the overall relative performance of the specific FEI seen in Figure 6.1. In contrast, the specific FEI is able to rise above the generic FEI in the efficiency vs. purity curves for the semi-leptonic tagging modes as tighter `SignalProbability` cuts retain a higher number of correctly reconstructed $\Upsilon(4S)$ candidates than the generic FEI, whilst also excluding enough incorrect candidates to surpass the generic FEI in purity.



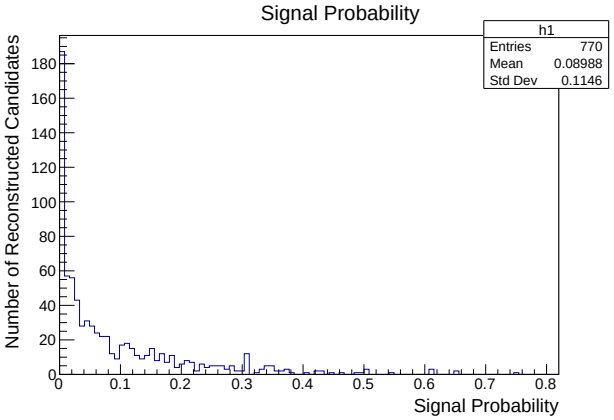
(a) Specific FEI, hadronic tagging



(b) Generic FEI, hadronic tagging



(c) Specific FEI, semi-leptonic tagging



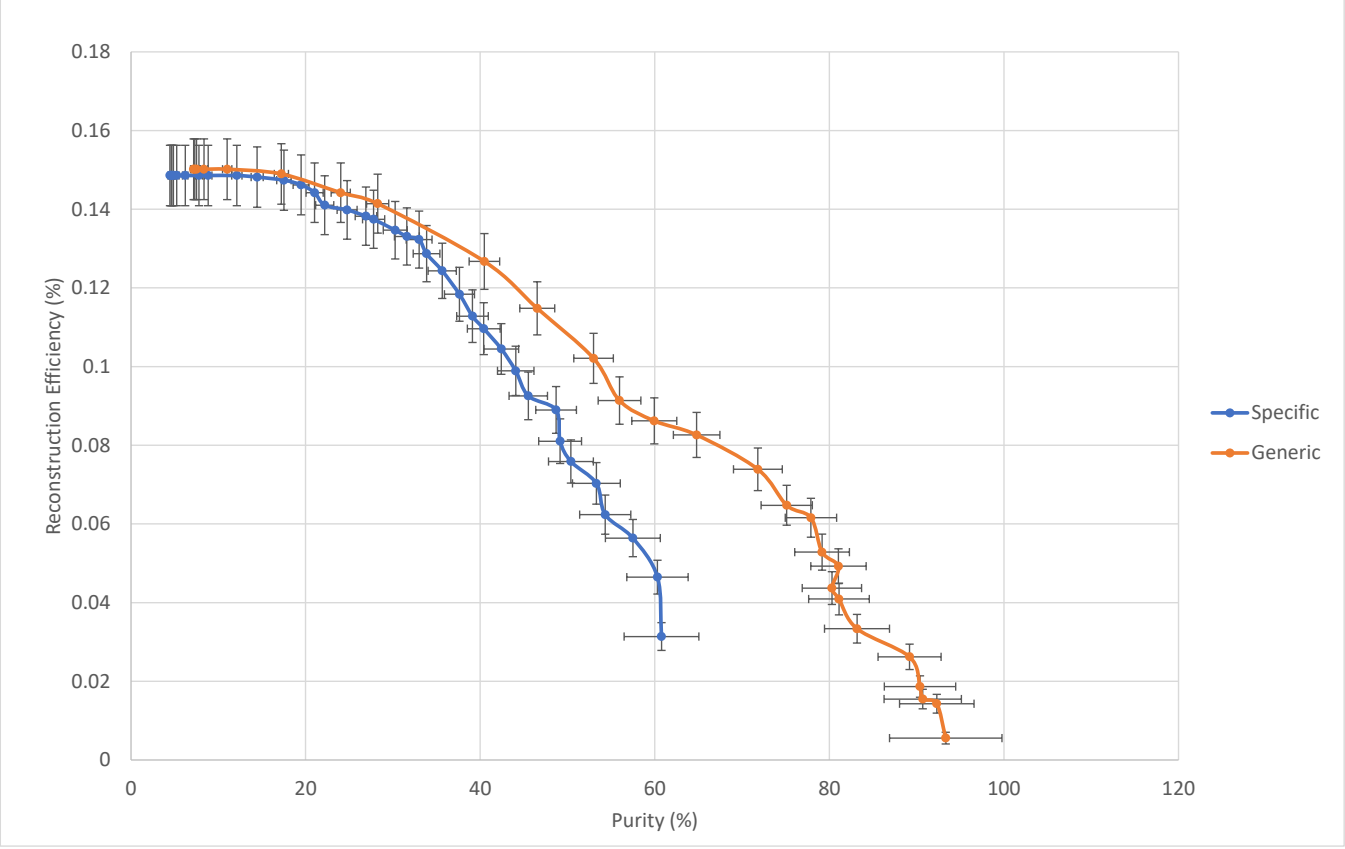
(d) Generic FEI, semi-leptonic tagging

Figure 6.2: Distributions of the `SignalProbability` classifier output for correctly reconstructed $\Upsilon(4S)$ candidates built from B_{tag} daughters with `SignalProbability` > 0, in BGx1 signal MC.

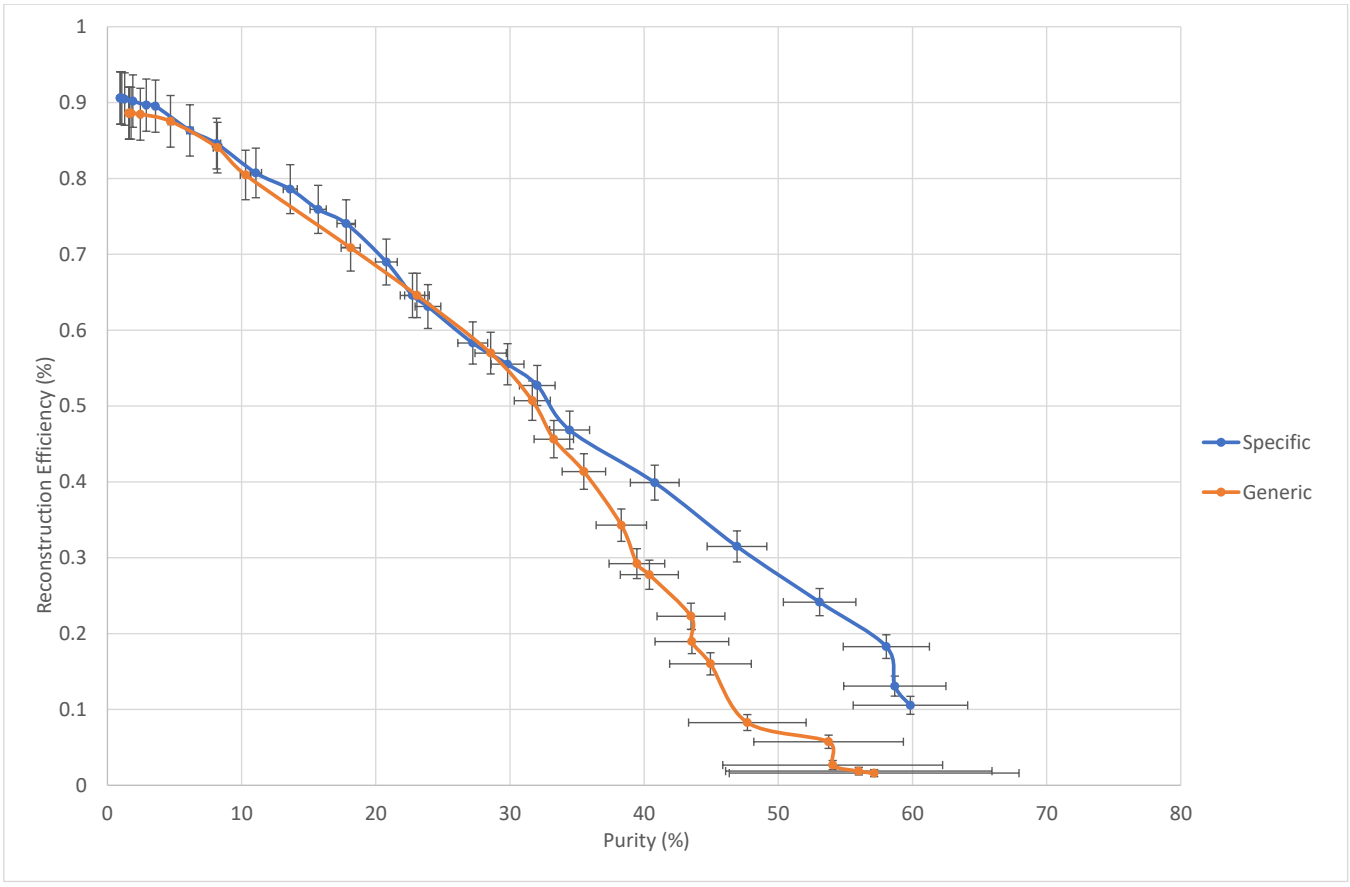
In addition, for all common purities, the reconstruction efficiencies obtained using semi-leptonic tagging exceed the equivalent efficiencies achieved via hadronic tagging by at least a factor of 2 with the specific FEI. This is an unsurprising feature given the nature of hadronic tagging described in Section 4.0.1, with the absence of tag-side neutrinos requiring a complete reconstruction of the B_{tag} decay, resulting in a higher purity at the cost of efficiency. However, this trend exists for the generic FEI only up to purities of approximately 40% where the specific FEI performance begins to exceed that of the generic FEI in Figure 6.1(b), after which the reconstruction efficiencies are comparable between the hadronic and semi-leptonic tagging cases. This suggests a semi-leptonic tagging capability of the specific FEI superior to that for hadronic tagging, when compared with the generic FEI.

The efficiency vs. purity curves in Figure 6.1 represent the level of performance capable from the specific and generic FEI when trained using the aforementioned MC sample sizes. Since $\Upsilon(4S)$ candidates built from all tagging channels failing the FEI application are absent from these curves due to the enforced `SignalProbability` > 0 condition, the distributions for the generic FEI are constructed using a larger number of daughter B_{tag} channels than present for the specific FEI, in both the hadronic and semi-leptonic cases. It was subsequently decided to investigate the relationship between the $\Upsilon(4S)$ reconstruction efficiency and purity for each method if only the common B_{tag} channels passing the FEI training and application stages for both the specific and generic FEI were included. This would provide a rough estimate of the relative performance of the specific and generic FEI if each were trained using the statistics necessary to result in the same proportion of channels successfully being trained and passing the FEI application. For B_{had} daughters, all successful channels in the specific FEI were likewise successful in the generic FEI, and including only common B_{had} channels would thus simply exclude additional channels from the generic FEI, reducing its efficiency. Figure 6.3(a) illustrates the efficiency vs. purity curves for $\Upsilon(4S)$ candidates built from B_{had} daughters reconstructed within only the ten channels common between the specific and generic FEI. Despite the generic FEI efficiency drop due to the exclusion of multiple channels, it remains superior to the specific FEI over the majority of the observed purity range, with the two methods comparable below purities of about 40%. Hence, a greater level of performance for the specific FEI with respect to the generic FEI is not obtained even when considering only mutually successful B_{had} tagging channels.

For the semi-leptonic modes, as per Table 6.2, all successful B_{SL} channels after application via the specific FEI were similarly successful in the application of the generic FEI, save for a single channel, $B^+ \to D^{*-}\pi^+e^+$, that failed the generic FEI application. This failure was investigated and found to be the result of a software bug in the FEI package whereby the output `SignalProbability` was neither zero nor non-zero, but rather was assigned an unphysical value. Re-defining the efficiency vs. purity curves for only common B_{SL} channels thus only slightly modified the specific FEI distribution, whilst excluding 12 additional channels used to reconstruct $\Upsilon(4S)$ candidates in the generic FEI. The updated distributions for the semi-leptonic modes are shown in Figure 6.3(b). The performance of the specific FEI still exceeds the generic FEI only at higher purities above approximately 35%. However, below this range, the specific FEI no longer falls behind the generic FEI and the two methods are consistent.



(a) Hadronic tagging



(b) Semi-leptonic tagging

Figure 6.3: Efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed in BGx1 signal MC, including only channels mutually successful between the specific and generic FEI.

6.3.3 Background Rejection

The application of the FEI to signal MC and consequent determination of the $\Upsilon(4S)$ reconstruction efficiencies and purities demonstrated the capability of the trained specific and generic FEI to discriminate between $\Upsilon(4S)$ candidates that could be matched to particles in the truth, and the background arising from incorrectly reconstructed candidates. Equally as important, however, is the ability of the `SignalProbability` classifier output to reject fake $\Upsilon(4S)$ candidates reconstructed from the various background samples. The ultimate goal of the FEI lies in its application to real collision data containing both signal and background events, in which the optimal configuration would retain as many $\Upsilon(4S)$ candidates correctly reconstructed from B_{sig} and B_{tag} daughters as possible whilst rejecting enough background candidates such that the signal is not overwhelmed.

The background rejection was investigated through observing the number of $\Upsilon(4S)$ candidates reconstructed in each of the given background samples when scanning over the same range of `SignalProbability` cuts quoted in the previous section, with 36 points in total sampled between `SignalProbability` > 0 and `SignalProbability` > 0.95. The numbers of fake $\Upsilon(4S)$ candidates delivered by the specific and generic FEI at the `SignalProbability` cut of > 0 are shown in Table 6.4, demonstrating a larger number of fake candidates associated with the specific FEI consistent with all background MC samples. Ultimately, however, the background rejection capabilities lie in the effectiveness of successive `SignalProbability` cuts at the removal of these candidates. The background rejection achieved by a certain `SignalProbability` cut was thus explicitly defined as the following:

$$\textbf{Background Rejection} \equiv \frac{\text{Number of } \Upsilon(4S) \text{ candidates after } \texttt{SignalProbability} \text{ cut}}{\text{Number of total } \Upsilon(4S) \text{ candidates with non-zero } \texttt{SignalProbability}}$$

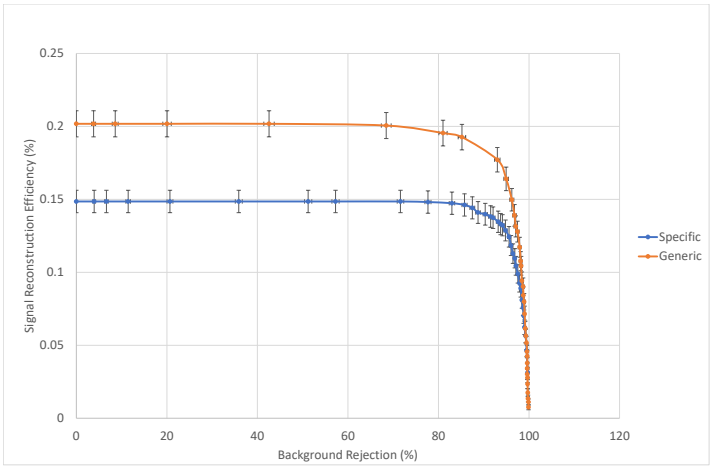
`SignalProbability` > 0	Charged		Mixed		$c\bar{c}$		uds	
Number of reconstructed $\Upsilon(4S)$ candidates	S	G	S	G	S	G	S	G
B_{had}	3888	1842	2760	1215	7613	3621	5841	3259
B_{SL}	26764	12752	22829	10235	32307	14161	21769	11542

Table 6.4: The number of fake $\Upsilon(4S)$ candidates reconstructed with the application of the specific (S) and generic (G) FEI to the various BGx1 background MC samples, for a single chosen `SignalProbability` cut.

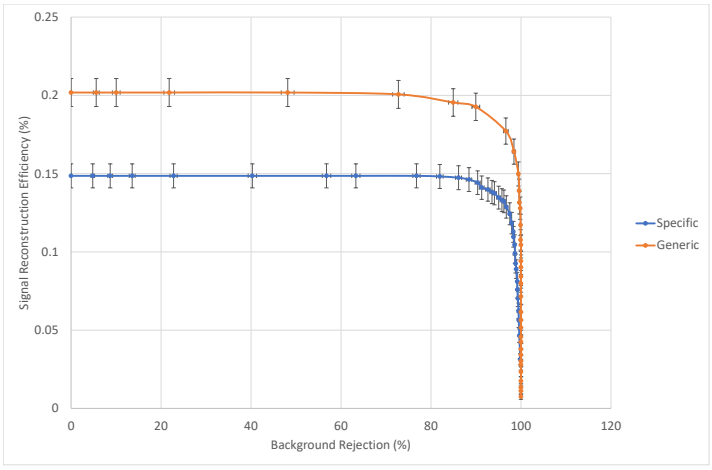
Here, only the $\Upsilon(4S)$ candidates built from B_{tag} daughters with decay modes passing both the FEI training and application conditions are included in the denominator, analogous to the efficiency vs. purity curves which similarly exclude the failed channels. Figures 6.4 and 6.5 portray the $\Upsilon(4S)$ reconstruction efficiencies from signal MC against the background rejection evaluated for each sampled cut on the `SignalProbability`, for hadronic and semi-

leptonic tagging modes, respectively. In each case, these distributions are plotted for both the specific and generic FEI applied to the chosen charged, mixed, $c\bar{c}$ and uds background samples.

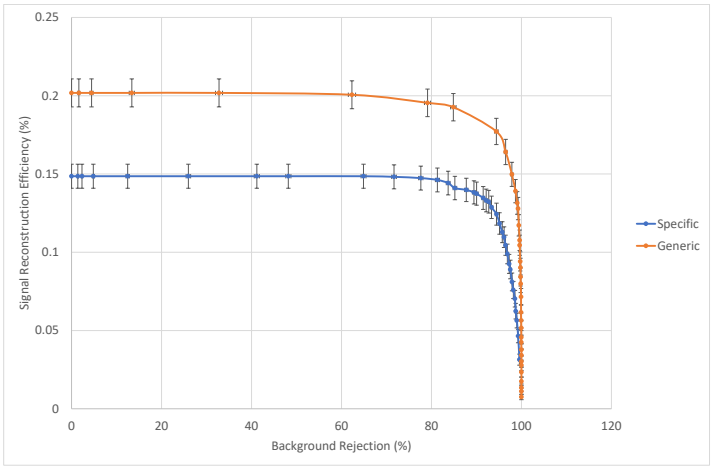
It can be clearly observed that the generic FEI consistently delivers a higher degree of performance than the specific FEI, with larger reconstruction efficiencies retained by the generic FEI at values of the background rejection equivalent between the two methods. Moreover, this trend is apparent across all backgrounds studied, considering both hadronic and semi-leptonic tagging modes. In an attempt to further investigate the relatively poor background rejection demonstrated by the specific FEI, the `SignalProbability` distributions for the B_{tag} daughters of the $\Upsilon(4S)$ candidates reconstructed from background MC were also examined. These distributions are presented for $\Upsilon(4S)$ candidates with B_{had} and B_{SL} daughters in Figures 6.6 and 6.7, respectively, with only modes possessing non-zero values of the `SignalProbability` included. Despite the substantially smaller number of tagging channels available from which to build $\Upsilon(4S)$ candidates in the specific FEI, significantly more of these fake candidates are reconstructed from background in the specific FEI when compared to the generic FEI, with `SignalProbability` values also skewed further towards 1.0, another trend observed consistently across all backgrounds. A possible explanation for this may be that the B_{sig} selections chosen (as described in Chapter 5) were somewhat deficient in their discriminating power between signal and background, as the specific FEI trained in light of these selections in many instances attributed high `SignalProbability` values to both correct and fake candidates.



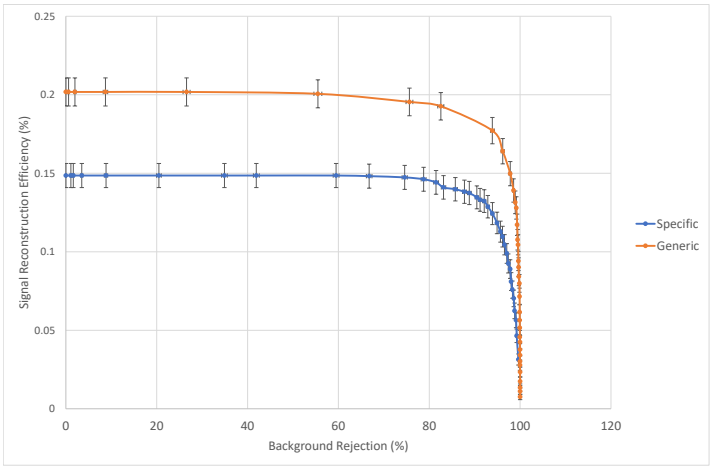
(a) Charged



(b) Mixed

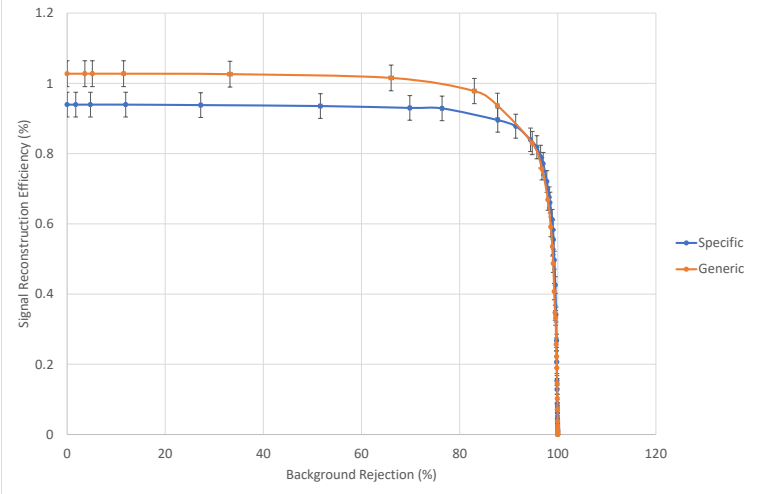


(c) $c\bar{c}$

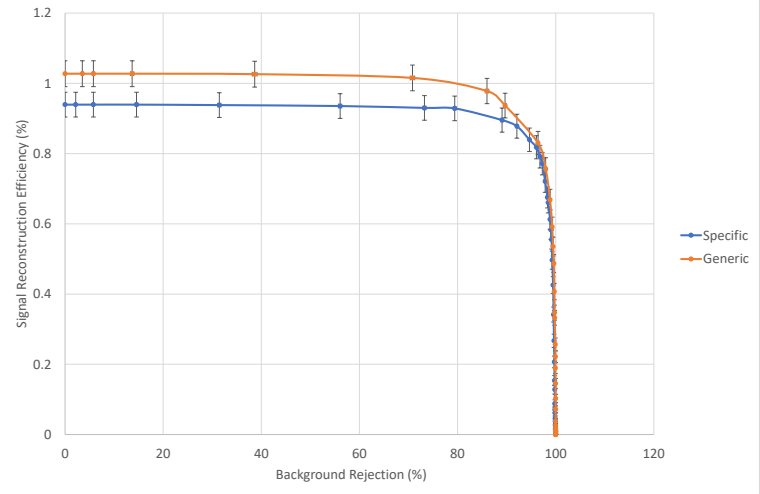


(d) uds

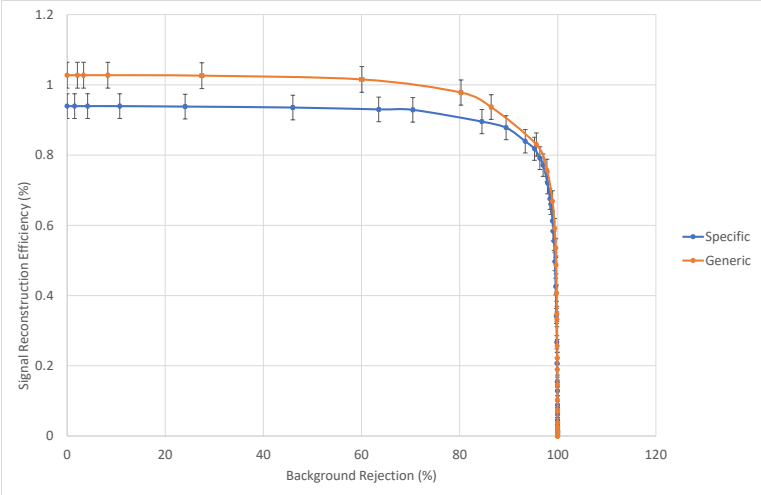
Figure 6.4: Background rejection curves for $\Upsilon(4S)$ candidates built from B_{had} daughters, in BGx1 background MC.



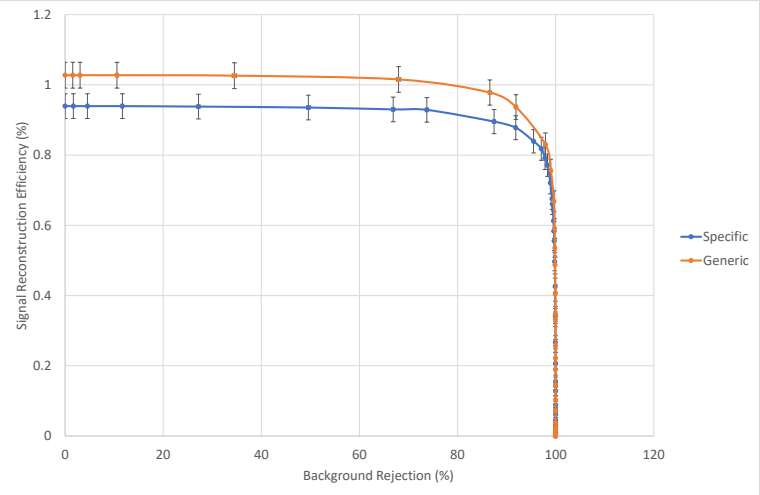
(a) Charged



(b) Mixed

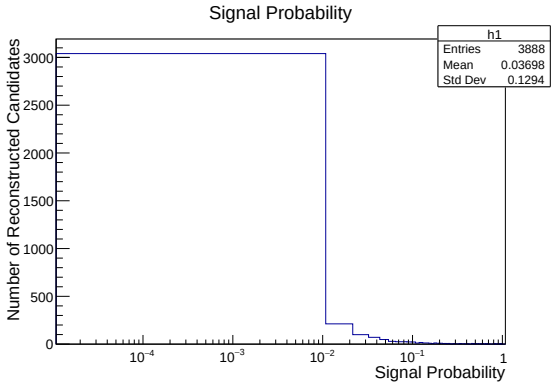


(c) $c\bar{c}$

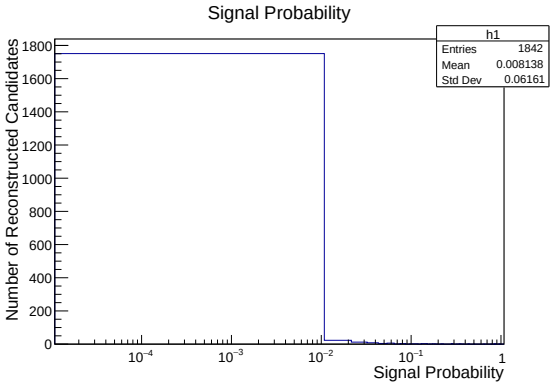


(d) *uds*

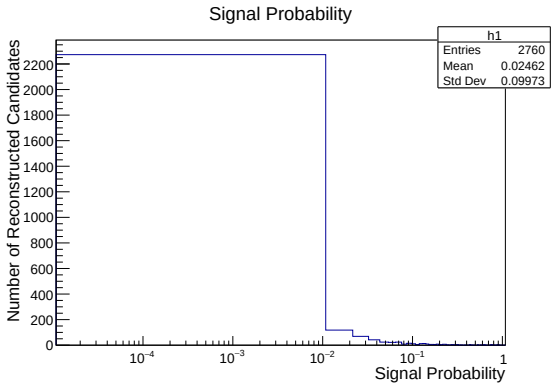
Figure 6.5: Background rejection curves for $\Upsilon(4S)$ candidates built from B_{SL} daughters, in BGx1 background MC.



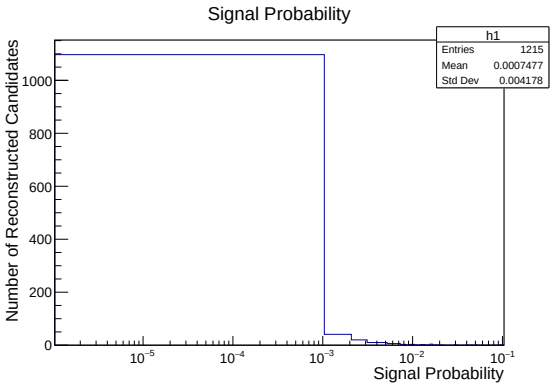
(a) Specific FEI, charged



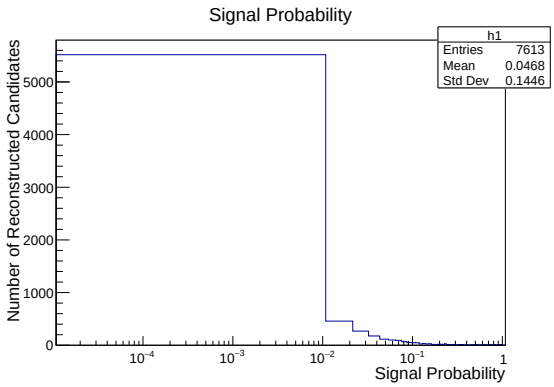
(b) Generic FEI, charged



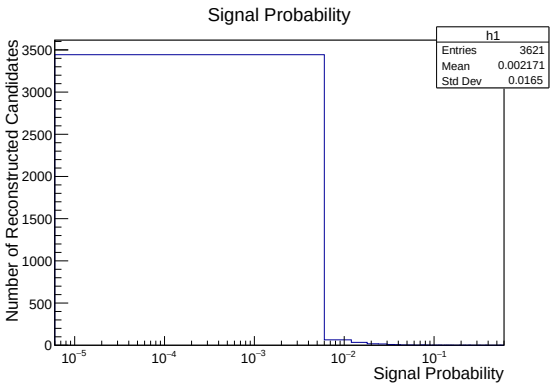
(c) Specific FEI, mixed



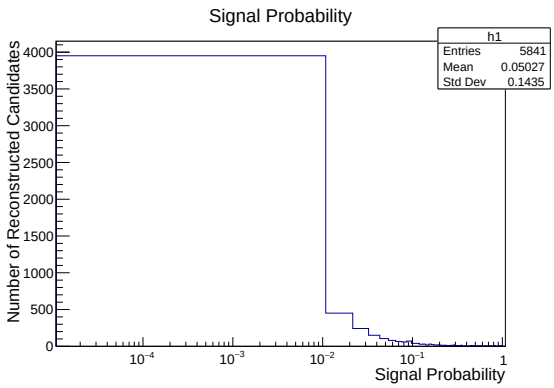
(d) Generic FEI, mixed



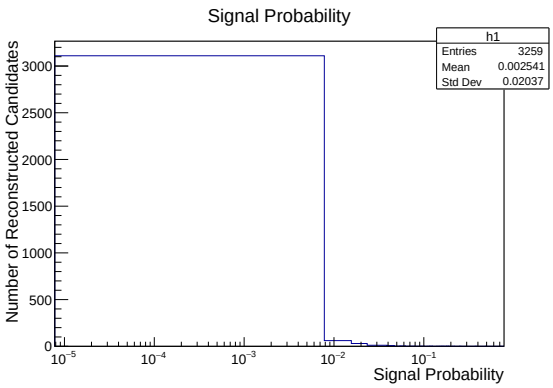
(e) Specific FEI, $c\bar{c}$



(f) Generic FEI, $c\bar{c}$

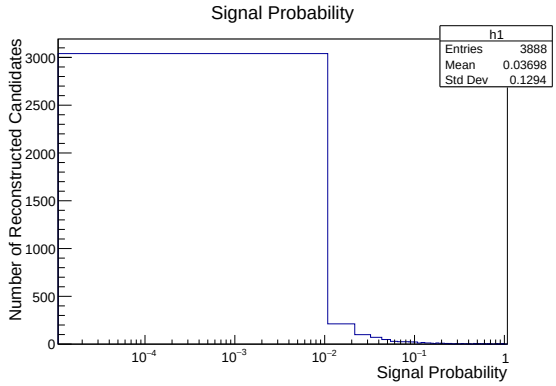


(g) Specific FEI, *uds*

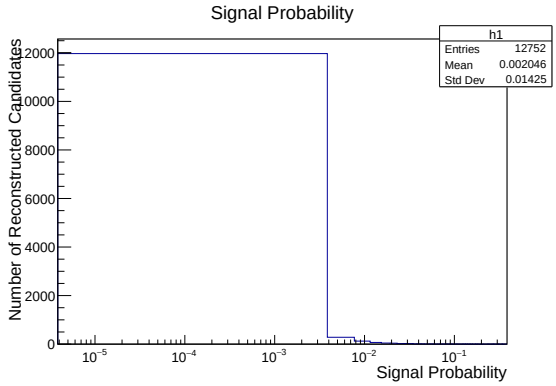


(h) Generic FEI, *uds*

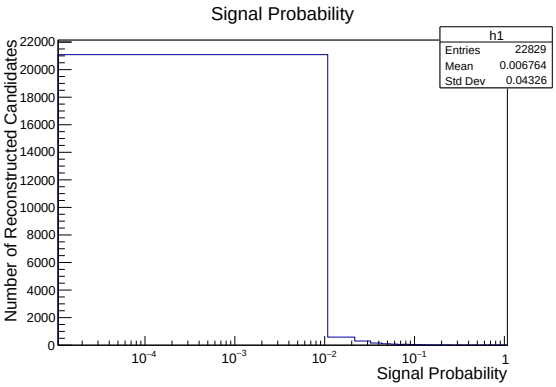
Figure 6.6: Distributions of the `SignalProbability` for fake $\Upsilon(4S)$ candidates built from B_{had} daughters with `SignalProbability` > 0, in BGx1 background MC.



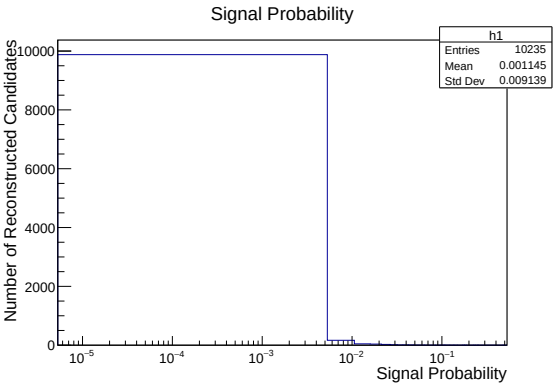
(a) Specific FEI, charged



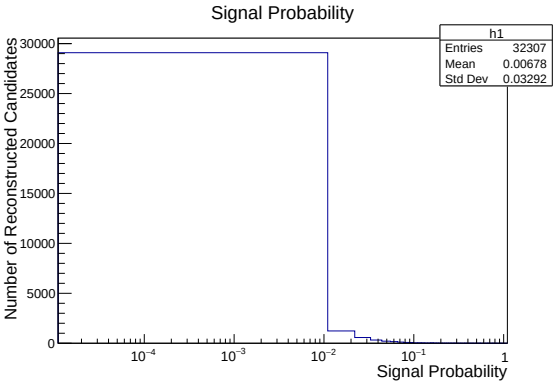
(b) Generic FEI, charged



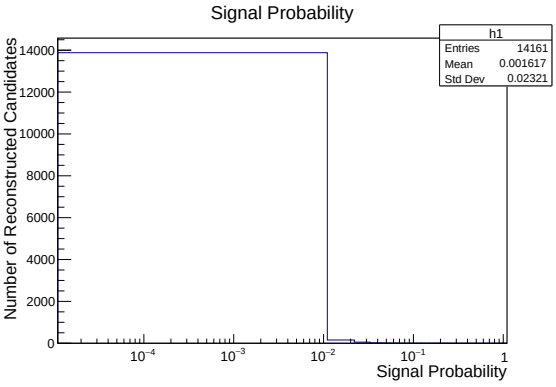
(c) Specific FEI, mixed



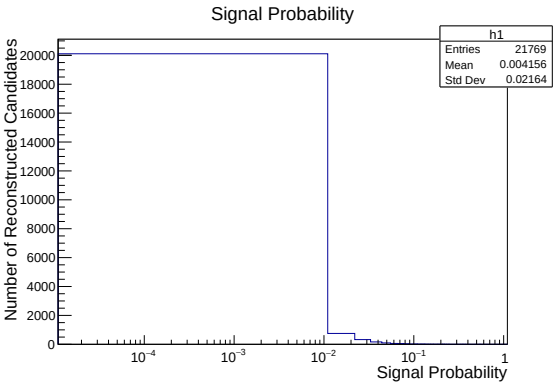
(d) Generic FEI, mixed



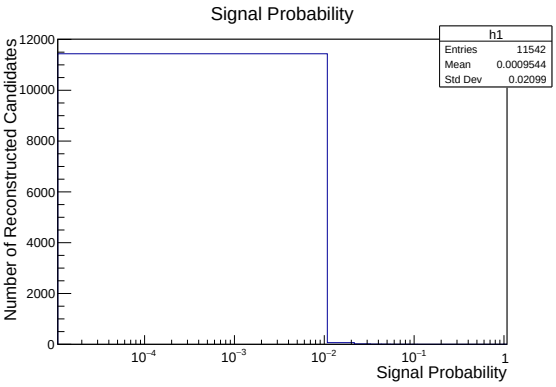
(e) Specific FEI, $c\bar{c}$



(f) Generic FEI, $c\bar{c}$



(g) Specific FEI, *uds*



(h) Generic FEI, *uds*

Figure 6.7: Distributions of the `SignalProbability` for fake $\Upsilon(4S)$ candidates built from B_{SL} daughters with `SignalProbability` > 0, in BGx1 background MC.

6.4 Performance Without Beam Background

The analysis presented thus far has described a specific FEI tailored towards the semi-leptonic decay mode of interest, $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$, that, despite being trained using significantly more statistics than the generic FEI, failed to reach its performance in a number of key areas. As mentioned, all MC used in this study was generated alongside the standard estimate of the Belle II beam-background, termed BGx1. Whilst MC produced in the absence of beam background, referred to as BGx0, cannot accurately represent the reality of the Belle II environment, analyses conducted on these samples can be meaningful in portraying the behavior expected under ideal conditions. As such, it was decided to repeat the above investigation with the omission of beam background in order to discern whether the relative performance of the specific FEI would be improved under these conditions.

The training and application of both the specific and generic FEI were conducted in an identical manner to the beam-background inclusive study, on BGx0 MC samples of the same sizes save for a few exceptions. Namely, the cross-feed backgrounds from $B^+ \to X^0_u \ell^+ \nu_\ell$ and $B^0 \to X^-_u \ell^+ \nu_\ell$ decays were excluded from the specific FEI training as only a very small proportion of these events without beam background were available in the official MC7 Monte Carlo production. The size of the signal $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$ MC sample used for the application of the FEI was also increased from 326 673 events to 1 million events, now generated personally, in order to reduce the uncertainties observed when producing the earlier efficiency vs. purity distributions.

Re-visiting the definitions of the B_{sig} reconstruction efficiency and purity introduced in Chapter 5, in the absence of beam background with identical selections on the B_{sig} decay chain applied, these quantities were calculated to be 21.14 $\pm$ 0.13 % and 42.78 $\pm$ 0.07 % respectively. Unsurprisingly due to the cleaner environment, both values exceeded those determined in the presence of beam background, a reconstruction efficiency of 17.28 $\pm$ 0.07 % and purity of 39.96 $\pm$ 0.13 % as quoted in the previous chapter. By extension, it was anticipated that the reconstruction efficiencies and purities of the individual tagging modes incorporated within the FEI would also increase, allowing each classifier to be trained on larger statistics thus leading to a greater number of channels passing both the training and application stages. It was hypothesised that this effect would be strengthened in the specific FEI due to the fact that the proportion of the number of events supplied to the training stage that subsequently get used in the training of classifiers is dependent on the B_{sig} reconstruction efficiency, which is slightly higher in the presence of no beam background.

Tables 6.5 and 6.6 display the number of respective B_{had} and B_{SL} decay channels for which a multi-variate classifier was successfully trained, together with the proportion of these channels passing the FEI application with non-zero values of the `SignalProbability`. The specific FEI remained unable to train a classifier for the latter four hadronic channels, and the total number of channels trained successfully for the specific and generic FEI were consistent with and without beam background. However, the number of B_{had} and B_{SL} channels subsequently passing the application was increased for both the specific and the generic FEI. In total, the BGx0-specific FEI successfully reconstructed one hadronic and

four semi-leptonic channels more than the BGx1 case, with the generic FEI boasting an additional three hadronic and six semi-leptonic channels. Thus, the removal of beam background ultimately improved the proportion of usable tagging channels for both methods, but this improvement was enhanced in the generic FEI, only widening the gap in performance between the two methods.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0\pi^+$	$(4.81 \pm 0.15) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0\pi^+\pi^0$	(1.34 ± 0.18) %	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^0\pi^0$	< 0.98 % *	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-$	$(5.7 \pm 2.2) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-\pi^0$	$> 1.5\%$ *	✓	✓	✓	✓
$B^+ \to \bar{D}^0 D^+$	$(3.8 \pm 0.4) \times 10^{-4}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^+ K_S^0$	$(1.55 \pm 0.21) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^+ K_S^0$	$(2.1 \pm 0.5) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^{*+} K_S^0$	$(3.8 \pm 0.4) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^{*+} K_S^0$	$(9.2 \pm 1.2) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^0 K^+$	$(1.45 \pm 0.33) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^0 K^+$	$(2.26 \pm 0.23) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^{*0} K^+$	$(6.3 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^{*0} K^+$	(1.12 ± 0.13) %	✓	✓	✗	✓
$B^+ \to D_S^+ \bar{D}^0$	$(10.0 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}\pi^+$	$(5.18 \pm 0.26) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^0$	$(9.8 \pm 1.7) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^0\pi^0$	5×10^{-4} *	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^+\pi^-$	(1.03 ± 0.12) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^+\pi^-\pi^0$	(1.8 ± 0.4) %	✓	✓	✓	✓
$B^+ \to D_S^{*+} \bar{D}^0$	$(7.6 \pm 1.6) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to D_S^+ \bar{D}^{*0}$	$(8.2 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 K^+$	$(3.65 \pm 0.33) \times 10^{-4}$	✓	✓	✗	✓

cont.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to D^- \pi^+ \pi^+$	$(1.07 \pm 0.05) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to D^- \pi^+ \pi^+ \pi^0$	0.20 % *	✓	✓	✓	✓
$B^+ \to J/\psi K^+$	$(1.016 \pm 0.033) \times 10^{-3}$	✗	✓	-	✗
$B^+ \to J/\psi K^+ \pi^+ \pi^-$	$(8.1 \pm 1.3) \times 10^{-4}$	✗	✓	-	✓
$B^+ \to J/\psi K^+ \pi^0$	1×10^{-4} *	✗	✓	-	✓
$B^+ \to J/\psi K_S^0 \pi^+$	0.11 % *	✗	✓	-	✓
	Total channels:	25/29	29/29	11/29	26/29

Table 6.5: The hadronic tagging channels incorporated by the FEI, with their associated branching fractions as recorded by the Particle Data Group (PDG) [12]. * indicates channels for which the branching fraction was not found in the PDG, with the quoted values taken or estimated to be those used in the generation of the standard Belle II MC7 production. The channels that passed the training and application stages of the specific and generic FEI on BGx0 MC are also indicated.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0 e^+$	(2.26 ± 0.11) %	✓	✓	✓	✓
$B^+ \to \bar{D}^0 \mu^+$	(2.26 ± 0.11) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} e^+$	(5.70 ± 0.19) %	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} \mu^+$	(5.70 ± 0.19) %	✓	✓	✓	✓
$B^+ \to D^- \pi^+ e^+$	$(4.2 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^- \pi^+ \mu^+$	$(4.2 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^{*-} \pi^+ e^+$	$(6.1 \pm 0.6) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^{*-} \pi^+ \mu^+$	$(6.1 \pm 0.6) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+$	0.8×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^0$	0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^0 \pi^0$	< 0.16 %	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^+ \pi^-$	0.9×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^+ \pi^- \pi^0$	> 0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 D^+$	0.6×10^{-4}	✓	✓	✗	✓
$B^+ \to \bar{D}_{SL}^0 D^+ K_S^0$	0.2×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}_{SL}^{*0} D^+ K_S^0$	0.3×10^{-3}	✓	✓	✗	✗

cont.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0_{SL} D^{*+} K^0_S$	0.6×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} D^{*+} K^0_S$	1.5×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^0 K^+$	0.2×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} D^0 K^+$	0.4×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^{*0} K^+$	1.0×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} D^{*0} K^+$	0.2 %	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^+_{SL}$	1.0×10^{-4}	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^+_{SL} K^0_S$	0.4×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^+_{SL} K^0_S$	0.5×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^{*+}_{SL} K^0_S$	0.7×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^{*+}_{SL} K^0_S$	1.7×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^0_{SL} K^+$	0.2×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^0_{SL} K^+$	0.4×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^{*0}_{SL} K^+$	1.0×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^{*0}_{SL} K^+$	0.2 %	✓	✓	✗	✗
$B^+ \to D^+_S \bar{D}^0_{SL}$	1.6×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+$	0.8×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^0$	1.6×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^0 \pi^0$	0.8×10^{-4}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^+ \pi^-$	0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^+ \pi^- \pi^0$	0.3 %	✓	✓	✓	✓
$B^+ \to D^{*+}_S \bar{D}^0_{SL}$	1.2×10^{-3}	✓	✓	✗	✓
$B^+ \to D^+_S \bar{D}^{*0}_{SL}$	1.3×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^0_{SL} K^+$	0.6×10^{-4}	✓	✓	✗	✓
$B^+ \to D^-_{SL} \pi^+ \pi^+$	0.3×10^{-3}	✓	✓	✓	✓
$B^+ \to D^-_{SL} \pi^+ \pi^+ \pi^0$	0.1%	✓	✓	✓	✓
	Total channels:	42/42	42/42	20/42	34/42

Table 6.6: The semi-leptonic tagging channels reconstructed by the FEI. Undetected neutrinos are omitted from the above descriptions of the former 8 channels, with the associated branching fractions shown obtained from the PDG [12]. The remaining 34 channels are paired with approximate branching fractions calculated from the values taken from the PDG or estimated to be those used in the generation of the standard Belle II MC7 production. The channels that passed the training and application stages of the specific and generic FEI on BGx0 MC are also indicated.

The efficiency vs. purity curves were also re-examined under these idealised conditions, and are displayed in Figure 6.8 with Table 6.7 depicting a breakdown of the candidates used for the determination of a single plotted point. As expected by the increased discrepancy observed between the number of successful channels available for the reconstruction of $\Upsilon(4S)$ candidates in the specific and generic FEI, the relative performance of the specific FEI is only worsened by the absence of beam background. The distance between the specific and generic curves for the hadronic modes seen in Figure 6.8(a) is only broadened, with the specific FEI efficiency similarly unable to reach that of the generic FEI at any purity value. In the case of the semi-leptonic modes in Figure 6.8(b), the specific FEI remains secondary to the generic FEI for an even lengthier purity range than seen with beam background, only overtaking the generic FEI at purities of above 60% compared with the earlier 40%. Continuing the trend observed in the beam background case, the reconstruction efficiencies for semi-leptonic tagging are at least twice as high than for hadronic tagging over all purities for the specific FEI, with this reflected in the generic FEI only up to the purity of $\approx$ 60% at which the specific FEI performance becomes superior.

`SignalProbability` > 0	B_{sig}	B_{had}	$\Upsilon(4S)_{had}$	B_{SL}	$\Upsilon(4S)_{SL}$
Number of events in MC truth	1000000	769528		230472	
Specific FEI					
Number of reconstructed candidates	995795	315525	26892	11383708	277366
Number of correctly reconstructed candidates	255034	6438	1790	4962	3252
$\Upsilon(4S)$ reconstruction efficiency			**(0.23 ± 0.01) %**		**(1.41 ± 0.02) %**
Purity			**(6.66 ± 0.15) %**		**(1.17 ± 0.02) %**
Generic FEI					
Number of reconstructed candidates	995795	1424379	21174	12154582	190705
Number of correctly reconstructed candidates	255034	17114	2339	18692	3629
$\Upsilon(4S)$ reconstruction efficiency			**(0.30 ± 0.01) %**		**(1.57 ± 0.03) %**
Purity			**(11.05 ± 0.22) %**		**(1.90 ± 0.03) %**

Table 6.7: The number of B_{sig}, B_{tag} and $\Upsilon(4S)$ candidates reconstructed and truth-matched with the application of the FEI to signal BGx0 MC for a single chosen `SignalProbability` cut. The subsequent values for the $\Upsilon(4S)$ reconstruction efficiency and purity are also included.

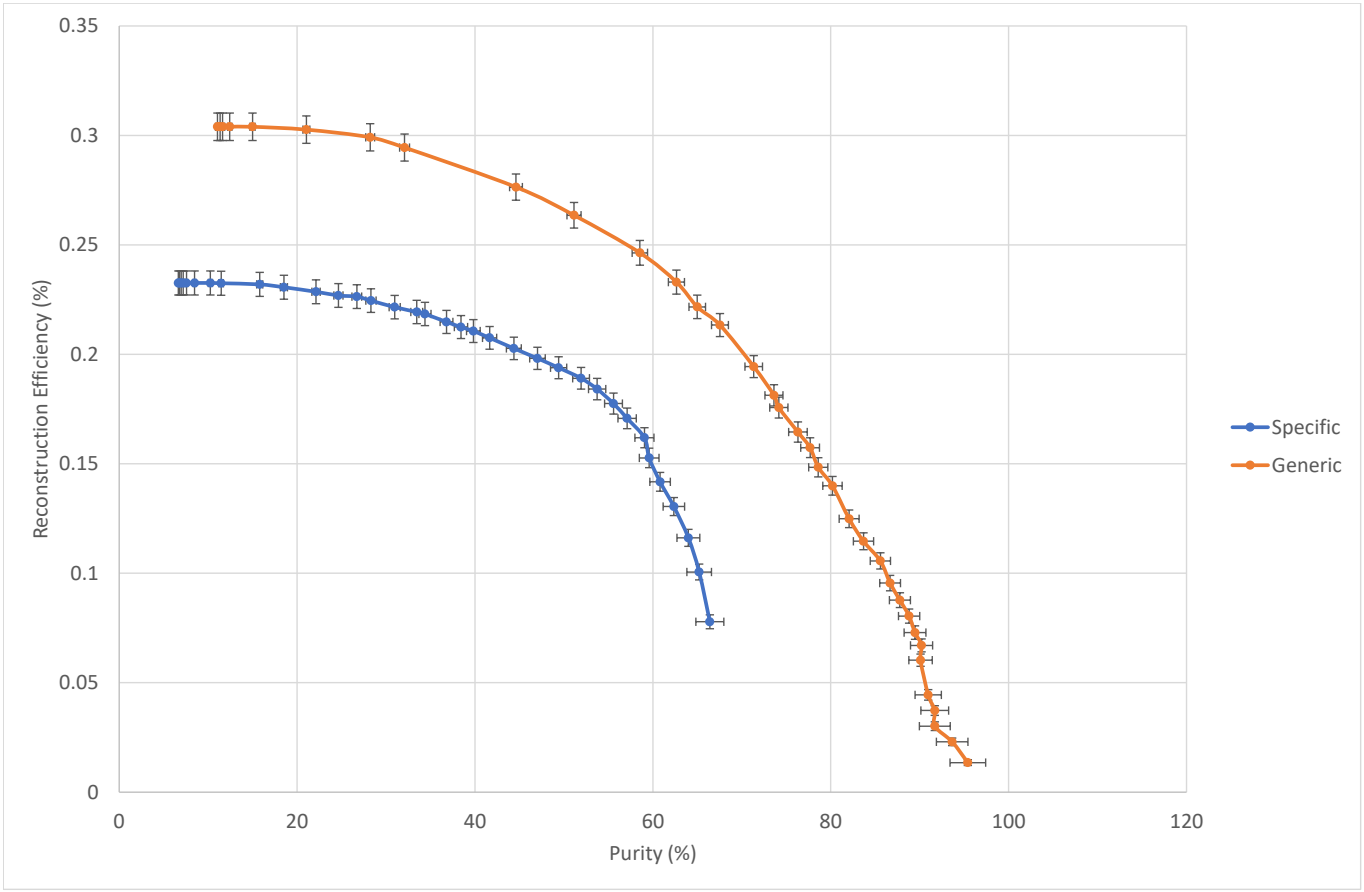
It was also decided to investigate the relative performance when including only common B_{had} and B_{SL} channels passing the FEI training and application for both methods, as was done for BGx1 conditions in Section 6.3.2. The efficiency vs. purity curves including a total of 11 hadronic and 20 semi-leptonic channels are shown in Figure 6.9, where the efficiency of the generic FEI is significantly reduced by the exclusion of channels whilst the specific FEI remains unchanged. With hadronic tagging, both methods are now consistent up to purities of about 50%, a moderate improvement over the equivalent BGx1 case, after which the spe-

cific FEI once again diverges towards lower efficiencies. For $\Upsilon(4S)$ candidates built from B_{SL} modes, however, the specific FEI outperforms the generic FEI over the entire purity range. These trends strongly suggest the sensitivity of the specific FEI to under-training. Given that more tagging channels possessed non-zero `SignalProbability` values in both the BGx0 specific and generic FEI implies that the classifiers as a whole were more adequately trained with the larger statistics and cleaner environment. Thus, when considering only the channels found to be successful with both methods, it appears as if the classifiers corresponding to a number of tagging modes in the BGx1 specific FEI were under-trained with respect to the generic, and the omission of beam background allowed for the training of these classifiers to be sufficient for the specific FEI to reach and even exceed the generic FEI in performance on signal MC.

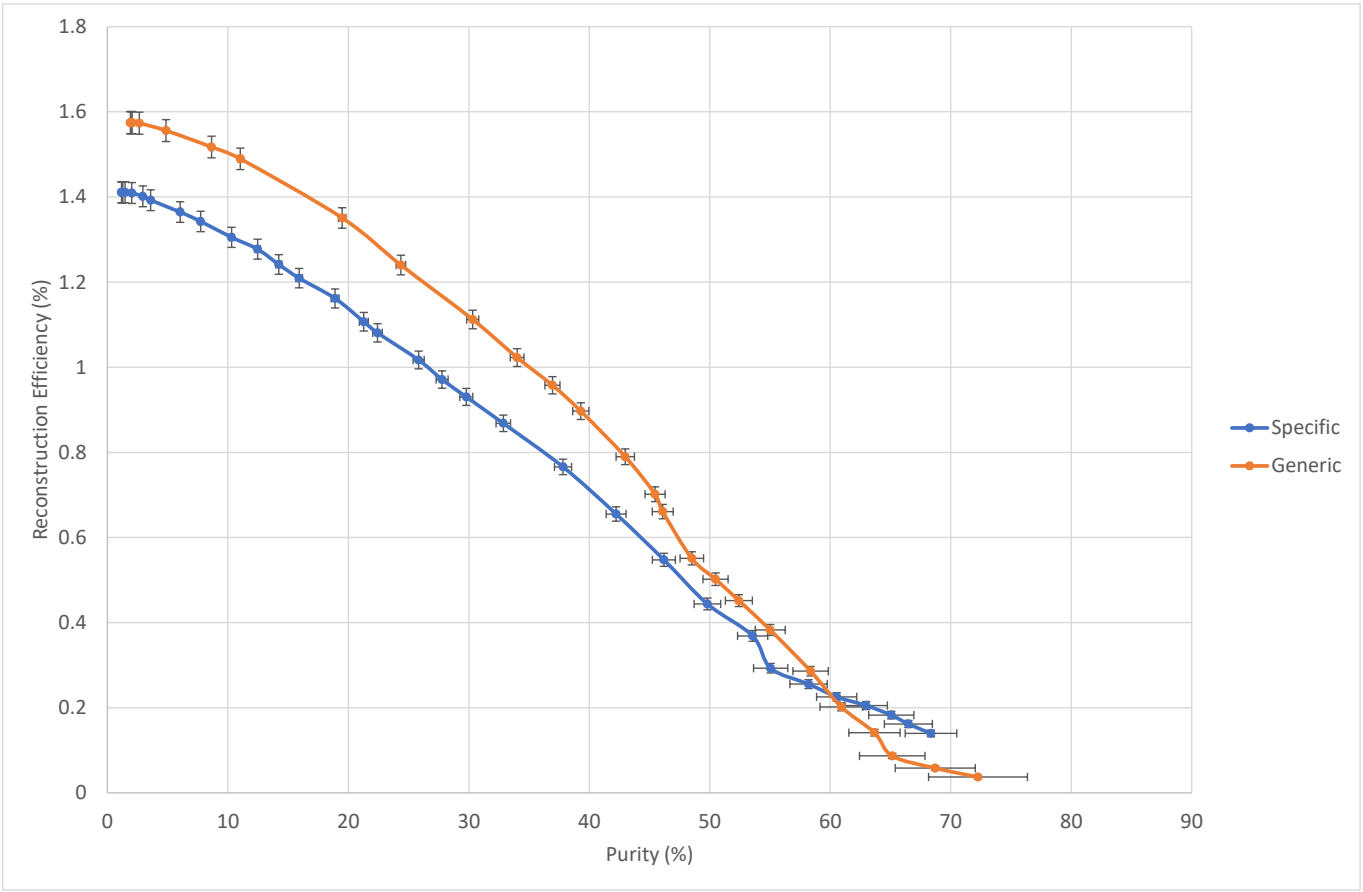
`SignalProbability` > 0	Charged		Mixed		$c\bar{c}$		*uds*	
Number of reconstructed $\Upsilon(4S)$ candidates	S	G	S	G	S	G	S	G
B_{had}	4027	1983	2729	1222	6893	3519	4733	2713
B_{SL}	33177	16438	28447	13400	34892	16428	21713	11032

Table 6.8: The number of fake $\Upsilon(4S)$ candidates reconstructed with the application of the specific (S) and generic (G) FEI to the various BGx0 background MC samples, for a single chosen `SignalProbability` cut.

Lastly, the relationship between the $\Upsilon(4S)$ reconstruction efficiency in signal and the background rejection capability of the specific and generic FEI under BGx0 conditions was similarly explored. Incorporating all successful channels for each method in order to compare their performance at full capacity, the background rejection curves analogous to those produced in 6.3.3 are depicted in Figures 6.10 and 6.11, with the number of fake $\Upsilon(4S)$ candidates associated with no additional `SignalProbability` cuts displayed in Table 6.8. For all studied backgrounds and both hadronic and semi-leptonic tagging modes, the generic FEI consistently reconstructs a higher number of correct $\Upsilon(4S)$ candidates than the specific FEI with the same background suppression. This is the same trend as was observed in the presence of beam background, but with the performance gap only further reinforced. In summary, when taking into account both the reconstruction efficiency of correct $\Upsilon(4S)$ candidates and the exclusion of fake candidates from background events, the specific FEI failed to reach the level of performance delivered by the generic FEI under both standard assumptions of the beam background and idealised conditions. These findings provided the motivation for the upcoming chapter, which recounts the attempt to replicate this analysis for the leptonic decay $B^+ \to \tau^+\nu_\tau$, associated with a lesser physical complexity in the absence of hadronic decay products such as the ρ^0 resonance, in order to explore any potential improvements in performance.

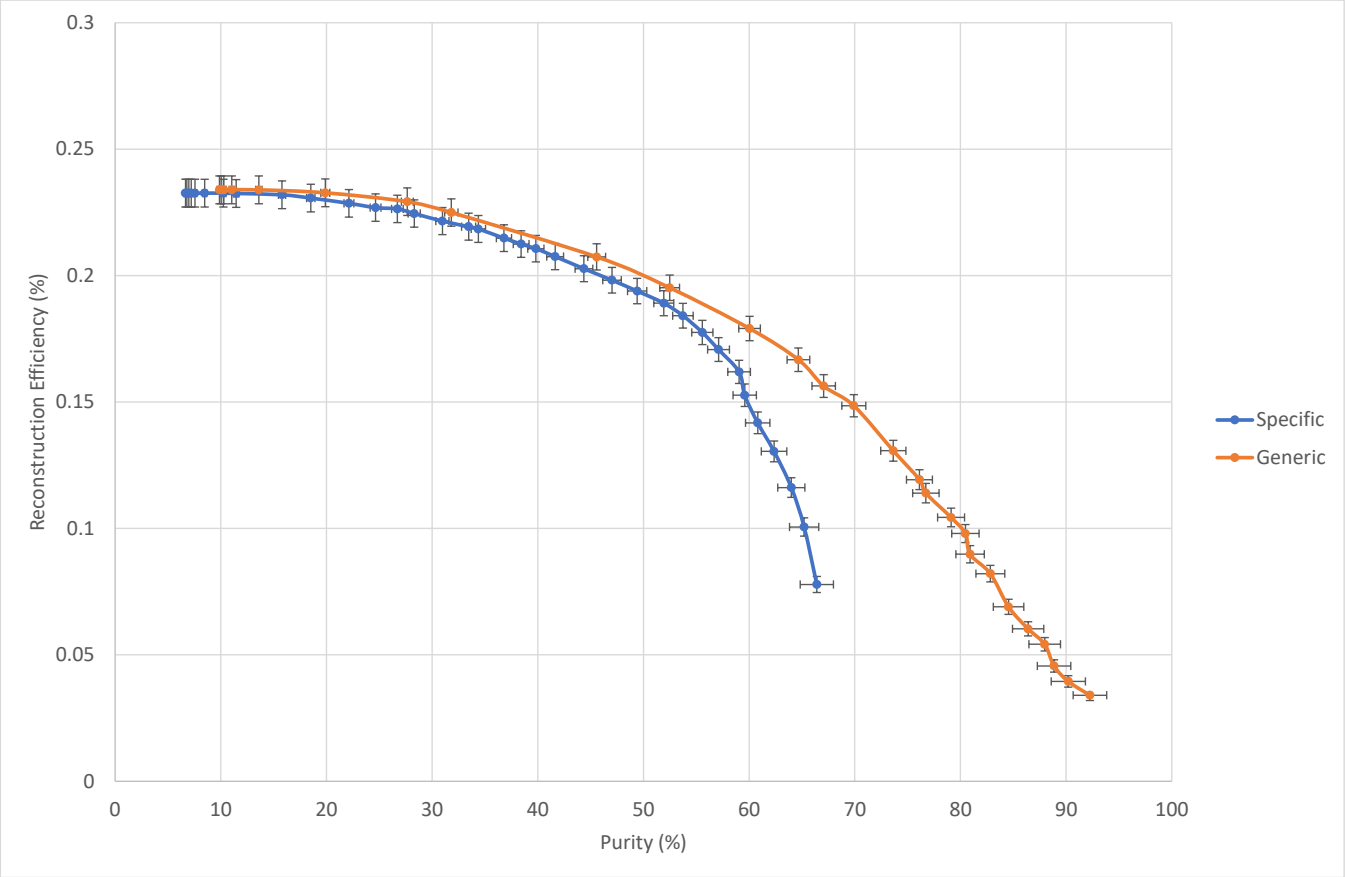


(a) Hadronic tagging

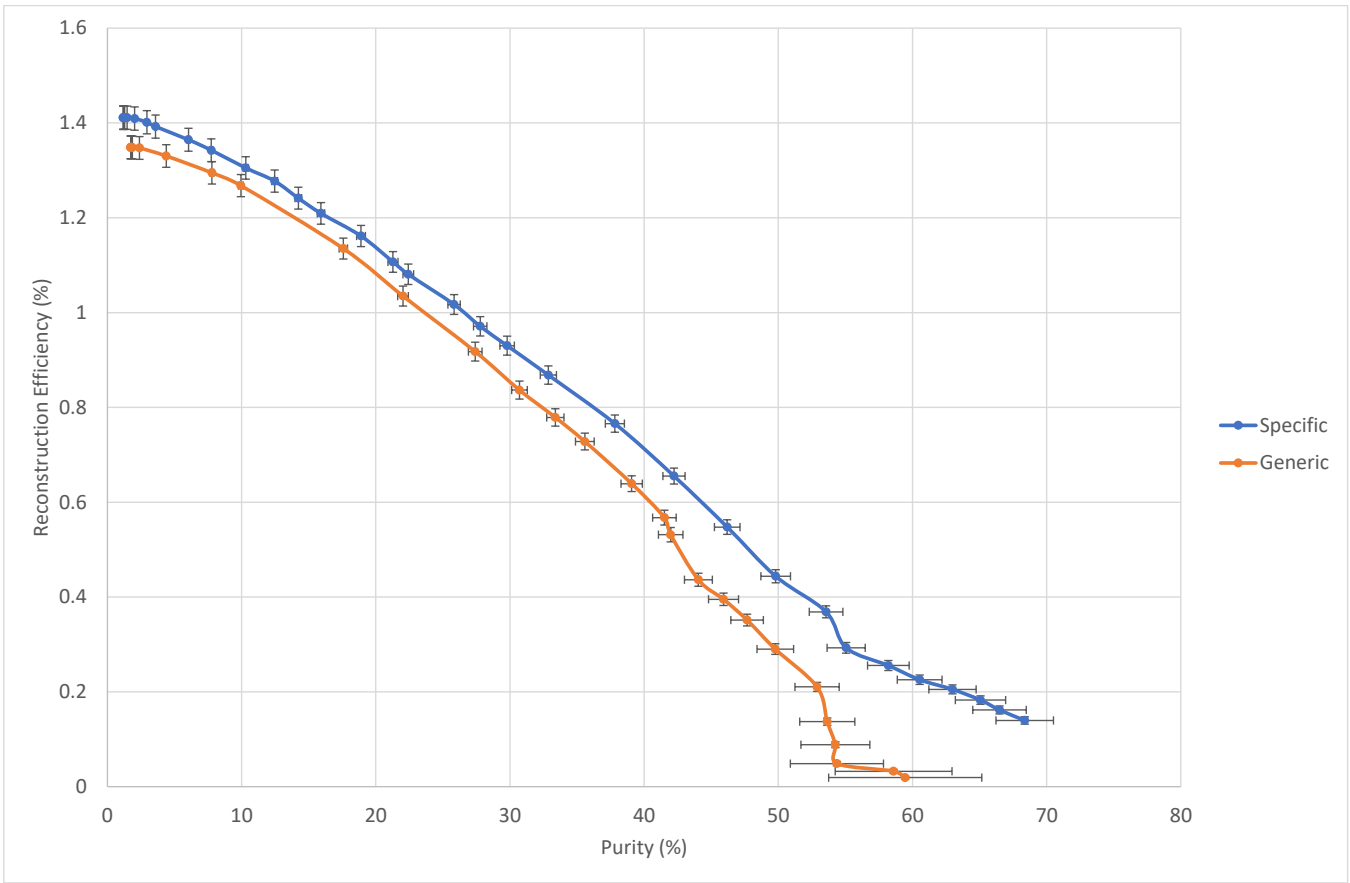


(b) Semi-leptonic tagging

Figure 6.8: Efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed from B_{sig} and B_{tag} daughters in BGx0 signal MC.

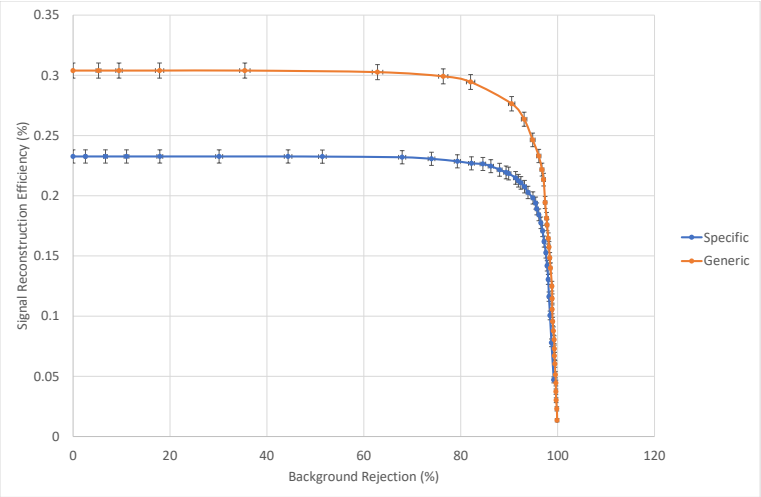


(a) Hadronic tagging

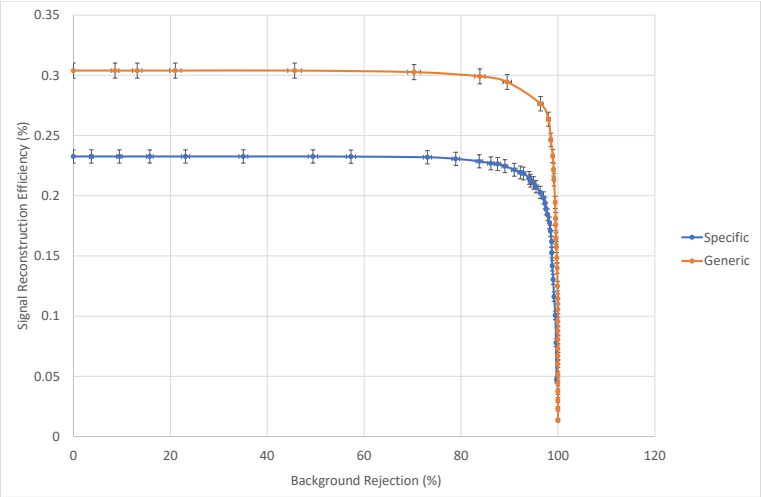


(b) Semi-leptonic tagging

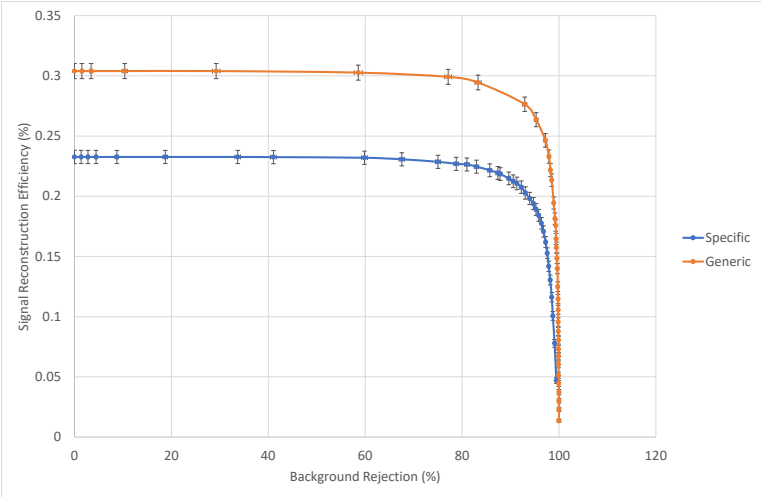
Figure 6.9: Efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed in BGx0 signal MC, including only channels mutually successful between the specific and generic FEI.



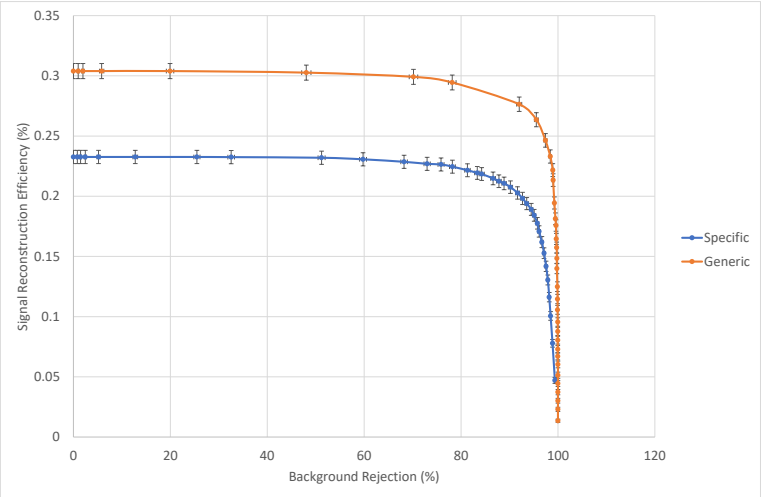
(a) Charged



(b) Mixed

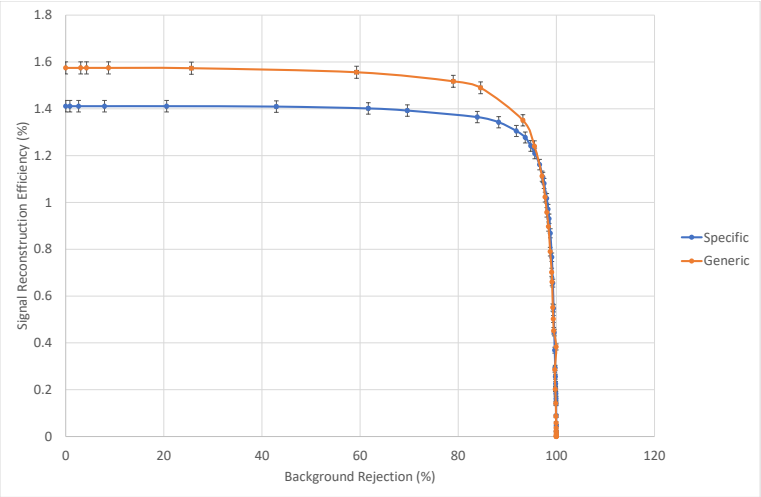


(c) $c\bar{c}$

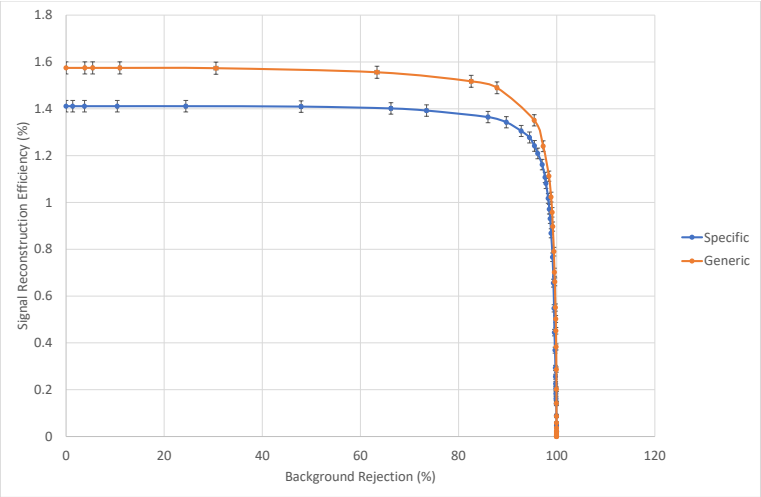


(d) *uds*

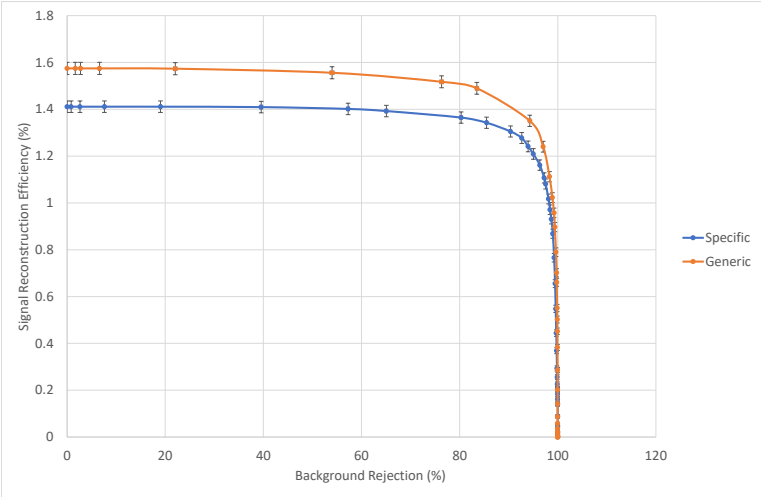
Figure 6.10: Background rejection curves for $\Upsilon(4S)$ candidates built from B_{had} daughters, in BGx0 background MC.



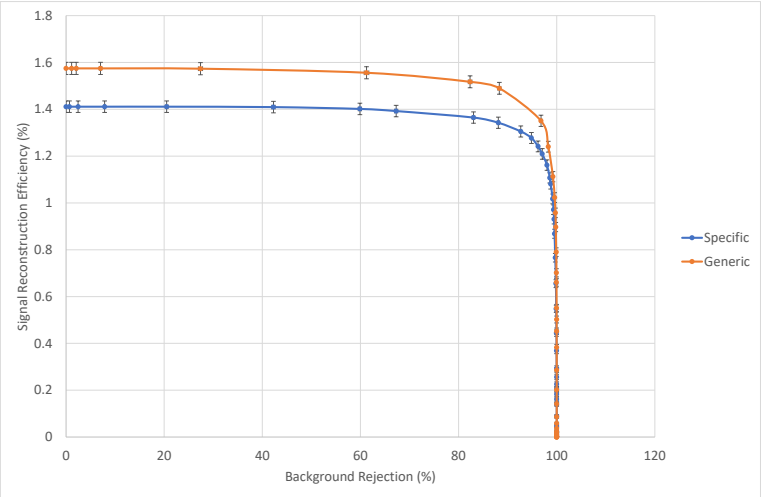
(a) Charged



(b) Mixed



(c) $c\bar{c}$



(d) *uds*

Figure 6.11: Background rejection curves for $\Upsilon(4S)$ candidates built from B_{SL} daughters, in BGx0 background MC.

Chapter 7

FEI Performance: $B^+ \to \tau^+\nu_\tau$

The previous chapter detailed the results of an investigation into the relative performance of the specific and generic FEI in the context of rare semi-leptonic decays, with $B^+ \to \rho^0\mu^+\nu_\mu$ chosen as a working example. Ultimately, the generic FEI was found to exhibit a higher degree of performance when considering a number of key output features including the $\Upsilon(4S)$ reconstruction efficiency and background rejection capabilities. As mentioned, existing analyses employing the use of the specific FEI were conducted on leptonic decays of interest including $B^+ \to \tau^+\nu_\tau$ [34] and $B^+ \to \ell^+\nu_\ell\gamma$ [11], with decays of this nature being physically less complex than semi-leptonic decays such as $B^+ \to \rho^0\mu^+\nu_\mu$ which involve the inclusion of additional hadronic daughters. It was therefore decided to compare the effectiveness of the specific and generic FEI when applied to one such leptonic mode, $B^+ \to \tau^+\nu_\tau$, with the aim of investigating any potential improvement in the relative performance of both methods from that observed in the earlier semi-leptonic case. It was also decided to carry out this analysis in the absence of beam background, for several reasons. Firstly, the $B^+ \to \tau^+\nu_\tau$ specific FEI analysis described in [34] was implemented on MC without beam background having been performed several years prior, before beam background MC samples were readily available for use. Additionally, as discovered in the BGx0 $B^+ \to \rho^0\mu^+\nu_\mu$ study, the larger statistics used for the training of classifiers for each B_{tag} channel resulting from BGx0 conditions allowed for a greater number of channels subsequently passing the application stage of the FEI. As such, choosing to omit beam background would allow for the comparison between the specific and generic FEI with both trained on the maximum statistics possible given the input training sample sizes and the chosen selections.

7.1 Signal Side Selection

For the $B^+_{sig} \to \tau^+\nu_\tau$ reconstruction, the decay chain studied was restricted to a single mode for the subsequent decay of the τ lepton, $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$. This decay has an associated branching fraction of 17.39 ± 0.04 % [12], and is an example of a 1-prong τ decay whereby a single charged daughter is produced. Thus, the full B_{sig} decay chain involved three neutrinos, and only one track belonging to the final-state muon was required by the reconstruction. The selection therefore chosen for the B_{sig} reconstruction, guided by the analysis in [34], was simply

a moderate `ParticleID` cut on the tracks used to build the final-state muon `ParticleList`, `muonID` > 0.6. No mass windows were applied to the tau or B_{sig} meson given the three total sources of missing 4-momentum, and no impact parameter cuts were particularly necessary for the muon with no tracks expected to be substantially displaced from the interaction point in the absence of beam background events.

Following the definitions introduced in Chapter 5, the B_{sig} reconstruction efficiency and purity were similarly calculated through the application of the B_{sig} reconstruction to a sample of 1 million signal MC $\Upsilon(4S) \to B^+_{sig}B^-$ events, with B^- decaying via any allowed mode. The reconstruction efficiency was evaluated to be 47.96 ± 0.05 %, more than double that of the BGx0 value for the semi-leptonic $B^+_{sig} \to \rho^0\mu^+\nu_\mu$. Given that the effective training sample size used in the specific FEI is largely dependent on the reconstruction efficiency of the B_{sig} decay mode, $B^+_{sig} \to \tau^+\nu_\tau$ would therefore provide the largest effective training sample for the specific FEI utilised throughout this study thus far, allowing for the relative performance of the specific and generic FEI to be probed with a lower level of discrepancy between the effective training samples. The $B^+_{sig} \to \tau^+\nu_\tau$ purity obtained also exceeded the equivalent value for $B^+_{sig} \to \rho^0\mu^+\nu_\mu$, with this calculated to be 52.14 ± 0.05 %.

7.2 FEI Training

The BGx0 **generic FEI** was trained centrally with the weight files obtained from KEKCC, these being the same as were used in the BGx0 $B^+ \to \rho^0\mu^+\nu_\mu$ study. The composition of the MC samples used as input to the training was as follows:

- 90 million charged $\Upsilon(4S) \to B^+B^-$ events, with B^+/B^- decaying via any allowed modes.
- 90 million mixed $\Upsilon(4S) \to B^0\bar{B}^0$ events, with $B^0/\bar{B}^0$ decaying via any allowed modes.

Conversely, the **specific FEI** was trained personally on MC sample sizes identical to those employed in the training of the BGx0-specific FEI on $B^+ \to \rho^0\mu^+\nu_\mu$. This corresponded to the following signal and background MC events:

- 100 million signal $\Upsilon(4S) \to B^+_{sig}B^-$ events, with $B^+_{sig} \to \tau^+\nu_\tau$ and B^- decaying via any allowed mode. These events were generated in order to achieve the desired sample size, larger than was available from the official MC7 prodution.
- 90 million charged $\Upsilon(4S) \to B^+B^-$ background events, with B^+/B^- decaying via any allowed modes, taken from MC7.
- 90 million mixed $\Upsilon(4S) \to B^0\bar{B}^0$ background events, with $B^0/\bar{B}^0$ decaying via any allowed modes, taken from MC7.

As was done with the training of the specific FEI for $B^+ \to \rho^0\mu^+\nu_\mu$ in Section 6.1.2, an initial skim to exclude any events with more than 12 tracks satisfying the impact parameter conditions $d_r < 2$ cm and $|d_z| < 4$ cm was included in order to minimise the computing time. As mentioned, most tracks likely possessed impact parameters within these ranges due to the

absence of beam background, making these selections somewhat redundant, but they were nevertheless included in the 12-track skim in mimicking the $B^+ \to \rho^0 \mu^+ \nu_\mu$ configuration. An additional cut on the maximum number of tracks assigned to B_{sig} was similarly employed as an FEI user cut on the B_{tag} `ParticleLists`. As only a single track arising from the B_{sig} decay was expected, a broad cut allowing up to three such tracks, `nRemainingTracksInEvent` $\leqslant$ 3, was chosen primarily for further reduction of the computing time.

7.3 FEI application

Both the trained specific and generic FEI were subsequently applied to the following signal and background MC samples:

- 1 million generated signal $\Upsilon(4S) \to B^+_{sig} B^-$ events, with $B^+_{sig} \to \tau^+ \nu_\tau$.
- 1 million MC7 charged $\Upsilon(4S) \to B^+ B^-$ background events.
- 1 million MC7 mixed $\Upsilon(4S) \to B^0 \bar{B}^0$ background events.
- 1 million MC7 $c\bar{c}$ continuum background events, with $e^+e^- \to c\bar{c}$.
- 1 million MC7 uds continuum background events, with $e^+e^- \to u\bar{u}$, $d\bar{d}$ or $s\bar{s}$.

The selections enforced within the application of the FEI included the lone B_{sig} reconstruction selection on the `muonID` along with the initial 12-track skim and B_{tag} user cut described in the previous section. All B_{sig} candidates were combined with B_{had} and B_{SL} tag candidates to reconstruct the full $\Upsilon(4S)$ event, after which a rest-of-event (ROE) object was built for each $\Upsilon(4S)$ candidate. One further selection was applied whereby only events with no tracks remaining in the $\Upsilon(4S)$ ROE were retained. This is a fairly harsh condition that considers only events in which the total number of tracks exactly equals the amount expected from the B_{sig} and B_{tag} decay modes. Previous BGx0 analyses [10, 34] had found the inclusion of this condition feasible in the absence of contamination due to beam background, and it was ultimately incorporated within this study with the aim of rejecting a significant number of fake and incorrectly reconstructed $\Upsilon(4S)$ candidates. After all selections were applied, only the $\Upsilon(4S)$ candidate built from a B_{had} or B_{SL} daughter with the highest value of the FEI classifier output, the `SignalProbability`, was kept per event.

7.4 Comparison of Performance

7.4.1 Decay Channel Breakdown

Following the analysis workflow of the previous chapter, the relative performance of the specific and generic FEI is first presented in the context of the number of B_{tag} channels failing the training and application conditions. The FEI automatic reporting provided information on the individual channels for which a multi-variate classifier was successfully trained, with these results displayed in Tables 7.1 and 7.2 for hadronic and semi-leptonic modes, respectively. For each of these channels, the results of the FEI application to signal MC are also shown, with channels regarded as having passed this application if all reconstructed candidates possessed non-zero values of the `SignalProbability`. The individual successful channels for the generic FEI are identical to those seen in the equivalent tables (6.3, 6.4) for the BGx0 $B^+ \to \rho^0 \mu^+ \nu_\mu$ case, as both studies utilised the same generic FEI.

Despite the change in the chosen B_{sig} decay to a leptonic mode with more than double the reconstruction efficiency, the specific FEI was still unable to train a classifier for the last four hadronic channels as was observed for $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$. Additionally, the specific FEI maintained a total of 11 B_{had} channels subsequently passing the application stage, remaining significantly overshadowed by the generic FEI with 26 out of 29 total channels found to be successful.

However, a substantial improvement in the semi-leptonic tagging functionality of the specific FEI was observed, with nine further B_{SL} channels passing the FEI training and application than was accomplished by the specific FEI trained for BGx0 $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$. This corresponded to 29 usable B_{SL} channels, an amount still below the 34 provided by the generic FEI but with the slightest discrepancy between the two methods achieved thus far of all analyses conducted. In fact, each of the five additional B_{SL} channels present in the generic FEI contained only either one or two $\Upsilon(4S)$ candidates that were reconstructed correctly, out of a total of 230 256 signal events in which the daughter B_{tag} decayed semi-leptonically. Thus, the channels found to have failed the specific FEI application represented only an insignificant amount of the total sample size.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0\pi^+$	$(4.81 \pm 0.15) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0\pi^+\pi^0$	(1.34 ± 0.18) %	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^0\pi^0$	$< 0.98\%$ *	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-$	$(5.7 \pm 2.2) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-\pi^0$	> 1.5 % *	✓	✓	✓	✓
$B^+ \to \bar{D}^0 D^+$	$(3.8 \pm 0.4) \times 10^{-4}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^+ K^0_S$	$(1.55 \pm 0.21) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^+ K^0_S$	$(2.1 \pm 0.5) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^{*+} K^0_S$	$(3.8 \pm 0.4) \times 10^{-3}$	✓	✓	✗	✗
$B^+ \to \bar{D}^{*0} D^{*+} K^0_S$	$(9.2 \pm 1.2) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^0 K^+$	$(1.45 \pm 0.33) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^0 K^+$	$(2.26 \pm 0.23) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^0 D^{*0} K^+$	$(6.3 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^{*0} K^+$	(1.12 ± 0.13) %	✓	✓	✗	✓
$B^+ \to D^+_S \bar{D}^0$	$(10.0 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}\pi^+$	$(5.18 \pm 0.26) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^0$	$(9.8 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^0\pi^0$	5×10^{-4} *	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^+\pi^-$	(1.03 ± 0.12) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}\pi^+\pi^+\pi^-\pi^0$	(1.8 ± 0.4) %	✓	✓	✓	✓
$B^+ \to D^{*+}_S \bar{D}^0$	$(7.6 \pm 1.6) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to D^+_S \bar{D}^{*0}$	$(8.2 \pm 1.7) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to \bar{D}^0 K^+$	$(3.65 \pm 0.33) \times 10^{-4}$	✓	✓	✗	✓

cont.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to D^- \pi^+ \pi^+$	$(1.07 \pm 0.05) \times 10^{-3}$	✓	✓	✗	✓
$B^+ \to D^- \pi^+ \pi^+ \pi^0$	0.20 % *	✓	✓	✓	✓
$B^+ \to J/\psi K^+$	$(1.016 \pm 0.033) \times 10^{-3}$	✗	✓	-	✗
$B^+ \to J/\psi K^+ \pi^+ \pi^-$	$(8.1 \pm 1.3) \times 10^{-4}$	✗	✓	-	✓
$B^+ \to J/\psi K^+ \pi^0$	1×10^{-4} *	✗	✓	-	✓
$B^+ \to J/\psi K_S^0 \pi^+$	0.11 % *	✗	✓	-	✓
	Total channels:	25/29	29/29	11/29	26/29

Table 7.1: The hadronic tagging channels incorporated by the FEI, with their associated branching fractions as recorded by the Particle Data Group (PDG) [12]. * indicates channels for which the branching fraction was not found in the PDG, with the quoted values taken or estimated to be those used in the generation of the standard Belle II MC7 production. The channels that passed the training and application stages of the specific and generic FEI on BGx0 MC are also indicated.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0 e^+$	(2.26 ± 0.11) %	✓	✓	✓	✓
$B^+ \to \bar{D}^0 \mu^+$	(2.26 ± 0.11) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} e^+$	(5.70 ± 0.19) %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} \mu^+$	(5.70 ± 0.19) %	✓	✓	✓	✓
$B^+ \to D^- \pi^+ e^+$	$(4.2 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^- \pi^+ \mu^+$	$(4.2 \pm 0.5) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^{*-} \pi^+ e^+$	$(6.1 \pm 0.6) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to D^{*-} \pi^+ \mu^+$	$(6.1 \pm 0.6) \times 10^{-3}$	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+$	0.8×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^0$	0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^0 \pi^0$	$< 0.16\%$	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^+ \pi^-$	0.9×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 \pi^+ \pi^+ \pi^- \pi^0$	> 0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}_{SL}^0 D^+$	0.6×10^{-4}	✓	✓	✗	✓
$B^+ \to \bar{D}_{SL}^0 D^+ K_S^0$	0.2×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}_{SL}^{*0} D^+ K_S^0$	0.3×10^{-3}	✓	✓	✗	✗

cont.

		Classifier Trained (✓/✗)		Passed Application (✓/✗)	
Decay Channel	Branching Fraction	Specific FEI	Generic FEI	Specific FEI	Generic FEI
$B^+ \to \bar{D}^0_{SL} D^{*+} K^0_S$	0.6×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0}_{SL} D^{*+} K^0_S$	1.5×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^0 K^+$	0.2×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} D^0 K^+$	0.4×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0_{SL} D^{*0} K^+$	1.0×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} D^{*0} K^+$	0.2 %	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^+_{SL}$	1.0×10^{-4}	✓	✓	✓	✓
$B^+ \to \bar{D}^0 D^+_{SL} K^0_S$	0.4×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^+_{SL} K^0_S$	0.5×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^{*+}_{SL} K^0_S$	0.7×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^{*0} D^{*+}_{SL} K^0_S$	1.7×10^{-3}	✓	✓	✗	✗
$B^+ \to \bar{D}^0 D^0_{SL} K^+$	0.2×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^0_{SL} K^+$	0.4×10^{-3}	✓	✓	✗	✓
$B^+ \to \bar{D}^0 D^{*0}_{SL} K^+$	1.0×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0} D^{*0}_{SL} K^+$	0.2 %	✓	✓	✗	✗
$B^+ \to D^+_S \bar{D}^0_{SL}$	1.6×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+$	0.8×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^0$	1.6×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^0 \pi^0$	0.8×10^{-4}	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^+ \pi^-$	0.2 %	✓	✓	✓	✓
$B^+ \to \bar{D}^{*0}_{SL} \pi^+ \pi^+ \pi^- \pi^0$	0.3 %	✓	✓	✓	✓
$B^+ \to D^{*+}_S \bar{D}^0_{SL}$	1.2×10^{-3}	✓	✓	✓	✓
$B^+ \to D^+_S \bar{D}^{*0}_{SL}$	1.3×10^{-3}	✓	✓	✓	✓
$B^+ \to \bar{D}^0_{SL} K^+$	0.6×10^{-4}	✓	✓	✓	✓
$B^+ \to D^-_{SL} \pi^+ \pi^+$	0.3×10^{-3}	✓	✓	✓	✓
$B^+ \to D^-_{SL} \pi^+ \pi^+ \pi^0$	0.1 %	✓	✓	✓	✓
	Total channels:	42/42	42/42	29/42	34/42

Table 7.2: The semi-leptonic tagging channels reconstructed by the FEI. Undetected neutrinos are omitted from the above descriptions of the former 8 channels, with the associated branching fractions shown obtained from the PDG [12]. The remaining 34 channels are paired with approximate branching fractions calculated from the values taken from the PDG or estimated to be those used in the generation of the standard Belle II MC7 production. The channels that passed the training and application stages of the specific and generic FEI on BGx0 MC are also indicated.

7.4.2 Performance on Signal $B^+ \to \tau^+\nu_\tau$

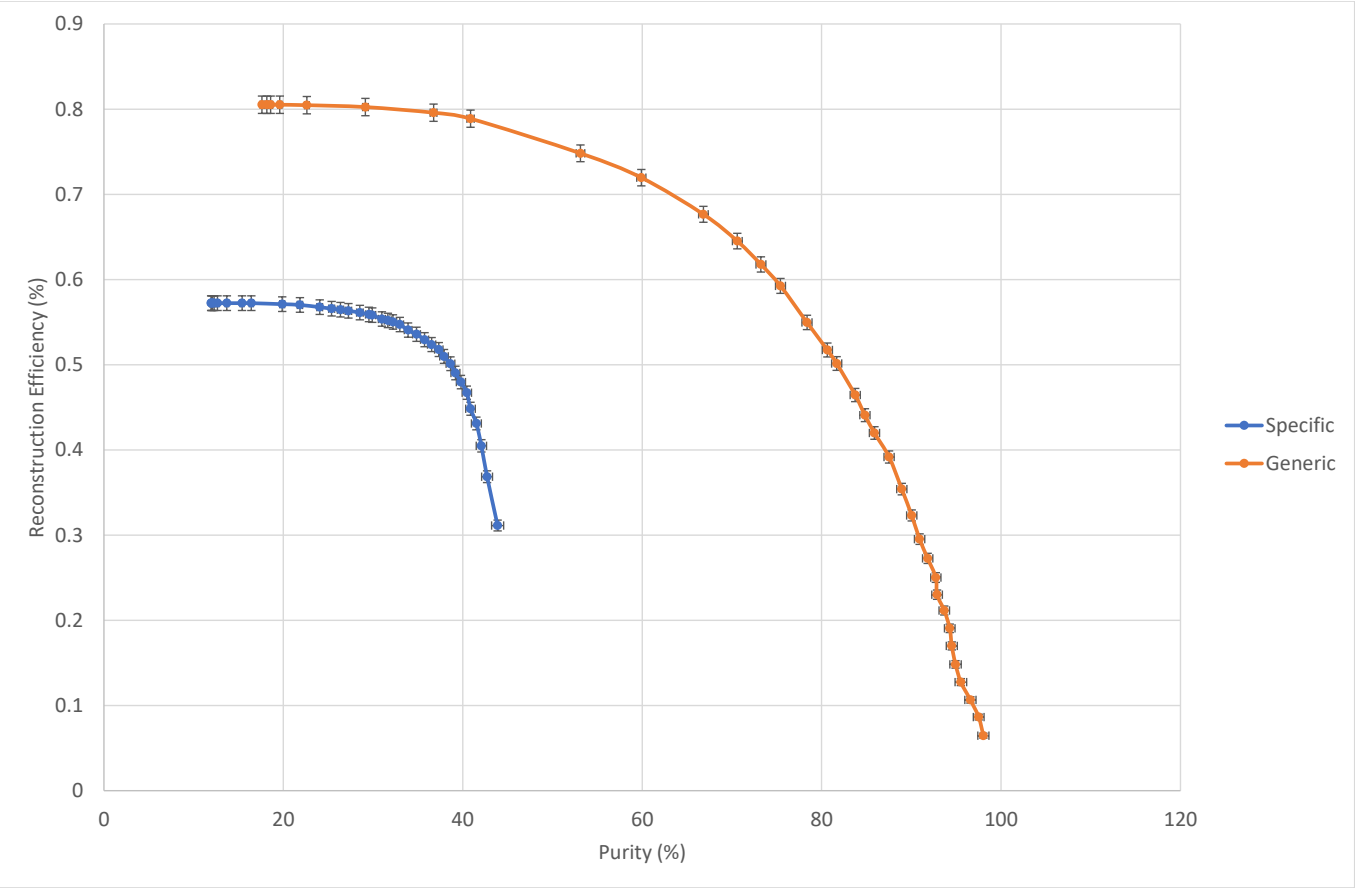
Scanning over a range of 36 B_{tag} `SignalProbability` cuts between `SignalProbability` $>$ 0 and `SignalProbability` > 0.95 in the application of the FEI to signal MC, the $\Upsilon(4S)$ reconstruction efficiencies and purities as defined in Section 6.3.2 were calculated. The details of one such calculation, for the first `SignalProbability` cut of > 0, are presented in Table 7.3, with the full efficiency vs. purity curves illustrated in Figure 7.1(a) and 7.1(b) for $\Upsilon(4S)$ candidates built from B_{had} and B_{SL} daughters respectively. Unsurprisingly given that the ratio between the number of usable B_{had} channels in the specific and generic FEI was equivalent to the BGx0 $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ study, the specific and generic FEI curves were similarly displaced by a considerable performance gap, with this only enhanced for $B^+_{sig} \to \tau^+\nu_\tau$. This reinforced the trend observed over the course of the multiple studies conducted whereby the generic FEI consistently proved more effective in the treatment of hadronic tags, with improvements to the statistics used in training classifiers for both methods only increasing the performance of the generic FEI relative to the specific FEI. The equivalent efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed from only the 11 usable B_{had} channels shared between the specific and generic FEI are presented in Figure 7.2(a), with only the generic FEI modified by the exclusion of channels, reducing its efficiency. The specific FEI was found to be lacking even with this configuration, further suggesting the superiority of the generic FEI in the training of hadronic modes, given the aforementioned sizes of the MC samples used for the training of both FEI methods.

An additional feature that can be discerned from Figure 7.1(a) is the dense population of the specific FEI at low purities, in contrast to the generic FEI which extended over a greater purity range. This was due to the fact that the `SignalProbability` values of the B_{had} daughters arising from those $\Upsilon(4S)$ candidates that were correctly reconstructed by the specific FEI were skewed largely towards 1.0. As such, consecutive `SignalProbability` cuts retained many of these candidates whilst rejecting successively more incorrect combinations, improving the purity without significant cost to the efficiency. However, even the harshest cut applied of `SignalProbability` > 0.95, whilst preserving more correct candidates than the generic FEI, maintained a significant number of these incorrect candidates such that the maximum purity achievable by the specific FEI was only approximately 44 %. As the B_{sig} reconstruction forms the basis of the training of the specific FEI, the nature of the selections chosen for this reconstruction has a direct effect on the maximum achievable purity of the $\Upsilon(4S)$, an effect that is entirely absent from that of the generic FEI. The sharp drop-off with purity observed in the specific FEI curves with respect to the generic FEI was thus likely due to the fact that the signal-side selections described in Section 7.1 were fairly minimal and not optimised for the removal of a large number of incorrect B_{sig} candidates.

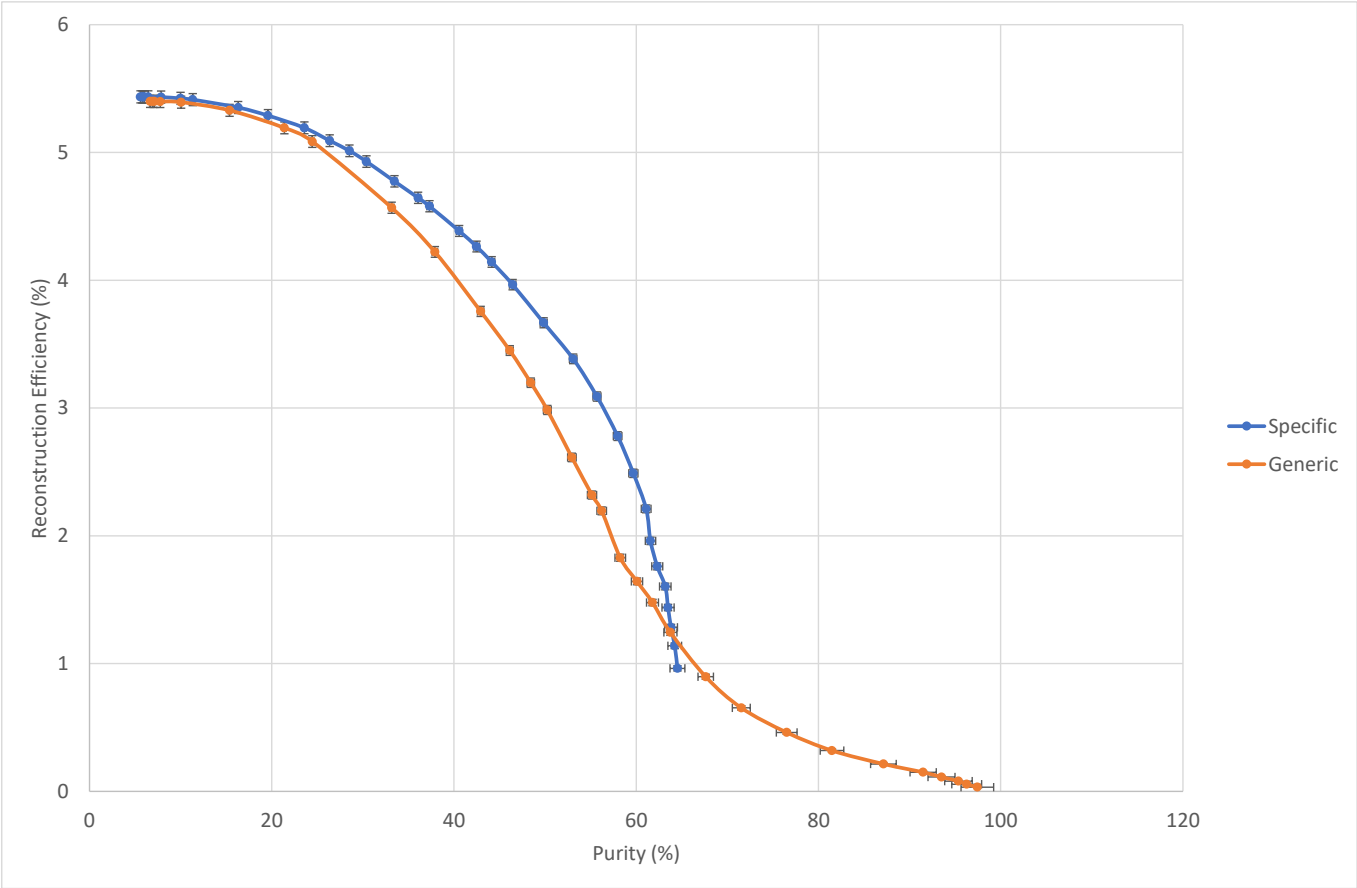
SignalProbability > 0	B_{sig}	B_{had}	$\Upsilon(4S)_{had}$	B_{SL}	$\Upsilon(4S)_{SL}$
Number of events in MC truth	1000000	769744		230256	
Specific FEI					
Number of reconstructed candidates	1965608	236706	36875	6011743	223687
Number of correctly reconstructed candidates	750093	15357	4405	16886	12515
$\Upsilon(4S)$ reconstruction efficiency			**(0.57 ± 0.01) %**		**(5.44 ± 0.05) %**
Purity			**(11.95 ± 0.17) %**		**(5.59 ± 0.05) %**
Generic FEI					
Number of reconstructed candidates	1965608	340489	35178	5935974	186876
Number of correctly reconstructed candidates	750093	20103	6198	21517	12433
$\Upsilon(4S)$ reconstruction efficiency			**(0.81 ± 0.01) %**		**(5.40 ± 0.05) %**
Purity			**(17.62 ± 0.20) %**		**(6.65 ± 0.06) %**

Table 7.3: The number of B_{sig}, B_{tag} and $\Upsilon(4S)$ candidates reconstructed and truth-matched with the application of the FEI to signal BGx0 MC for a single chosen SignalProbability cut. The subsequent values for the $\Upsilon(4S)$ reconstruction efficiency and purity are also included.

In contrast to the FEI hadronic tagging, the investigation of the semi-leptonic modes revealed a specific FEI consistent with or exceeding the generic FEI in performance over the majority of the sampled purity range, as shown in Figures 7.1(b) and 7.2(b) for all and common usable channels, respectively. No noticeable difference between these two figures can be perceived due to the aforementioned minimal contribution of the five B_{SL} channels excluded from the generic FEI curve in Figure 7.2(b). Of the studies conducted on $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$ and $B^+_{sig} \to \tau^+ \nu_\tau$, this was the first instance observed in which the overall performance of the specific FEI on signal MC was higher than the generic FEI when considering all channels utilised by each method.

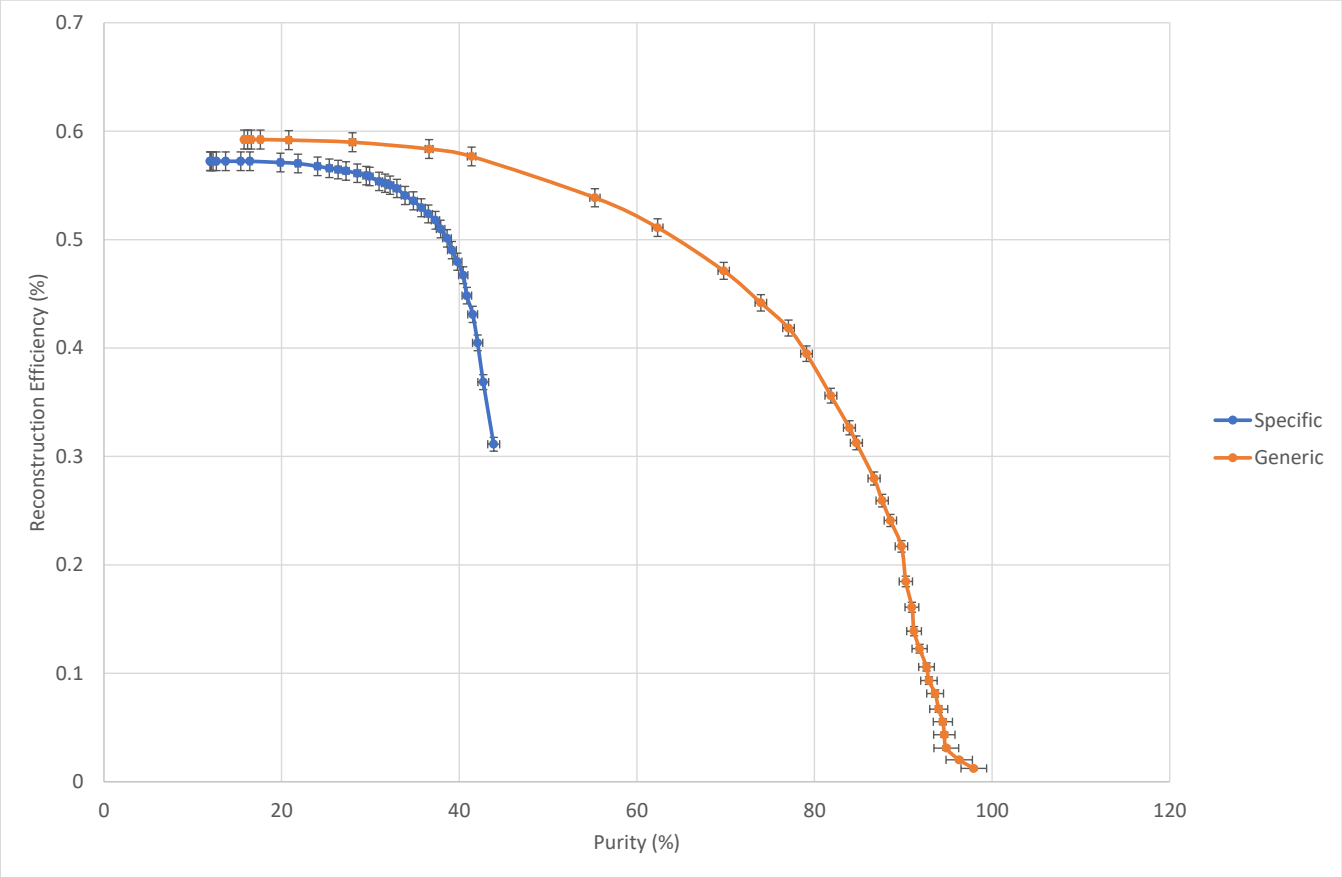


(a) Hadronic tagging

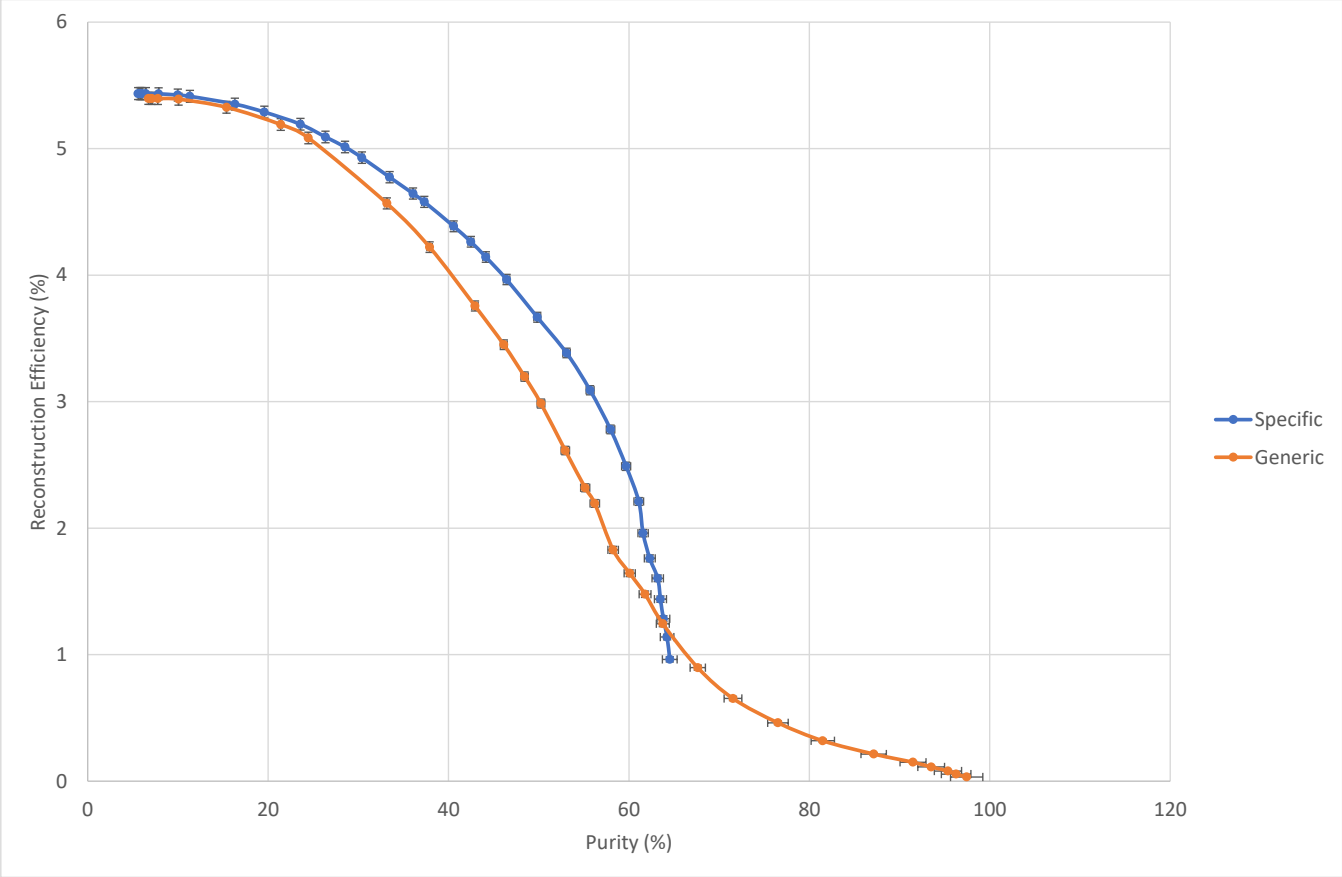


(b) Semi-leptonic tagging

Figure 7.1: Efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed from B_{sig} and B_{tag} daughters in BGx0 signal MC, for $B^+ \to \tau^+\nu_\tau$.



(a) Hadronic tagging



(b) Semi-leptonic tagging

Figure 7.2: Efficiency vs. purity curves for $\Upsilon(4S)$ candidates reconstructed in BGx0 signal MC, for $B^+ \to \tau^+\nu_\tau$, including only channels mutually successful between the specific and generic FEI.

7.4.3 Background Rejection

Finally, the effectiveness of successive B_{tag} `SignalProbability` cuts on the rejection of fake $\Upsilon(4S)$ candidates reconstructed from the various background MC samples was explored in a manner analogous to that described in Sections 6.3.3 and 6.4 for $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$. Table 7.4 displays the number of $\Upsilon(4S)$ candidates reconstructed in the application of the specific and generic FEI to these backgrounds for all successful channels without any further cut on the `SignalProbability`. Consistent with the trends observed in the previous chapter, the specific FEI was associated with a greater number of fake candidates across all backgrounds, despite possessing fewer usable B_{had} and B_{SL} channels from which these candidates could be built. The background rejection was then calculated for each of the 36 `SignalProbability` cuts sampled in the determination of the $\Upsilon(4S)$ reconstruction efficiency, with Figures 7.3 and 7.4 depicting the relationship between these quantities for B_{had} and B_{SL} modes, respectively.

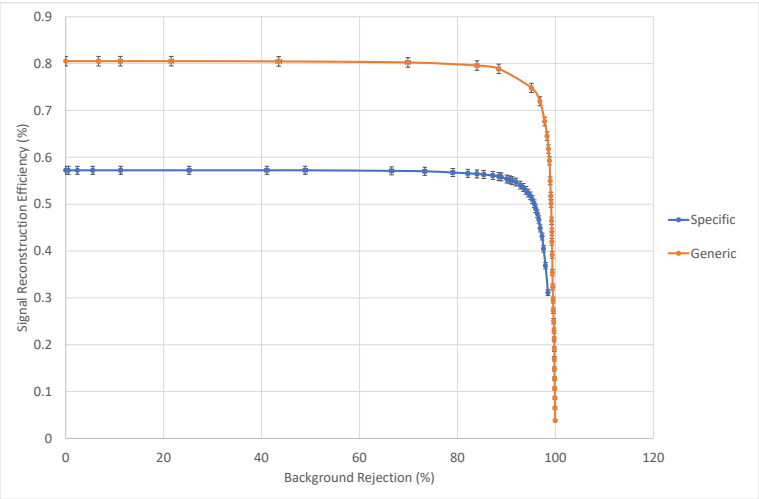
`SignalProbability` > 0	Charged		Mixed		$c\bar{c}$		uds	
Number of reconstructed $\Upsilon(4S)$ candidates	S	G	S	G	S	G	S	G
B_{had}	12520	8065	9980	6239	17253	12002	12205	8868
B_{SL}	35910	23298	30212	20298	75039	51776	52107	35751

Table 7.4: The number of fake $\Upsilon(4S)$ candidates reconstructed with the application of the specific (S) and generic (G) FEI to the various BGx0 background MC samples, for a single chosen `SignalProbability` cut.

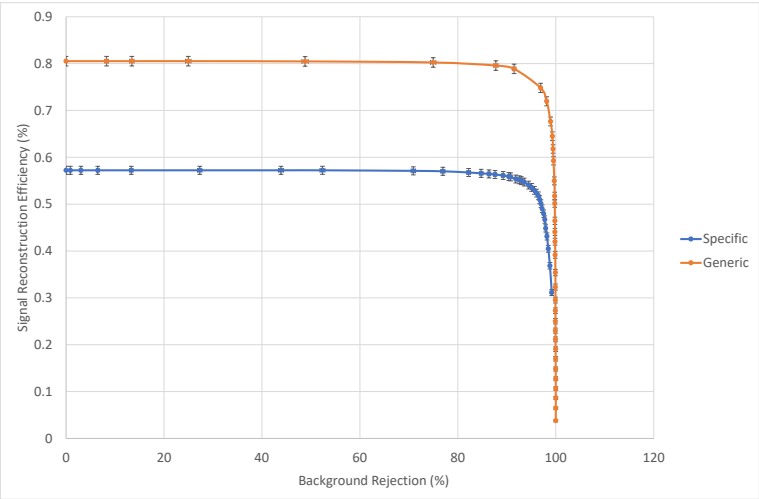
The behaviour of the hadronic modes reflected the inconsistency in the performance of the specific and generic FEI demonstrated by the discrepancy between the efficiency vs. purity curves constructed from all successful B_{had} channels. For each background MC sample examined, the $\Upsilon(4S)$ reconstruction efficiency achieved by the generic FEI noticeably surpassed that of the specific FEI for equivalent levels of background rejection. In considering the semi-leptonic tagging modes, however, the ratio between the reconstruction efficiency and background rejection for any given `SignalProbability` cut was found to be largely consistent between the two FEI configurations save for minor fluctuations, with the exception of the application to the mixed MC sample which demonstrated a slightly enhanced generic FEI performance at rejection levels approaching 100 %. This observation was once again unsurprising given the comparability of the proportion of usable B_{SL} channels provided by the specific and generic FEI algorithms, and was likewise the only instance noted thus far in which the specific and generic FEI displayed a similar level of performance when simultaneously considering the reconstruction efficiency and background rejection.

In summary, despite the increase in effective training statistics due to the absence of beam background and the substitution of the semi-leptonic B_{sig} decay mode with a leptonic mode possessing more than double the reconstruction efficiency, the overall relative performance of

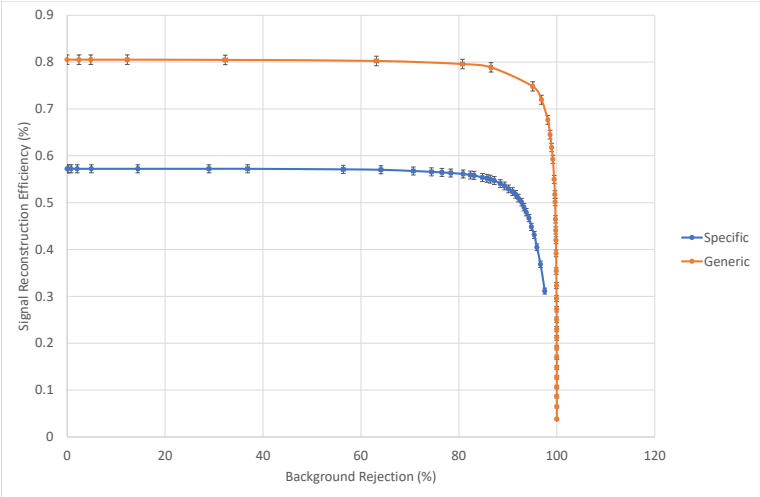
the specific FEI compared to the generic FEI only worsened for B_{had} channels, but improved for B_{SL} modes. This suggests that the semi-leptonic tagging capability of the specific FEI is considerably higher than that of hadronic tagging, at least in the context of B_{sig} decays involving missing energy as was investigated in this and the previous chapter. The upcoming final chapter will expand on these observations, providing some insight into the implications of this study on the practicality of utilising the specific FEI in future related analyses.



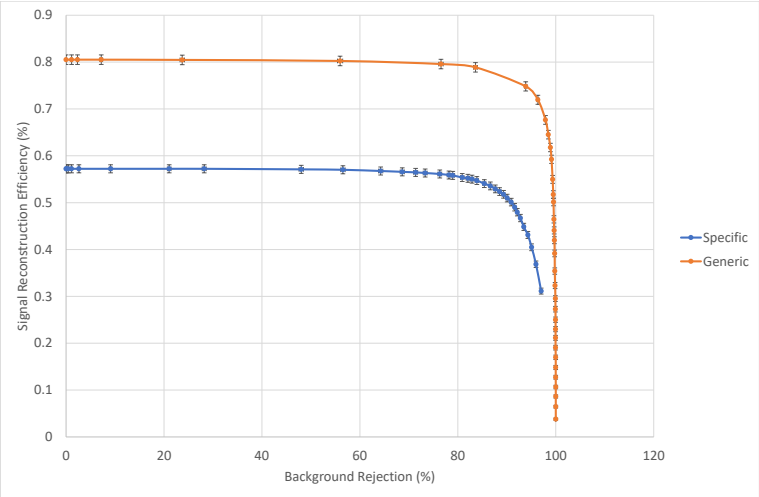
(a) Charged



(b) Mixed

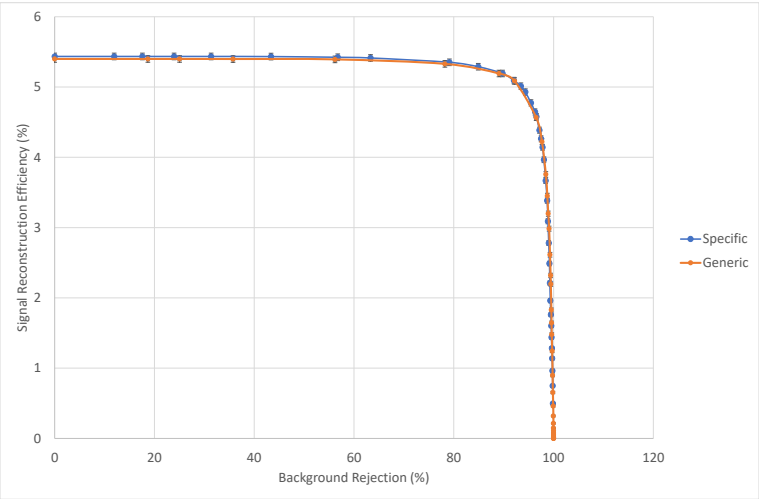


(c) $c\bar{c}$

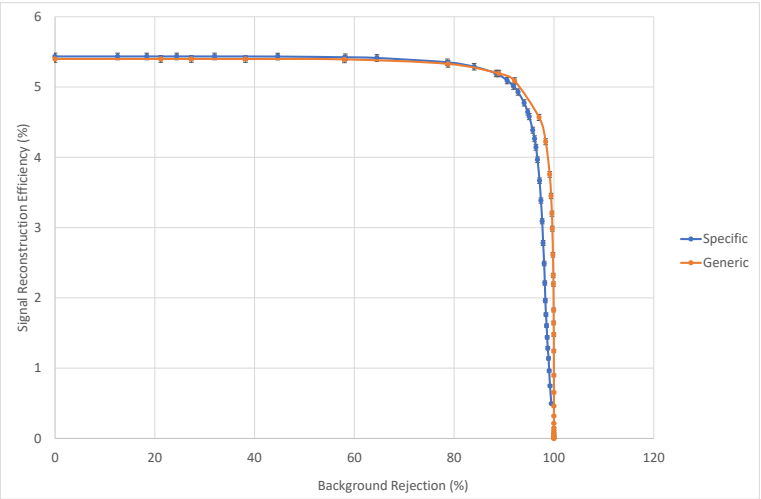


(d) uds

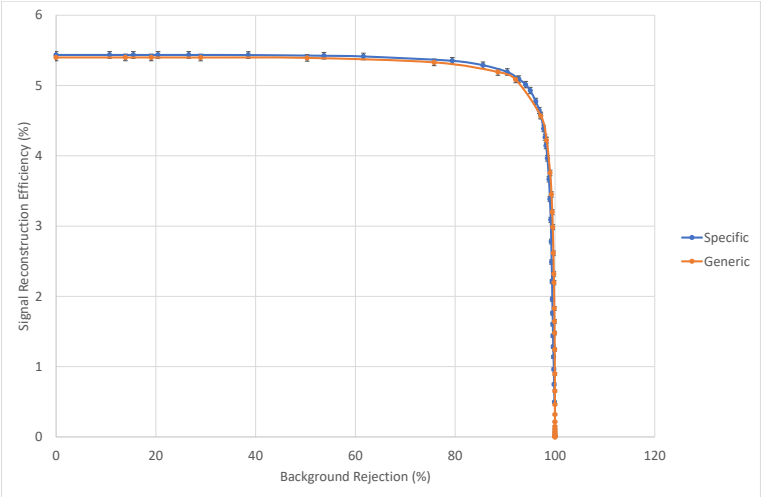
Figure 7.3: Background rejection curves for $\Upsilon(4S)$ candidates built from B_{had} daughters, in BGx0 background MC, for $B^+_{sig} \to \tau^+\nu_\tau$.



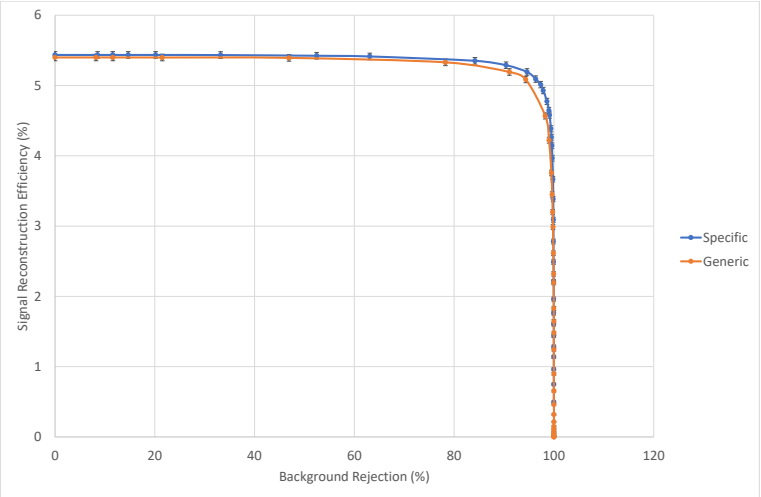
(a) Charged



(b) Mixed



(c) $c\bar{c}$



(d) uds

Figure 7.4: Background rejection curves for $\Upsilon(4S)$ candidates built from B_{SL} daughters, in BGx0 background MC, for $B^+_{sig} \to \tau^+\nu_\tau$.

Chapter 8

Conclusions and Future Work

The Belle II Experiment, currently in preparation for long-term data collection scheduled to commence early 2019, encompasses a broad physics program centered around the decays of pairs of B-mesons resulting from high energy electron-positron collisions. Its primary goal is to become a global contender in the search for physics Beyond the Standard Model, an aim it hopes to accomplish through the analysis of rare B-meson decays for which the luminosity was designed to produce roughly 50 times the amount achieved by its highly successful predecessor, the Belle Experiment.

The techniques utilised in the reconstruction of B-meson decays can be categorised into two distinctive groups, namely tagged and untagged analyses that differ on the basis of whether the remaining B-meson in an $\Upsilon(4S)$ event containing a signal decay mode of interest is reconstructed via an exclusive channel. For such analyses in which a B_{tag} is explicitly reconstructed, machine learning approaches can be adopted whereby an algorithm is trained to effectively reconstruct multiple B_{had} and B_{SL} modes in unseen data using detector information, by first supplying it with substantial amounts of simulated data known as Monte Carlo (MC) from which the the true decay chain can be accessed. The Full Event Interpretation is one such machine learning technique, available in the Belle II Analysis Software Framework (BASF2), that utilises a hierarchical approach to train a series of multi-variate classifiers for consecutive steps in the decay chains of the B_{tag} mesons. The FEI is modeled closely after the equivalent tagging algorithm employed by Belle, the Full Reconstruction (FR), but designed to be faster and more user-friendly whilst incorporating a number of additional features.

Arguably the most innovative of these features and of particular relevance to this study is the option available to execute the classifier training within the rest-of-event associated with a reconstructed B_{sig} decay mode of interest, a method referred to as the specific FEI. This is in contrast with the signal-independent generic FEI which is simply trained on MC samples consisting of pairs of B-mesons arising from the decay of an $\Upsilon(4S)$ in a manner analogous to the Belle FR. The specific FEI was conceptualised for the purpose of removing the bias associated with applying various signal-side selections to a centrally trained sample of reconstructed B_{tag} mesons, thereby allowing users to tailor the training of the B_{tag} reconstruction to their own individual analyses. This is primarily achieved through the execution of the

training on both signal MC samples, in which the $\Upsilon(4S)$ always possesses a B_{sig} daughter, and background samples containing none such B_{sig} daughters, with the classifier training subsequently carried out only on the rest-of-event objects of B_{sig} candidates correctly reconstructed from signal, and incorrectly reconstructed from background. One disadvantage associated with this present configuration of the specific FEI, however, is that the size of the MC samples required as input to the training significantly exceeds that of the generic FEI, and is dependent upon the reconstruction efficiency of the chosen B_{sig} decay mode. At present, these sizes are largely only recommended and in need of validation, with the outcomes obtained from the conducted analysis detailed in the following section.

With the development of the FEI on the whole being relatively new and part of a software framework undergoing constant updates in anticipation of the upcoming collision data collection, the specific FEI in particular has been utilised thus far in very few studies, primarily in the context of rare leptonic decays. These included analyses for which the specific FEI was trained for the signal decay modes $B^+_{sig} \to \tau^+\nu_\tau$ [34], a MC study conducted in the absence of Belle II beam-induced backgrounds, and $B^+_{sig} \to \ell^+\nu_\ell\gamma$ [11], using Belle MC and data converted to a BASF2-compatible format with Belle beam background conditions substantially cleaner than those expected for Belle II. As such, the performance of the specific FEI both on Belle II MC including the standard estimate of beam background, BGx1, and within the context of semi-leptonic decays introducing hadronic daughters into the decay products was largely unexplored territory at the commencement of this analysis.

Given that sensitivity to new physics is also attributed to rare semi-leptonic decays, it was decided thus to compare the relative performance of the specific and generic FEI when applied to one such mode, $B^+_{sig} \to \rho^0\mu^+\nu_\mu$. Furthermore, as the ultimate aim of the FEI lies in its application to real data produced under heavy beam background conditions, this investigation was designed for MC samples inclusive of the standard BGx1 estimate, with an examination of the behaviour of both FEI methods in the absence of beam background also conducted in order to study any potential change in the relative performance under idealised conditions. Finally, these results were compared to the treatment of the FEI to the simpler leptonic mode, $B^+_{sig} \to \tau^+\nu_\tau$, possessing a higher reconstruction efficiency and purity. These three studies conducted culminated in the determination of a number of conclusions related to the effectiveness and feasibility of the specific FEI in relation to the generic FEI based on a number of key performance indicators, and these are summarised and discussed in detail below.

8.1 Conclusions

8.1.1 Discrepancy in Performance

The charged B-meson tagging modes incorporated within the framework of the FEI consist of 29 individual hadronic and 42 semi-leptonic channels, covering a range of branching fractions of the order of 10^{-4} up to a few percent. As mentioned in Section 6.3.1, for a classifier

to be trained for a particular channel, nominally a minimum of 500 correctly reconstructed and 500 incorrectly reconstructed instances of the channel must be present in the MC events passed to the FEI training. Furthermore, in the application of the FEI to the signal samples utilised within this analysis, a trained channel was found to be usable only when the reconstructed B_{tag} candidates possessed non-zero values of the `SignalProbability` classifier output. In all cases explored, the generic FEI successfully trained a classifier for every B_{had} and B_{SL} channel, whilst the specific FEI failed to do so for four given B_{had} modes, even with the statistics passed to the classifier training more than doubling with the change to the leptonic B_{sig} decay mode. The total number of channels subsequently passing the application stage was also consistently lower in the specific FEI, though increases in the B_{sig} reconstruction efficiency did result in a greater number of successful channels in the majority of cases.

Given the eventual application of the FEI to real data, signal events, particularly those in which the B_{sig} decay mode is rare, will be significantly under-represented when compared to the number of events pertaining to the various backgrounds. Thus, the trained algorithms must be such that they are able to effectively reconstruct correct B_{tag} candidates from signal events for combination with B_{sig} candidates forming an $\Upsilon(4S)$, whilst also rejecting both $\Upsilon(4S)$ candidates built from incorrect combinations in signal and the far more frequent fake candidates arising from background events, to such a degree that the signal is not overwhelmed. The reconstruction efficiency must therefore be considered in conjunction with both the purity and background rejection capabilities in making any determinations about the relative effectiveness of the specific and generic FEI. When considering the BGx1 study conducted for $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ in Chapter 6, the $\Upsilon(4S)$ efficiency vs. purity curves obtained (illustrated in Figure 6.1) for all successful channels generated from signal MC demonstrated the superiority of the generic FEI in the reconstruction efficiency achieved across the entirety of the sampled purity range for hadronic modes, and the vast majority of this range for semi-leptonic modes, being overtaken by the specific FEI only at higher purities associated with tighter cuts on the `SignalProbability`. This discrepancy in performance was only enhanced when considering background rejection (see Figures 6.4 and 6.5), with the specific FEI consistently corresponding to a lower signal reconstruction efficiency for equivalent levels of background rejection.

8.1.2 The Effect of Beam Background

The omission of beam background in the $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ analysis amounted to higher reconstruction efficiencies both for B_{sig} as calculated, and the individual B_{tag} channels, implied by the increases observed in the total number of usable channels provided by the specific and generic FEI indicating the use of greater statistics in the training of the classifiers. However, this increase was not uniform across the two methods, due to the fact that the number of events passed to the classifier training was first modified by the B_{sig} reconstruction efficiency in the specific FEI, which in this case represented only a moderate improvement from the BGx1 case. As such, the proportional increase detected in the number of successful channels was greater in the generic FEI, a result reflected in the BGx0 efficiency vs. purity curves (see

Figure 6.8) and the background rejection plots (see Figures 6.9 and 6.10), in which an even greater discrepancy in performance can be observed for both hadronic and semi-leptonic tag modes.

The observations above, detailing the effect of the number of usable B_{tag} channels on the resultant efficiency distributions for $B^+_{sig} \to \rho^0\mu^+\nu_\mu$, demonstrate the relative under-training of the specific FEI despite the MC samples supplied as input to the training stage consisting of ≈100 million more events than the generic FEI. The only observed case thus far in which the specific FEI delivered a higher degree of performance when simultaneously considering the reconstruction efficiency, purity and background rejection corresponded to the Υ(4S) candidates reconstructed from semi-leptonic tags and B_{sig} candidates with decay mode $B^+_{sig} \to \tau^+\nu_\tau$, in MC containing no beam background (BGx0). In this case, the background rejection plots (see Figure 7.4) portray relationships largely comparable between the specific and generic FEI, with the equivalent efficiency vs. purity curves (see Figure 7.2(b)) revealing specific FEI reconstruction efficiencies higher or consistent with those of the generic FEI over the majority of the purity range sampled. This was evaluated in light of the fact that the difference in the number of usable B_{SL} channels between the specific and generic FEI was the smallest observed of all cases studied, with the additional channels present in the generic FEI contributing insignificantly to the number of correctly reconstructed Υ(4S) candidates, in turn making the number of successful channels more or less equivalent between the two methods.

The success of the specific FEI apparent for the semi-leptonic tagging modes was not mirrored in the investigation of the hadronic modes, however, with a similar performance gap maintained between the specific and generic FEI as seen in the previous $B^+_{sig} \to \rho^0\mu^+\nu_\mu$ analyses. This represented a single facet of a trend noticeable across all studies performed, whereby the discrepancy between the specific and generic FEI was consistently higher for hadronic modes as opposed to semi-leptonic modes. This was established firstly through the proportion of B_{SL} channels passing the FEI application exceeding that for B_{had} channels in the specific FEI recorded for each of the three studies. Given this relationship, the specific FEI hence failed to reach the performance of the generic FEI at any given purity in the efficiency vs. purity curves (see Figures 6.1(a), 6.8(a) and 7.1(a)), whilst some degree of overlap was observed for the equivalent semi-leptonic curves (see Figures 6.1(b), 6.8(b) and 7.1(b)).

8.1.3 Hadronic vs. Semi-leptonic Tagging

When considering the efficiency vs. purity curves for cases involving only channels mutually successful between the two FEI methods, with the interpretation of such being a rough estimation of the relative performance expected if the MC sample sizes supplied to the FEI training were such that both methods produced similarly trained classifiers, some serious implications can be ascertained. For all studies, these curves corresponded to a specific FEI performing either consistently with (such as for lower purities in Figure 6.3(b)), or better than the generic FEI for semi-leptonic tagging. However, this effect was reversed for hadronic

tags, with the specific FEI either reaching consistency with the generic FEI at restricted purity ranges (as in Figures 6.3(a) and 6.9(a)), or remaining at performance levels below the generic FEI (as seen in Figure 7.2(a)). This observation exacerbates the issues pertaining to the practicality of the training statistics required for the specific FEI, as even if input samples large enough to train the specific FEI classifiers in a way that surpasses those of the generic FEI for B_{SL} channels could practically be used, no guarantee can be made that the hadronic classifiers will likewise achieve superiority without a further increase to these statistics.

A probable explanation for this preferred treatment of the specific FEI towards the B_{SL} tag channels is that, given the semi-leptonic and leptonic B_{sig} decay modes studied, sources of missing 4-momentum in the form of neutrinos existed on both the tag and signal sides of the $\Upsilon(4S)$ decay. In the reconstruction of B_{sig} detailed in Chapter 5, the chosen selections on the beam-constrained mass and energy difference were corrected for the neutrino 4-momentum, resulting in the final list of B_{sig} candidates constructed taking this neutrino into account. As the B_{sig} reconstruction was performed first in the specific FEI, with the classifiers trained upon the rest-of-event objects of each B_{sig} candidate, the resultant B_{tag} candidates would also be constructed in light of this consideration. Thus, the specific FEI in principle is in a better position to distinguish the sources of missing energy between the signal and tag sides in cases where both the B_{sig} and B_{tag} decay modes involve neutrinos. In the generic FEI, however, B_{tag} candidates are reconstructed independently of the B_{sig} decay mode and thus are devoid of any prior information relating to the 4-momentum of neutrinos associated with the signal side, suggesting the cause of the efficiency vs. purity curves for common successful channels favouring the specific FEI. The hadronic FEI does not suffer from cases in which both B-mesons from an $\Upsilon(4S)$ parent decay to neutrinos, with B_{had} candidates capable of being reconstructed from tracks and clusters associated with every particle involved in the decay chain. As such, the well-trained classifiers characteristic of the generic FEI presented in these studies, despite being signal-independent, were likely better able to distinguish between correct and incorrect candidates in the hadronic channels to a higher degree of performance than the specific FEI.

8.1.4 Summary of Findings

When all is considered, three individual studies revealed a specific FEI unable to compare with the generic FEI trained using a significantly smaller number of events, in terms of the proportion of usable channels and subsequent relationships between the $\Upsilon(4S)$ reconstruction efficiency, purity and background rejection, with the exception of the semi-leptonic tagged $B^+_{sig} \to \tau^+\nu_\tau$ analysis. This exception, being a study performed in the absence of beam background, provided useful information on the effectiveness of the FEI under idealised conditions. However, neither cases involving the inclusion of beam background mirrored this result, and the conditions most closely representing the expected Belle II environment for which the FEI should be optimised did not result in a greater performance of the specific FEI. The total number of events supplied to the specific FEI training amounted to between 280 and 290 million across the three instances of the training conducted, and corresponded to

a considerable run-time, with each training taking between 2-5 days to complete running on the KEKCC cluster. Moreover, at present, the official Belle II MC campaigns incorporate the generation of charged, mixed and continuum background samples as well as signal events for a variety of B_{sig} decay modes, including $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$ and $B^+_{sig} \to \tau^+ \nu_\tau$. However, whilst the background samples available in the 7th official production were large enough for use in the training, the signal samples had to be generated personally to achieve the desired sample size, which proved to be another CPU-intensive task that took between a few days and a week for each set of 100 million signal samples generated. Thus, a sizable amount of time and effort was required for the execution of the specific FEI on the aforementioned statistics, with these ultimately proving insufficient for the training of a specific FEI comparable to the generic FEI, for which the training output could be readily accessed from the KEK computing system. It can therefore be concluded that the specific FEI in its current form, at the very least in the context of the semi-leptonic and leptonic decays studied, has proven to be impractical for implementation in analyses. Until such a point at which the specific FEI has undergone modifications necessary for it to be feasible for use, the generic FEI is seemingly the better choice for the reconstruction of tagged $\Upsilon(4S)$ events.

8.2 Future Directions

This study, whilst seeming to be relatively straightforward in its execution at the time of conception, ultimately was found to raise as many questions as it answered. In considering the broader implications of the comparative study between the specific and generic FEI performance, numerous lines of inquiry can be identified by which the current FEI framework stands to benefit from further investigation. An obvious example is the need for further validation of the optimal sample sizes utilised in the training of the FEI and their resultant effect on the number of usable channels at the application stage. A meaningful extension to such an examination could lie in revisiting the candidate thresholds defined by the FEI for the successful training of multi-variate classifiers in order to analyse any relative changes in performance. The depth of exploration into these issues allowed for within the scope of the project undertaken was fairly limited. In turn, the findings presented in this thesis have been communicated to those responsible for the FEI framework with the aim of encouraging further inquiry into potential ways of improving the current configuration of the specific FEI.

In terms of potential extensions to the analysis presented on the signal decay modes $B^+_{sig} \to \rho^0 \mu^+ \nu_\mu$ and $B^+_{sig} \to \tau^+ \nu_\tau$, one possibility considered but ultimately not realised was the training of an independent neural network or alternative multi-variate analysis technique for the suppression of continuum background. This is a common technique employed in both tagged and untagged analyses and relies upon the inherent differences between the nature of continuum events, in which the electron-positron collisions do not result in the creation of an $\Upsilon(4S)$ meson, and charged or mixed background events in which they do, resulting in the subsequent decays of pairs of B-mesons.

In addition, during the course of this project, newer official Belle II MC campaigns were

produced with software updated for modifications to the parameters including the magnetic field and detector geometry. These MC samples were not implemented in the ongoing analysis for the purpose of consistency with earlier measurements obtained from the 7th official production. Thus, potential exists for the presented analysis to be replicated on these updated samples.

To conclude, the Full Event Interpretation technique, being the primary method available for tagged analysis at Belle II, is currently being utilised in a number of analyses, with this only expected to increase as the experiment progresses towards long-term data collection. As such, its functionality must be well-understood and its performance thoroughly validated. The studies conducted were one such contribution to this goal, with further investigation anticipated within the greater collaboration.

Bibliography

[1] T. Keck et al., The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment, arXiv:1807.08680 [hep-ex], 2018. Submitted to Computing and Software for Big Science.

[2] F. Abe et al., Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab, Phys. Rev. Lett., 74:2626–2631, 1995.

[3] DONUT Collaboration, Physicists Find First Direct Evidence for Tau Neutrino at Fermilab, Fermilab Press Release, PRESSCUT-2000-037, 2000.

[4] ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys.Lett. B716, 1-29, 2012.

[5] A. D. Sakharov, Violation of CP Invariance, C Asymmetry and Baryon Asymmetry of the Universe, Journal of Experimental and Theoretical Physics, 5, 24-27, 1967.

[6] M. Kobayashi, T. Maskawa, CP-Violation in the Renormalizable Theory of Weak Interaction, Progress of Theoretical Physics, Volume 49, Issue 2, 652–657, 1973.

[7] B. A. Schwartz, Belle II Collaboration, The Belle II Experiment, Nuclear and Particle Physics Proceedings, 260, 233-237, 2015.

[8] K. Abe et al., Belle Collaboration, Observation of Large CP Violation in the Neutral B Meson System, Phys.Rev.Lett.87:091802, 2001.

[9] B. Aubert et al., BABAR Collaboration, Observation of CP Violation in the B^0 Meson System, Phys. Rev. Lett. 87, 091801, 2001.

[10] T. Keck, Machine learning algorithms for the Belle II experiment and their validation on Belle data, EKP-2017-00067, 2017.

[11] M. Gelb, Search for the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$ at Belle and Belle II, Presented at the Cracow Epiphany Conference on Advances in Heavy Flavour Physics, Krakow, Poland, 2018.

[12] M. Tanabashi et al., Particle Data Group, Phys. Rev. D 98, 030001, 2018.

[13] Public domain image.

[14] S. Descotes-Genon, P. Koppenburg, The CKM Parameters, Annual Review of Nuclear and Particle Science, 67, 97-127, 2017.

[15] A. J. Bevan et al., The Physics of the B Factories, Eur. Phys. J. C74 3026, 2014.

[16] J. H. Christenson et al., Evidence for the 2π decay of the K_2^0 meson, Phys. Rev. Lett. 13, 138, 1964.

[17] T. Browder, B Mesons, Prog.Part.Nucl.Phys.35:81-220, 1995.

[18] M. K. Sundaresan, Handbook of Particle Physics, CRC Press, 2001.

[19] O. Eberhardt et al., Status of the Two-Higgs-Doublet Model of Type II, JHEP 07 118, 2013.

[20] N. Toge (ed.)., KEKB B-Factory Design Report, KEK Report 95-7, KEK, 1995.

[21] K. Ikado et al., Belle Collaboration, Evidence of the Purely Leptonic Decay $B^- \to \tau^- \bar{\nu}_\tau$, Phys.Rev.Lett. 97 251802, 2006.

[22] T. Abe et al., Belle II Technical Design Report, KEK Report 2010-1, arXiv:1011.0352 [physics.ins-det], 2010.

[23] C. Z. Yuan, Belle II Collaboration, The Belle II Experiment at the SuperKEKB, Presented at Flavor Physics and CP Violation, Hefei, China, 2012.

[24] B. Wang, Belle II PID Group, The Assembly of the Belle II TOP Counter, Nuclear Instruments and Methods in Physics Research A 766, 204-207, 2014.

[25] S. Sandilya, Belle II Collaboration, Particle Identification with the TOP and ARICH detectors at Belle II, arXiv:1706.08515 [physics.ins-det], 2017.

[26] T. Auschev et al., A scintillator based endcap K_L and muon detector for the Belle II experiment, Nuclear Instruments and Methods in Physics A 789, 134-142, 2015.

[27] A. Abashian et al., Muon identification in the Belle experiment at KEKB, Nuclear Instruments and Methods in Physics Research A 491, 69–82, 2002.

[28] Y. Iwasaki et al., Level 1 Trigger System for the Belle II Experiment, IEEE Transactions on Nuclear Science, 58:4, 2011.

[29] T. Kuhr et al., The Belle II Core Software, arXiv:1809.04299 [physics.comp-ph], 2018.

[30] R. Itoh et al., Belle Analysis Framework, J. Phys.: Conf. Ser. 396 022026, 2012.

[31] R. Brun, F. Rademakers, ROOT - An Object Oriented Data Analysis Framework, Nucl. Inst. & Meth. in Phys. Res. A 389, 81-86, 1997.

[32] A. Ryd et al., EvtGen: A Monte Carlo Generator for B-Physics, EVTGEN-V00-11-07, 2005.

[33] S. Agostinelli et al., Geant4 — a simulation toolkit, Nuclear Instruments and Methods in Physics Research A 506, 250-303, 2003.

[34] T. Keck. The Full Event Interpretation for Belle II, IEKP-KA-2014-18, 2014.

[35] S. Neuhaus et al., A neural network z-vertex trigger for Belle II, J. Phys.: Conf. Ser. 608 012052, 2015.

[36] M. Feindt et al., The NeuroBayes neural network package, Nuclear Instruments and Methods in Physics Research A 559, Issue 1, 190-194, 2006.

[37] A. Hoecker et al., TMVA: Toolkit for Multivariate Data Analysis, arXiv:physics/0703039 [physics.data-an], 2007.

[38] T. Keck, FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification, arXiv:1609.06119 [cs.LG], 2016.

[39] J. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis Volume 38, Issue 4, 367-378, 1999.

[40] C. Pulvermacher, Analysis Software and Full Event Interpretation for the Belle II Experiment, BELLE2-PTHESIS-2016-003, 2015.

[41] J. Schwab, Calibration of the Full Event Interpretation for the Belle and the Belle II experiment, EKP-2017-00058, 2017.

[42] A. Sibidanov et al., Belle Collaboration, Study of Exclusive $B \to X_u \ell \nu$ Decays and Extraction of $|V_{ub}|$ using Full Reconstruction Tagging at the Belle Experiment, Phys. Rev. D 88, 032005, 2013.