Just a little bit of flavour: A search for $B^+ \to \mu^+\nu_\mu$ using semileptonic-tagging at Belle II

Shanette De La Motte

Supervisors:
Professor Paul Jackson

A thesis submitted towards the degree of Doctor of Philosophy

School of Physics, Chemistry and Earth Sciences
The Faculty of Sciences, Engineering and Technology
The University of Adelaide

November 28, 2024

Abstract

While all fundamental mechanisms predicted by the Standard Model have been observed, there still remains universal phenomena that cannot be characterised using this current formulation of particle physics. A natural way of extending Standard Model theory to include unexplained phenomena like dark matter and gravitation is to search for theoretical anomalies in precision measurements of particle physics parameters, such as those seen in B-meson decays. One such decay is $B^+ \to \mu^+\nu_\mu$, a rare leptonic decay that has not yet been seen in high energy physics experiments, but is expected to take place due to its similarity to the observed $B^+ \to \tau^+\nu_\tau$ decay. Additionally, measurements of the branching fraction of $B^+ \to \mu^+\nu_\mu$ can also be used to constrain the key flavour physics input parameter V_{ub}, which has shown consistent disagreement when measured in exclusive and inclusive B decays. A measurement of $B^+ \to \mu^+\nu_\mu$ with sufficient statistical significance could possibly confirm Standard Model expectations and resolve V_{ub} tensions. Likewise, it could also support said tensions, indicating the presence of new physics mechanisms which potentially link to the unexplained universal phenomena.

This thesis presents a search for $B^+ \to \mu^+\nu_\mu$ using $362\,\mathrm{fb}^{-1}$ of electron-positron collisions at the $\Upsilon(4S)$ resonance, recorded by the Belle II experiment between 2019-2021. To restrict the search to events most likely to have coalesced as $B\overline{B}$ pairs, a machine learning exclusive B-tagging algorithm called the Full Event Interpretation is used to identify and reconstruct semileptonic B mesons that may decay alongside $B^+ \to \mu^+\nu_\mu$. To further reject non-$B^+ \to \mu^+\nu_\mu$ decays, selection criteria are adopted in a Monte Carlo-based optimisation procedure of the signal-to-background ratio $\frac{S}{\sqrt{S+B}}$. Due to the bias-reduction standards adopted, a final $B^+ \to \mu^+\nu_\mu$ branching fraction pertaining to the Belle II dataset is left for future work. However, after a thorough systematic uncertainty analysis in Monte Carlo and signal-depleted collision events, a 90% upper limit on the branching fraction of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29 \times 10^{-9}$ is established using predicted signal yields. This predicted result does not improve upon the current limit, a Belle search using inclusively-tagged B mesons. When instead comparing with the previous semileptonic-tagged limit, having been measured by the BaBar experiment in a much smaller dataset of B decays, this thesis has been able to produce a more constrained result.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Shanette De La Motte

"For (Dr.) Caitlin Elizabeth MacQueen (1991-2021): I wish you were given the chance to finish this before I did."

"I put my heart and my soul into my work, and have lost my mind in the process"
— Vincent Van Gogh

Acknowledgements

In some ways, completing this thesis has made me proud with how many concepts I now understand that I thought I never would. In other ways, completing this thesis has left me the most deflated I've ever been in my life. The following is an acknowledgement and thank you to those who have done their best to keep me inflated for just a little bit longer throughout these 5 years.

Thank you to Prof. Paul Jackson, for his guidance as supervisor and for taking a chance on yet another Melbourne Physics refugee! The choice to move to Adelaide was questioned by many, including myself, but when COVID struck Melbourne as hard as it did, I became immensely grateful for his choice to accept me onto Team Jack!

Similarly, thank you to the CoEPP/CSSM/CDMPP students and academics in the UoA Physics department, for their welcome and acceptance. This is especially directed towards the admin staff, Sharon Johnson, Emily Campbell and Silvana Santucci. Silvana, you are an angel amongst us mere mortals, thank you for being an Adelaide-mum to me!

Speaking of Adelaide-mums (though she'd probably prefer I say Adelaide-sister or Adelaide-*hala*!), I must also thank Dr. Shagufta Perveen. When we lived together in Toorak Gardens, she made sure to keep me fed with incredible vegan curries! I am incredibly thankful to have both her and her son Ashras in my life!

Thank you to Dylan Morgan, for keeping me filled with pasta, memes, podcasts and support. This is also extended to the Morgan family, who housed me when I visited Melbourne and needed to stay within 'Zone 1' myki fare!

Thank you also to fellow Apocry-friend Greg Peiris, for helping me learn more about myself and culture. Eight hours away and five years later, I am proud to have you as a friend.

Thank you to Dr. Tomas Howson, who has gone above and beyond to help me meet this milestone. Tom, you are one of a kind, in how selflessly you set aside your own work to whole heartedly help not just me, but everyone in this department. His assistance has included (but not limited to): providing the laptop that this thesis has been written on, BibTeX assistance, bank-rolling Grill'd feasts and the removal of splinters. People often say that certain accomplishments could not have been achieved without a key supports, but I *know* I could not have done this without you.

The most thanks goes to my parents. I am sorry it has taken me so long to fully realise how instrumental you have been in my education. To Dad, for instilling a love of science (and science-fiction!) in me. For the sacrificing of his time for a job with a two-hour daily commute so that he could support all of us, even as I decided to stay in university until I was 30! To Mum, for adapting her cooking to support me in my choice to become vegan. But most of all, for sacrificing her proximity to her family and taking a chance on coming to Australia, so that myself, my sisters, my cousins have the best possible education.

Finally, thank you to all the people along the way who have helped me smile. This is almost certainly an incomplete list but in this thesis-writing haze I attempt to list them all here!

- Abhishek Sharma
- Alec Gunn
- Ali Simpson
- Andre Scaffidi
- Annette De La Motte
- Arwa Nawaz
- Bill Loizos
- Braden Moore
- Cameron Harris
- Cate MacQueen
- Charles Grant
- Curtis Abell
- David Liptai
- Dona Kireta
- Dorita De La Motte
- Emily Filmer
- Filip Rajec
- Geoff McNulty
- Hannah Vuong
- Haylea Purnell
- Hitarthi Pandya
- James Biddle
- Jason Oliver
- Josh Charvetto
- Josh Crawford
- Judith Kull
- Julia McCooey
- Kate Atwell
- Katey Chrisp
- Kimmy Cushman
- Kristy McNamee
- Lissette De La Motte
- Meera Deshpande
- Melissa McIntyre
- Michael Chrisp
- Mischa Batelaan
- Nick Hunt-Smith
- Paul Gill
- Phil Grace

- Red Steele
- Riley Patrick
- Rob Perry
- Rose Smail
- Ryan Bignell
- Sangeeta Sathyanath
- Sarah Watzdorf
- Scott Williams
- Seb Matthews
- Skye Platten
- Tarini Fernando
- Theo Motta
- Tommi Kabelitz
- Tristan Ruggeri
- Zach Matthews

Preface

This work is a summary of my last 5 years as a postgraduate researcher with the Belle II experiment. One (of many) of the challenges in writing a thesis is capturing the tone correct for those who will ultimately read it. The primary goal of this writing is to convey what I have learned to leaders in the high-energy physics community, people who would most likely be familiar with most of the jargon, techniques and phenomena I have endeavoured to define. However, I have also kept in mind that this may also be read by people who may not have the same background, such as the students who take on my work in the future (*Hi Cameron!*) and my parents (*Hi Dad, hello Mum if you make it this far!*). From what I've seen with my supervisor, I may even be able to bring myself to consult this work in the future, when I'm (many) years removed from this experience.

For this reason, I have done my best to make minimal assumptions of the reader. This may result in some points seeming laboured, but this was done with the intention of avoiding similar frustrations I ran into when attempting to reproduce work of others (*not you, Red!*). There certainly will be places where I will have been less successful in this goal, given the complexity of this work (*Sorry Dad!*). Thus, for future students in particular, I welcome any clarifying questions about what I've written! You can address all questions to `saphdlm@gmail.com` and I will be happy to discuss!

The structure of this thesis is the following:

- Chapter 1, the introduction of this thesis, summarises the properties of the $B^+ \to \mu^+\nu_\mu$ decay, as a flavour physics process, as well as previous $B^+ \to \mu^+\nu_\mu$ searches.
- Chapter 2 describes how the data used for this search was produced, recorded and prepared by the Belle II experiment.
- Chapter 3 presents the foundation and the results of the B-tagging algorithm essential to this work, the Full Event Interpretation.

- Chapter 4 reconstructs the full $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ (and charged conjugate) process. Selection criteria in Monte-Carlo simulated events are chosen to optimise the number of $B^+ \to \mu^+\nu_\mu$ events yielded. After they are applied, agreement between simulation and data in key variable distributions is investigated.

- Chapter 5 includes an analysis of the uncertainties in the measurement of $B^+ \to \mu^+\nu_\mu$ events, concluding with the evaluation of a predicted $B^+ \to \mu^+\nu_\mu$ branching fraction and confidence limit for potential signal yields.

- Chapter 6 concludes this thesis, by summarising what has been performed in the analysis as well as the key derived results.

Charge conjugation is assumed throughout this thesis, unless specified otherwise. References to $B^+ \to \mu^+\nu_\mu$ are generally inclusive of both $B^- \to \mu^-\overline{\nu}_\mu$ and $B^+ \to \mu^+\nu_\mu$.

Contents

List of Figures

List of Tables

Chapter 1.

Introduction: The Physics of $B^+ \to \mu^+\nu_\mu$

In Chapter 1, the rare, high-energy particle decay $B^+ \to \mu^+\nu_\mu$ (and its charge conjugate) is introduced as the decay of interest for this thesis.
The goal of the Chapter is to motivate that precision measurements of this decay can reveal tensions in the current modelling of particle physics. Such tensions can provide a natural way of integrating new theories of unexplained universal phenomena.
The primary behaviour of the fundamental forces and particles of the Standard Model of particle physics will be discussed as it pertains to the decay $B^+ \to \mu^+\nu_\mu$.
The secondary effects of this sector of the Standard Model, that is the flavour phenomenology, will be presented in detail. This includes CKM flavour mixing, flavour oscillation and charge-parity violation, as applied to B-physics.
The properties of the flavour sector will be extrapolated to the kinematics of the $B^+ \to \mu^+\nu_\mu$ decay, as an illustration of the decay signatures that can be exploited in their detection.
A discussion of how the measured parameters of $B^+ \to \mu^+\nu_\mu$ can reveal new physics phenomena will be presented.
The previous attempts at detecting the rare $B^+ \to \mu^+\nu_\mu$ decay in high-energy physics colliders will be studied, especially how they leverage the kinematic signatures to perform their measurements.
These previous attempts go on to inform this search for $B^+ \to \mu^+\nu_\mu$ within the Belle II experiment, where the searchs outline will be summarised in the final section of this Chapter.

$\bowtie \bowtie \bowtie$

Particle physics experiments perform some of the most precise measurements of natural phenomena in science. Quantities such as the mass of the electron are known within 1 part in 100 trillion [1]. 20$^{\text{th}}$ century particle physics was primarily concerned with formulating a self-contained theory of subatomic matter and its interactions, in a period when new particles were frequently being added. With the most recent fundamental particle discovery being in 2012, that of the Higgs boson in proton-proton collisions by the CERN experiments, particle physics research enters a new paradigm. In the 21$^{\text{st}}$ century, the focus shifts to testing the boundaries of existing assumptions via precision testing the consequences of the model in larger and larger collision datasets. Through these tests, physicists look to either confirm the rigour of theory or discover tensions between predictions and observations. Just as discovering new particles in the 1950's challenged the understanding of nature at the time, identifying tensions in measurement offer valuable opportunities to be linked to universal phenomena not yet explained by particle physics. Such phenomena include gravitation and dark matter, an invisible, gravitationally attractive substance inferred to surround galaxies and impact their rotation.Thus, today's particle physicists are not merely performing high precision measurements for measurements sake. With no particle model of either gravitation or dark matter currently realised in nature, it is clear that the current particle physics formulation requires further efforts to be a complete theory of everything.

The decays of B particles are of interest to a number of large scientific collaborations, due to long-standing tensions in theoretically calculated observables and their observed quantities. This thesis will endeavour to perform precision measurement of the rare B decay, $B^+ \to \mu^+\nu_\mu$. This decay that has not yet been observed in any collider experiment and thus provides an excellent laboratory for uncovering new physics.

In order to search for potential anomalies in this decay, it must first be understood how it is theoretically modelled. The theory that underpins this B decay, as well particle physics more generally, is referred to as the *Standard Model*.

1.1. Understanding $B^+ \to \mu^+\nu_\mu$ via the Standard Model

The subatomic physics decay $B^+ \to \mu^+\nu_\mu$ is that of an electrically-charged B, a composite particle. The B decays to the electron-like particle, a muon (μ), alongside the electrically neutral neutrino (ν), a particle with near zero mass. The process can be drawn as a Feynman diagram, as in Figure 1.1.

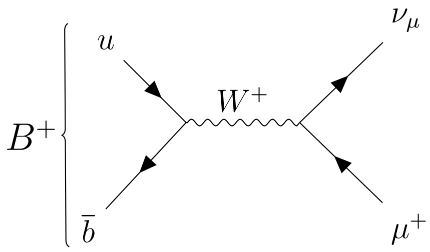


Figure 1.1. A Feynman diagram of the decay $B^+ \to \mu^+\nu_\mu$, indicating the effect of Standard Model forces that facilitate this process.

The three stages, the parent B state, the decay and the two final state particles, neatly highlight the involvement of three mechanisms of the Standard Model (SM) in the $B^+ \to \mu^+\nu_\mu$ process. These mechanisms are the three forces observed in nature, excluding gravity. The SM mathematically encodes these forces and the matter they act on as between as quantum fields. From this theory, particle physicists combined the known quanta of matter and force mediators into a finite set of particles that underpin natural phenomena. These particles are presented in Figure 1.2

SM theory has also been used to mathematically predict and later experimentally confirm the existence of further ones. A complete explanation of the matter, mediator and resulting mathematical mechanisms of SM is not the focus of this thesis, and thus the scope is limited to how SM explains and predicts $B^+ \to \mu^+\nu_\mu$ process.

1.1.1. Standard Model Particles and their Interactions

Muons and neutrinos

The μ and ν_μ in the final state of $B^+ \to \mu^+\nu_\mu$ are a pair of matter particles referred to as *leptons*. As seen in Figure 1.2, there are six leptons in the SM, occurring in three generations of pairs between a neutral and a charged lepton.

Much like electrons, muons are one of the electrically charged leptons. Muons are thus susceptible to the electromagnetic force, via photon interactions. Unlike electrons, the larger-massed muons are unstable particles and therefore do not naturally occur

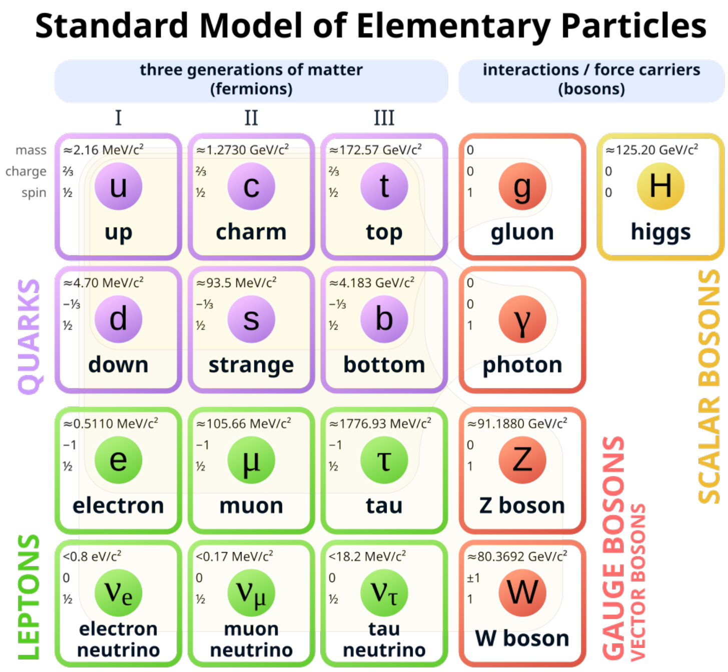


Figure 1.2. The twelve fundamental fermions and five fundamental bosons of the Standard Model of Particle Physics [2]. Each particle is uniquely defined by its mass, charge and intrinsic spin.

in atomic structures. Instead, muons are created as a secondary decay product in the cosmic ray interaction of upper Earth atmosphere nuclei. Muons produced in the decay $B^+ \to \mu^+\nu_\mu$ have lesser momentum than that of atmospheric ones [3], though they are still some of the most energetic produced in any B decay. This momentum will be derived in Section 1.2.1.

Due to their relativistic velocity, the lifetime of high momentum muons are time dilated to observers. This results in a longer than expected trajectory before decaying to electrons. Particle physics experiments (such as the one discussed in this thesis), can thus approximate muons as stable, provided their detector dimensions are within the mean muon decay length of around 660 m.

When muons are created, in nature or in experiment, they are often accompanied by a muon-neutrino (ν_μ). The production of neutrinos, as uncharged leptons, created alongside charged leptons is also seen in radioactive beta decay. Specifically, beta decay produces electron-type neutrinos, which are emitted alongside electrons. As neutrinos do not interact via the electromagnetic force, they are notoriously challenging to detect. Most general purpose high-energy physics experiments thus infer the presence of neutrinos via missing energy, a large deficit in the final decay products energy when compared to initial state energy. Section 1.2.3 will present how the muon-neutrino of the $B^+ \to \mu^+\nu_\mu$ decay will impact the search within this thesis.

The challenges in neutrino detection also makes it difficult to investigate phenomena such as neutrino mass. The SM has been developed with the assumption of massless neutrinos. Despite this, neutrinos of different generations have been observed to spontaneously change into each other as they propagate. This is called *neutrino oscillation* and has no counter part in charged leptons (i.e. a muon cannot directly become an electron, without additional products). The mechanism of neutrino oscillation is defined using *beyond Standard Model* (BSM) theory. The theory motivates that oscillations can only take place if there are non-zero mass differences between the different kinds of neutrinos. The need for neutrino mass models to explain oscillation also motivates the need for precision measurements of decays containing neutrinos. The link between measuring $B^+ \to \mu^+\nu_\mu$ and constraining neutrino masses is presented in Section 1.3.4.

All physics process concerning leptons, either neutral like the neutrinos or charged like a muon, conserve *lepton number*. Lepton number is the difference between the number of lepton particles and lepton antiparticles. In the $B^+ \to \mu^+\nu_\mu$ final state, there is one regular neutrino and one anti-muon, yielding a lepton number of zero. This is matched by there being zero leptons in initial state. This symmetry is an observed phenomenon of SM. An additional symmetry of leptons is *lepton flavour symmetry*, where leptons can only directly interact with leptons of their own generation. This can be see in $B^+ \to \mu^+\nu_\mu$, where no other kind of neutrino can be created alongside the muon in the SM.

The third and final pair of leptons in the SM consists of τ (sometimes referred to as tauon or just tau) and the tau-neutrino. All interactions with each generation of lepton pairs are identical, called *lepton universality*. This results in decay rates being identical, when ignoring the effects of the charged lepton masses. As is seen within all generations of fermions, the third generation contains the heaviest particles. The tau is the most massive of the leptons, at almost double the mass of the proton. This generational mass ordering is not yet confirmed in the neutrinos, due to ongoing measurements of neutrino mass and their origins. With such a large mass, the tau lepton can decay to produce the two lighter leptons, alongside their neutrinos. However, unlike the muon, taus can also decay to yield non-leptonic particles, namely *hadrons*

Table 1.1. Table of spin-0 B mesons and their masses [4].

B	Structure	Mass GeV/c^2	Antiparticle
B^+	u$\bar{b}$	5.279	B^-
B^0	d$\bar{b}$	5.280	$\overline{B}^0$
B_s^0	s$\bar{b}$	5.367	$\bar{B}_s^0$
B_c^+	c$\bar{b}$	6.275	B_c^-

B: hadrons

Unlike the SM leptons, the B is a composite particle, as indicated by Figure 1.1. As a *hadron*, it has a substructure of fundamental particles. Hadrons are strictly composed of *quarks* in the SM. There are six quarks in the SM, also occurring in three generations of differently-charged pairs.

The structure of the specific B^+ hadron is a combination of *up*-type quark with a *bottom*-type antiparticle quark, while the charge conjugate B^- is the opposite with structure $\bar{u}b$. The u-quark and $\bar{b}$ antiquark are bound by the attractive *strong force*. This mechanics of the strong force is referred to as *quantum chromodynamics* or QCD. The charge of QCD are *colour charges*. Quarks are encoded in the SM as having one of three colours, labelled as 'red', 'green' or 'blue'. Antiquarks have 'anticolour' (i.e. 'anti-red', 'anti-green', 'anti-blue'). They can bind together, provided the net colour of the group adds to zero colour (i.e. even numbers of red, green and blue quarks, or each colour quark is cancelled by an anticolour quark). The colour quantity dictates whether particles can interact via the strong force. Quarks are the only fermions with colour. Additionally, no free propagating quarks have been observed, only existing within hadrons. This is described as *confinement*. All quarks are additionally charged under electromagnetism, with u holding +2/3 the unit electron charge and b holding -1/3. B^+, with a $u\bar{b}$ substructure, has total charge $+\frac{2}{3} +^- \frac{2}{3} = 1$, and therefore is a singly-charged B meson.

The B hadrons are more commonly referred to as *mesons*, owing to their quark-antiquark structure. An alternate hadron is a *baryon*, consisting of three quarks or three antiquarks. Mesons and baryons of larger combinations of quarks have been observed, though their properties are less well known [5]. A B meson, generally, is any meson which contains exactly one b or $\bar{b}$. The four types of ground-state B mesons are listed in Table 1.1. The t quark has the largest mass of any particle in the SM, meaning they decay before hadronising. There is thus no B meson containing a t-quark. The ground-state mesons are *pseudoscalar* particles, meaning they have zero units of angular spin. Excited B mesons of spin 1 and 2 also exist, though their decay paths are not as well known as ground-state B mesons. If a b and $\bar{b}$ are combined, it is referred to as *bottomonium*. Bottomonia vary according to their mass, excitation, spin, and parity. B mesons do not occur naturally in the universe, so in this thesis, $B^+ \to \mu^+\nu_\mu$ events will be sourced from the decays of a spin-1 bottomonium state, $\Upsilon(4S)$.

Understanding the nature of the strong force within hadrons such as B mesons is non-trivial, owing to properties of the strong force mediator. The gauge boson of QCD is the *gluon*. Unlike photons, which have no electric charge themselves, gluons are in fact charged under the force they mediate. Gluons posses a pair of colour and anti-colour charges that do not cancel each other out. This means that pure gluonic interactions can take place. As per their name, QCD interactions between quarks and/or gluons are stronger than any other in the SM. To first order, the attraction between the constituents of B^+ can be thought of as u, $\bar{b}$ and a single gluon mediator. More complicated scenarios can also take place within the B^+ meson structure, including additional gluon mediators as well as spontaneous quark and gluon creation and annihilation. When attempting to sum these scenarios together to understand the full hadronic behaviour, the large QCD strength results in higher order diagrams contributing more and more to observables. As such a method cannot converge, non-analytical techniques are needed to evaluate observables of QCD. A QCD observable concerning the $B^+ \to \mu^+\nu_\mu$ decay, is the B decay constant f_B, which is described in Section 1.2.5. Lattice QCD is one such technique that can be used to calculate f_B numerically [6].

For now, having discussed the initial state and the final state of $B^+ \to \mu^+\nu_\mu$, focus proceeds to the theory of how SM models the most complex aspect, the weak decay between the B meson to its leptonic final states

The decay, as a weak interaction

The transition from the initial state to the final state in $B^+ \to \mu^+\nu_\mu$ is governed by the *weak interaction*. Instead of creating atomic structures like electromagnetism or hadronic structures like QCD, the short-ranged weak interaction can mediate particle change. It is the same interaction which facilitates the transmutation of neutrons to protons in radioactive decays.

The weak interaction is mediated by two kinds of mediator gauge bosons, the electrically charged $W^\pm$ and the neutral Z^0. Within $B^+ \to \mu^+\nu_\mu$, it is the W^+ gauge boson that facilitates the u-$\bar{b}$ annihilation, as well as the production of the muonic lepton pair. It is the only boson within the SM which allows for interactions or decays between different types of fermions, provided they have a charge difference of ± 1. The interactions of $W^\pm$ gauge bosons are also referred to as *charged current* interactions. This is to distinguish with those of the Z^0 gauge boson, which is a *neutral current* interaction. The Z^0 gauge boson cannot change particle type in fermions, instead functioning similar to a photon to mediate scattering events with no net particle change. Where the Z^0 differs from a photon is its ability to interact with neutrinos.

An additional difference between Z^0 and photons is that it has mass, an attribute shared with the W gauge bosons. The Z^0 and W bosons are the third and fourth most massive particles of SM. This seemingly contradicts the quantum field theory formalism of SM, in which gauge bosons cannot have mass. The Higgs mechanism was put forward to resolve the observed mass of the weak mediators. Here, a Higgs field permeates space, where Higgs bosons are excitations of the underlying quantum field. Particles of SM acquire their mass proportional to how strongly they interact with the Higgs field. At high enough energies, such as that of the early universe after the Big Bang, the *electroweak interaction* dominates. This interaction is a unification of the electromagnetic and weak interactions, mediated by the same four massless gauge bosons. Once energy reduces to below a threshold of $\mathcal{O}(100\,\mathrm{GeV})$, the Higgs field develops a non-zero expectation value in the vacuum. This allows the four mediators to interact with the Higgs field. In acquiring mass, the bosons separate into respective phenomena of the electromagnetic force and the weak interaction. Thus the massive Z^0 and W boson can be considered gauge bosons, as they are quantum superpositions of the massless electroweak bosons. The Higgs mechanism was confirmed with the Higgs bosons discovery by the CERN experiments in 2012.

The weak interaction is also a *chiral* interaction. Chirality is a mathematical property, indicating if the particle's field is an eigenstate of the left-handed or right-handed projection operator. As a chiral interaction, the weak interaction can only takes place in left-handed particles or right-handed antiparticles. The opposite chirality cases (i.e. right-handed particles and left-handed antiparticles) are otherwise identical to particles with the weakly-interacting chirality, when it comes to other SM processes. The chirality of the weak interaction has a great impact on the rarity of the $B^+ \to \mu^+\nu_\mu$ process, which will be presented in Section 1.2.2. Chirality is closely linked to *parity*. A particle process respects parity symmetry if replacing it with spatially inverted version does not yield a change in its measured observables. As a result of chirality, the weak interaction is the only interaction to break parity symmetry, first observed in the biased emission of beta radiation in cobalt-60 atoms [7].

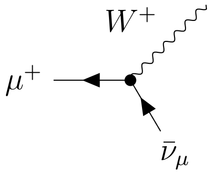
The basic phenomena of the weak interaction have been presented, though arguably the most important characteristic has not yet been discussed. The particle identity changing of the weak interaction, or the *flavour change* of fermions, and will be elaborated upon in the following section.

1.1.2. Flavour Physics in the Standard Model

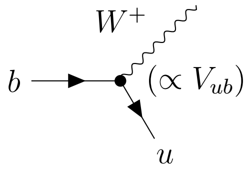
The general SM phenomena have been described. The focus is now narrowed to the *flavour sector* of SM, as the decay $B^+ \to \mu^+\nu_\mu$ is an excellent probe of flavour physics phenomena. The term flavour is used to describe the different varieties of fundamental fermions, such as the b-flavoured quarks or the muon-neutrino-flavoured leptons. There are 6 flavours of quarks and an additional 6 flavours of lepton.

While all SM forces can interact with flavoured particles, it is only the weak interaction that can affect a net flavour-change. Example flavour changes are given in Figure 1.3, where both a lepton flavour change and quark flavour change take place. Both are mediated by a W^+ boson, as it is only the charged boson that can create flavour change. Transitions are only allowed if there is a ± 1 difference in electric charge, thus there are no quark flavour to lepton favour direct transitions. The Z^0 cannot be substituted in Figure 1.3 to create a direct transition, such as $\mu^+ \to e^+$ or $b \to s$. Net neutral flavour change, would require additional W interactions for quarks, and beyond SM physics for leptons. A neutral Z^0 process can only facilitate scattering processes (like electron-neutrinos scattering off of electrons of an atom) or annihilation/production of fermions in flavour-anti-flavour pairs (like neutrino production via $e^+e^- \to \nu_\ell\, \bar{\nu}_\ell$, for a given lepton ℓ).

As was mentioned in the previous discussion of leptons, W bosons couple equally to each generation of leptons, and only to leptons of the same family i.e. $W^+ \to \mu^+\nu$, where ν cannot be electron- or tau- flavoured. Within quarks, W bosons can couple (i.e. allow direct interactions between) both within and between generations, provided they have an electric charge difference of 1. This results in a limit to coupling between charged-neutral leptons of the same type and up-type and down-type quarks. The phenomenology of



(a) Lepton transition: must be between charged leptons and neutrinos of the same generation, or vice versa.



(b) Quark transitions: must be between up-type and down-type, or vice versa, between any generations. Coupling strength of this transition is proportional to CKM matrix element V_{ub}.

Figure 1.3. Example allowed flavour-change transitions for Standard Model fermions.

this direct quark flavour change between generations arises from *quark flavour mixing*, and is summarised below. A similar framework defines neutrino oscillations, which is not facilitated by Z^0 bosons, nor SM. The discussions of this is not relevant to the scope of this $B^+ \to \mu^+\nu_\mu$ search, and so only the quark phenomenology is discussed.

Quark Flavour Mixing and the CKM Mechanism

In the SM formalism, quark flavour mixing occurs due to a mismatch between the quantum eigenstates of definite interacting-flavour versus the state of definite mass [8]. Mathematically, this can be expressed as the following.

$$|b^{'}\rangle = V_{td}|d\rangle + V_{ts}|s\rangle + V_{tb}|b\rangle \tag{1.1}$$

Here, $|b^{'}\rangle$ denotes a quark with the weak-interaction properties of a b quark and $|q\rangle$ as a quark that with the characteristic mass of quark q, provided it has the same charge as b. V_{tq} is a complex number that denotes the amplitude of each component and whose nomenclature will be made clear soon. Quarks do not propagate freely due to QCD confinement. However, Eq. (1.1) can be thought of as a quark that is created in a weak interaction with flavour b, that is in fact a superposition or 'mix' of all quark states that would propagate with down-type quark masses. The reverse is also true; a would-be propagating quark of fixed b mass identity can collapse to interact as quarks with d, s or b flavour. It is convention to express quark-mixing in the down-type quarks; a similar derivation could have been performed in the up-type quarks. The term quark-mixing is now justified.

From Eq. (1.1), this b quark mixing can be generalised to expressions for all down-type quarks of flavour q'. To complete the expression, the second quark of fixed flavour in the weak interaction is included. Expressing the second quark as U', a vector of up-type quarks in fixed flavour basis and D' as the corresponding flavour vector of down-type quarks, as defined in Eq. (1.1), the flavour component of the interaction can be constructed as $\langle U'|D'\rangle$ via the following.

$$\begin{bmatrix} \bar{u}' & \bar{c}' & \bar{t}' \end{bmatrix} \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} \tag{1.2}$$

The flavour structure of $\langle U'|V|D\rangle$ has been obtained, with $|D\rangle$ in mass basis. Thus, a quark q_D with definite mass can interact with the quark flavour q_U of any generation with opposing charge sign, through collapsing to one of its underlying flavour states with the corresponding amplitude from V. The amplitudes in V are used in the evaluation of the strength of a flavour coupling between two quarks, which can then be used to evaluate the probability that a given flavour process takes place. In the process $B^+ \to \mu^+\nu_\mu$, the u and $\bar{b}$ quarks annihilate to a W^+, with coupling proportional to V_{ub}.

The matrix V is defined as the *Cabbibo-Kobayashi-Maskawa matrix*, also known as the CKM matrix or quark-mixing matrix. It is named first after Nicola Cabbibo, for their initial work on representing flavour change as a rotation between strange and non-strange hadrons, which was then only compatible only up to a two quark generations [9]. This rotation matrix was extended to three quark generation theory by Makoto Kobayashi and Toshihide Maskawa [10].

As generalisation of Cabbibo's real-valued rotation matrix, the CKM flavour-mixing matrix takes on dimension 3×3 and becomes unitary with complex elements. The magnitudes of the elements are quoted below, reproduced from the Particle Data Group averages [4]. The largest values are along the main diagonal, suggesting that couplings within the same generation are preferred. Thus, transitions between quarks of the same generation are more probable. Uncertainties on these measurements vary from fractions of a percent for couplings between lighter quarks, to more than 2% in $|V_{ub}|$. This interaction between first and third generations has the smallest magnitude of any amplitudes in the matrix.

$$|V_{\text{CKM}}| = \begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97435 & 0.22501 & 0.003732 \\ 0.22487 & 0.97359 & 0.04183 \\ 0.00858 & 0.04111 & 0.999118 \end{bmatrix} \tag{1.3}$$

Unitarity (i.e. $V_{\text{CKM}}^{\dagger} V_{\text{CKM}} = I_3$) is an important feature of the CKM matrix, as it limits quark mixing to the six known flavours of quarks. The following equation can be derived from the unitarity of the CKM matrix.

$$|V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 = 1 \tag{1.4}$$

Here, the sum of squares of CKM matrix elements for b quarks is unity, limiting the b decay transitions to only the three up-type quarks of SM . Thus if it were found that Eq. (1.4) failed, or its counterpart in d or s quarks, it would suggest down-type quarks can decay to to something other than up-type quarks. Such a failure would be a signal of BSM mechanisms contributing to weak quark transitions. It is therefore important to make precision measurements of CKM matrix amplitudes.

The CKM matrix elements are free parameters within the SM, so their values cannot be derived theoretically. With the assumption of unitarity however, the nine complex amplitudes can be reduced to a set of four independent free parameters. Measuring the CKM matrix in particle physics experiments, in either its parameterisations or the direct matrix amplitudes, is necessary to understand particle transitions concerning quarks. While all nine complex amplitudes can be derived from a set of four parameters, measuring all possible magnitudes and phases is an important method for constraining non-unitarity and thus BSM mechanisms. Commonly, what is measured are the nine magnitudes and three combined complex phases

The magnitude of the matrix amplitudes can be quantified by measuring observables of particle processes proportional to the desired CKM amplitude. For example, the branching fraction of $B^+ \to \mu^+\nu_\mu$ is proportional to $|V_{ub}|$, as will be seen in Eq. (1.6), and thus measuring $B^+ \to \mu^+\nu_\mu$ events in experiment could constrain $|V_{ub}|$. Currently, the state of the art in precise measurement of $|V_{ub}|$ is to create partial branching fractions of semileptonic B decays (i.e. B decays where the final state includes hadrons and a lepton pair) for differing q^2 bins (i.e. the four momentum delivered to the lepton pair, squared). Accurately measuring V_{ub} is of great interest. The Heavy Flavour Averaging

Group (HFLAV) has reported inconsistent values of $|V_{ub}|$ depending on if the hadron of the semileptonic B decay is fixed or left unspecified. This will be discussed further in Section 1.2.6.

The main diagonal of $V_{\mathrm{CKM}}^{\dagger}V_{\mathrm{CKM}}$ yields unitarity constraints such as Eq. (1.4). The other six elements yield expressions such as the following equation.

$$V_{ud}^{*}V_{ub} + V_{cd}^{*}V_{cb} + V_{td}^{*}V_{tb} = 0 \tag{1.5}$$

Considering each term as a vector in complex space, the sum's equality to zero suggests the vectors are coplanar and form the sides of a triangle. This construction is referred to as a *unitary triangle*. While there is one expression per the remaining six elements of $V_{\mathrm{CKM}}^{\dagger}V_{\mathrm{CKM}}$, the triangle of the equation above is the most commonly referenced. The three angles of Eq. (1.5)'s unitary triangle are denoted as (α, β, γ) or (ϕ_2, ϕ_1, ϕ_3), respectively. As the triangle is derived under the assumption of unitarity, any measured inconsistencies in the angles or sidelengths can break the triangle and indicate the presence of new physics. The angle $\beta = \arg\left(-\frac{V_{td}V_{tb}^{*}}{V_{ud}^{*}V_{ub}^{*}}\right)$ is especially important within B physics. It's measurement lead to the discovery of charge-parity violation in B mesons.

Flavour oscillation and charge-parity violation

Flavour oscillation and charge parity violation are both features of the weak interaction and the flavour sector. While not directly linked to the $B^+ \to \mu^+\nu_\mu$ process, they are briefly summarised as a key phenomenon in B-physics.

Flavour oscillation is a property of neutral mesons, arising from the quark mixing described above. A neutral meson can be created in a state of definite flavour and later be found to have opposite flavour after travelling some distance, without any additional products. For example, a meson created as B^0 could be later found as $\overline{B}^0$. This arises due to the mixing between the quark flavour and the propagating mass states, such that for B mesons, quark flavour changes of $b \to t \to d$ and $\overline{d} \to \overline{t} \to \overline{b}$ have taken place, as in Figure 1.4. From the quark couplings, measuring B^0 oscillations can constrain V_{tb} and V_{td}. Such a flavour oscillation could not take place for a B^+ meson, as this would violate electric charge.

The measurement of B^0-$\overline{B}^0$ oscillations is closely linked to the phenomenon of charge-parity violation (CPV). Charge conjugation refers to replacing a particle with its antiparticle, or vice versa. As parity is violated in the weak interaction, it was thought

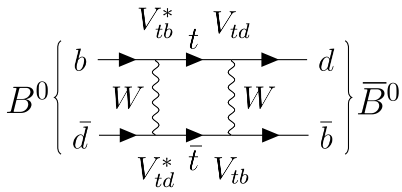


Figure 1.4. Feynman diagram of B^0-$\overline{B}^0$ oscillation, with relevant CKM matrix parameters indicated.

combining a parity inversion with a charge conjugation could restore the symmetry, such that there will be no difference in observables measured for any weak interaction process replaced with its charge-parity conjugate. This was found to not be true.

It is only in certain weak interactions where the charge-parity symmetry is violated. There are three categories of CPV. The first is CPV due to mixing (or *indirect CPV*), occurring in meson-antimeson pairs that can oscillate into each other before decaying. For example, if the rate of $B^0 \to \overline{B}^0$ oscillation is out of sync with the rate of $\overline{B}^0 \to \overline{B}^0$ oscillations, this can result in an asymmetry in the number of B mesons recorded and their observables. The second is CPV in the decay, or *direct CPV*, occurring if a process and its conjugate have decay amplitudes that differ by a complex phase. Decay amplitudes will be proportional to the CKM matrix elements relevant to quark interaction of the decay. The phase of this matrix element combining with that of the complex decay amplitude may interfere constructively in a particle and destructively in its charge-parity conjugate, resulting in different decay probabilities. This kind of CPV can occur in charged mesons as well, as oscillation is not involved. The final kind of CPV is due to interference between mixing and decaying. If the final state of a particle decay is a CP eigenstate (i.e. the decay state is possible for both the particle and its conjugate, like $B^0 \to \pi^+\ \pi^-$ and $B^0 \to \overline{B}^0 \to \pi^+\ \pi^-$), there can be a similar interference in the decay amplitudes, resulting in different decay rates.

CPV via interference between decays with and without mixing is the mechanism behind the first observation of CPV in any mesons other than K mesons, specifically in B mesons. The difference in the normalised decay rates between B mesons that begin as B^0 and those that begin as $\overline{B}^0$ in B^0-$\overline{B}^0$ oscillations is called the *time dependent CP-violating asymmetry*. With both β and the B^0-$\overline{B}^0$ oscillation Feynman diagram containing a $V_{td}V_{tb}^*$ term, the asymmetry quantity is directly proportional to $\sin(2\beta)$. This was the first measurement of this angle of the unitary triangle. It was also the first proof that CPV was a general property of SM physics, rather than a niche property of kaons.

As was mentioned in the description of the CPV categories, the CKM matrix contains complex amplitudes, which can contribute a complex phase to decay amplitudes. Not only did the observation of CPV in B^0-$\overline{B}^0$ oscillations demonstrate that CPV could occur in the SM, but that it was the complex phase from the CKM mechanism that led to this instance of CPV. Moreover, it also validates the existence of (at least) three generations of fermions in nature, as no complex CP-violating phase is possible in the real-valued, two-generation Cabbibo matrix.

The only source of CPV in the SM is within the CKM mechanism. However, more sources of CPV are required to explain nature, such as the current prevalence of matter over antimatter in the universe. This is called the *baryon asymmetry of the universe*. To obtain a matter-dominated universe, physics must show a preference interactions with baryons over antibaryons, which in turn breaks charge-parity symmetry. Taking into account the cosmological evolution of the universe, the magnitude of the CP-violating phase of the CKM mechanism is insufficient to have created the baryon asymmetry today. This suggests there must be additional, BSM CPV phases explain this phenomenon. New CPV phases may be within the quark sector like in the CKM matrix, or within the lepton sector via mixed baryon-lepton transitions. Neutrino oscillation, a BSM phenomenon introduced when discussing leptons in Section 1.1.1, can be modelled via a similar mixing matrix similar to that of CKM with its own CPV phase [11]. This phase is not yet measured with high enough precision to determine how such a phase could have affected the matter-antimatter evolution of the universe.

The flavour sector is a rich area of SM physics. As has been discussed, flavour-changing decays can be important probes into the assumptions of nature. Discovering discrepancies to flavour observables, such as CKM magnitudes and CP-violating phases, can provide a natural way of incorporating new physics phenomena. With the physics of SM and the physics of flavour presented, the discussion can now turn to how these mechanisms lead to the decay of interest for this thesis, $B^+ \to \mu^+\nu_\mu$.

1.2. The features of $B^+ \to \mu^+\nu_\mu$, according to the Standard Model

Branching fractions give a measure of how often a given particle decay occurs, relative to all other modes of the particle. Using the quantum field theory of SM, the theoretical branching fraction for $B^+ \to \mu^+\nu_\mu$ can be expressed with the following equation.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{G_F^2 m_{B^+} m_{\mu^+}^2}{8\pi}\left(1 - \frac{m_{\mu^+}^2}{m_{B^+}^2}\right)^2 \tau_{B^+} f_{B^+}^2 |V_{ub}|^2 \tag{1.6}$$

Table 1.2. Parameter inputs to $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ in Eq. (1.6). Uncertainty contributions are the percentage uncertainty in the SM evaluation of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$, propagated from each parameters uncertainty. $|V_{ub}|$ is substituted as the average of inclusive and exclusive measurements, except for in the evaluation of the uncertainties due to each value of $|V_{ub}|$, where they are used individually.

Param.	Value	Source	Definition	Uncert. contrib.(%)
G_F	$(1.16 \pm 8.48 \times 10^{-7}) \times 10^{-5}$ GeV^{-2}	PDG	Fermi coupling [4] constant	1.54×10^{-4}
m_{B^+}	5279.41 ± 0.07 MeV/c^2	PDG	Mass of B^+	1.33×10^{-3}
m_{μ^+}	$105.7 \pm 2.3\times10^{-6}$ MeV/c^2	PDG/CODATA	Mass of μ^+ [1]	4.35×10^{-6}
τ_{B^+}	$(1.638 \pm 0.004)\times10^{-12}$ s	PDG/HFLAV	Mean lifetime of B^+ [12]	0.24
f_{B^+}	190.0 ± 1.3 MeV	FLAG 2021	Leptonic B decay constant (using combined f_B) [6]	1.37
$\|V_{ub}\|_{\text{excl.}}$	$(3.51 \pm 0.12)\times10^{-3}$	HFLAV 2021	CKM matrix element magnitude (from exclusive decays)	6.84
$\|V_{ub}\|_{\text{inc.}}$	$(4.19 \pm 0.17)\times10^{-3}$	HFLAV 2021	CKM matrix element magnitude (from inclusive decays)	8.11

Eq. (1.6) is a succinct expression, consisting of a number of precisely measured quantities, such as particle masses and lifetimes. The least well know parameters are V_{ub} and f_B, the relevant CKM matrix element and the B leptonic decay constant. Both are B-physics parameters whose measurements are active areas of research in high energy physics experiments and in Lattice QCD respectively. Evaluating Eq. (1.6) using the parameters listed in Table 1.2, (where $|V_{ub}|$ is substituted as the average of the inclusive and exclusive result), yields result.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (4.24 \pm 0.24) \times 10^{-7} \tag{1.7}$$

This is equivalent to a rate of 4.24 $B^+ \to \mu^+\nu_\mu$ events occurring in every ten million B^+ decays, assuming only SM mechanisms contribute to the process. From Table 1.2, it is seen that $|V_{ub}|$ contributes the largest uncertainty to this result, contributing 6.84 and 8.11 % uncertainty depending on the method by which V_{ub} is evaluated. It is noted that the inclusive and exclusive methods of measuring V_{ub} disagree substantially, with the uncertainty on their difference being 30%. The impact of this will be discussed further in Section 1.2.6).

With such a small branching fraction, $B^+ \to \mu^+\nu_\mu$ is considered a rare decay. By comparison, the decay $B^+ \to \bar{D}^0\pi^+$ has a measured branching fraction of $4.61{\times}10^{-3}$ Given the rarity of such an event taking place, it will be important to know the key features of the decay, so that $B^+ \to \mu^+\nu_\mu$ events can be differentiated from other more common events. These features will be discussed in the following section. Additionally, the impact that the B parameters V_{ub} and f_B have on the $B^+ \to \mu^+\nu_\mu$ will also be discussed as a feature of the decay.

1.2.1. Feature 1: $B^+ \to \mu^+\nu_\mu$ is a two-body decay

The kinematics of particle decays can be derived via conservation of energy and momentum. As $B^+ \to \mu^+\nu_\mu$ is a two-body decay, the momentum of the muon can be uniquely determined by considering the decay in the centre-of-mass frame of the B.

$$p_B^B = p_\mu^B + p_\nu^B \tag{1.8}$$
$$\implies (m_B, \vec{0}) = (E_\mu^B, \vec{p}_\mu^B) + (E_\nu^B, \vec{p}_\nu^B) \tag{1.9}$$

From conservation of momentum in Eq. (1.9), the relationship between the muon and neutrino momentum in this frame is produced.

$$\vec{0} = \vec{p}_\mu^B + \vec{p}_\nu^B \tag{1.10}$$
$$\implies \vec{p}_\mu^B = -\vec{p}_\nu^B \tag{1.11}$$
$$\implies |\vec{p}_\mu^B| = |\vec{p}_\nu^B| \tag{1.12}$$

Thus, the muon and the neutrino are created back-to-back in the B-frame i.e. with momenta of equal magnitude and opposite directions. From conservation of energy in Eq. (1.9) the rest-mass energy of the B is converted to kinetic energy of the leptons.

$$m_B = E_\mu^B + E_\nu^B \tag{1.13}$$

After some algebra, the energy-momentum relation of the muon and the neutrino can be substituted.

$$(m_B - E_\mu^B)^2 = (E_\nu^B)^2 \tag{1.14}$$

$$m_B^2 + (|\vec{p}_\mu^B|^2 + m_\mu^2) - 2m_B\sqrt{|\vec{p}_\mu^B|^2 + m_\mu^2} = (|\vec{p}_\nu^B|^2 + m_\nu^2) \tag{1.15}$$

Substituting in Eq. (1.12) and the fact that the SM neutrinos are massless, an expression for the muon momentum in the B-frame is produced.

$$m_B^2 + m_\mu^2 - 2m_B\sqrt{|\vec{p}_\mu^B|^2 + m_\mu^2} = 0 \tag{1.16}$$

$$\implies |\vec{p}_\mu^B| = \sqrt{\frac{(m_B^2 + m_\mu^2)^2}{4m_B^2} - m_\mu^2} \tag{1.17}$$

$$|\vec{p}_\mu^B| = \frac{m_B^2 - m_\mu^2}{2m_B} \tag{1.18}$$

$$|\vec{p}_\mu^B| = 2.64\,\mathrm{GeV}/c = |\vec{p}_\nu^B| \tag{1.19}$$

Up to detector-based smearing and losses from radiated photons, the B-frame momentum magnitude is a fixed value. From Eq. (1.18), muon momentum is predominantly due to the large mass difference between the B meson ($5.28\,\mathrm{GeV}/c^2$) and the muon ($0.106\,\mathrm{GeV}/c^2$).

One could attempt a similar derivation with $B^+ \to \pi^0\mu^+\nu_\mu$ as a three-body decay, by setting $X = \pi^0\nu_\mu$. Now, $\vec{p}_X^B$ and $\vec{p}_\mu^B$ would be back-to-back. In this case, however, Eq. (1.16) would now contain a m_X^2 term on the right-hand side, unlike with what occurs with $m_\nu^2 = 0$. This term still has free parameters, depending on the B-frame momenta of π^0 and ν_μ, and the angle between them. The lepton momentum magnitude in this scenario would not be fixed in the parent particle's frame. Moreover, the numerator of Eq. (1.17) would become $(m_B^2 + m_\mu^2 - m_X^2)^2$, i.e. characteristically less than the fixed value of the two-particle case, $|\vec{p}_\mu^B|$=2.64 GeV/c. This difference in behaviour between three-body and two-body decays, where one yields a broad distribution when measured and the other relatively narrow, is thus a hallmark of the $B^+ \to \mu^+\nu_\mu$ decay.

Experimentally, an observer will not be co-moving with the B meson, so as to directly measure the muon momentum in the B-frame. In the external observer's frame, the muon momentum distribution will appear 'smeared', no longer with a singular value. The distributions of the muon momenta as compared between the lab and various parent frames are presented in Figure 1.5. To be able to apply the singular muon momentum constraint, the boost of the muon provided by the B meson must be derived. Techniques

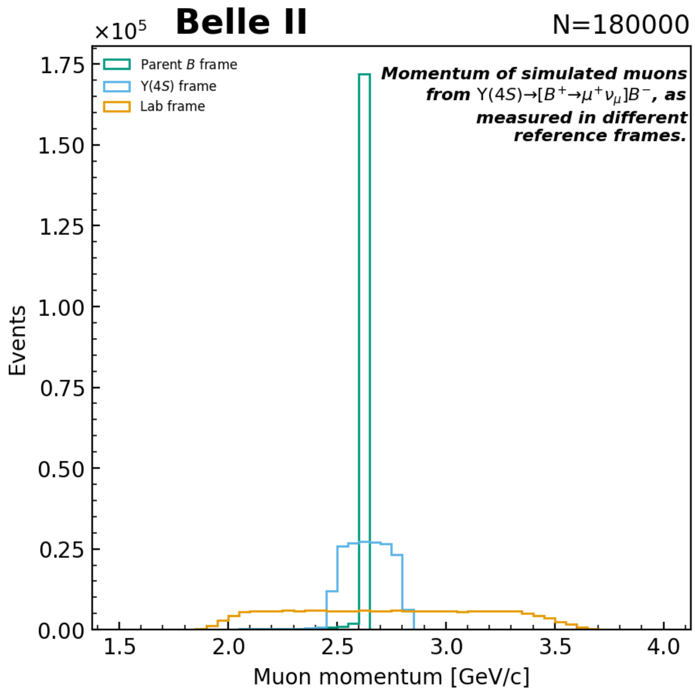


Figure 1.5. Distributions of simulated muon momenta from a $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu]B^-$ (and charge-conjugate) process, as measured in differing reference frames. Smearing of the muon momentum is most significant in the lab frame, while it is singular in the parent B-frame, up to losses due to radiated photons.

of doing so will be presented when studying previous experimental $B^+ \to \mu^+\nu_\mu$ searches in Section 1.4.

1.2.2. Feature 2: $B^+ \to \mu^+\nu_\mu$ is a helicity suppressed decay

From Eq. (1.6), it is seen that the SM theoretical branching fraction is proportional to $m_\mu^2(1 - m_\mu^2/m_{B^+}^2)^2$. Given the mass difference between the muon and the B meson, the $m_\mu^2/m_{B^+}^2$ can be assumed to be negligible, resulting in $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ being directly proportional to the relatively small value of m_μ^2. This encodes the rarity of $B^+ \to \mu^+\nu_\mu$, owing mostly to the *helicity suppression* phenomenon, occurring as a result of the small muon mass.

Helicity is a kinematic quantity obtained by projecting the intrinsic spin onto the direction of motion of a particle. Helicity is not a Lorentz-invariant quantity; in a reference frame moving faster than the particle, the direction of motion reverses, causing the sign of the helicity to do the same. Assuming the neutrino of $B^+ \to \mu^+\nu_\mu$ is a SM

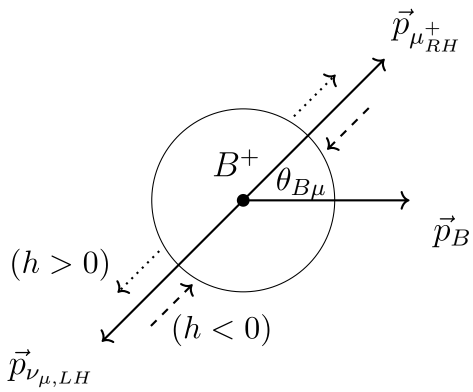


Figure 1.6. A schematic of the back-to-back decay of $B^+ \to \mu^+\nu_\mu$, with potential helicities, in B-frame. The angle $\theta_{B\mu}$ between the muon momentum in the B frame, versus B momentum in an external frame is also designated. Smaller arrows indicate potential spin-projection alignments. Either both particle spins are aligned to the linear momentum (dotted) or anti-aligned (dashed). This corresponds to the particles having negative or positive helicity respectively.

neutrino and massless, it has no faster-moving reference frame. Thus, helicity becomes Lorentz-invariant for only neutrinos.

Considering the two-body dynamics discussed previously, the weak decay of $B^+ \to \mu^+\nu_\mu$ occurs such that neutrino and muon are produced back-to-back in the B-frame. This conserves linear momentum. To conserve angular momentum, the spin-0 B meson must create the muon and neutrino such that their spin-1/2 intrinsic spins cancel to zero. There are two cases to consider: either the lepton spins are aligned to the linear momentum (i.e. pointing outwards) or aligned against (i.e. pointing inwards). In helicity terms, either both leptons are positive helicity (spin parallel to momentum) or negative helicity (spin anti-parallel to momentum). A schematic of this scenario is presented in Figure 1.6

As a weak decay, it is important to also discuss the chirality of the $B^+ \to \mu^+\nu_\mu$ products. For the positively charged B meson decaying via $B^+ \to \mu^+\nu_\mu$, the resulting muon is an antiparticle and thus right handed. The accompanying neutrino is not an antiparticle, and thus is left-handed. The opposite holds for $B^- \to \mu^-\bar{\nu}_\mu$.

Helicity and chirality are related, but not identical concepts. Mathematically, a chiral eigenstate can be expressed as a linear combination of positive and negative helicity eigenstates. For a left-handed chirality state with relativistic energy, the amplitude of the positive helicity component is proportional to mass-energy (m/E) ratio of the particle [8]. The negative helicity component of the left chiral state will be unaffected. The opposite

Table 1.3. Approximate amplitudes of different helicity components in left and right chiral particles.

	L (particle)	R (antiparticle)
$h > 0$	$\frac{m}{E}$	1
$h < 0$	1	$\frac{m}{E}$

is true for a right-handed chirality state; m/E applies to the negative helicity and the positive helicity component is unaffected. Therefore, the less massive a particle, the less overlap there is between chirality and helicity. A massless neutrino with left-handed chirality guarantees it must have negative helicity. This is summarised in Table 1.3

Within the $B^+ \to \mu^+\nu_\mu$ decay, it was reasoned earlier that the lepton pair are either both positive or both negative helicity, so as to conserve angular momentum in the back-to-back decay. Moreover, the anti-muon has right-handed chirality while the neutrino is left-handed. A left-handed neutrino must also have negative helicity. Thus, the anti-muon must also have negative helicity, to conserve angular momentum. However, recall the factor of m/E applying to the negative helicity of this right-handed anti-muon. From the fixed $|\vec{p}_\mu^{\,B}|$=2.64 GeV/c derived previously, this m/E factor is will be small for the muon in this process. The resulting small amplitude of the negative helicity results in a lower likelihood that the $B^+ \to \mu^+\nu_\mu$ process can take place. Thus, to preserve both angular and linear momentum, the muon of $B^+ \to \mu^+\nu_\mu$ is forced to be in a helicity mismatched to the chirality required for the weak decay, resulting in a lower probability or *helicity suppression* of the mode.

While this logic was derived in the B-reference frame, helicity is Lorentz-invariant for massless particles, meaning the $B^+ \to \mu^+\nu_\mu$ decay in any frame is helicity suppressed. It is noted that if the neutrino was not massless, the impact of helicity suppression would be lessened. Models of the $B^+ \to \mu^+$ transition that could be accompanied by a BSM neutrino, which could also be detected in the search for SM $B^+ \to \mu^+\nu_\mu$, will be discussed in Section 1.3.4. Additionally, if a third particle were present, such as a photon emitted by one of the quarks, this breaks the helicity requirements and the suppression no longer applies. This results in the decay of $B^+ \to \mu^+\nu_\mu\gamma$ having a larger SM branching fraction than its non-radiative counterpart, and could be considered a background to the identification of $B^+ \to \mu^+\nu_\mu$ events in experiment.

Table 1.4. Table of $B^+ \to \ell\nu_\ell$ Standard Model Branching Fractions and Related Parameters.

Mode	m_ℓ [GeV/c]	$m_\ell^2\left(1-\frac{m_\ell^2}{m_B^2}\right)^2$ [GeV^2/c^4]	$\lvert\vec{p}_\ell^{\,B}\rvert$ [GeV/c]	$\mathcal{B}_{\rm SM}(B^+ \to \ell^+\nu_\ell)$	$\mathcal{B}_{\rm exp}(B^+ \to \ell^+\nu_\ell)$
$B^+ \to e^+\ \nu_e$	0.511×10^{-3}	2.611×10^{-7}	2.640	9.94×10^{-12}	$< 9.8\times10^{-7}$
$B^+ \to \mu^+\ \nu_\mu$	0.105	0.011	2.639	4.24×10^{-7}	$< 8.6\times10^{-7}$
$B^+ \to \tau^+\ \nu_\tau$	1.777	2.482	2.341	9.45×10^{-5}	$(1.1\pm0.2)\times10^{-4}$

The helicity in other leptonic B decays, $B^+ \to \tau^+\nu_\tau$ and $B^+ \to e^+\nu_e$ can also be discussed. For the tau mode, the m/E factor is larger than what it is for the muon case, thus there is less suppression of the negative helicity mode. For the electron mode, m/E is much smaller than the muon case and thus experiences more helicity suppression. Table 1.4 shows how this effects the branching fractions.

1.2.3. Feature 3: $B^+ \to \mu^+\nu_\mu$ is a 'missing energy' decay

The presence of a neutrino in a two-body B decay dramatically impacts its dynamics, as was presented in the prior discussion of helicity suppression. This was mostly attributed to masses of the particles in the decay and the chirality of the weak interaction. The next feature of $B^+ \to \mu^+\nu_\mu$ is owed to the limited interactions of the neutrino.

Not only are neutrinos modelled with zero mass in the Standard Model, but they are also modelled as largely non-interacting particles. Being massless, (electric) chargeless and colourless, it is limited how neutrinos can be detected in particle physics experiments. At dedicated neutrino experiments like Super-Kamiokande in Japan, neutrinos are measured via inverse beta decay, occurring within the large water tank that constitutes the detectors volume. Facilitated by the charged weak interaction, neutrinos incident with neutrons of the water molecule results in the neutron becoming a proton, while the neutrino transforms into a charged lepton of the same flavour. The charged lepton is much simpler to measure, signalling the neutrino's presence. An alternate neutrino detection method is via the neutral weak interaction, where the neutrinos scatter off nucleons in the water molecule, bestowing enough kinematic energy to free the nucleon which is consequently measured. The cross-section of such a process from upper atmosphere neutrinos was found by Super-Kamiokande to be 1.01 fb [13]. This is six orders of magnitude smaller than accelerator processes which will be presented in Section 2.2.3.

Considerable infrastructure is required to detect neutrinos via this method in neutrino-only experiments. The sequel experiment, Hyper-Kamiokande is planning for a 74m diameter, 60m tall tank for neutrino detection [14]. For general purpose experiments, where a decay like $B^+ \to \mu^+\nu_\mu$ would be measured, it is not yet feasible to include such large neutrino detection units alongside those for other particles. Thus, neutrinos remain undetected within most particle physics experiments. While the nominal neutrino momentum is known to be equal and opposite to the muon momentum in the B-frame, this requires knowledge of the B momentum boost first, something that is generally only accessible if all products (including the neutrino) were known. The energy and momentum carried away by the neutrino in a $B^+ \to \mu^+\nu_\mu$ decay is seemingly lost, and the presence of the B meson must be inferred from the muon alone. The neutrino is sometimes referred to as *invisible* or unreconstructable for this reason. However, if the $B^+ \to \mu^+\nu_\mu$ decay takes place in an experiment where the initial energy of the process is known, the *missing energy* can be derived. Under the assumption that all particles of the event have been recorded, the missing energy is derived by subtracting the recorded energies from the initial energy of the event. This can then be attributed to the neutrino, where the missing momentum can also be derived in a similar fashion. Thus any particle physics experiment intending to measure $B^+ \to \mu^+\nu_\mu$ should be one that has well-known initial energies and momenta and the ability to record all particles produced in a given event. Further methods of identifying $B^+ \to \mu^+\nu_\mu$ without the information of the neutrino are explored in both Section 1.4 and Chapter 3.

1.2.4. Feature 4: $B^+ \to \mu^+\nu_\mu$ is a leptonic decay

While both helicity suppression and missing energy makes $B^+ \to \mu^+\nu_\mu$ a difficult decay to identify, properties such as the exact B-frame muon momentum are signatures of the event. Another signature of the $B^+ \to \mu^+\nu_\mu$ event is that the final state is composed of leptons only, which can make event identification a simpler process.

The $B^+ \to \mu^+\nu_\mu$ decay takes place via a charged current interaction, with no intermediate hadron states that may have additional products. Leptons (excluding the tau) either do not decay, or are long-lived relative to particle detector dimensions for muons. This avoids complications of their reconstruction from other particle products, as well as the combinatoric issues that can come from attempting to choosing the correct daughter particles amongst background particles.

Lepton processes in experiments generally leave 'clean' signals in detectors. Electrons are usually identified via electromagnetic showers in detector media. As highly ionising particles, they often deposit all their energy in materials. Moreover, muons are identified via specialised subdetector units due to their long lifetime relative to the size of the detector. As the lighter leptons do not hadronise or decay, they are often isolated from other particles, instead of appearing in collimated jets of particles produced from complex QCD fragmentation processes. This allows singular leptons to be chosen directly, avoiding difficulties with discerning individual hadrons from jets to use in reconstruction. This also reduces backgrounds when choosing daughter particles.

As a leptonic decay, lepton universality dictates that the decay properties of $B^+ \to \mu^+\nu_\mu$ are similar to that of $B^+ \to \tau^+\nu_\tau$ and $B^+ \to \ell^+\nu_e$. Thus, the branching fraction of Eq. (1.6) is of the same form for $B^+ \to \tau^+\nu_\tau$ and $B^+ \to e^+\nu_e$. Any differences between leptons are due to different masses. This explains how the electron and muon modes can be suppressed while the tau mode is not. The large mass additionally allows the tau lepton to decay to hadrons (alongside a tau neutrino). Performing simultaneous measurements of decays to different generations of leptons, such as the ratio $R(D^{(*)}) = \mathcal{B}(B^+ \to \bar{D}^{(*)}\tau^+\nu_\tau)/\mathcal{B}(B^+ \to \bar{D}^{(*)}\ell^+\nu_\ell)$ (where ℓ^+ is e^+ or μ^+) is a natural way to test the assumption of lepton universality. Anomalies in such ratios may indicate the presence of BSM mechanisms with preference for certain lepton generations.

The impact of QCD is also largely avoided in a leptonic decay like $B^+ \to \mu^+\nu_\mu$. As leptons do not interact via QCD, the complexity of gluon coupling can be restricted to the B meson only. Even for a decay such as $B^+ \to \pi^0\mu^+\nu_\mu$, where only a single hadron is added, the effects of QCD become much more complicated. Here, the gluon interactions need to be accounted for in the B meson, the final π^0 meson, as well as how the hadronisation takes place when the $\mu\nu_\mu$ lepton pair is produced. With no QCD in the final state, the QCD interactions are condensed into a single value, f_B, the decay constant. The role of f_B in the $B^+ \to \mu^+\nu_\mu$ process is discussed in the next section.

1.2.5. Feature 5: $B^+ \to \mu^+\nu_\mu$ measurements can assist in measuring the decay constant, f_B

The decay constant of B mesons, f_B, encodes the QCD effects occurring between the quarks of the B meson prior to their annihilation. They are traditionally measured using Lattice QCD, where numerical techniques evaluate the non-analytical features of QCD, due to the large QCD coupling constants for higher-order Feynman diagrams. Lattice QCD calculations are challenged by the difficulties in simulating b-quarks, whose larger mass requires greater granularity and thus larger computational power. Precision of the f_B calculation varies depending on the methods used. Allowing for more flavours of sea quarks within the B meson has generally led to lower uncertainties, at the cost of higher computation required. Moreover, most results do not distinguish between B^+ and B^0 mesons, instead using an average mass for the non-b quark. While using a decay constant inclusive of any electromagnetic effects between the quarks would better simulate the conditions within the B meson, reported results of f_{B^+} and f_{B^0} are not yet competitive with the B-averaged values [6].

The B meson decay constant is a key input to the $B^+ \to \mu^+\nu_\mu$ branching fraction expression in Eq. (1.6). Aside from $|V_{ub}|$, all other parameters are known with high accuracy. Thus a high precision measurement of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ could provide an alternate method of deriving f_{B^+}, as a charged B meson decay constant. However, such an approach may still be some time away. The previous $B^+ \to \mu^+\nu_\mu$ searches to be discussed in Section 1.4, uncertainty on $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ has significant uncertainties compared to those of the Lattice QCD derivations f_B. This scenario could change, if the reduction of statistical uncertainties coming from new data in high-luminosity experiments occurs quicker than that of computational and calculation improvements within Lattice QCD. An alternate approach would be to use the related $B^+ \to \tau^+\nu_\tau$ branching fraction, that could set a better experimental limit on f_B. From Table 1.4, it was seen that the experimental branching fraction of $B^+ \to \tau^+\nu_\tau$ is currently at the precision such that a central value with an uncertainty of 20% can be quoted, thus providing a tighter constraint than what the current upper limit of $B^+ \to \mu^+\nu_\mu$ could offer. However, $B^+ \to \mu^+\nu_\mu$ still provides an complimentary method to a $B^+ \to \tau^+\nu_\tau$ result, especially considering the complexity of a signal τ reconstruction in the latter where further neutrinos and missing energy is produced.

1.2.6. Feature 6: $B^+ \to \mu^+\nu_\mu$ measurements is Cabbibo suppressed and can assist in measuring V_{ub}

From Figure 1.1, it can be seen that decay of the B^+ meson requires the direct coupling of a u and $\bar{b}$ quark. As was discussed in Section 1.1.2, the amplitude of this coupling is proportional to the CKM matrix parameter, V_{ub}, hence its inclusion in the $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ equation. This parameter is the smallest of all CKM matrix parameters, as a coupling between the first and third generation of quarks. Due to the intergenerational coupling, the $B^+ \to \mu^+\nu_\mu$ process can be described as *Cabbibo suppressed*. As V_{ub} appears as a square within the branching fraction, this small value has considerable impact on the overall branching fraction. Combined with the role the helicity suppression plays, the $B^+ \to \mu^+\nu_\mu$ process is therefore extremely rare, requiring a large dataset to be measured precisely.

Like f_B, $|V_{ub}|$ can also be calculated by measuring the $B^+ \to \mu^+\nu_\mu$ branching fraction in experiment. Unlike f_B, there is no first principles method of deriving $|V_{ub}|$, as it is solely dependent on free parameters of the Standard Model formalism. There is great interest in measuring $|V_{ub}|$ due to the current disagreement in the values produced in *inclusively* measured decays (i.e. via $B^+ \to X_u\ell^+\nu_\ell$, where X_u can be any meson) and *exclusive* decays (i.e. where X_u is fixed). This was summarised in Table 1.2, where the two values disagree within their error. Moreover, $|V_{ub}|$ has the largest percentage uncertainty of any other CKM matrix magnitudes. Further studies of such modes are required to ascertain if this is a statistical anomaly or if there may be underlying BSM mechanisms causing this difference, such as unaccounted for new physics processes enhancing the inclusive measurement. BSM mechanisms that can impact that the $b \to W^+u$ coupling is discussed in Section 1.3.

Currently, HFLAV obtains their average of $|V_{ub}|$ from $B \to \pi\ell^+\nu_\ell$ decays, where ℓ^+ is either an electron or a muon. This is derived from differential decay rates of this process, where $|V_{ub}|$ appears alongside form-factors obtained via Lattice QCD or other techniques. Instead of a single decay constant, a momentum-dependent form factor is required to describe the process, changing according to how much of the B momentum is transferred to the $\overline{D}^0$. Such an expression is considerably more complicated than the branching fraction expression of Eq. (1.6). While $B \to \pi\ell^+\nu_\ell$ is the state of the art as it is currently known to higher precision than either the tau or muon leptonic decays, using a leptonic decay like $B^+ \to \mu^+\nu_\mu$ or $B^+ \to \tau^+\nu_\tau$ could provide a more direct calculation of $|V_{ub}|$. This was done in Ref. [6], where FLAG evaluates $|V_{ub}|$ using the branching fraction measurement of $B^+ \to \tau^+\nu_\tau$ from different experimental collaborations and using average f_B, where each were separated by the number of sea quarks modelled. The largest variation in $|V_{ub}|$ came from using different experimental results, rather than the variations in sea quarks, further demonstrating that experimental variables are the largest source of uncertainty in flavour observables. Similar to the discussion in Section 1.2.5, at the current precision of the leptonic branching fractions, $B^+ \to \tau^+\nu_\tau$ provides a better constraint on than using the upper limit on $B^+ \to \mu^+\nu_\mu$, though the latter still provides an alternate supporting channel. Thus, with more data and thus better experimental precision, leptonic B-decays could become more competitive in precise measurements of $|V_{ub}|$.

1.3. Beyond the Standard Model prospects for $B^+ \to \mu^+\nu_\mu$

The SM properties of the $B^+ \to \mu^+\nu_\mu$ decay have been discussed. Particular focus was given to how measuring the $B^+ \to \mu^+\nu_\mu$ branching fraction could access key Standard Model properties, such as CKM matrix amplitudes V_{ub} and the decay constant f_B. With a Standard Model branching fraction of 4.24×10^{-7}, $B^+ \to \mu^+\nu_\mu$ remains unobserved in previous B-physics experiments, as will be discussed in Section 1.4. However, some BSM theories suggest that the experimental branching fraction could be larger due to BSM mechanisms. This makes high precision measurements of $B^+ \to \mu^+\nu_\mu$ events a valuable undertaking - if an excess of events is seen, this could provide evidence against being a SM-only process. If no events are seen, a more precise limit on the BSM mechanisms can be obtained, in addition to a limit on the $B^+ \to \mu^+\nu_\mu$ Standard Model branching fraction. The following is a discussion of BSM physics and how they could impact the $B^+ \to \mu^+\nu_\mu$ process.

1.3.1. $B^+ \to \mu^+\nu_\mu$ as a probe of charged Higgs models

From Figure 1.2, it can be seen that the Higgs boson is the sole scalar boson of the Standard Model formalism. Mathematically, the Higgs field is a complex valued, two-component object called a doublet. The Higgs boson discovery in 2012 by LHC experiments confirmed the existence of one such doublet [15] [16].

The inclusion of a second doublet is presented in Ref. [17], as a model that allows for the breaking of CP symmetry in the early universe. Within this *two-Higgs-doublet model* (2HDM) extension to SM, additional scalar bosons emerge, including an electrically charged Higgs boson. Which particles acquire mass from which of the two Higgs fields varies according to different implementations of the theory. Regardless of implementation, a key parameter of this theory is the ratio of each of the Higgs' vacuum expectation values, expressed as tanβ.

The decay $B^+ \to \mu^+\nu_\mu$ is facilitated by the charged W boson in the SM. In Ref. [18], it was posited that the B decay to the muon neutrino pair could also be mediated by the charged Higgs boson of the 2HDM. While the amplitude of the SM contribution is proportional to m_μ^2 and thus helicity suppressed, the amplitude of the charged Higgs contribution is proportional to $\tan^2\beta$. Provided $\tan^2\beta$ has a large enough value, the $B^+ \to \mu^+\nu_\mu$ process would have a non-helicity suppressed mode that could enhance the measured branching fraction. Specifically, the $B^+ \to \mu^+\nu_\mu$ branching fraction would be modified by a factor $r_{H^\pm} = (1 - (m_B^2/m_{H^\pm}^2)\tan^2\beta)^2$. It is noted that the two processes could also destructively interfere, if $\tan^2\beta$ was measured to be small.

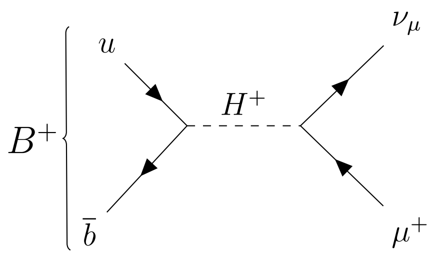


Figure 1.7. A Feynman diagram of the decay $B^+ \to \mu^+\nu_\mu$, with the inclusion of a charged Higgs propagator, H^+.

To probe the existence of a charged Higgs boson, the allowed values of charged Higgs mass and $\tan^2\beta$ must be explored. From the proposed r_H factor measurements of the $B^+ \to \mu^+\nu_\mu$ branching fraction could constrain the values of these parameters. The most recent $B^+ \to \mu^+\nu_\mu$ search at Belle, whose methods will be discussed in Section 1.4, used the limits placed on $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ to determine excluded regions of the charged Higgs mass/$\tan^2\beta$ parameter space. A χ^2 test was performed to determine the agreement between the experimental $B^+ \to \mu^+\nu_\mu$ branching fraction and the charged Higgs branching fraction for varying values of $\tan^2\beta/m_H^2$. From this, 68% and 95% excluded regions were determined. In the range of $\tan\beta \in$[0,300] and $m_H \in$[0,1000] GeV/c^2, the $B^+ \to \mu^+\nu_\mu$ experimental branching fraction was not compatible with low masses of the charged Higgs and large values of $\tan\beta$. This result specifically pertained to the type II 2HDM, where up-type and down-type quarks masses are determined uniquely by one of the two Higgs bosons. Large exclusion regions were also derived for the couplings of type III 2HDM, which can facilitate tree-level flavour changing neutral currents.

1.3.2. $B^+ \to \mu^+\nu_\mu$ as a probe of leptoquark models

Just as the electromagnetic force and the weak interaction unite at high energies to give the electroweak interaction, there is a desire within particle physics to also unite the strong and gravitational forces. A mechanism which does so is referred to as a *Grand Unified Theory* (GUT).

One such GUT is the Pati-Salam model, where lepton pairs are treated as a fourth colour of quark [19]. Here, the charged leptons would be considered symmetrically alongside the down-type quarks according to their generation, while the neutrinos would be grouped with the up-type quarks. For example, the u quark in Standard Model QCD is actually a superposition of (u_{red}, u_{green}, u_{blue}). Within the Pati-Salam extension, the u quark becomes a quartet object of (u_{red}, u_{green}, u_{blue}, ν_e), the d quark becomes (d_{red}, d_{green}, d_{blue}, e), the c quark becomes (c_{red}, c_{green}, c_{blue}, ν_μ) and so on. With this new definition, couplings between fermions that are not possible in the Standard Model can be mediated by a Pati-Salam boson, such as directly between a quark and lepton or between leptons of differing generations. Being able to do so makes the Pati-Salam boson a leptoquark, a particle that can interact with both quarks and leptons. They therefore have both baryon and lepton number. While direct couplings with quark and leptons can violate the observed conservation of baryon number and lepton number in Standard Model processes, the difference between the two as $n_B - n_L$ is observed in this model.

The coupling between leptons of different generations is where the $B^+ \to \mu^+\nu_\mu$ decay may be impacted. As was discussed in Section 1.2.3, neutrinos are not detectable at general collider experiments. Within the Standard Model, the weak interaction can only couple neutrinos with their corresponding charged lepton. Thus, within experiment if a $B^+ \to \mu^+$ transition is seen with missing energy, it is assumed that the missing energy is a Standard Model muon-neutrino. If Pati-Salam extensions to the Standard Model are allowed, there are then additional processes that could enhance the number of events measured, being $B^+ \to \mu^+\nu_\tau$ and $B^+ \to \mu^+\nu_e$. Moreover, such processes may not be helicity or Cabbibo suppressed as like Standard Model $B^+ \to \mu^+\nu_\mu$, motivating excesses in a $B^+ \to \mu^+\nu_\mu$ branching fraction measurement.

In Ref. [20], constraints on Pati-Salam boson masses according to 90% confidence limits were placed on various decays of spin-0 meson to two leptons. Figure 1.8 gives an example process which could impact $B^+ \to \mu^+\nu_\mu$ measurements, with a B^+ transition via a Pati-Salam boson. Such a process is a slight deviation from the theory above, for the case where the second generation and third generations of leptons swap with their

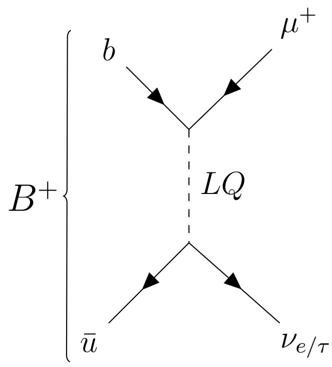


Figure 1.8. A Feynman diagram of a $B^+ \to \mu^+\nu$ decay that would appear identical to SM $B^+ \to \mu^+\nu_\mu$ decay in experiment, mediated by a Pati-Salam boson propagator, LQ.

corresponding generations of quarks. Thus, a b quark is coupled to a muon while an $\overline{u}$ is coupled to a non-muon-neutrino. Such a process is not allowed as a standard W^+ transition. To obtain a bound of the Pati-Salam boson that could give rise to this process, the experimental upper bound of $B^+ \to \mu^+\nu_\mu$ branching fraction is used, as there are no neutrino flavour assumptions used in $B^+ \to \mu^+\nu_\mu$ searches. The lower bound on the boson mass of 16 TeV is obtained, though it is noted that the branching fraction limit used in Ref. [20] is from the 1995 CLEO experiment search, a bound which has been improved by two orders of magnitude since publishing. A similar bound is also derived for $B^+ \to \mu^+\nu$ with a tau-neutrino, where phenomenologically the third generation leptons join the first generation quarks, in addition to the second generation leptons with the third generation quarks.

More recently, Pati-Salam models have also been invoked for various B meson anomalies, such as in lepton universality ratios of $B^+ \to D^{(*)}\ell\nu_\ell$ and rare decays like $B \to K\nu\nu$ and $B \to K\mu\tau$ [21]. The latter decay is allowed if allowing for neutrino oscillations, though it still would be extremely rare. If these ratios and rare decays are found to be anomalously large, this could be confirmation of charge lepton flavour violation.

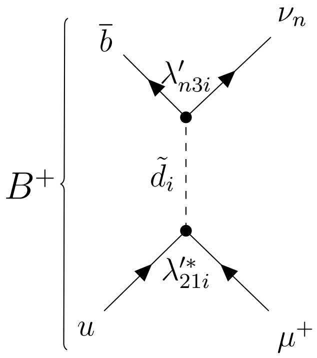
1.3.3. $B^+ \to \mu^+\nu_\mu$ as a probe of supersymmetry

An additional SM extension to look at is one that incorporates *super-symmetry* (SUSY). SUSY began as an attempt to construct a quantum field theory that was invariant under exchange of bosons with fermions and vice versa [22]. When extended to SM, SUSY predicts that for each SM fermion, a new bosonic partner particle exists, as well as a new fermionic partner for fundamental bosons. This effectively doubles the number of fundamental particles expected in nature. While the SUSY partners masses are expected to exist on the TeV scale, no such particles have been seen in the collider experiments at the energy frontier, casting doubts on the theory. Regardless, SUSY is desirable as it provides a dark matter candidate, the fermion partner of weak interaction bosons, the neutralino.

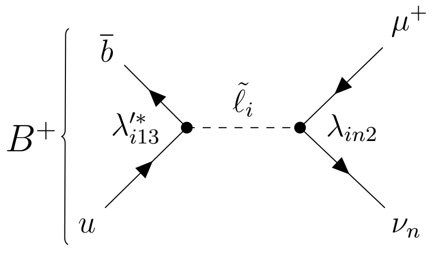
The simplest of SUSY extensions to SM is called the *Minimal Supersymmetric Standard Model* (MSSM). Within MSSM, the two-Higgs doublet model is required to make the theory consistent. Additionally, in MSSM formalism the baryon number and lepton numbers can be completely violated (unlike the partial breaking via $n_B - n_L$), unless another conserved quantity is introduced. This quantity is called R-parity, where regular particles have R-parity 1 and SUSY particles have R-parity -1. However, while baryon number and lepton number conservation is an observed phenomena, it is not a foundational SM (nor MSSM) principle. This motivates theories of R-parity violating MSSM.

In Ref. [23], the constraints on R-parity violating MSSM interactions were derived from the $B^+ \to \mu^+\nu_\mu$ branching fraction upper limit. Taking into account SUSY interactions in $B^+ \to \mu^+\nu_\mu$, new mediators between the B^+ initial state and lepton pairs are possible. The new mediators take the form of *sleptons*, the charged lepton superpartners, and the down-type *squarks*, as the superpartners of quarks. Notably, as a R-parity violating theory, such mediators allow for couplings between charged leptons and neutrinos from different generations. As was mentioned in the previous section, as discriminating between flavours of neutrino is impossible in a missing-energy experimental approach, such modes can enhance a measurement of $B^+ \to \mu^+\nu_\mu$ branching fraction. Additionally, helicity effects would be broken with SUSY mediators, further increasing the number of $B^+ \to \mu^+\nu_\mu$ events expected.

The paper specifically investigated the couplings affecting a $B^+ \to \mu^+\nu_\mu$ measurement, which were both three-point couplings. The coupling λ (between leptonic-only super-fields) and λ' (between one lepton and two quark superfields) were examined. Constructing a $B^+ \to \mu^+\nu_\mu$-like mode using only λ' couplings (i.e. with squark mediator presented in Figure 1.9a), the constraint on the couplings from $B^+ \to \mu^+\nu_\mu$ bounds were not competitive with those derived from bounds on $\mu \to e\gamma$. The case of using both λ and λ' (i.e. using a slepton mediator, as shown in Figure 1.9b) were instead presented. This was done by summing over all possible combinations of SUSY sleptons and neutrino flavours



(a) Squark exchange. Summation is implied over each generation of $\tilde{\ell}_i$ and ν_n.



(b) Slepton exchange. Summation is implied over each generation of $\tilde{d}_i$ and ν_n.

Figure 1.9. Feynman diagrams of the decay $B^+ \to \mu^+\nu$ facilitated by SUSY particles.

and relating this to the branching fraction. The result was an improvement of over a magnitude on the product coupling $|\lambda\lambda'|$.

Notably, the bounds were achieved by also using the outdated CLEO $B^+ \to \mu^+\nu_\mu$ branching fraction. Despite this, R-parity violating MSSM and the constraints that B physics can place on λ-λ' has continued to be investigated, as recently as 2020 [24].

1.3.4. $B^+ \to \mu^+\nu_\mu$ as a probe of sterile neutrinos

The identity of the invisible particle produced alongside the muon in $B^+ \to \mu^+\nu_\mu$ searches is assumed to be a muon neutrino, as is required in the SM. Previous BSM contributions discussed have included neutrinos from other generations as well as SUSY particles. An additional BSM invisible particle contribution could be *sterile neutrinos*

A sterile neutrino is defined in opposition to 'active' neutrinos. Within the SM, only left-handed neutrinos and right-handed anti-neutrinos actively participate in weak interaction. Right-handed anti-neutrinos and left-handed neutrinos cannot exist in the SM. If they did, they would be 'sterile', meaning they do not interact via SM forces. However, with the observation of neutrino flavour oscillation indicating non-zero neutrino mass, the hypothetical sterile neutrinos would be able to interact gravitationally. This oscillation taking place means neutrinos in a state of definite flavour are in a superposition of states with unique masses, similar to CKM quark mixing. Difficulties in measuring neutrinos means that the mechanism by which neutrinos acquire mass is still unknown, whether it be via the Higgs like other SM fermions or a new mechanism entirely.

The simplest theory extension of the Standard Model to include massive neutrinos, both active and sterile, is called the *Neutrino Minimum Standard Model* (νMSM). In this model, sterile neutrinos are assigned a mass model such that they are free from chirality and are their own antiparticles (referred to as *Majorana mass* model). Meanwhile, the active neutrinos are derived from those of the SM. They are extended by using a fermion mass model (referred to as *Dirac mass*), though this is altered slightly to allow mass contributions from the sterile neutrino. The terms are related by the *see-saw mechanism*, where the sterile neutrinos are assigned heavy masses, which results in the active neutrinos having light masses to match experiment.

From its definition in Ref. [25], it is shown that the inclusion of sterile neutrinos that can mix with active neutrinos can contribute to observations of neutrino flavour oscillation. Additionally, the lightest sterile neutrino has keV/c^2 mass with a long lifetime, functioning as a dark matter candidate that could match the dark matter density of the universe. In the related Ref. [26], equipping sterile neutrinos of the ν-MSM with a CP-violating phase from their own mixing matrix can contribute to baryon asymmetry, specifically prior to interaction freeze-out as the universe evolves. Additionally, the amount of CPV generated by the sterile neutrino mixing is compatible with the baryon asymmetry of the universe today. It is compelling that the three phenomena, oscillation, dark matter and baryon asymmetry, all arise when assuming three generations of sterile neutrinos, the same number as SM fermions.

For the process $B^+ \to \mu^+\nu_\mu$, there is potential for the missing energy particle to be a sterile neutrino. Not only did the previously discussed 2020 Belle $B^+ \to \mu^+\nu_\mu$ search constrain charged Higgs model parameters, they also placed limits on sterile neutrino contributions [27]. It is noted that precision of the search is only sensitive to the MeV/c^2 and GeV/c^2 mass ranges, thus such a sterile neutrino candidate would require an extension to the νMSM described above. The analysis strategy was altered to fix the measured yield of of $B^+ \to \mu^+\nu_\mu$ to match the predicted SM branching fraction and then quantify how well a sterile neutrino contribution matched the residual yield. Effectively, a simultaneous fit was performed to both SM and sterile neutrino contributions in $B^+ \to \mu^+\nu$, via a ratio of branching fractions. This allowed for limits to be placed on sterile neutrinos, according to its coupling with the muon. Ultimately, no excess was consistent with a sterile neutrinos between the evaluated range of 0 to 1.5 GeV/c^2. The branching fraction upper bounds on $B^+ \to \mu^+N$ varied from 1×10^{-6} to $6{\times}10^{-6}$, according to neutrino mass. This bound was not competitive against results from other experiments, though it has improved bounds derived from a dedicated $B^+ \to \mu^+ X^0$ search by Belle, where X^0 denotes an non-Standard Model particle [28].

1.4. Previous searches for $B^+ \to \mu^+\nu_\mu$ in high-energy physics experiments

The features of $B^+ \to \mu^+\nu_\mu$ events in experimental searches according to SM have been reviewed. Such motivations to observe $B^+ \to \mu^+\nu_\mu$ have remained largely unchanged over the years. Not only is the search important so as obtain a statistically significant branching fraction of a light-lepton analogue to $B^+ \to \tau^+ \nu_\tau$, but also to constrain the

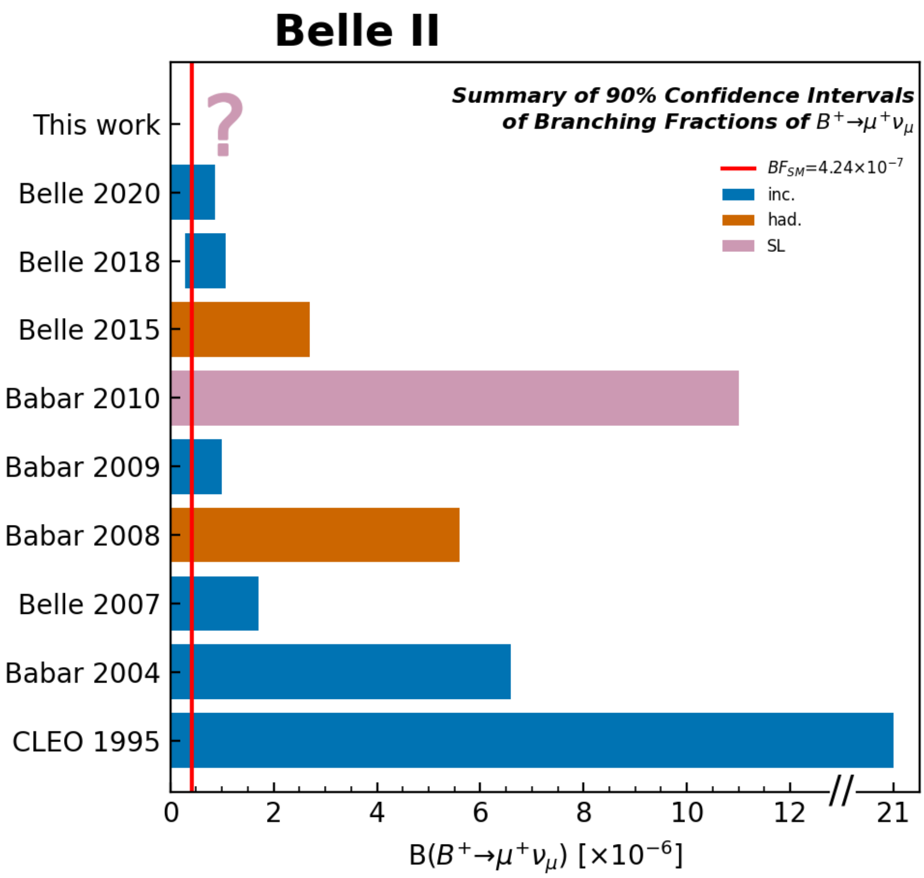


Figure 1.10. Summary of previous $B^+ \to \mu^+\nu_\mu$ 90% confidence limits, separated by B_{tag} type. Sourced from Ref. [4].

Table 1.5. Table of the number of $\Upsilon(4S) \to B\overline{B}$ events used in previous $B^+ \to \mu^+\nu_\mu$ searches. The CLEO result quotes N_{BB}=2.2×10^6 for charged pairs only, thus the full number of charged and neutral $B\overline{B}$ events listed below is taken to be double this values.

Search	B_{tag} mode	$N_{B\overline{B}}$
1995 CLEO	Inclusive	4.40×10^{6}*
2004 BaBar	Inclusive	8.84×10^{7}
2007 Belle	Inclusive	2.77×10^{8}
2008 BaBar	Hadronic	3.78×10^{8}
2009 BaBar	Inclusive	4.68×10^{8}
2010 BaBar	Semileptonic	4.68×10^{8}
2015 Belle	Hadronic	7.72×10^{8}
2018 Belle	Inclusive	7.72×10^{8}
2020 Belle	Inclusive	7.72×10^{8}
2024 Belle II (this work)	Semileptonic	3.87×10^{8}

large number of BSM theories discussed in the previous section that predict enhanced branching fractions on this mode.

In order to construct a search in a modern high-energy physics experiment, it is important to study the previous 20 years of attempts to establish an experimental $B^+ \to \mu^+\nu_\mu$ branching fraction. Three particle phyics collaborations have performed searches for the $B^+ \to \mu^+\nu_\mu$ mode: the CLEO Collaboration [29], the BaBar Collaboration [30] and the Belle Collaboration [31]. The experiments have aimed to measure the branching fraction of $B^+ \to \mu^+\nu_\mu$. Currently, only 90% confidence limits (as either upper bounds and two-sided intervals) have been placed on the branching fraction.

Prior searches for $B^+ \to \mu^+\nu_\mu$ have all been performed in circular e^+e^- collision experiments. The advantage of experiments in fundamental particle colliders for a $B^+ \to \mu^+\nu_\mu$ search is that the initial collision energy is completely known, such that the missing momentum from the neutrino can be compensated for. Further advantages of electron-positron experiments, particularly at centre-of-mass energy 10.58 GeV/c (i.e. the $\Upsilon(4S)$ resonance, $\sqrt{s}$=10.58 GeV), will be discussed in Section 2.2.2.

The structure of the detector used for measuring $B^+ \to \mu^+\nu_\mu$ processes remains broadly unchanged, despite vast improvements in the specific technologies used within each subdetector. All three experiments to be discussed used a barrel-shaped detector with further recording elements in the endcaps. This was done to fully enclose the location of the electron-positron collision, allowing for all particle of the event to be measured. Subdetector units were nested in concentric cylinders, where each unit specialised role in identifying particles. From closest to furthest away from the collision, this included an inner charged particle tracker and vertexing unit, a drift chamber, a time of flight counter and/or Cherenkov radiation detector for charged particle identification, an electromagnetic crystal calorimeter, a superconducting solenoid and muon detection chamber. The function of each subdetector will be discussed in further detail when introducing the Belle II detector in Section 2.3.1.

All previous searches have studied $\Upsilon(4S) \to B\overline{B}$ processes with B-tagging. The B that decays via $B^+ \to \mu^+\nu_\mu$ is referred to as the signal B meson or the B_{sig}. In B-tagging techniques, properties of B mesons that decay alongside the B decay of interest are used. Doing so provides additional features to ensure the decay of interest is identified correctly. For *inclusive* B-tagging, it is verified that all energies and momenta that remain after B_{sig} identification are consistent with the B mass. Here, the partner B (or B_{tag}) is not fixed to decay to any mode (i.e. inclusive of all decay modes). Conversely, *exclusive* B-tagging involves restricting the dataset to B_{tag}'s of easy to identify decay modes. This provides more specific criteria to determine if the signal B candidate is correct, at the expense of a lower retention of candidates. An in-depth discussion of B-tagging techniques will take place in Chapter 3. B-tagging can be used to derive the four-momentum of the signal B meson from the tagged B. This is especially useful for the $B^+ \to \mu^+\nu_\mu$ analysis, as the full signal B four momentum cannot be determined due to the neutrino. Almost all previous $B^+ \to \mu^+\nu_\mu$ searches have derived the signal B four-momentum from the tagged-B, so as to use the signal muon's parent-frame momentum as a key measure of correctness.

The analysis of prior searches have been divided into three previous 'eras' of $B^+ \to \mu^+\nu_\mu$ studies: the inclusively-tagged era, the BaBar era and the Belle era. The techniques unique to each era are described in the following section.

1.4.1. 1995-2007: The first inclusive searches of $B^+ \to \mu^+\nu_\mu$ at CLEO, BaBar and Belle

1995: CLEO

The first search for $B^+ \to \mu^+\nu_\mu$ was published in 1995 [32] by the CLEO experiment. It was performed in a dataset of 2.2×10^6 $\Upsilon(4S) \to B^+B^-$ decays with inclusive B-tagging. The CLEO experiment was located on the equal-energy collider CESR, at Cornell University in New York state. The experiment had also performed the first measurement of a B meson via $e^+e^- \to \Upsilon(4S)$ $B\overline{B}$ processes 14 years earlier [33]. The ALEPH collaboration of CERN, had indirectly constrained the $B^+ \to \tau^+\nu_\tau$ branching fraction through $b \to \tau^-$ $\overline{\nu}_\tau$ X processes [34], thus motivating CLEO to perform a direct search for $B^+ \to \tau^+\nu_\tau$, alongside the electronic and muonic versions.

Events chosen were those containing a muon and with remaining energies and momenta consistent with originating from a B-meson. This was done by restricting the analyses to physics values of two key B quality variables, $M_{\rm bc}$ and ΔE. The role of these variables is unique to electron-positron colliders and will be derived in Section 2.2.2. For now, they can be understood as the B mass and the deviation of the tagged-B's energy from the nominal value in the $\Upsilon(4S)$ frame. These variables were so important that the yield of the companion search $B^+ \to \tau^+\nu_\tau$ used a two-dimensional fit to $M_{\rm bc}$ and ΔE to evaluate the signal yield. The $B^+ \to \mu^+\nu_\mu$ yield was evaluated as a 'cut-and-count' approach, to the much lower expected number of events from the SM branching fraction.

To remove events which were likely to have originated from non-$\Upsilon(4S)$ decays, the cosine angle between the derived missing momentum and accelerator beamline in the labframe was restricted. Additional limitations were placed on the evaluated Fox-Wolfram moments, to prioritise the unique spherical $B\overline{B}$ event topology. A tight selection ($\pm 100\,\mathrm{MeV}/c$ of the nominal value) was also placed on the derived signal muon momentum in the parent B frame, created from the inclusive $B_{\rm tag}$. This resulted in a 13% selection efficiency in simulated $B^+ \to \mu^+\nu_\mu$ events. Three events remained in data, with 1.5 of them attributed to non-$\Upsilon(4S)$ $e^+e^- \to q\overline{q}$ events and 0.4 from $b \to u\ell\nu_\ell$ processes. The 90% upper limit on the $B^+ \to \mu^+\nu_\mu$ yield was 5.7 events. The corresponding branching fraction of $B^+ \to \tau^+\nu_\tau$ was also assigned an upper limit. It was found to be inconsistent with contributions from a small-mass and large $\tan\beta$ model of a charged Higgs from two-Higgs doublet model, which was introduced in Section 1.3.1. This result has been the only published search for $B^+ \to \mu^+\nu_\mu$ from the CLEO Collaboration. The rate of B

meson production at CLEO was not able to compete with the new dedicated B-physics experiments taking over the efforts in the 2000's.

2004: BaBar

Nine years after CLEO's result, the BaBar experiment published its own search for $B^+ \to \mu^+\nu_\mu$ with inclusive tag [35]. The search was performed with 88.4 million $B\overline{B}$ events, a larger dataset than what CLEO had used previously and but a small fraction of what BaBar would go on to produce. Around half of this could be assumed to be B^+B^- events in which $B^+ \to \mu^+\nu_\mu$ could take place.

The BaBar experiment took place at SLAC National Accelerator Laboratory, which was operated by Stanford University in California. The experimental design varied from CLEO, implementing asymmetric energy electron-positron collisions. This was designed to produce boosted $B\overline{B}$ pairs in which to precisely measure each B's time to decay, thus probing time-dependent CPV, as was described in Section 1.1.2. Moreover, BaBar was one of the two new *B-factory experiments*, with capability to produce one of the largest $B\overline{B}$ datasets ever created. Beginning data-taking in 1999, BaBar had competed alongside the Belle B-factory experiment in Japan to obtain the first measurement of CPV in B decays. With both experiments publishing results in 2001 [36] [37], the collaboration could now dedicate efforts to other B-physics phenomena, namely $B^+ \to \mu^+\nu_\mu$ searches.

The analysis procedure was similar to that of CLEO's, but with a much larger dataset of 88.4 million $B\overline{B}$ events. BaBar began with events containing a high-quality muon and ensured that the remaining detected energies and momenta had realistic values of M_{bc} and ΔE. To obtain more accurate energies and momenta of the remaining particles, the particle identities were taken into account. Whereas CLEO used a scintillating unit to measure the time of flight of charged particles, which could aid in particle identification, the BaBar detector introduced a Cherenkov-light based approach. In this unit, the angle of emitted radiation could be used to demarcate between charged pions and kaons.

Key backgrounds were semileptonic B decays containing the $b \to u\mu\nu_\mu$ process and non-$\Upsilon(4S)$ $e^+e^- \to q\overline{q}$ processes. A restriction on the previously discussed Fox-Wolfram moment and the cosine polar angle of missing momentum was also implemented to restrict the impact of $q\overline{q}$ events. The derived B-frame momentum of the signal muon was restricted to be within $100\,\mathrm{MeV}/c$ from the nominal value. New techniques include bounding values were determined from the combination of selection criteria that optimised the ratio of signal and background events in simulated collisions, $\frac{S}{\sqrt{S+B}}$. Additionally, a cross-check of the inclusive tagging technique was performed for a *control mode* with a well known branching fraction, $B^+ \to \bar{D}^{(*)0}\pi^+$. Here, a different signal B is identified before creating the B_{tag}, and the yield is compared against what is predicted by control mode branching fraction. This technique will also be implemented for the analysis within this thesis, to be discussed in Section 5.2.2. The signal selection strategy resulted in a 2.24% efficiency of simulated $B^+ \to \mu^+\nu_\mu$ events.

To estimate the signal yield in collision data, the number of events attributed to background processes is first estimated, by fitting collision data to an ARGUS distribution model. The ARGUS distribution is a probability density function that models backgrounds in kinematically bounded quantities, such as in invariant mass distributions [38]. The fit is performed in the region of $5.10\ \text{GeV}/c^2 < M_{\text{bc}} < 5.24\ \text{GeV}/c^2$, below the nominal B mass, where misreconstructed and background events are expected to dominate. This is then extrapolated to $M_{\text{bc}} < 5.27\ \text{GeV}/c^2$, where correctly reconstructed B_{tag} associated with $B^+ \to \mu^+\nu_\mu$ would occur. This resulted in 5.2 ± 0.5 events attributed to background, 5.7 events in total found, where the difference could be associated with $B^+ \to \mu^+\nu_\mu$ events. Thus an 90% upper limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 6.6\times10^{-6}$ was obtained. Relative to CLEO, this result was approximately a third of the previous boundary, though still an order of magnitude larger than the SM evaluation.

2007: Belle

In 2007, the Belle experiment published its own inclusive $B^+ \to \mu^+\nu_\mu$ search, performed alongside a $B^+ \to e^+\nu_e$ search [39]. The result came after Belle had presented the first evidence of $B^+ \to \tau^+\ \nu_\tau$, measuring the branching fraction of 1.79×10^{-4} with a significance of 3.5σ [40]. Belle had acquired a dataset of 276.6 million $B\overline{B}$ pairs, exceeding both CLEO's and BaBar's dataset. Moreover, the Belle detector used both a scintillating time of flight counter and a Cherenkov radiation detector.

Belle generally followed the previous inclusive tag analyses, with selections to optimise muon quality, B_{tag} quality as well as prioritising events consistent with the narrow B-frame muon momentum. Instead of using a single Fox-Wolfram momentum selection to reject $q\bar{q}$ backgrounds, a linear combination of moments was constructed. The coefficients of each moment is optimised to create a distribution that maximises discrimination between for signal and background simulated collisions. This combination is called a Fischer discriminant.

After this selection was incorporated, the signal efficiency was found to be 2.18%. The remaining events are shown in Figure 1.11, in a distribution of derived B-frame lepton momentum. Main backgrounds were $B^+ \to X_u\mu^+\nu_\mu$. Signal yield was similarly estimated by fitting the simulated collisions to recorded collisions with an ARGUS distribution to $M_{\rm bc}$. This resulted in 4.1 ± 3.1 $B^+ \to \mu^+\nu_\mu$ events, with 7.4 events measured altogether and a 90% confidence interval of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.7 \times 10^{-6}$. This was the tightest upper bound on the branching fraction yet.

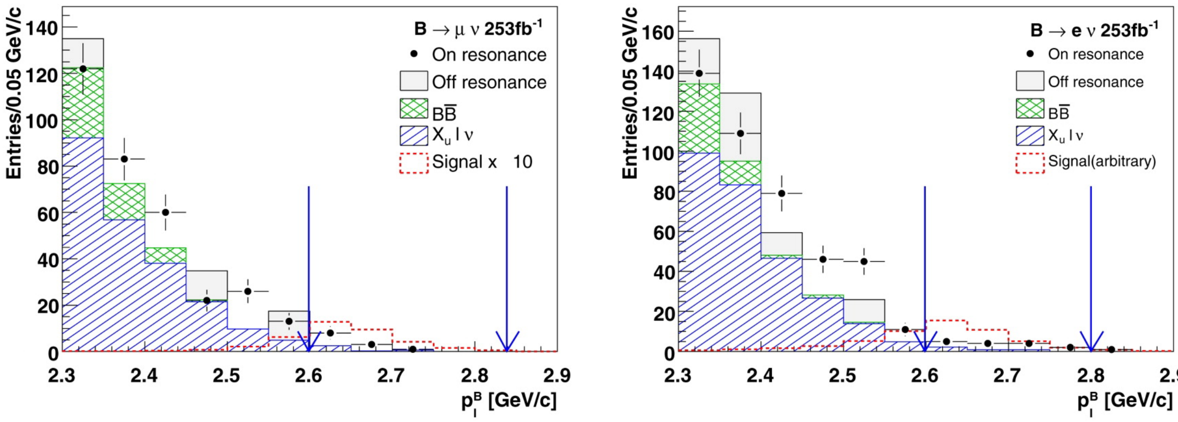


Figure 1.11. Distributions of B-frame momentum of the signal lepton, for $B^+ \to \mu^+\nu_\mu$ (left) and $B^+ \to e^+\nu_e$ (right) candidate events in the 2007 Belle search. The regions which were used to estimate the signal yield is indicated by the arrows. Images taken from Ref. [39].

1.4.2. 2008-2010: BaBar performs the first exclusive tagged searches

The next three searches for $B^+ \to \mu^+\nu_\mu$ were performed by the BaBar experiment. One was an improvement on the previous inclusive tag, while the other two were completely new and unique exclusive tagged searches.

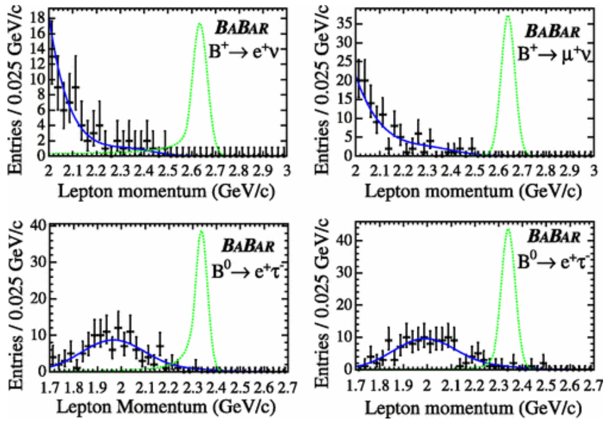
2008: BaBar (with hadronic $B_{\rm tag}$)

In 2008, BaBar published their second search for $B^+ \to \mu^+\nu_\mu$, this time as $B^+ \to \ell^+ \nu_\ell$ for ℓ=e,μ [41]. A companion search for $B^0 \to \ell^\pm\tau^\mp$ was also performed. This mode is a lepton-flavour violating decay forbidden within the SM, though technically possible if allowing for neutrino flavour oscillations within a charged current loop. This was performed in 378 million $B\overline{B}$ pairs, 100 million more than the Belle study and 200 million more than the original BaBar inclusive result.

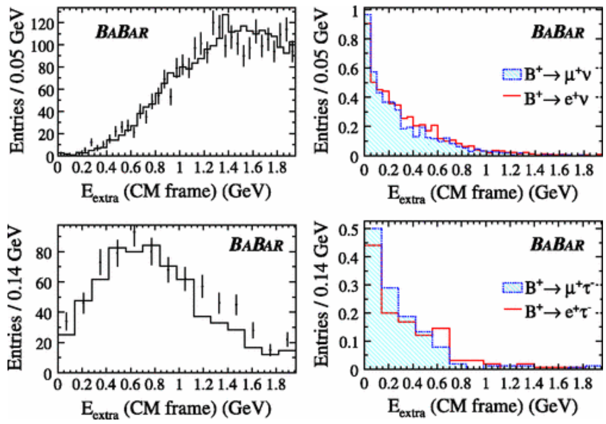
This study of $B^+ \to \mu^+\nu_\mu$ was the first hadronic tagged search, meaning the B_{tag} decay modes were restricted to obtain a higher purity of B_{tag} candidates. The B_{tag} was constructed by first reconstructing $D^{(*)0}$ and $D^{(*)+}$ mesons, via fixed kaon and pion modes. Then, additional kaons and pions were added. Whichever combination yielded the lowest ΔE was chosen as the B_{tag} candidate of the event, provided it was in an appropriate range of M_{bc}. This could be considered a 'semi-exclusive' search, as there were no constraints placed on the resulting particle identity after the additional pions and kaons were added. The signal muon was taken to be the fastest muon in the event, leveraging the large momentum compared to that of the muon in other common backgrounds i.e. $B^+ \to X_u\ell^+\nu_\ell$.

To prioritise $B^+ \to \mu^+\nu_\mu$ events, selections were made on the derived B frame muon momentum, the second Fox-Wolfram moment, cosine polar angle of missing momentum, as well as the cosine angle between the thrust axes of the B_{tag} and all other remaining particles in the event. This selection also exploits the spherical $B\overline{B}$ event topology, whereas $q\overline{q}$ events have jet-like events. Later on in the analysis of this thesis, the variable will be referred to as $\cos\theta_{T_B T_{\text{ROE}}}$. The variable E_{ECL} is also as a selection (also referred to as E_{extra}). This variable is the amount of energy recorded in the calorimeter that is not associated with the B_{tag} or the signal muon. For correctly identified B_{tag} and signal muon, there will be little unused energy, as all particles of the $\Upsilon(4S) \to B\overline{B}$ process will have been identified. While the neutrino remains missing, it is not able to interact electromagnetically with the calorimeter. For false $B\overline{B}$ events, it is very likely that there are particles remaining after the event is reconstructed, thus making E_{ECL} a highly discriminating variable. Such a selection cannot be used in inclusive tagged searches, as it is assumed that all unused energy belongs to the B_{tag}. The comparison of E_{ECL} for signal and background events is presented in Figure 1.12b.

Instead of estimating the signal yield via counting events in regions of M_{bc} and ΔE, or fitting to M_{bc}, the BaBar analysis performed an unbinned fit to the B-frame muon momentum. The fit is presented in Figure 1.12a. Notably, the signal shape is centred on lower values of B-frame lepton momentum for the $B^0 \to \ell^+\tau^-$ leptonic decays than $B^+ \to \ell^+\nu_\ell$. This is owed to the heavy τ lepton receiving a greater proportion of the B momentum than in the neutrino case, as was mentioned in the two-body kinematics derivation of Section 1.2.1. This is the first analysis to be able to do so, owing to the higher accuracy in an exclusive B_{tag} four-momentum, thus yielding a narrow enough



(a) B-frame lepton momentum. The signal shape is arbitrarily scaled. Note the varied x-axes.



(b) Energy leftover in the electromagnetic calorimeter after candidate identification. Background shapes are shown in the left column, while signal shape is shown in the right column. The $B^+ \to \ell^+\nu_\ell$ results are in the top row, while the $B^0 \to \ell^+\tau^-$ are in the bottom row.

Figure 1.12. Distributions for $B^+ \to \ell^+\nu_\ell$ and $B^0 \to \ell^+\tau^-$ (for $\ell = \mu$, e) candidate events in the 2008 hadronic B-tagged $B^+ \to \mu^+\nu_\mu$ search. Images taken from Ref. [41].

distribution that fitting is possible. However, the accuracy of the exclusive B_{tag} sacrifices efficiency, with the signal procedure retaining 0.12% of simulated $B^+ \to \mu^+\nu_\mu$ events. 5.67 events were derived from background and -0.11 $B^+ \to \mu^+\nu_\mu$ events derived from the fit, consistent with zero events. Thus a 90% confidence limit is placed at $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ $< 5.6 \times 10^{-6}$. This work placed a stricter limit than the previous BaBar result, but did not supersede the inclusive Belle result.

2009: BaBar (with inclusive B_{tag})

In 2009 BaBar sought to improve the inclusive tagged result from 2004 [42]. To do so, they used their updated dataset of 468×10^6 $B\overline{B}$ pairs, and also included a search for

$B^+ \to e^+ \nu_e$. With the BaBar experiment ending in 2008, the dataset utilised was the final set of $B\overline{B}$ events resulting from e^+e^- collisions at the $\Upsilon(4S)$ energy. Selections remain similar to the previous BaBar inclusive search, with the inclusion of a new Fischer discriminant constructed from Fox-Wolfram moments and other topology variables to reject the characteristically shaped $q\overline{q}$ events. An additional Fischer discriminant was constructed by combining both the centre-of-momentum and B-frame momentum, specifically optimised to reject the common non-$B^+ \to \mu^+\nu_\mu$ $B\overline{B}$ backgrounds, such as semileptonic B decays. Signal estimation was performed via a simultaneous unbinned fit of the muon momentum Fischer discriminant and M_{bc}. In simulated collisions, the signal efficiency was 6.1% and with a single signal event. Thus a 90% upper limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.0 \times 10^{-6}$ was derived. The tagging strategy was verified against control mode $B^+ \to \bar{D}^{(*)0}\pi^+$, where tagging provided the largest source of uncertainty.

2010: BaBar (with semileptonic B_{tag})

In 2010, the final BaBar search for $B^+ \to \mu^+\nu_\mu$ was also the first semileptonic-tagged $B^+ \to \mu^+\nu_\mu$ result [43]. This search was done in the full BaBar dataset alongside the $B^+ \to \ell^+ \nu_\ell$ and $B^+ \to \tau^+ \nu_\tau$ modes. The specific tag modes used were of the form $B^- \to \bar{D}^0\ell^- \nu_\ell$, where D^0 was reconstructed from hadrons and ℓ is either an electron or muon. This is the first $B^+ \to \mu^+\nu_\mu$ search where the whole B_{tag} four-vector is not available, as there is now a missing neutrino in the tagged B, in addition to the neutrino from the signal B. Despite this loss, semileptonic B decays can have higher efficiencies as they generally have higher branching fractions than hadronic decays. This is to be discussed Section 3.4.1. The previous tagging variables, M_{bc} and ΔE, are also no longer useful as the invariant B_{tag} mass is incomplete. Despite this, there are alternative constraints that can be placed on the semileptonic B_{tag} to achieve a higher purity than what is possible from an inclusive tag.

A key variable suitable for restricting semileptonic tagged events is $\cos\theta_{BY}$. This is the cosine angle between the visible products of the B_{tag} (i.e. $Y = D^0\ell^-$) and the nominal B direction. This quantity is not created by measuring the angle directly and computing the cosine. Instead, an expression for $\cos\theta_{BY}$ can be derived from four-momentum conservation and the fixed dynamics of the $\Upsilon(4S) \to B\overline{B}$ process. If the visible products of the B_{tag} candidate have originated from a B meson, $\cos\theta_{BY}$ will be within the allowed range of $|\cos\theta_{BY}\,| \leq 1$ i.e. θ_{BY} is a physical value. Otherwise, $\cos\theta_{BY}$ may be calculated outside of this range, indicating tagging has not been performed correctly for the event.

The full derivation is presented in Section 3.4.4. The variable of $\cos\theta_{BY}$ can also be used in a technique to approximate the muon momentum in the parent B frame. With the true B system known to be θ_{BY} away from the Y system and the nominal B momentum in the $\Upsilon(4S)$ frame known, trial B vectors can be constructed relative to Y for varying azimuthal angles. After reversing the direction to obtain the signal B dynamics, each trial vector is used to boost the muon momentum. The final estimate of the B frame muon momentum is then the average over all boosted muon momenta.

To reject background events, two likelihood ratios are constructed - one to reject $q\overline{q}$ events and a second to reject non-$B^+ \to \mu^+\nu_\mu$ $B\overline{B}$ events. These were constructed from probability density functions (PDF), created using key discriminating variables from simulated $B^+ \to \mu^+\nu_\mu$ and simulated background events i.e. normalised histograms. For each variable, a new PDF was created via $S/(S+B)$, where S is the signal PDF and B is the background PDF. A final multi-dimensional PDF was created by taking outer product of all the variable PDF's. This is the completed likelihood ratio. If combining sufficiently discriminating variables, the events with likelihoods closer to 1 will be less likely to be $q\overline{q}$ or non-$B^+ \to \mu^+\nu_\mu$ $B\overline{B}$ events. Example variables used are the Fox-Wolfram moments, cosine polar angle of the missing momentum, lepton momentum from the semileptonic B_{tag}, as well as particle identification probabilities and vertexing variables. The variable $\cos\theta_{BY}$ is omitted as it was used to create the B frame muon momentum.

The final key variable used was E_{ECL}, which was also used in the 2008 hadronic tagged BaBar search. This also includes rejecting any events with additional charged particles after a suitable signal muon is identified are rejected. Again, this is only possible because a tagged approach is taken.

Selections to the B frame muon momentum, E_{ECL}, and the two likelihood ratios of events are chosen to optimise the significance metric $\frac{S}{\sqrt{S+B}}$, a ratio of the number of signal and background simulated events. The efficiency of the analysis quoted separately for muon efficiency and tagging efficiencies, which combined gives 0.27%. This resulted in a signal efficiency of 28.65%. Signal yields were estimated by a 'cut-and-count' method in the region of $E_{\text{ECL}} < 0.72$ GeV with the selections applied, as opposed to fitting. In this region, the number of background events in data was estimated by scaling the yield predicted in background simulation by the of ratio data-to-simulation for E_{ECL} outside $E_{\text{ECL}} < 0.72$ GeV. This distributions of E_{ECL} for $B^+ \to \mu^+\nu_\mu$ and the other leptonic decays in Figure 1.13. This estimated 13 background events, whereas 12 events were seen, consistent with no signal seen. The control mode $B^+ \to \overline{D}^0\ \ell^+\ \nu_\ell$ was reconstructed with semileptonic B_{tag} in the same data, as a verification of the analysis method. The mode was chosen as this reconstructed mode would also have low left over energy in the calorimeter after being identified. After this, a upper limit on the branching fraction was placed as $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 11{\times}10^{-6}$. This is currently the second-least precise upper bound of the $B^+ \to \mu^+\nu_\mu$ branching fraction, second only to the CLEO measurement in 1995. However, it holds merit as a measurement in an independent dataset to the previous results, which had been performed using inclusive or hadronically-tagged B mesons. The requirement of having a lepton originate from the tagged B meson selects a distinct subset of data to the other tagged methods, thus providing an independent probe of $B^+ \to \mu^+\nu_\mu$.

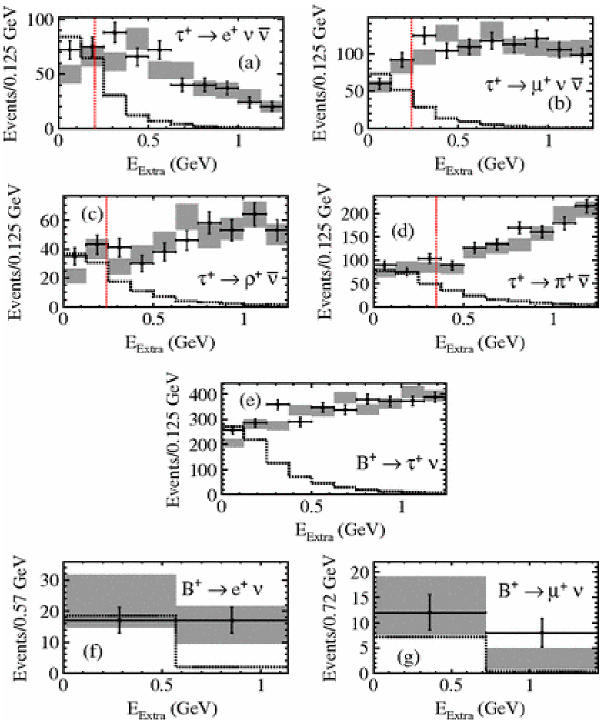


Figure 1.13. Distributions of energy remaining in the electromagnetic calorimeter after $B^+ \to \ell^+\nu_\ell$ ($\ell = e$,μ, τ) and semileptonic B_{tag} candidates are identified, from the 2010 BaBar search. The first four histograms are separated by individual τ decay modes, and are combined to give the fifth histograms. Images taken from Ref. [43].

1.4.3. 2015 onwards: Belle searches for $B^+ \to \mu^+\nu_\mu$ in largest ever dataset of $B\overline{B}$ events

The Belle experiment finished data-taking in 2010, with a final $B\overline{B}$ dataset of 772 million pairs. Using this dataset, Belle published three $B^+ \to \mu^+\nu_\mu$ searches. These consisted of one hadronic tagged search and two inclusive searches.

2015: Belle (with hadronic B_{tag})

In 2015, Belle attempted its own measurement of hadronic tagged $B^+ \to \mu^+\nu_\mu$ [44]. This was also the first search to use a machine learning algorithm to identify the hadronic B_{tag}. This is the *Full Reconstruction*(FR) method, which uses neural network learning to build the B meson decay chain with particles most probably to descend from a B [45]. This effectively reconstructed candidates in one of 615 hadronic decay modes involving combinations of pions, kaons, $D^{(*)}$ mesons and J/ψ mesons. The method could produce multiple B_{tag} candidates per event when only one B_{tag} can be used. Thus the 'best' B_{tag} candidate was decided based on the highest value of the output FR classifier for each candidate. To verify this method, a control mode of $B^+ \to \overline{D}^0\ \ell^+\ \nu_\ell$ was used with FR-based hadronic tagging. In the analysis of this thesis, an updated version of the FR method will be implemented in tagging, to be presented in Chapter 3.

Event restrictions were made on the FR classifier output, the M_{bc} and ΔE of the B_{tag}, the signal muon vertex, and the thrust angle between signal muon and B_{tag}. No events with charged particles remaining after the signal muon and B_{tag} are identified, and the remaining energy is limited to $E_{\text{ECL}} < 0.5$ GeV. Events were considered to be from $B^+ \to \mu^+\nu_\mu$ in data if the derived B-frame muon momentum was between 2.6 GeV/c and 2.7 GeV/c. Common backgrounds that remained after selection were $B^+ \to X_u\ell^+\ \nu_\ell$, $B^+ \to K\pi$ modes and $B^+ \to \ell^+\nu_\ell\gamma$. The selections ensured that $q\overline{q}$ events had been effectively removed from the analysis due to the selections. The resulting signal efficiency was 10.2%. An unbinned fit was performed to the data outside the signal estimation region. The fit was also verified by performing the analysis to $B^+ \to \overline{D}^0\ \pi^+$, where the pion was treated like the signal muon to be boosted via the hadronic B_{tag}. Ultimately no events in collision data were found in the signal region and a 90% confidence limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 2.7\ \times 10^{-6}$. This is an improvement on the 2008 BaBar hadronic tagged results, but still more than double the bound of the 2009 BaBar inclusive tagged search.

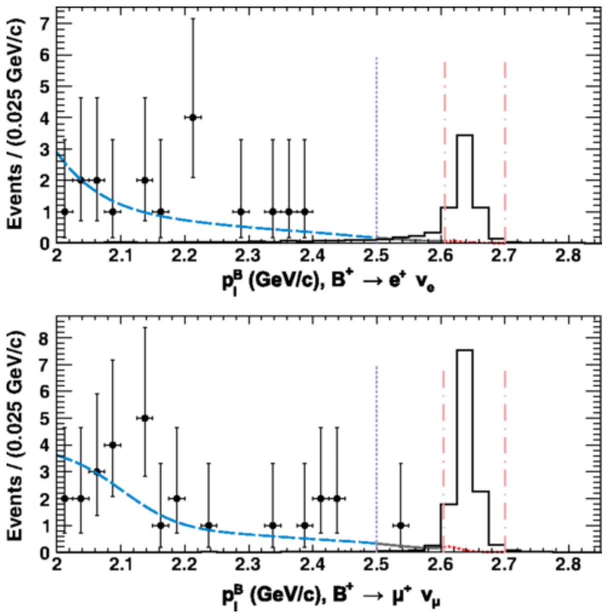


Figure 1.14. Distribution of B-frame momentum of the signal lepton, for $B^+ \to \mu^+\nu_\mu$ (bottom) and $B^+ \to e^+\nu_e$ (top) candidate events in the 2015 Belle search. Signal shape has arbitrary scaling. Images taken from Ref. [44].

2018: Belle (with inclusive B_{tag})

In 2018, Belle revisited their $B^+ \to \mu^+\nu_\mu$ inclusive tagged search, using a dataset of more than double the number of $B\overline{B}$ events than what was used previously [46]. This analysis was one of the only methods in which no attempt was made to derive the signal B-frame momentum of the signal muon from the B_{tag}. Instead, the muon momentum in $\Upsilon(4S)$ frame utilised as the key discriminating variable, despite the distribution smearing from centering on 2.6 GeV/c in the parent B-frame, to a larger range of 2.45 GeV/c to 2.85 GeV/c. The resultant range was treated as the interval in which signal yield would be estimated and was thus kept 'blind' until the analysis method was finalised.

Similar to preceding searches, requirements are made to the inclusive tag energy and mass via M_{bc} and ΔE. Moreover, the number of additional leptons in the event was limited, to avoid including any B_{tag} events that had missing energy themselves, such as semileptonic B or resultant semileptonic D decays. A great deal of $q\overline{q}$ events still remained after these selections. Thus, a range of variables that discriminated between $q\overline{q}$ and $B\overline{B}$ events were used to train a neural network, with intention of producing a classifier that could separate signal and background events. Some of these variables included event shape variables, polar angle of missing momentum vector, event thrust angles and B_{tag} vertex information. Both this neural network classifier and the $\Upsilon(4S)$ frame muon momentum were used to extract the signal yield. This was performed as a template fit, where the binned histograms provided the fitting PDF's, rather than other continuous functions that have been used previously. The projected distributions are shown in Figure 1.15. Included in the fit were dedicated components for $B \to \pi\ell\nu_\ell$ to account for this process as a key background. To obtain the branching fraction of $B^+ \to \mu^+\nu_\mu$, the fit was performed to construct the ratio of $B^+ \to \mu^+\nu_\mu$ events to $B \to \pi\ell\bar{\nu}_\ell$, a ratio in which a number of systematic errors would cancel out. From this, 195 events in collision data were predicted to be from $B^+ \to \mu^+\nu_\mu$. No signal efficiency is quoted. The derived branching fraction of 6.46×10^{-7}, with a systematic uncertainty of 25%, resulting in a significance of 2.4 standard deviations. To be comparable with previous results, the 90% confidence limit is found to be $\mathcal{B}(B^+ \to \mu^+\nu_\mu) \in[0.29,1.07]\times10^{-6}$, thus consistent with an excess over background. This is the only search currently to have obtained a two sided interval rather than an upper limit.

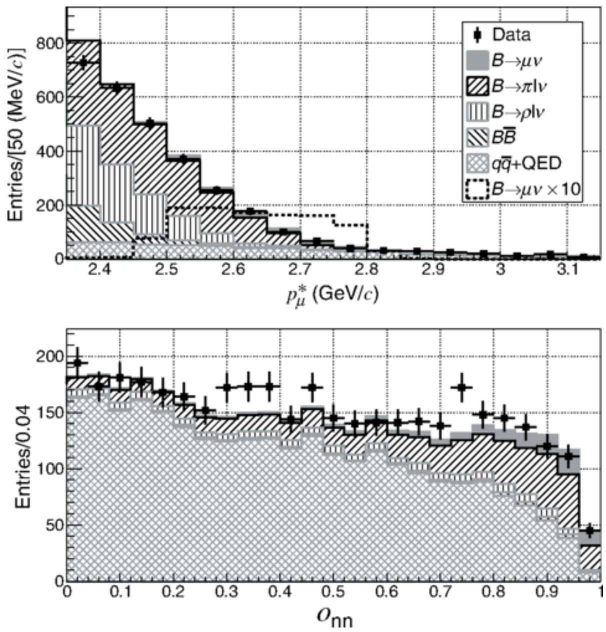


Figure 1.15. Distributions of $\Upsilon(4S)$-frame momentum of the muon (top) and the neural-network classifying variable (bottom), for $B^+ \to \mu^+\nu_\mu$ candidate events in the 2018 Belle search. Image taken from Ref. [46].

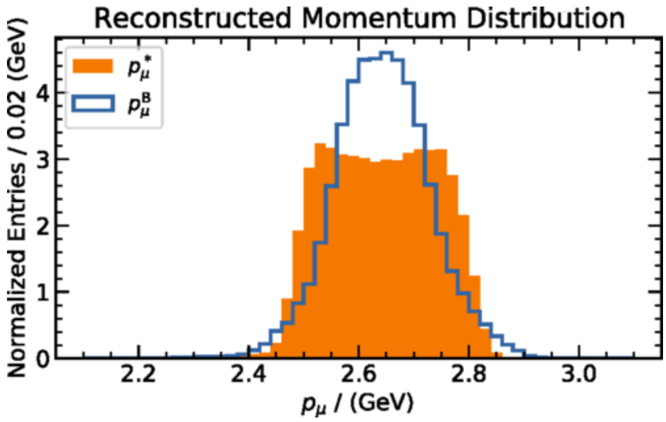
2020: Belle (with inclusive B_{tag})

The 2020 Belle search for $B^+ \to \mu^+\nu_\mu$ is the most recent $B^+ \to \mu^+\nu_\mu$ study published by any collaboration [27]. It is also the best 90% upper limit on the $B^+ \to \mu^+\nu_\mu$ branching fraction so far. This result also was used to place bounds on the contributions of charged

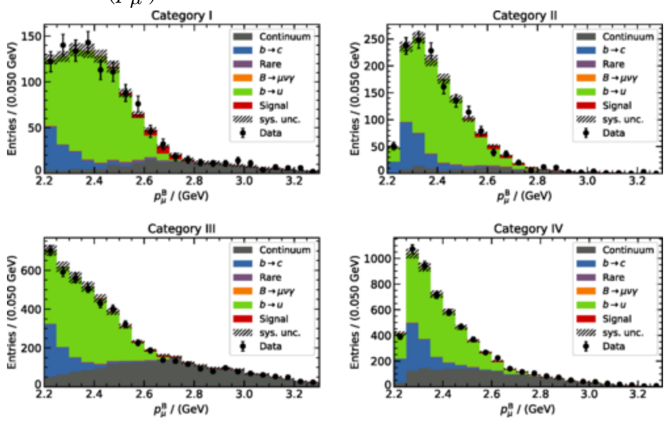
Higgs and sterile neutrino contributions to the mode, which was discussed in Section 1.3.1 and Section 1.3.4.

The inclusive tagged $B^+ \to \mu^+\nu_\mu$ search was performed on the full Belle $B\overline{B}$ dataset. Considerable effort is dedicated to accurately modelling backgrounds. This includes updating simulations of $B^+ \to X_u\ell\bar{\nu}_\ell$ and $B^+ \to X_c\ell\bar{\nu}_\ell$ processes with latest branching fractions and hadronisation models. A boosted decision tree classifier was trained to better align simulated $q\overline{q}$ events with real non-$B\overline{B}$ events from e^+e^- collisions below the $\Upsilon(4S)$ resonance. Like almost all previous searches, the signal muon momentum in the signal B-frame is derived from the inclusive B_{tag}. Unlike previous searches, a study was performed to improve the resolution of the derived B-frame muon momentum. The B_{tag} momentum was replaced with the inclusive B_{tag} unit vector scaled to the nominal value constrained by the two-body dynamics in the $\Upsilon(4S)$ frame. Moreover, an additional calibration is performed on the B_{tag} momentum to account for missing particles in the inclusive tag. The difference between the $\Upsilon(4S)$-frame and B-frame muon momenta is presented in Figure 1.16. A boosted decision tree classifier is trained to distinguish $B^+ \to \mu^+\nu_\mu$ from the considerable $q\overline{q}$ background. This was trained on Fox-Wolfram moments, event shape variables, M_{bc}, ΔE, as well as the number of particles used in the tag.

After the selections, the signal efficiency was 28%. Signal estimation is performed as a binned template fit to the derived B-frame momentum of the muon, simultaneously to four unique regions of the classifier output and the cosine angle between the signal muon in the B-frame and the B in the $\Upsilon(4S)$ frame. Figure 1.16b contains the results of this fit. This technique was verified using $B^+ \to \bar{D}^0\pi^+$, where the pion boosted into the signal B frame as an analogue for the signal muon. Moreover, due to the careful modelling of the background component $B \to X_u\ell^+\nu$, a measurement of its branching fraction was made by loosening selections on the classifier output. These events peaked at B-frame muon momentum 2.25 GeV/c, outside the consideration of $B^+ \to \mu^+\nu_\mu$ events. Restoring the classifier selection, the $B^+ \to \mu^+\nu_\mu$ branching fraction was measured to be $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ =$(5.3 \pm 2.0 \pm 0.9)\times10^{-7}$, where uncertainties are statistical and systematic respectively. The largest sources of error were due to $B \to X_u\ell\nu_\ell$ modelling and MC statistics. Using this systematic error, a 90% confidence limit is placed at 0.86×10^{-6}.



(a) Comparison of normalised distribution shape of $\Upsilon(4S)$-frame muon momentum (p_μ^*) and the derived B-frame muon momentum (p_μ^B).



(b) B-frame muon momentum, in each of the four fitting regions. The top and bottom rows are signal-enhanced or signal-depleted according to the $q\bar{q}$ classifier, while the left and right columns contain the signal-like or background-like cosine angle of the muon-momentum from the B meson boost.

Figure 1.16. Muon momentum distributions from the 2020 Belle $B^+ \to \mu^+\nu_\mu$ search.

1.5. Summary and Outline of this $B^+ \to \mu^+\nu_\mu$ search

In this Chapter, the $B^+ \to \mu^+\nu_\mu$ decay was introduced as the decay of interest for this thesis. The decay is predominantly explained by the flavour sector within the SM. Since $B^+ \to \mu^+\nu_\mu$ is a two-body decay, the momentum of the muon is fixed to be 2.64 GeV/c in the parent B-frame, a relatively large momentum compared to other B meson processes. As a weak decay, $B^+ \to \mu^+\nu_\mu$ is heavily suppressed due to the low probability of producing the light lepton pair with the appropriate helicity to conserve both linear and angular momentum. In addition to this rarity, it can be difficult to reconstruct the $B^+ \to \mu^+\nu_\mu$ decay in particle physics experiment, as the final state neutrino is very difficult to detect. This can be remedied by purposefully looking for particle events with a large amount of missing energy that can be attributed to the presence of the neutrino. Such a technique is only possible if the experiment performing the $B^+ \to \mu^+\nu_\mu$ search has known initial and final energy conditions. Despite these difficulties, a search for $B^+ \to \mu^+\nu_\mu$ in experiment can benefit from its simple reconstruction, a result of being a purely leptonic decay with no intermediate particles to identify. Leveraging this simple reconstruction could allow for high precision measurement of the $B^+ \to \mu^+\nu_\mu$ branching fraction, and thus the calculation of related B^+ meson observables, such as the decay constant f_B and CKM matrix element $|V_{ub}|$. The CKM matrix element is especially important, as $|V_{ub}|$ is the least precisely known parameter, potentially pointing to broken unitarity and new physics in B decays.

In addition to measuring the B meson observables, $B^+ \to \mu^+\nu_\mu$ holds potential to constrain BSM physics. BSM physics must exist, in order to explain phenomena such as the baryon asymmetry of the universe. The decay could be affected by charged Higgs models, leptoquark models, supersymmetry and sterile neutrino models. While a rare decay according to SM, the branching fraction could be enhanced by these models, leading to a higher probability of observing a $B^+ \to \mu^+\nu_\mu$-like event. If an excess in events is seen, it could validate the presence of new physics. If no excess is seen, the search still can improve the precision of upper bounds on this previously unseen B decay.

The previous $B^+ \to \mu^+\nu_\mu$ searches performed at the CLEO, BaBar and Belle experiments were reviewed, so that the analysis strategy of this thesis can be planned. The searches have also given a preliminary look at the shapes of distributions from key variables in the decay, such as muon momenta and E_{ECL}.

This search will take place using B decays sourced from the sequel to the Belle experiment, the Belle II experiment. Belle II takes place on an upgraded accelerator, with the intention to produce a dataset of 50 times that of the original Belle experiment. It is thus well suited to a precision measurement, to be discussed in Chapter 2.

Unfortunately, the currently available Belle II dataset is small, with around half the number of $B\overline{B}$ events as what was in the original Belle dataset. For this reason, the semileptonic B-tagging approach will be implemented to balance the event retention with the accuracy of reconstruction, similar to what was used in the 2010 BaBar search. The hadronic mode will also be trialled in Section 3.4.3, to ensure the best tagging technique is chosen. Like the 2015 Belle search, a machine learning algorithm will be implemented to obtain a B_{tag} from a large number of possible decay modes, called the *Full Event Interpretation* [47]. Doing so will be the focus of Chapter 3.

Without the full B_{tag} information as a semileptonic tagged search, the signal variable for this search will be the $\Upsilon(4S)$-frame muon momentum, which was also used in the 2018 Belle study. Efforts towards estimating the muon momentum in the signal B frame will be found to not be as discriminating and thus not utilised. In order to reject background processes from being included in the estimated signal yield, the previous studies used a variety of techniques: standard rectangular selections, linear Fischer discriminants, constructing profile likelihoods, training of multivariate classifiers etc. Fisher discriminants have fallen out of favour, as new machine learning techniques have emerged. Despite this, such techniques do bring in added complexity and so standard selections will be implemented. The decision of such selections will be performed in a systematic way, As in the 2004 BaBar search, the ratio $\frac{S}{\sqrt{S+B}}$ will be used as to derive appropriate event selections to obtain a subset of data most likely to contain $B^+ \to \mu^+\nu_\mu$ events. Variables that had appeared in previous searches will inform those chosen in this search. This includes E_{ECL}, cosine polar angle of missing momentum ($\cos\theta_{\vec{p}_{\text{miss}}}$), cosine angle between B_{tag} thrust axis and particles of event ($\cos\theta_{T_B T_{\text{ROE}}}$), cosine angle between the visible B_{tag} daughters and the nominal B direction ($\cos\theta_{BY}$), number of unreconstructed charged particles in the event etc. The signal reconstruction and selection criteria optimisation will be performed in Chapter 4.

Most studies used a control mode to validate B-tagging techniques, using either $B^+ \to \bar{D}^0\ell^+\nu_\ell$ like the BaBar 2010 search or $B^+ \to \bar{D}^0\pi^+$ like the Belle 2020 search. For this analysis, the $B^+ \to \bar{D}^0\pi^+$ will be implemented, so as to minimise reconstruction difficulties between a double semileptonically decaying event. Additionally, the hadronic control mode matches the two-body kinematics of $B^+ \to \mu^+\nu_\mu$, where it would be expected that the signal pion in the B-frame has higher momentum spectrum than that of the lepton from a three-body decay. The control mode will also be used to derive corrections to the signal efficiency. The 'cut-and-count' approach of the 2010 BaBar search will be followed, as well as their technique of estimating a 90% confidence interval. This will complete the analysis in Chapter 5.

This work will be the first time a semileptonic B_{tag} has been used to identify $B^+ \to \mu^+\nu_\mu$ decays at either Belle or Belle II experiments. Before diving into the search for $B^+ \to \mu^+\nu_\mu$, the method of producing the B meson dataset will be presented, by discussing the Belle II experiment.

$\bowtie\bowtie\bowtie$

Chapter 2.

The Belle II Experiment

In Chapter 2, Belle II is detailed as a collider physics experiment which focusses on the measurement of B mesons. The Belle II dataset of B mesons will be used to potentially measure rare $B^+ \to \mu^+\nu_\mu$ *decays in this thesis.*
The wider goals of Belle II, as a "flavour factory" of high-energy physics research will be introduced.
The physics principle by which a large dataset of B mesons are created in electron-positron collisions will be presented, especially as it pertains to the asymmetric energy collider, SuperKEKB, that serves the Belle II experiment.
Belle II's detection hardware, a general purpose spectrometer located at the electron-positron interaction point, will be outlined with focus on how the various subdetectors function to indicate the presence of a $B^+ \to \mu^+\nu_\mu$ *decay.*
The transition from signals in hardware to an analysable dataset is facilitated by high-performance computing, of which Belle II's approach will be reviewed.
The use of simulated Belle II collisions, referred to as Monte-Carlo (MC) data, will be discussed in its role to examine properties of the rare $B^+ \to \mu^+\nu_\mu$ *decay within the Belle II detector, including the specification of the MC samples utilised.*

$\bowtie \bowtie \bowtie$

2.1. The Belle II physics program

2.1.1. Origins

Belle II is a high-energy physics experiment located at the SuperKEKB collider in Tsukuba, Japan. SuperKEKB is an asymmetric energy electron-positron collider, operating at a centre-of-mass energy equivalent to the $\Upsilon(4S)$ resonance i.e. $\sqrt{s}$ =10.58 GeV. Belle II differentiates itself from other high-energy physics experiments by focusing on the *luminosity frontier*, with goals of collecting 50 ab^{-1} of collision data over the next 10 years. This integrated luminosity is equivalent to tens of billions of B mesons, charm hadrons and tau leptons. With this dataset, world-leading precision measurements of SM, as well as BSM physics are possible [48]. In the search for $B^+ \to \mu^+\nu_\mu$, charged B decays measured at the Belle II experiment are used.

The Belle II experiment is based largely on its predecessor, the original Belle experiment. In the 1990's, the Belle experiment was purpose built to measure charge-parity violation (CPV) in B meson pairs from the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ is a spin-1 bottomonium meson (i.e. has structure $b\bar{b}$), which was found to almost always decay to $B\overline{B}$ pairs by the CLEO experiment a decade earlier [33]. Belle improved upon CLEO's experimental design by using asymmetric energy electron/positron beams, such that the $B\overline{B}$ pairs created would be produced with a Lorentz boost. The boost dilates the lifetime of the B mesons when observed in the lab-frame. This allows for higher precision in measurements of decay time differences in each B, a necessary addition to measure CPV. Belle also improved upon CLEO by increasing the number of electron-positron collisions that could be performed per second, the *instantaneous luminosity*. This earned Belle the moniker of being a *B-factory* experiment.

Belle was not the only B-factory experiment at the time. The BaBar experiment was similarly designed, taking place at the Stanford Linear Accelerator (SLAC) facility using the circular electron-positron collider, PEP-II [30]. Both experiments successfully measured CPV in B decays, with their measurements of $\sin(2\beta)$, a parameter which was introduced in Section 1.1.2. Both measurements were consistent with each other and with CPV in B decays, with papers published simultaneously in Physical Review Letters [36] [37]. These results provided further evidence for Kobayashi and Maskawa theory, that having six-quarks within the Standard Model could allow for CPV without new physics mechanisms [10].

The Belle and BaBar experiments each conducted diverse physics programs throughout the 2000's, investigating further into CPV in B decays, as well as measuring CKM matrix parameters, B meson oscillations, quarkonium (i.e. $q\bar{q}$ mesons) spectroscopy, charm and tau physics, as well studies outside of the $\Upsilon(4S)$ energy. BaBar finished data taking in 2008, without renewed funding. Belle finished data taking in 2010 after acquiring a 1 ab^{-1} dataset (711 fb^{-1} at $\Upsilon(4S)$), with the funding to improve the accelerator and the detector. The design and construction of Belle II began soon after Belle completed. Preliminary data-taking began in 2018 with a partial detector, with the official 'Physics Run 1' beginning in April 2019. The outline of data-taking so far will be presented in Section 2.4.4. With vast improvements in detector and B meson production rate, Belle II has commenced the *Super-B factory era* [49].

Belle was designed to measure CPV as a probe of the Standard Model. With CPV in B mesons now confirmed, Belle II shifts towards probing Beyond Standard Model (BSM) physics. While particle physics experiments at the *energy frontier* can conduct direct searches for BSM particles at the TeV energy scale, Belle II can take a complementary approach through precision measurement of flavour physics anomalies. The anomalies in different measured flavour physics parameters can hint towards new physics phenomena. They may also arise as a purely statistical effect. Focusing on using Belle II's large datasets to precisely measure flavour physics parameters can definitively determine if the anomalies are a statistical error or if they are potentially a result of BSM mechanisms. The specific current physics efforts of Belle II are described in the next section.

2.1.2. Belle II Physics Working Groups

Semileptonic and Missing Energy

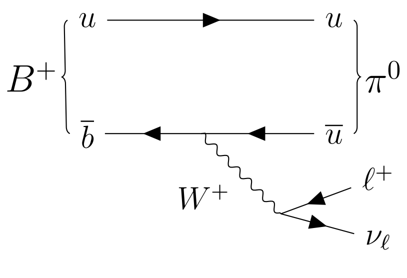


Figure 2.1. Feynman diagram of an example semileptonic decay, $B^+ \to \pi^0 \ell^+ \nu_\ell$, where ℓ is either an electron or muon.

The Semileptonic and Missing Energy (SLME) group focuses on semileptonic (and pure leptonic) B decays that occur via tree-level and include neutrinos. Such decays are of the form $B \to X\ell\nu$, where ℓ is e, μ, τ. X is then a meson in the semileptonic decays, or not present for the purely leptonic case. The missing energy from the neutrino in such decays can make B mesons in SLME decays incomplete and difficult to identify. The SLME group extensively uses the B-tagging algorithms, as will be outlined in Chapter 3, in order to assist in identifying such events. Special focus is given to SLME B decays that include a $b \to u$ or $b \to c$ charged weak transitions. Not only is it important to focus on these decays to measure their previously unseen branching fractions, but they can also be portals to new physics. Examining these decays allow for measurements of the relevant CKM matrix amplitudes, V_{cb} and V_{ub}. Interest is also given to measurements which probe the assumption of lepton universality, R($D^{(*)}$), which compares semileptonic decays to light-leptons versus the τ system. Studies into CKM amplitudes, as well as lepton universality are methods of searching for new physics. A recent publication by the SLME group tests lepton universality by measuring the ratio of inclusive semitauonic B decays to light-lepton inclusive semileptonic B decays, as a complementary approach to R($D^{(*)}$) [50]. New physics mechanisms can include new CPV phases to explain baryon-asymmetry or leptoquark theories that propose new particles more strongly coupled to the heavier tau lepton. The studies of $B^+ \to \mu^+\nu_\mu$ also fall into the jurisdiction of this group.

Radiative and Electroweak Penguin

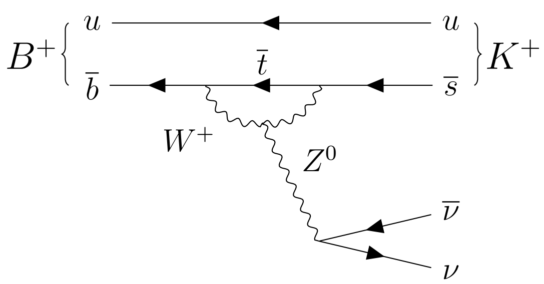


Figure 2.2. Feynman diagram of the $B^+ \to K^+\nu\overline{\nu}$ which takes place via a penguin diagram.

The Radiative and Electroweak Penguin (EWP) group focuses on B decays that cannot occur at tree level in the Standard Model. Specifically, they are those that contain a $b \to s$ or $b \to d$ transition, alongside a spectator quark. As there are no SM direct transitions between quarks without a change in electric charge, these can only occur via 'loop' diagrams, where a W boson is emitted and reabsorbed. A radiative penguin transition also emits a final-state photon in this process. The name is earned based on the 'arms' and 'legs' of the incoming and outgoing quarks, as shown in the Feynman diagram of Figure 2.2. As a neutral transitions, decays involving leptons are to exact lepton pairs (i.e. dileptons). With branching fractions being inversely proportional to the minimum number of loops, EWP decays are typically heavily suppressed and require large amounts of data to measure with precision. The search for the rare decay $B^+ \to K^+ \nu\overline{\nu}$ was published by the EWP group in 2024 [51]. The motivations behind measuring this decay mode are similar to that of rare decay $B^+ \to \mu^+\nu_\mu$. As process which only has 90% upper limits quoted on the branching fraction, there are BSM theories such as those concerning dark matter scalar particles, that suggest that the it could be enhanced by new physics contributions [52]. Such a decay also features missing energy through the undetectable neutrinos, benefiting from B-tagging algorithms like in the SLME working group.

Time-Dependent Charged-Parity Violation

The Time-Dependent Charged-Parity Violation group (TDCPV) seeks to measure CPV in weak B processes. They are primarily concerned with the α and β (also denoted as ϕ_2 and ϕ_1) angles of the CKM unitarity triangle, which was introduced in Section 1.1.2. These involve neutral B decays to to CP eigenstates, like $B^0 \to J/\psi\ K_S^0$ and $B^0 \to \pi^+ \pi^- \pi^0$. Here, their final states could originate from either B^0 or $\overline{B}^0$ and there is interest in measuring differences in decay rates as a function of propagation time, thus their likelihood for the B^0 to oscillate flavour. This group also relies on flavour-tagging to perform their analyses, where events with B decays of definite flavour (i.e. are uniquely B^0 or uniquely $\overline{B}^0$) are identified on the basis of their decay products [53]. This can be used to then identify the original flavour of a CP eigenstate decay. The TDCPV group at Belle II recently revisited the $\sin(2\beta)$ result using an graph-neural-network flavour tagger, which was an improvement on the original algorithm [54].

Hadronic *B*

The Hadronic B working group (BHAD) focuses on B decays to purely hadronic final states, namely 2- to 3- body mesonic decays. The efforts are divided into B to charm transitions, and B to charmless transitions. For the hadronic decays involving $B \to D$ transitions, emphasis is on measuring CPV through unitary angle γ (i.e. ϕ_3). Here, CPV takes place not due to B mixing, but due to interference between D^0 and $\overline{D}^0$ modes, such as in $B^- \to \overline{D}^0\ K^-$ and $B^- \to D^0\ K^-$. For the charmless B decays (i.e. B decays to light-quark hadrons), a b to u transition is usually featured. These have small branching fractions due to the Cabbibo suppression and $|V_{ub}|$. Studies of the kinematics of N-body decays are of focus, as well as measurements of direct CP violation. The latest Belle II results from this group include a measurement of the branching fraction of $B^- \to D^0\ \rho^-$ and measurement of the branching fractions of $B^0 \to K\pi$ and $B^0 \to \pi\pi$ for a number of different charge combinations [55] [56].

Charm

The charm group aims to perform precision measurements of processes concerning charm-flavoured hadrons. Charm hadrons can be obtained from B mesons, thanks to the large number of $b \to c$ transitions. One can also obtain c-type hadrons, such as D mesons, via the continuum process $e^+e^- \to c\bar{c}$, as will be discussed in Section 2.2.3. Many of the same phenomena studied in B mesons with previously discussed working groups can be extended to D mesons, due to their structure as pseudoscalar mesons. This includes CP violation in $D^0/\overline{D}^0$ processes, D lifetime measurements, measuring $D/\overline{D}$ oscillations, $c \to u$ penguin decays, leptonic decays of D and other rare decays such as $D^0 \to \gamma\gamma$. Unlike the B meson, this can be performed for a variation of D mesons at $\sqrt{s}$ =10.58 GeV, including doubly-excited D states and strange D_s mesons. Additionally, studies of charm baryons can be performed, including of Λ_c, Ω_c and Ξ_c. Ξ_c in particular was studied using data combined from Belle and Belle II, to measure relative branching fractions of $\Xi_c \to \Xi$ transitions [57].

Quarkonium

The quarkonium working group focuses on the properties of $b\bar{b}$ and $c\bar{c}$ mesons. This is also inclusive of quarkonium-like states, hadrons with similar properties to $q\bar{q}$ states, but potentially novel structure, such as quark-gluon hybrids, tetraquark mesons etc. The original Belle experiment studied quarkonia extensively, including spectroscopy, measurement of quarkonium quantum numbers, transitions between quarkonia such as $\Upsilon(1S) \to \eta_b\gamma$, energy scans outside of Υ resonances, as well observing exotic quarkonia states like $X(3872)$. For Belle II, such topics are revisited and extended. Focus is currently on the energy scans that have been performed above $\Upsilon(4S)$, resulting in observation of an $\Upsilon(10753)$ state, which is between $\Upsilon(4S)$ and $\Upsilon(5S)$ [58]. The structure of this state is not yet certain, whether it be one of a number of different exotic quarkonium structures or a D-wave conventional $b\bar{b}$ state.

Low multiplicity and Dark sector

Low multiplicity processes are generally those that do not produce hadrons. Instead, when an e^+e^- collision takes place, a low multiplicity state may be produced, usually a combination of light leptons. Measuring the rate of processes such as Bhabha scattering (i.e. scattering via $e^+e^- \rightarrow e^+e^-$) and digamma processes ($e^+e^- \rightarrow \gamma\gamma$) can be used to evaluate the number of $B\overline{B}$ events taking place. Cases when e^+e^- collisions do not result in $B\overline{B}$ events will be discussed further in Section 2.2.3. The first three papers by the Belle II collaboration originated from low multiplicity and dark sector processes. The first was measurement of integrated luminosity of data recorded during Phase II of the experiment, when data-taking was performed with a partial detector [59]. The second paper was a search for a Z' particle in low multiplicity processes, a theorised new vector boson that only couples to leptons and decays preferentially to dark matter particles [60]. The third paper was a search for axion-like particles (ALP's) that are created in e^+e^- processes and decays to $\gamma\gamma$. ALP's are a BSM explanation for why the strong force respects CP symmetry, unlike the weak interaction. The low multiplicity and dark sector group have published a number of other results concerning new physics, including dark photon and dark Higgs searches, further Z' boson searches and searches for spin-0 mediators of direct $s \rightarrow b$ transitions

Tau

The final Belle II working group is the tau group. Formerly considered alongside low-multiplicity processes, it differs from light leptons due to its ability to produce hadrons in its decays. Tau studies are usually performed in $e^+e^- \rightarrow \tau^+\tau^-$ processes, rather than as a product of B decays. Studies include the tau lepton mass measurement, tau lifetime measurement, and tests of lepton flavour universality in ratios of $\tau \rightarrow \ell\nu_\ell\nu_\tau$ decays. Tau decays are also useful in performance studies, as a reliable source of pions in their decays. They have been used to test muon/pion identification, as well as how well final-state charged particles are identified (*tracking*). The studies will be discussed in Section 5.2, while the role of tracking will be described in Section 2.3.2.

2.2. SuperKEKB and B production at electron-positron colliders

An overview of the Belle II experiment and its scientific goals have been discussed, particularly with respect to B-physics. It is now necessary to discuss how B mesons are produced for analysis within the Belle II collaboration. This will include a description of the SuperKEKB collider and how the electrons and positrons are accelerated to the energies required for the Belle II experiment. The uniquely constrained kinematics of B mesons produced at e^+e^- collision within the Belle II detector will also be presented, motivating why particle decays featuring missing energy are well suited to Belle II. Unfortunately, not all e^+e^- collisions will result in B mesons, so the possible non-$B\overline{B}$ backgrounds that can be produced will be highlighted, whether they take place at the design energy of $\sqrt{s}$=10.58 GeV or alternate collision energies.

2.2.1. The SuperKEKB collider

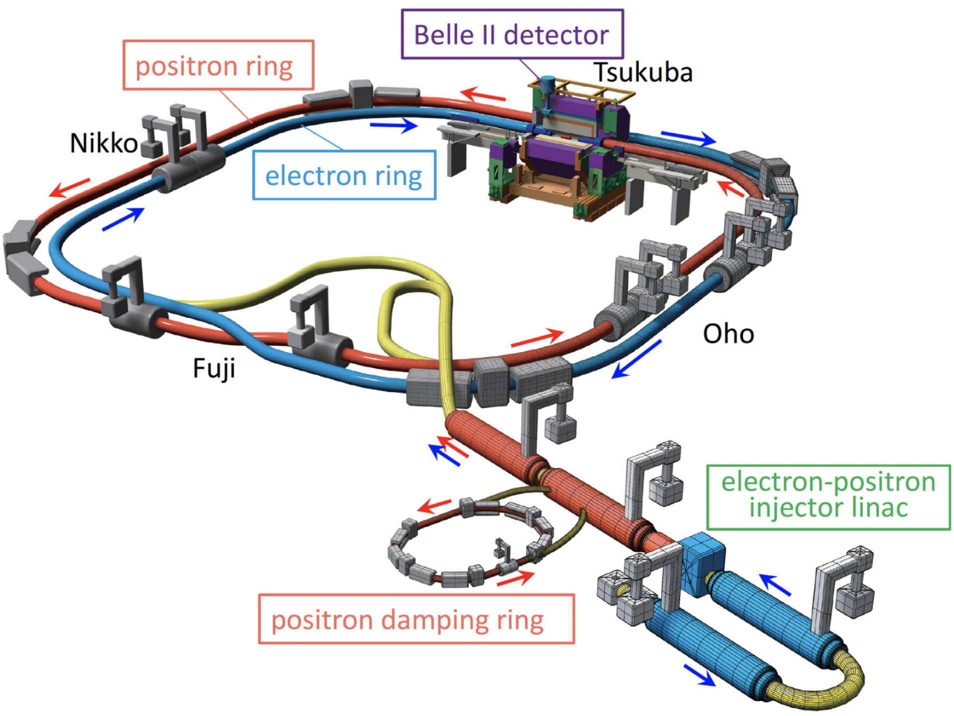


Figure 2.3. The SuperKEKB collider, consisting of the radio-frequency electron gun, the injection linac, positron damping ring, electron and positron storage rings and the Belle II detector. Image taken from Ref. [61].

SuperKEKB is the circular e^+e^- collider which services the Belle II experiment [61]. Inheriting the majority of machinery used in KEKB, the collider of the original Belle experiment, SuperKEKB has been improved with plans to operate at top instantaneous luminosity of $6\times10^{35}\text{cm}^{-2}\text{s}^{-1}$ [62]. This is equivalent to approximately 1400 B mesons per second and around 30 times that achieved by KEKB.

Located in Tsukuba, Japan, SuperKEKB is managed by the High Energy Accelerator Research Organization, more commonly known as KEK (from Japanese, *Kō Enerugī Kasokuki Kenyū Kikō*). In addition to the original KEKB collider, the KEK laboratories have housed a number of accelerators, including a 12 GeV proton synchrotron and neutrino beamline. These facilities were previously used to create and direct muon neutrinos towards the Super-Kamiokande observatory, in order to measure neutrino oscillation [63]. Currently, alongside SuperKEKB, KEK also operates the 'Photon Factory' electron synchrotrons and a linear accelerator, to be used as a testing facility for the planned International Linear Collider (ILC).

SuperKEKB consists of an injection linear accelerator (or *linac*), a damping ring and two counter propagating storage rings, one for each of the electron and positron [64]. These are indicated in Figure 2.3.

A radio-frequency particle gun acts as the source of electrons for SuperKEKB. In this apparatus, laser light is incident on a photocathode, such that 4 nC of electrons are liberated. Radio-frequency (RF) quasi-standing waves are implemented to to accelerate the emitted electrons towards the beginning of the linac beamline. An alternate thermionic particle gun is also used as an additional source, whose electrons will be used in the production of positrons. In this gun, 10 nC of electrons are freed due to heat from current passing through the cathode. Each gun alternates in 'firing' along the same beamline. Facilitated by bending magnets, the beamline has a 180° turn in it. This is referred to as the 'J-arc', a result of space restriction during reconstruction [65]. After this, a pulsed corrector magnet deviates the path of the 10 nC grouping of electrons towards a tungsten target. The interaction between the electrons and the tungsten target results in a reliable source of positrons. The magnet is pulsed off for 4 nC electrons, so they can pass beside the target, unaffected. Throughout the linac, the electrons and positrons are accelerated through up to 60 specialised units. These units use static radio-frequency standing waves to give the electrons and positrons their energy appropriate for collisions. The maxima of these standing waves also coax the electrons into periodic bunches of uniform velocity. The design is such that a low emittance, (i.e. a narrowed electron beam of uniform momenta) is prioritised. The positrons are created with a high emittance, and so they are corrected by passing through the damping ring.

The damping ring works similarly to a general electron storage ring. In the curved sections, dipole magnets and quadrupole magnets are alternated. The dipole magnets curves the trajectory of the positron bunches, where faster positrons in the bunch lose excess energy as synchrotron radiation. In the straight sections, RF acceleration units replenish energies of those positrons that are below what is desired. Quadrupole magnets are placed throughout to narrow the bunches. Unlike in storage rings, the positrons are kept in the damping ring for a short duration, until the emittance is lowered sufficiently. After leaving the ring, the energy spread of the positrons has reduced by a third [64].

The electrons that are guided passed the positron target, once accelerated to 7 GeV, are injected from the linac into a dedicated storage ring referred to as the high energy ring (*HER*). Similarly, the positrons leaving the damping ring, once accelerate to 4 GeV, are injected into their own storage ring, the low energy ring or *LER*. These rings are counter propagating and both cover the same circumference of 3km. Electrons travel around the HER in a clockwise direction and defines the 'forward' direction at the point of collision. In actuality, the 'rings' are a rounded square, with four straight sections and four curved sections. As was the case in the damping ring, dipole bending magnets are used to redirect the beam in the curved sections, RF accelerating units are used in the straight sections to maintain the energies of the electrons and positrons, with quadrupole magnets to focus the beam throughout.

The *interaction point* (IP), where the electrons and positrons will finally collide, is located on the straight directly opposite the point of injection. This straight takes place in Tsukuba experimental hall, named for its proximity to Mount Tsukuba. Prior to the collision, the electrons and positrons both travel through their own final focusing superconducting electromagnets on either side of the interaction point, referred to as the QCS (from the reversed acronym of *Super Conducting Quadrupole*). This is a system of multiple magnets consisting of 2-12 poles that narrow the electron and positron beams, so as to increase the instantaneous luminosity required for the Belle II experiment.

To achieve a high rate of collision between electron and positron bunches and thus higher instantaneous luminosity, two key improvements were made when transforming the KEKB collider into the SuperKEKB collider. The first is to increase the current of the electron and positron beams, while the second is to reduce the vertical beta function of the beam at the IP. The vertical beta function is a quantity related to the bunch height,

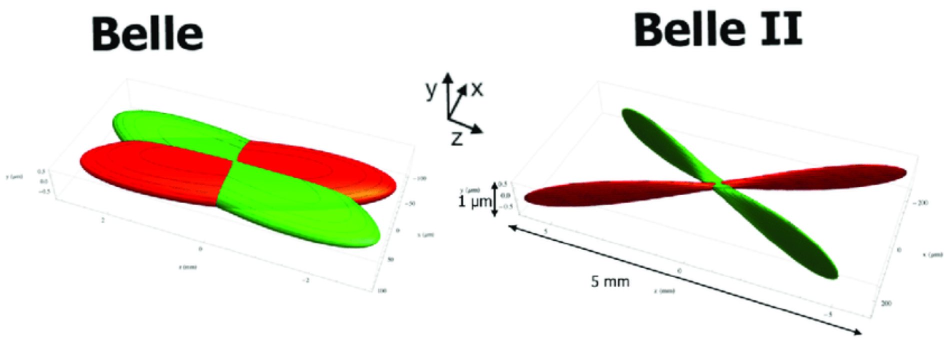


Figure 2.4. A comparison of the electron-positron bunch crossing between the original Belle experiment (left) and Belle II (right, with the nano beam scheme).

when modelling a particle beam with Gaussian shape. These improvements were made as the instantaneous luminosity is directly proportional to the current, and inversely proportional to the vertical beta function. To improve the current, the electron guns of the linac were upgraded to maximise the number of electrons/positrons in each collided bunch, resulting in an overall increase in the beam current. To narrow the beam height, the *nano-beam scheme* is implemented. This collision method was originally conceived for an Italian B-factory experiment that was never built. Traditionally beam collisions are performed head-on, where the bunches have large physical cross-sections. With the nano-beam design, electron and positron bunches are squeezed into much smaller cross-sections, increasing the probability of interaction. A schematic of this is shown in Figure 2.4. The nano-beam shape is facilitated by the QCS prior to collision, and is made easier by the lower emittance that is prioritised throughout the accelerator.

As of June 2022, SuperKEKB achieved a peak instantaneous luminosity $4.65 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ [62]. By comparison, the peak luminosity achieved by KEKB was $2.11 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ [66]. This SuperKEKB record surpasses the record set by the B factories and LHC experiments. Achieving this record was assisted by incorporating the *crab waist scheme* in to beam production, which was not included in the original plans for SuperKEKB. Here, additional sextupole magnets are implemented in the QCS to better align the centres of the narrowed beams. It is expected that new records in instantaneous luminosity will be set by SuperKEKB, as the goal luminosity is still a factor of 13 higher than current record [62]. Achieving this goal luminosity will require further careful tuning of beam currents and vertical beta function, as well as an upgrade to the machinery of around the IP in 2027-2028.

A drawback of the increased instantaneous luminosity is the effects of *beam background*. This can occur electrons/positrons of the beam are scattered before collision, due to their narrowed bunching or radiative losses [67]. In such cases, these particles can interact with the detector medium or gases if the beam pipe is not successfully vacuumed. Such particles can then go on to mimic the products of a successful collision. Difficulties in managing beam background events as the operating instantaneous luminosity increased resulted in the initial luminosity goal being reduced, originally being $8{\times}10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ instead of $6{\times}10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.As beam background was less prevalent in the original collider, additional shielding and collimating was added around the SuperKEKB IP, to reduce its impact in the Belle II detector.

The Belle II detector is located at the site of the electron-positron IP, such that any particle process is almost fully contained and recorded. With the production electron and positron beams with energies appropriate for exploring B-physics with the Belle II experiment, it is now natural to discuss the physics processes that come about when colliding electrons and positrons.

2.2.2. Kinematics of the $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ process in electron-positron colliders

Within the Belle II detector, electrons produced with 7 GeV of energy are collided with 4 GeV positrons, where values are relative to the *lab frame*. This operating energy is designed such that the production of $\Upsilon(4S)$ resonance is favoured. The $\Upsilon(4S)$ is a bottomonium meson with mass 10.58 GeV/c^2. It is the fourth radial excitation of the $J^{PC} = 1^{--}$ Υ family of bottomonia. From the combined energy of the leptons and the $\Upsilon(4S)$ mass, the $\Upsilon(4S)$ momentum is evaluated to be ≈ 3 GeV/c, in the direction of electron momentum. The *centre-of-momentum frame* of the e^+e^- collision is a useful frame in which to analyse the dynamics of $\Upsilon(4S)$. This frame is equivalent to the $\Upsilon(4S)$ rest frame. It is also a useful frame to analyse the products of $\Upsilon(4S)$. Such a transformation is possible due to the known collision energies of the leptons and thus known boost of $\Upsilon(4S)$ frame.

The process of $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ is presented as a Feynman diagram in Figure 2.5. Here, $\Upsilon(4S)$ exists momentarily as $b\bar{b}$ after the collision, before hadronising as a $B\overline{B}$ or B^+B^- pair. According to the Particle Data Group (PDG), the combined branching fraction of $\Upsilon(4S)$ to either B^+B^- or $B^0\overline{B}^0$ has a lower limit of 96% [4]. Thus, the vast

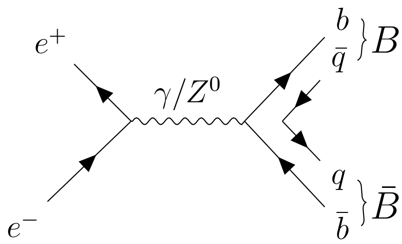


Figure 2.5. Feynman diagram of the $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ process. q can either $q = u$ or d.

majority of $\Upsilon(4S)$ decay to $B\overline{B}$ pairs. Within this thesis, the branching fraction will be approximated as 100% for $\Upsilon(4S) \to B\overline{B}$, a convention adopted within most analyses. This preference for $\Upsilon(4S)$ transitions to $B\overline{B}$ can also be motivated by the magnitude of the $\Upsilon(4S)$ mass. At only 20 MeV/c^2 above combined $B\overline{B}$ mass, the $\Upsilon(4S)$ is on threshold to produce $B\overline{B}$ with little leftover energy. The dynamics of the B mesons in the $\Upsilon(4S)$ $\to B\overline{B}$ decay has the following dynamics (where the asterisk denotes dynamics in in the $\Upsilon(4S)$ frame).

$$p^*_{e^+} + p^*_{e^-} = p^*_{\Upsilon(4S)} = p^*_B + p^*_{\overline{B}} \tag{2.1}$$
$$\implies (m_{\Upsilon(4S)}, \vec{0}) = (E_B, \vec{p^*_B}) + (E^*_{\overline{B}}, \vec{p^*_{\overline{B}}}) \tag{2.2}$$

By equating the momentum components, the following is obtained.

$$\vec{p^*_B} + \vec{p^*_{\overline{B}}} = \vec{0} \tag{2.3}$$
$$\implies \vec{p^*_B} = -\vec{p^*_{\overline{B}}} \tag{2.4}$$
$$\implies |\vec{p^*_B}| = |\vec{p^*_{\overline{B}}}| \tag{2.5}$$

Thus, the B mesons are produced back-to-back in the $\Upsilon(4S)$ frame, with equal momentum. This is a powerful constraint, in that the reconstructed B decay of interest can be cross-checked by comparison of its dynamics with its partner B. The use of the co-decaying B meson to assist analysis of the desired B decay will be explored in Chapter 3.

By equating the energy components in Eq. (2.2), the following is also obtained.

$$m_{\Upsilon(4S)} = E^*_B + E^*_{\overline{B}} \tag{2.6}$$
$$m_{\Upsilon(4S)} = \sqrt{m_B^2 + |\vec{p^*_B}|^2} + \sqrt{m_{\overline{B}}^2 + |\vec{p^*_{\overline{B}}}|^2} \tag{2.7}$$

Substituting Eq. (2.5) and the mass equivalence of the B mesons, the equation becomes:

$$m_{\Upsilon(4S)} = 2\sqrt{m_B^2 + |\vec{p}_B^{\,*}|^2} \tag{2.8}$$

$$\implies |\vec{p}_B^{\,*}| = \sqrt{\left(\frac{m_{\Upsilon(4S)}}{2}\right)^2 - m_B^2} \tag{2.9}$$

$$= 331\,\mathrm{MeV}/c \tag{2.10}$$

It is shown that B mesons have fixed momentum in the $\Upsilon(4S)$ frame, ignoring the impact of measurement smearing. This B momentum is a small relative to the B mass of $\approx$5.28 GeV/c. In the lab frame, the B momentum is a larger value, owing to the $\Upsilon(4S)$ boost. The $\Upsilon(4S)$ boost forces both B mesons to be emitted forward in this frame. As has been mentioned at the beginning of the Chapter, the larger B momentum in lab frame is integral to precision time-dependent CPV measurements to be feasible.

From Eq. (2.1), the total four-momentum of the electron-positron system is also equal to that of the $B\overline{B}$ pair system. For the $\Upsilon(4S)$ to be created with zero momentum in the e^+e^- collision frame, the collision must be head-on with equal magnitude momenta. Combining these two ideas results in the equality of the momentum magnitude and thus the energy of each electron/positron of the beam with that of the B (i.e. $E_B^* = E_{\mathrm{beam}}^*$). This provides an extra constraint on the mass of the B meson, with the following definition,

$$m_B = M_{bc} = \sqrt{E_{\mathrm{beam}}^{*2} - |\vec{p}_B^{\,*}|^2}. \tag{2.11}$$

M_{bc} denotes the *beam-constrained* calculation of the mass, where the B meson energy derived from B products is substituted for the energy of the electron beam. While the energy of the electron beam and the B meson energy are identical in principle, the B meson energy is susceptible to larger error if its products are incorrectly identified. Thus using E_{beam}^* to evaluate the B meson mass is less susceptible to systematic error. A related beam-constrained variable is ΔE, defined as the following equation.

$$\Delta E = E_B^* - E_{\mathrm{beam}}^*. \tag{2.12}$$

This gives a measure of the difference between the electron beam energy and the reconstructed B energy. For a correctly reconstructed B meson, ΔE will be closer to zero.

The properties discussed thus far highlight two of the three unique advantages of pursuing B-physics measurements at Belle II.

- *$B\overline{B}$ pair production*: B mesons are created in pairs, doubling the dataset of B meson studies. Moreover, the $B\overline{B}$ pairs are created with no additional particles in the $\Upsilon(4S) \to B\overline{B}$ process.
- *Fixed collision energy*: The energy of the e^+e^- collision is known, allowing energy, momentum and charge conservation constraints to improve reconstruction quality.
- *Hermetic detector*: The Belle II detector near-completely encloses the e^+e^- collision, ensuring almost all particles of the event are detectable.

The latter point will be expanded on in Section 2.3. The discussed unique dynamics and the three reasons above are considered some of the key conditions of B-factory experiments.

2.2.3. When e^+e^- doesn't produce $\Upsilon(4S)$: continuum events

It is generally safe to assume that $\Upsilon(4S)$ always decays to $B\overline{B}$. It is however not safe to assume that the e^+e^- collisions at the operating energy of $\sqrt{s}$ =10.58 GeV will always produce the $\Upsilon(4S)$ state. Despite the tuning of the collision energy to favour $\Upsilon(4S)$ production, there is still significant probability of other processes occurring. These are generally of the form $e^+e^- \to f\bar{f}$, where f indicates any SM fermion. This excludes the t quark, as there is insufficient energy to pair-produce this 173 GeV/c^2 particle. Other processes possible are $e^+e^- \to \gamma\gamma$, and four light lepton processes like $e^+e^- \to e^+e^-\mu^+\mu^-$. Key examples are given in Table 2.1.

Cross-sections quantify the probability of a particle interaction taking place, in units of barn (denoted by 'b'). From Table 2.1, it can be seen there are number of processes more probable than $e^+e^- \to \Upsilon(4S) \to B\overline{B}$. Belle II also positions itself more generally as a *flavour physics factory*, due to the high cross-sections of $b\bar{b}$, $c\bar{c}$ and $\tau^+\tau^-$ events.

In B physics studies, the non-$b\bar{b}$ modes are background events, as it is impossible to create B mesons in such an event. While there are no true B mesons, non-$b\bar{b}$ modes may mimic the presence of B mesons by producing particles that are common in real B-decays. For example, D mesons are a common product of B decays. The reason for this can be see in the CKM matrix magnitudes of Eq. (1.1.2), where $|V_{cb}|$ is the second-largest

Table 2.1. $e^+e^- \to f\bar{f}$ processes at $\sqrt{s}$=10.58 GeV at Belle II. Note that the cross-section of the $e^+e^- \to e^+e^-$ process has been calculated for restricted geometry and kinematics.

Process	Cross-section [nb]	No. of events in 100 fb^{-1}
$b\bar{b}$ (i.e. $\Upsilon(4S)$)	1.110	1.11×10^8
$u\overline{u}$	1.61	1.61×10^8
$d\overline{d}$	0.40	4.00×10^7
$s\overline{s}$	0.38	3.80×10^7
$c\overline{c}$	1.30	1.30×10^8
$\tau^+ \tau^-$	0.919	9.19×10^7
$\mu^+ \mu^-$	1.15	1.15×10^8
e^+e^-	300	3.00×10^{10}
$\nu\bar{\nu}$	0.25×10^{-3}	2.50×10^4

amplitude concerning b-quark transitions. D mesons can also be created in $c\bar{c}$ events, made worse for B physics analyses as $e^+e^- \to c\bar{c}$ has a larger cross-section than $e^+e^- \to \Upsilon(4S)$. Implementing criteria to reject the particles of non-$B\overline{B}$ events from the analysis will be a large focus in both Chapter 3 and Chapter 4.

The $e^+e^- \to c\bar{c}$ events, alongside the other $q\bar{q}$ background events, are referred to as *continuum events*. Their dynamics differ from $B\overline{B}$ processes, as they will hadronise as other particles. As mentioned above, $e^+e^- \to c\bar{c}$ can produce $D\overline{D}$ pairs, like D^+D^- or $D^0\overline{D}^0$. The $D\bar{D}$ pairs are created in collision energy of $\sqrt{s}$ =10.58 GeV, which is well above $D\bar{D}$ production threshold. Each D meson has a mass of 1.86 GeV/c^2, leaving left over energy after that which is used in $D\bar{D}$ masses. Through a derivation similar to that of the B meson momentum earlier, D mesons will also be produced back-to-back in the collision frame, though with much larger momentum due to the excess energy. From Eq. (2.9) (where $m_{\Upsilon(4S)}$ is now technically $E^*_{e+} + E^*_{e-}$, though they are the same total value), their momentum is $\approx$4 GeV. This momentum will be even larger for products $e^+e^- \to q\bar{q}$ with lighter quarks.

A key difference between continuum and $B\overline{B}$ processes is the event shape. Event shape refers to the geometry of the products of the event. As momentum must be conserved, resultant particles of the high momentum continuum events are highly collimated in their direction of travel, best seen in the centre-of-momentum frame of the collision. $B\overline{B}$ pairs are created with sub- GeV/c momenta, resulting in less direction bias of momenta. Products of B events are more spherical in geometry. A number of quantities have

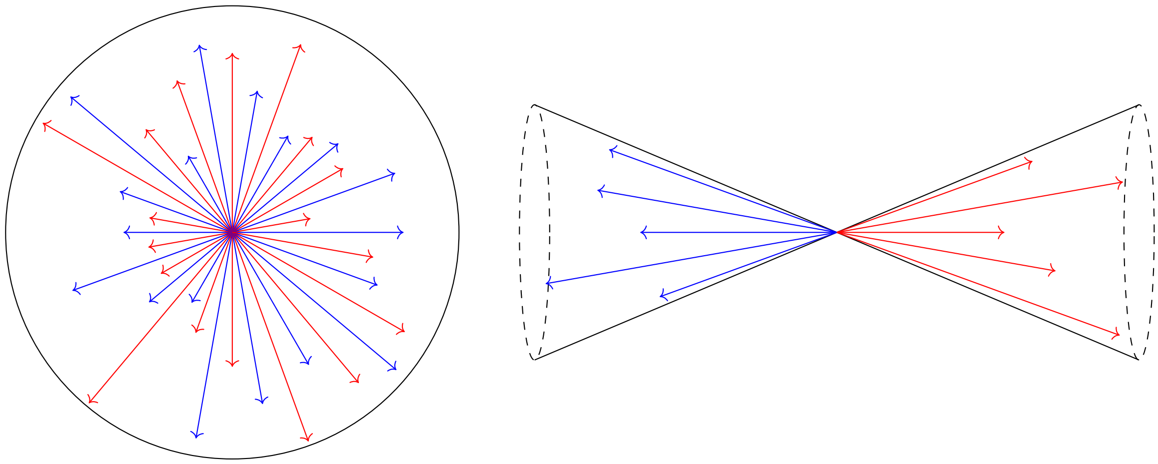


Figure 2.6. Schematic comparison of event-shapes in $B\overline{B}$ (left) versus $q\overline{q}$ (right) events at $\sqrt{s}$ =10.58 GeV. Particle products are highly collimated and directional for continuum events in the collision frame, whereas $B\overline{B}$ events overlap and are spherical.

been developed to measure how jet-like or spherical the event shape is, including the Fox-Wolfram moments [68], which will be explored in Section 4.1.5 The difference in geometry of the events analysed will be a key discriminator between identifying the signal B decay and background continuum events.

For B-physics searches, $q\overline{q}$ events are undesired. On the other hand, within charm physics the $c\overline{c}$ events become signal events. For example, in the measurement of the D_s^+ lifetime at Belle II $e^+e^- \to c\overline{c}$ events are the main source of the desired meson. However such searches can be made difficult due to fragmentation in $c\overline{c}$. As the centre-of-momentum energy is mismatched to pair-production threshold, occasionally the extra energy becomes additional hadrons, meaning the same constraints on the $B\overline{B}$ system cannot be assumed as readily. The BESIII experiment [69], a charm-physics experiment located in Beijing, avoids this by varying the energy of e^+e^- collisions to a range of charmonium ($c\overline{c}$) resonances, to favour charm hadron pairs.

Table 2.1 also contains leptonic processes. Due to their masses, the products of lepton events have large momenta, with highly directional event shape. As muons and electrons do not hadronise, they are rarely backgrounds to B-physics searches, though the electron scattering can be used to derive the number of B decays recorded (i.e. the *integrated luminosity*) [59]. As tau leptons can decay to hadrons, they can mimic products of B decays, though their distinct event shape make such processes simple to reject using Fox-Wolfram moments in B analyses. Conversely for dedicated tau studies, the opposite

Table 2.2. Υ family of spin-1, S-wave bottomonium mesons. The B meson production is only possible at (and after) energy equivalent to $\Upsilon(4S)$, while strange and excited B production is only at (and after) $\Upsilon(5S)$.

Υ	Mass [MeV/c^2]
$\Upsilon(1S)$	9460
$\Upsilon(2S)$	10023
$\Upsilon(3S)$	10355
$\Upsilon(4S)$	10850
$\Upsilon(5S)$	10860
$\Upsilon(6S)$	11020

selection on Fox-Wolfram moments may be used to reject the spherical B events and prioritise collimated tau ones.

2.2.4. When e^+e^- doesn't produce $\Upsilon(4S)$: other operating energies

Most of the discussion so far is of e^+e^- collisions at the $\Upsilon(4S)$ resonance. The original Belle experiment varied the energy of collisions to investigate the physics at other Υ resonances, for $\Upsilon(nS)$ for n=1-6. Below the $\Upsilon(4S)$ resonance, no B mesons can be created. However, the physics of bottomonium decays could be investigated as transitions from the Υ families. The BaBar experiment used data measured at $\Upsilon(2S)$ and $\Upsilon(3S)$ to make the first measurements of the spin-0 $\eta_b(1S)$ meson mass in $\Upsilon(1S, 2S) \rightarrow \eta_b(1S)\gamma$, which Belle later reproduced [70]. There are no immediate plans for Belle II to also take data at lower Υ resonances.

The $\Upsilon(5S)$ resonance is not only above threshold to produce up-type and down-type $B\overline{B}$ pairs, but also above threshold to produce strange-B pairs ($B_s^0\overline{B}_s^0$) and excited B pairs ($B^*\overline{B}^*$). Unlike the $\Upsilon(4S) \rightarrow B\overline{B}$ process, these new B pairs cover a much smaller fraction of $\Upsilon(5S)$ decays, with 20% of $\Upsilon(5S)$ decaying to either strange B or excited strange B pairs [4]. The original Belle experiment had a comprehensive program at the $\Upsilon(5S)$ resonance, recording 121.4 fb^{-1} of collisions at this energy. Studies included measuring CP violation parameters in neutral and charged B mesons from $\Upsilon(5S)$, B_s decays to light meson pairs, as well as decays to non-$q\overline{q}$ meson candidates [71] [72] [73].

Performing a dedicated $\Upsilon(5S)$ run is likely, once sufficient $\Upsilon(4S)$ data has been collected in the coming years.

While Belle II has not yet performed a dedicated $\Upsilon(5S)$ run, they did perform a scan where data-taking occurred at multiple energies near the $\Upsilon(5S)$ resonance. The 2021 scan at $\sqrt{s}$ =10.70, 10.75 and 10.81 resulted in a potential new $\Upsilon(10573)$ resonance, between that of $\Upsilon(4S)$ and $\Upsilon(5S)$ [74]. The term *off-resonance* is reserved for collisions very close to traditional resonances. Belle II has performed a number of off-resonance data-taking periods at an energy of $\sqrt{s}$ = 10.53 GeV, just below that of the $\Upsilon(4S)$ resonance. The goal of this is to probe conditions of energy near enough to $\Upsilon(4S)$ to be able to extrapolate the results to $\sqrt{s}$=10.58 GeV, but in absence of $B\overline{B}$ pairs. In this way, $e^+e^- \to q\overline{q}$ continuum events can be studied. Particularly, off-resonant data is used to improve simulations of continuum events, whose fragmentation processes can be difficult to model from first principles alone. Additional operating beam energies include scans between $\Upsilon(3S)$ and $\Upsilon(4S)$ energies, between the $\Upsilon(4S)$ and $\Upsilon(5S)$ energies, off-resonant $\Upsilon(5S)$ and at the $\Upsilon(5S)$ resonance. Moreover, the original Belle program also recorded at a multiple energies between the $\Upsilon(5S)$ and $\Upsilon(6S)$ as well as at the $\Upsilon(6S)$ resonance.

2.3. The Belle II Detector

The Belle II detector is centered at the interaction point and parallel to the electron-positron beams produced by SuperKEKB so as to record the products of the collision. It's cylindrical shape wraps around the collision event, allowing resultant particle processes that radially propagate to be measured. Additional front and end cap detectors ensure that all particles of the event are near-completely contained.

As a spectrometer, B-physics studies at Belle II does not measure B mesons directly. With lifetimes $\mathcal{O}(10^{-12}\text{s})$, both B^+ and B^0 mesons decay very quickly via one of hundreds of possible modes, each containing a variety of intermediate particles. Such particles will then go on to have their own additional decay processes. Instead of being directly detected, the properties of B mesons are instead inferred. This is done by studying the particles at the end of the decay chain, using the Belle II detector.

The detector consists of multiple *subdetectors* that are each designed to identify a finite set of particles from B decays. This finite set are particles most likely to be present at the end of B-decay chains and are unlikely to decay within the dimensions of the detector. These particles, are referred to as *final state particles* (FSPs). The particles considered in the final set are listed in Table 2.3.

This section will first describe the function of each of the Belle II subdetectors and broadly how they measure particle observables. How specific FSPs are identified and reconstructed will be discussed afterwards.

2.3.1. The subdetectors of the Belle II detector

The detection of particles originating from e^+e^- collisions takes place within a number of subdetectors at Belle II. Each subdetector measures particle observables, so that the kinematic information at the point of production can be derived, as well as to contribute to an overall particle identity hypothesis. The interaction region, located between the two QCS systems, is separated from the subdetectors by a cylindrical beryllium beam pipe held at vacuum. The inside of the pipe begins 1cm radially from the IP with a thickness of 2mm. As was presented in the discussion about the nano-beam collision scheme, the Belle II experiment is subject to a large number of beam-induced background processes. Thus the primary role of the beam pipe is to provide shielding to the subdetectors from beam backgrounds.

The subdetectors will now be described from closest to the beampipe radially, to the furthest. This also includes those that extend in both the barrel and endcaps, as well as the one subdetector that is solely located in endcaps.

1. *Pixel Detector (PXD)*: The first subdetector experienced by particles as they leave the IP is the PXD. The PXD consists of two layers of silicon pixel components, designed with 8 'ladder' segments in the first layer and 12 in the second. For the data-taking period used in this analysis, the PXD was only partially installed. There were only two of 12 ladders used in the second layer, as construction was not complete by the time collisions were schedule to begin. The full PXD has since been installed, as of second physics run period in early 2024. The ladders are laid in an overlapping 'windmill' pattern, cylindrically around the beampipe. The two layers are 14 and 22 mm from the IP. Each pixel component is a DEPFET (depleted p-channel field-effect transistor), with the smallest being $50{\times}50$ μm^2. As transistors, ionising particles from the IP can induce an electric field in the depleted silicon material that allows current flow proportional to the number of incident particles. This constitutes a signal in the pixel, where the location of the activated pixel gives

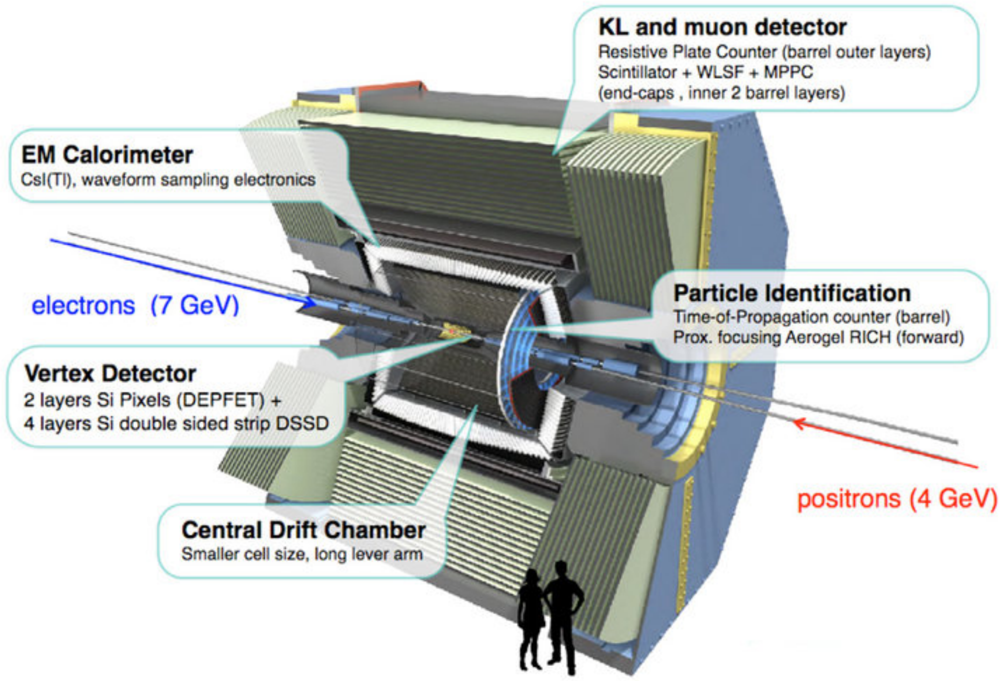


Figure 2.7. The Belle II detector, with subdetectors labelled. Image taken from Ref. [75].

a 2D measurement of the particle's position. Obtaining the positions of the incident particles with subdetectors is referred to as *vertexing*. After combining vertices from the first and second PXD layers, the trajectory of the particle can begin to be traced. The PXD was added specifically for the Belle II experiment upgrade, to realise higher precision vertexing with the expected increased instantaneous luminosity, and therefore beam-induced backgrounds. It contains 10 million readout channels, the most readout channels of any subdetector at Belle II.

2. *Silicon Vertex Detector (SVD)*: The SVD is a secondary vertexing system located between the PXD and 140 mm from interaction point. Like the PXD, sensors are placed on ladders. Instead of small pixels, the largest silicon strip sensors used measure 12cm long, 6 cm wide and 300 μm thick. As silicon is a semiconducting material, incident ionising particles on the strip sensors create current flow which records the presence of a particle. There are four layers of ladders, overlapping cylindrically around the PXD in the 'windmill' arrangement. Whereas pixel detectors can provide precise 2D positioning of incident ionising particles, individual silicon strips provide 1D measurements. To remedy this, silicon strips are placed on either side of the ladder. On the underside, the length of the strips are directed parallel to the beamline along the ladder, while the lengths of the outer set are

laid perpendicular to the beamline. Thus, particles that pass through SVD layers provide two separate 1D measurements, which when combined, offer comparable 2D vertex precision to the pixel layer. The outer SVD ladders contain a bend towards the beamline in the forward direction. This is to increase the forward angular coverage of particles. The SVD is thus able to measure any particle originating from the IP travelling with polar angle between 17° and 150°. This asymmetric design accounts for the lab-frame boost that products of $\Upsilon(4S)$ are created with, such that particle products are preferentially forward emitted. Together, the PXD and SVD can be referred to as the *Vertex Detector* (VXD).

3. *Central Drift Chamber (CDC)*: The CDC is barrel shaped and radially extends between 16cm and 113cm from the IP. The CDC contains 14336 wires running parallel to the beam line, arranged in 56 layers within in a chamber of helium-ethane gas. Charged particles can ionise gas molecule as the particle travels through the chamber. This ionisation frees bound electrons, which can also ionise further gas molecules to produce an electron cascade. Current carrying wires are placed such that the electric field causes the cascade to 'drift' towards nearby sensing wires. Electrons from the cascade induce a current in the sense wires, which signals the presence of the incident particle. As the cascade will travel towards the nearest sense wire, the trajectory of the incident particle will be known to be within a fixed radius of the participating wire, somewhere along its length. To obtain position information about how far along the wire the incident particle's trajectory intersected, groups of sense wires alternate between three alignments: axially (i.e. parallel to the beam line), stereo aligned (i.e. with slight positive skew angle to the axis) and opposite stereo alignment (i.e. slight negative skew angle to the axis). With signals from multiple sense wires at multiple angles, the three-dimensional trajectory within the CDC can be traced. Combining this trajectory with vertexing information from the VXD, the full trajectory or *track* of the charged particle can be produced. Due to the magnetic field that is directed parallel to the beamline, the tracks of charged particles in the CDC begin to curve. This can provide particle identification information about the sign of the charge from the resulting Lorentz force- curved path. Crucially, the CDC also provides information about the energy loss of particles due to gas ionisation. Different FSPs lose energy at a different rates as they travel through a medium, according to the ratio of their momentum to their mass. Both tracking information and characteristic energy loss provide vital information on the particle identity and momentum.

4. *Time of Propagation counter (TOP)*: The TOP counter is another novel subdetector commissioned especially for the Belle II experiment. Instead of providing vertexing information like the previous subdetectors, the TOP is implemented to improve particle identification. It primarily functions to discriminate pion-based versus kaon-based tracks. The TOP counter is made of 16 fused silica bars placed 1.2m from the IP, azimuthally around the barrel of previous subdetectors. The silica acts as a Cherenkov radiator for the charged particles that pass through the medium. Silica has a high enough refractive index such that Cherenkov radiation induced by incident charged particles is totally internally reflected. The light is continuously reflected within the bars and guided towards electronics where it can be recorded. The time it takes for the light to propagate (hence the name *time of propagation* counter) varies with the Cherenkov angle, which itself varies with the charged particle's velocity. Thus, the time of propagation of the Cherenkov light becomes a characteristic quantity. The velocity of the particle, combined with the momentum-mass ratio derived from the energy losses in the CDC, allows the mass and thus particle identity to be derived. The more traditional measurement used in particle detectors, the *time of flight* (i.e. travel time from IP to subdetector) is reclaimed by using the Cherenkov travel time and the monitored time of collision.

5. *Aerogel Ring Imaging Cherenkov detector (ARICH)*: The TOP counter requires precise timing based measurements to perform particle identification. The ARICH instead requires precise location based measurements, with more traditional geometric-based particle identification in emitted Cherenkov radiation. The ARICH is also the only subdetector be located exclusively in the forward endcap of the barrel. It's angular acceptance is relatively small, recording particles with polar angle between 14° and 30°. This complements the TOP counter acceptance, which is between 31° to 128°. The ARICH begins 1.67 m in front of the IP. It consists of 248 aerogel tiles, each measuring 17cm by 17cm. The aerogel acts as the Cherenkov radiator for incident charged particles where similar to the silica bars in the TOP, the Cherenkov photons are emitted at an angle proportional to the charged particles's velocity. However in the ARICH, instead of being totally internally reflected as in the TOP, the Cherenkov radiation transmits through the aerogel and allowed to propagate before reaching a plane of photodetector electronics. From the geometry of the structure, the detected Cherenkov light can be traced back to the responsible aerogel tile to then derive the Cherenkov angle (i.e. angle of emitted radiation relative to the continued parent track). For two tracks with the same momentum, a heavier particle (such as charged kaons relative to charged pions) will have a lower

velocity, and thus smaller Cherenkov angle. This provides complementary particle identification information to assist the tracking based subdetectors. The ARICH is only in the forward direction due to the $\Upsilon(4S)$ boost, leading to very few particles emitted towards the backwards direction.

6. *Electromagnetic Calorimeter (ECL)*: The ECL is responsible for recording the energies of particles which can interact electromagnetically. The subdetector barrel extends from 1.25cm to 1.62cm radially away from the beam line. It also has units in the endcaps, 1.02m behind the IP and 1.96m in front. The ECL is constructed from over 8000 caesium iodide crystals. Incident particles such as electrons and photons interacting with the material leads to scintillation of the crystal. The scintillated light can be converted to electrical signal via photodetectors, where the amount of light is proportional to the amount of energy deposited by the particle. This can also provide locational information, such as if the particle was emitted forward or backwards, based on which crystals are scintillated. The amount of energy deposited, like in the CDC, is a characteristic of the particle identity. This means a relationship can be made between the amount of energy deposited and the total energy of the particle can be created for different particle species. This will be discussed further when discussing clustering and its role in particle identification in Section 2.3.2.

7. *K_L("K-long") and Muon detector (KLM)*: The KLM is the outermost subdetector of the Belle II detector. The solenoid that provides the magnetic field that curves charged tracks in the CDC is located between the KLM and ECL. The KLM barrel contains scintillator strips with steel plates separating them, where the steel plates provide the flux return for the magnetic field. In the outer KLM layers, resistive plate chambers (RPC) are used. In these units, incident particles ionize the gas in the chamber, similarly to in CDC. These ionized particles are the detected by electrodes in the plate, signalling an incident particle. As per the name, the KLM is well suited to muon and K_L meson identification. As an outer detector, only the longest lived particles survive this distance. To determine if a particle is a muon, tracking from the CDC is extrapolated to determine if it could have reached compatible signal in the RPC layers. K_L mesons are neutral and cannot be associated with a track. They will generally shower hadronically within the steel plates. Clustering algorithms are also applied within the KLM to group signals within the scintillators and RPC's. Thus K_L meson identification can be performed

using clusters within ECL and the KLM, both of which must not be associated to a track.

2.3.2. Reconstruction and Particle Identification

The subdetectors of Belle II are used in conjunction to form a hypothesis about the particle identity, based on the signal it leaves in each volume. These signals must be combined so as to reconstruct the particle that gave rise to them, which in turn are used to reconstruct the presence of B mesons, or any other particle being investigated. Up until this point, the term *reconstruction* has been used loosely, but it is now defined properly in the context of this thesis.

The first kind of reconstruction describes the grouping of multiple signals in the Belle II subdetectors with the hypothesis they were created by the presence of a single FSP. In such a case, the term signal is used to mean a hardware response, and not the signal mode of this analysis, $B^+ \to \mu^+\nu_\mu$. There are two key groupings: *tracks* and *clusters*. Tracks and clusters are pre-particle objects, with their own energy and dynamics. The origins of each will be discussed in this section.

The second kind of reconstruction, which will be used more often than the first in this thesis, relates to particle decay reconstruction. This is where FSPs are combined with the hypothesis that they originated from the decay of the same particle. Specifically, the four-momenta of the FSPs are summed to evaluate the invariant mass. This can then be evaluated against the nominal mass of the desired parent identity. The invariant mass is not the only variable that can be used to evaluate reconstruction quality. For B mesons, additional constraints such as M_{bc}, ΔE and $\cos\theta_{BY}$ can be used, as were implemented in previous $B^+ \to \mu^+\nu_\mu$ searches in Section 1.4.

For an analyst to perform particle decay based reconstruction of events taking place within the Belle II experiment, the FSP reconstruction must happen first. The method of creating the FSP from detector signals depends mainly on the particle's electromagnetic charge.

Table 2.3. Final State Particles of Belle II.

Particle	Symbol	Primary reconstruction objects
electron	$e^\pm$	clusters (ECL)
muon	$\mu^\pm$	tracks, clusters (ECL, KLM),
pion	$\pi^\pm$	tracks, clusters (ECL)
kaon	$K^\pm$	tracks, clusters (ECL)
kaon	$K^0/\bar{K}^0$ (as K_S)	V^0's, clusters (ECL)
kaon	$K^0/\bar{K}^0$ (as K_L)	clusters(ECL, KLM)
proton	$p/\bar{p}$	tracks, clusters (ECL)
neutron	$n/\bar{n}$	clusters(ECL,KLM)
deuteron	$d/\bar{d}$	tracks, clusters (ECL)
photon	γ	V^0s, clusters(ECL)
lambda	Λ^0	V^0's

Energy measurement, clustering and neutral particles

The energy of FSPs are generally derived from energy deposits in the electromagnetic calorimeter. When an ionising particle is incident with the ECL, the energy deposited can induce an electromagnetic shower in the medium. Hadronic FSPs may also interact with nuclei of the ECL medium, producing a hadronic shower whose ionising products can also go on to deposit their energy. Showers of either kind results in scintillation of multiple ECL crystals, leading to multiple signals. Algorithms are used to identify the groups of participating crystals and then to locate the coordinates of its centre i.e. where the shower began, at the point of incidence with particle. This group of energy deposits is called an *electromagnetic calorimeter cluster*, or simply a cluster. Clustering can be challenging if there are multiple incident particles and thus multiple showers in a similar region.

To be able to infer the energy of a particle from its cluster, two stages of calibration are required [76]. The first is to relate the electric signal response to energy deposited, according to the location of the crystal in the detector. The second is to relate the energy deposited to the total energy of the particle, as these quantities may not be equal. Particles such as electrons and photons are fully stopped by the ECL and deposit most of their energy, whereas muons are minimally ionising and leave the ECL with a small amount of energy deposited. This can create different shower shapes, which can provide additional information about the particle identity. For neutral FSPs (i.e. neutral kaons, neutrons and photons), the ECL is the first subdetector that they can leave a signal in, if they have not decayed already to the track-based particles described in the next paragraph. Specifically, K_L^0 mesons can either initiate hadronic showers in the ECL, or pass through without interacting. With both K_L^0 mesons and muons capable of leaving the ECL with minimal (or potentially none for K_L^0 case) signal, they are better identified by their more prominent energy deposits in the KLM. Clustering of deposited energies in the KLM is also used to identify takes particle identity.

As described when introducing the KLM, both ECL and KLM clusters can be combined with tracks from previous subdetectors to obtain both the energy and momentum of a particle. An ECL cluster without a track is deemed the result of a photon, while unmatched KLM clusters are treated as K_L^0 mesons.

Tracks and charged particles

Charged FSPs are identified primarily from their tracks, reconstructed from detector signals left in the PXD, SVD and CDC. Creating tracks occurs in three stages. The first is in the clustering of the signals in each of the subdetectors, also called 'hits'. The second concerns the track finding of the particle, where patterns of clusters that constitute the particle trajectories are further grouped to be from a single track. The final step is the track fitting. This refers to evaluation of the parameters that best define a helical trajectory on the track, such as the curvature or the point of closest approach to the IP. There are six charged FSPs that are expected to leave tracks in the Belle II detectors: electrons, muons, charged pions, charged kaons, protons, and rarely, deuterons. The fitting of the parameters is performed for each of the six possible FSP identities, as the differing masses of these particles impact the allowed track parameters of the helix. There then is one set of helix track parameters for each particle identity that could have led to the track. These parameters can then be used to obtain the momentum and the origin of the particle. The quality of this fit can then be used to determine which particle hypothesis best fits the observed track.

Charged FSPs within the Belle II detector can be assigned a particle identification probability or *particle ID* (PID) [67]. PID probabilities are ratios of particle likelihood profiles and measure how probable one particle hypothesis is over other ones. Each subdetector contributes its own likelihood profile for each particle type. These are based on probability density functions (PDF's) of the expected responses for each charged particle hypothesis. If possible, PDF's are derived analytically. Otherwise they are experimentally defined from detector response of particle decays with well constrained particle identity. PDF's may also be derived from simulated detector conditions. In this way, measurements made by the subdetectors of an unknown particle can be compared with the characteristic likelihood profile of the six possible charged FSPs. The particle ID captures how likely the unknown particle matches the given particle hypothesis, according the quantity measured. There is no analogue defined for neutral particles, though the attributes of ECL cluster can serve a similar purpose.

For analysts, particle ID is implemented in a selection variable, such that the four-momenta of the charged particles used have a minimum probability of originating from a different particle. As this analysis will rely on a muon to identify the $B^+ \to \mu^+\nu_\mu$ decay, the `muonID` variable will be important for ensuring the tracks chosen are from real muons. In particular this variable is based on a number of measurements from various subdetectors used to build the muon likelihood profile. This includes muon velocity derived via specific ionisation energy loss in the CDC, timing information from muon Cherenkov radiation in the TOP, energy deposits in the ECL, as well as the penetration depth and characteristic scattering of the muon within the KLM [77]. The particle velocity would usually also be derived using signals left by muons in the SVD However, for the set of data used within this analysis, Belle II recommends the use of the particle ID derived without the contributions of the SVD for muons. This is due to temporary issues with the construction of the SVD likelihood profiles that have since been resolved. Any references to the name of the variable for muon identification will be formatted as `muonID`, with the understanding that what is actually being implemented is the version without the SVD, `muonID_noSVD`.

V_0's: neutral particles from tracks

V_0 is the name given to the class of neutral particles which decay to two opposite-charged final state particles. V_0 are technically not FSPs, in that they are not created from a single track or single electromagnetic cluster. They can however, be considered alongside the other 'pre-particle' objects as the same algorithm used to identify single tracks also establishes if there are pairs of oppositely-charged tracks that share a vertex. Historically, V_0 particles were named due to their characteristic 'V' shaped tracks created from their decays in bubble chambers. The V_0's commonly seen in the Belle II experiment are K_S^0 mesons and Λ baryons, with decays $K_S^0 \to \pi^+ \pi^-$ and $\Lambda \to p^+\pi^-$ respectively. Such hadrons can decay outside the interaction region, resulting in the origin of their products appearing displaced from the IP. Thus they are also referred to as *displaced vertices*. Photons can also be identified from V^0's, if they are absorbed within matter and as a result produce an electron-positron pair. The K_S^0 mesons are common products of D mesons, which in turn are common products of B mesons. Thus effective V_0 identification will become integral to the reconstruction of the full $\Upsilon(4S) \to B\overline{B}$ event that contains $B^+ \to \mu^+\nu_\mu$ decay modes in this thesis.

2.4. From detections to data

With the method of creating and detecting B decays within the Belle II experiment presented, the next phase to consider is how the detector responses become analysable data. First to be reviewed is the data processing strategy, beginning from the Belle II detector and ending with a dataset in which a particle search can be performed. Next, the role of Monte Carlo simulated data in creating a particle search will be discussed. The software used by Belle II will be introduced, including how it can be used to provide lists of reconstructed particle attributes and events. Finally, the data and Monte Carlo samples used for the $B^+ \to \mu^+\nu_\mu$ analysis will be specified.

2.4.1. Data processing and computing

The output of the Belle II detector is a series of signals from subdetectors indicating the presence of final-state particles. To investigate the physics of the event, analysts require quantities such as momenta and energies of the final-state particles that are present. The Belle II Data Production group is in charge of transforming the raw subdetector information into physics observables. The pipeline from recording detector output to providing partially reconstructed files of particle information will now be discussed.

Data acquisition and triggers

A *run* refers to when collisions are turned on and data-taking occurs. This can last anywhere from minutes to hours, and multiple runs can occur each day. The collision energy can be varied between sets of runs. Multiple runs constitute an *experiment*. The set of runs in the experiment may cover multiple beam energies and periods.

As SuperKEKB aims to achieve a large instantaneous luminosity, the Belle II detector must be capable of rapidly recording data. However, the amount of diskspace and bandwidth required to store every event is unfeasible. Moreover, not all events will be of interest; from the cross-sections in Table 2.1, it is highly likely that the e^+e^- collision that takes place will produce an $e^+e^- \to e^+e^-$ event, which is of little interest to the B analyses. Thus it is necessary to implement *triggers*, a system that decides if an event of interest has taken place. The full subdetector response is saved based on the trigger systems response, thus effectively filtering out background e^+e^- processes.

The first trigger is referred to as the *Level 1 Trigger*(L1). The L1 trigger is a series of field-programmable gate arrays (a kind of integrated circuits) that execute logic to determine if the received signal sample matches the conditions for an 'interesting' event. This is based on raw signal received from key subdetectors, namely the CDC, ECL and KLM. If successful, the L1 trigger commands all subdetectors to record their signal. The L1 trigger is limited to 30 kHz, i.e. no more than 30000 events per second can be recorded.

A secondary trigger, the *high-level trigger* (HLT) is software-based, as opposed to the hardware-based L1. It consists of 10000 CPU cores. Additionally, while the L1 trigger considers individual subdetectors, the HLT response is based on reconstructed quantities from the full event, combining the signals from all subdetector signal. This includes combining raw subdetector information to produce clusters and track-fitting, to then determine the kinds of particles in the event and further filter out 'uninteresting' events for analysts. This reduces the number of stored events to 10 kHz (10000 events per second). Based on the preliminary reconstruction, the HLT assigns a label to events according to the particles in the event. Examples include if the event contains a hadron, dimuon, tau-tau, digamma etc. The data used in the $B^+ \to \mu^+\nu_\mu$ analysis will use events that are labelled hadron by the HLT. This is because it is desired that the B mesons that decays alongside $B^+ \to \mu^+\nu_\mu$ will contain hadrons. The choice of the specific partner B meson decay mode used will be discussed in Section 3.4. By choosing events that have been stored with the hadron label, a large number of events that cannot contain a true B meson, such as $e^+e^- \to \gamma\gamma$, are immediately rejected from the analysis without additional work. The rejection of events least likely to match a decay hypothesis is called *skimming*. The HLT is a trigger skim, while analysis-based skims will be described soon. All events that survive any of the categories of the HLT are then stored for offline analysis.

Overall, the trigger logic that signals the recording of an event prioritises high efficiency for hadronic products of $\Upsilon(4S) \to B\overline{B}$ and e^+e^- to continuum events. This is well-suited to the $B^+ \to \mu^+\nu_\mu$ search of this thesis, where the alternate B has been allowed to decay to a mix of lepton and hadrons. Thus, it is highly likely that if such a $\Upsilon(4S) \to B^-$ $[B^+ \to \mu^+\nu_\mu\ ]$ (and charged conjugated) event is produced, the trigger system allows it to be recorded. The efficiency of the trigger system in recording such an event is thus not considered to dramatically impact this analysis. Considering trigger efficiency becomes more important to analyses in the low multiplicity physics working group, which often require pure leptonic events that leave minimal signal in the detector, thus could be incorrectly rejected as e^+e^- background processes.

Calibration

The mathematical relationship between incident particles observables, their interactions with subdetectors and the resulting raw subdetector response is not always theoretically known. Thus, data-driven studies must be performed to establish how detector voltages map to quantities such as energy deposits. Such studies produce factors to transform detector response into the desired quantity. The application of these factors is referred to as data calibration, which is necessary to allow analysts to interface with more familiar particle observables, like energy and momentum. This is the stage that previously discussed tracking and clustering algorithms are performed, as well as the determination of particle ID variables.

An example of such a calibration in ECL crystals to photon events has already been discussed in Section 2.3.2. Ref. [76] describes how the response of ECL crystals are calibrated using a sample of $e^+e^- \to \gamma\gamma$ collision events, where the response is matched to the energy deposited in the crystal. Specifically, the energy of each crystal is predicted by simulating $e^+e^- \to \gamma\gamma$ interactions with the ECL, which is then compared to the electronic signal of the crystal from data, so as to establish the calibration factor. Once the energy deposited in the crystal has been calibrated, the total energies recorded (whether one or multiple crystals are scintillated) can then be related to the expected energy of the incident photon in simulation.

Such a calibration relies on accurate knowledge of the detector at the time of data-taking. An instance where the magnetic field within the Belle II detector was not accurately modelled, leading to incorrectly calibrated momenta, will be discussed in Section 5.2.1. Calibration also requires the monitoring of key variables during data-taking.

For example, the mass of the neutral pion reconstructed in the decay of $\pi^0 \to \gamma\gamma$ can be used as a check of photon energy in events. If the peak of the pion mass distribution deviates from its expected value ($135\,\mathrm{MeV}/c^2$), this may be an issue with the modelling of the photon energy or in the detector hardware.

Calibration is performed in stages. The *prompt* processing takes place in the days after the data is recorded, once a certain threshold of events is met [78]. The prompt processing is performed on groups of runs, called *buckets*. After a number of buckets is collected, a reprocessing (shortened to *proc*) is performed on all runs, including both promptly calibrated buckets and any previously reprocessed data. The reprocessing calibration is performed again with the larger set of data, allowing for more precise calibration studies to be performed and applied to the data.

Skims

Skimming has already been mentioned in the role that the HLT plays in selecting collision events with interesting physics to the Belle II experiment. More specific skimming is also performed on the basis of the calibrated observables, so as to reduce the size of the analysed dataset. As was discussed in the introduction of the Belle II physics program in Section 2.1.2, many categories of physics searches are possible using Belle II data. What is considered signal events by one analysis may be background events to another. Thus performing simple selections to reject the events least likely to contain the desired signal mode ensures that data analysis is manageable. Skims may also partially reconstruct the event, further rejecting events which cannot provide the candidate required before optimising the analysis further. This will become especially important as Belle II moves closer to the goal of recording $50\,\mathrm{ab}^{-1}$ of e^+e^- collisions, a factor of more than a 100 times what is currently available.

Skims of data are provided by the Belle II Data Production group. From the same collision dataset, multiple skimmed samples can be produced, according to the analysis they are required for. In the $B^+ \to \mu^+\nu_\mu$ search, the skim to be used is the Full Event Interpretation (FEI), a B-tagging algorithm to be discussed further in Chapter 3. In essence, this skim reconstructs a B meson in a set of predetermined decay modes, on the basis of the kinematic qualities of FSPs. In this way, a portion of the event is already assigned to a B_{tag}, such that the signal B-decay can be reconstructed from a condensed set of tracks and clusters. Using such a skim is useful as it ensures that considered events meet minimum conditions that mostly correlated with containing a B_{tag}. Such a skim is not perfect - as will be seen in Chapter 3, there is still retention of non-$B\overline{B}$ events. However, they do reduce the efforts required to construct such an event, so that an analyst can focus on optimising their B decay of choice, rather than determining the criteria to build a quality B_{tag}. Overall, the goal of analysis skims data selections are that at most 20% of events remain, and that these events are likely to provide candidates of interest for a given physics search. For example, the semileptonic FEI skim that will be applied in Chapter 3 has an event-based efficiency of 8.89% in on-resonance Belle II data that passes the 'hadron' trigger in the HLT, as recorded in the experiment 18 subset. Experimental numbering of the Belle II dataset is discussed in Section 2.4.4. This skimming results in a reduction to 16.07% of the original file size in the collision data sample.

Computing

After the analysis skims, the event files are made available to the analyst. This, as well as all previous steps described in the data processing workflow, is facilitated by high performance computing. As a collaboration of over 1000 physicists from many institutions around the globe, reliable storage and transmission of data is required to run a competitive physics program.

Belle II implements a distributed cloud computing network, referred to as 'the grid' [79]. From the detector, raw data is saved on disk and on tape at the KEK data center, with a copies at the Brookhaven National Laboratory data center. The raw data is calibrated before being sent to regional data centers across Asia and Europe. This calibrated data can then be sent to grid sites of the distributed computing network provided by institution. It is using these sites where further computationally complex processes such as skimming or creating simulated data takes place and additionally stored. An individual user can then use the grid computing software on their local computing resources to access these samples, or submit computing jobs to the grid sites.

Collision data is not the only data samples analysts have access to. Using *Monte Carlo* simulated data is an important method of making predictions about how analysis techniques will perform in particle physics measurements.

2.4.2. The role of Monte Carlo simulations

The term 'simulated data' has been referenced previously in this thesis. In this section, the use of Monte Carlo simulated data will be discussed.

Within high-energy physics, simulated collision data is created using the Monte Carlo method. Monte Carlo methods involve randomly sampling known probability distributions, where the obtained sample is used to calculate a result. In particle physics, it is invoked to model the probabilistic nature of particle decays and their detection. Monte Carlo methods are used at all stages of simulating Belle II data. This is generally whenever a stochastic result is required in the simulation: whether electron-positron bunches interact; whether they produce the $\Upsilon(4S)$ resonance or another mode; what momentum and energy particles should be created with; how long after being created should a particle decay; the selection of a given particle decay mode out of all allowed decay modes; whether a particle interacts with subdetectors; and how the subdetector responds to the incident particle. Thus, analyses can be performed in Monte Carlo simulated data (commonly referred to as simply 'MC' data or just 'MC') so as to understand what might be seen in collision data.

Designing analyses through MC has a number of advantages. There are very few particle decays which can be identified with high purity without the use of MC modelling. Large amounts of MC events can be created with sufficient computing resources, independent of the detector infrastructure. With the advent of sophisticated machine learning techniques used to identify particle decays, large amounts of data is required to teach such algorithms, a luxury not present in a finite collision dataset. Most importantly, the performance of analyses can be evaluated against the truth information of the simulation, which is impossible in collision data. Naturally, MC is not a one for one replacement for collision data. New phenomena cannot be 'discovered' in MC, as it is limited by the known physics embedded in the method. Moreover, measurements made in MC can differ from that of data. This can be due to idealised detector or accelerator conditions, poor modelling of the underlying physics, as well as limited knowledge of detector response in this early stage of the Belle II program. To best rectify this, the Belle II Performance group performs studies of data-MC discrepancies, so as to make recommendations at the data processing stage (i.e. calibration factors) or to produce corrections factors for use by analysts. Correction factors will be applied to both data and MC used in the $B^+ \to \mu^+\nu_\mu$ search, to be discussed in Section 5.2.

There are two key stages in the creation of Belle II MC: the event generation and the event simulation. The event generation involves the physics modelling, from the electron-positron collision to the FSP creation. In a Belle II analysis, multiple kinds of MC will be used. The first will be *signal MC*, where the electron-positron collision is fixed to produce the particle decay of interest. This can be used to learn about the key kinematic features of the desired mode. The second will be *generic MC*, where the electron-positron collision is free to proceed via any allowed mode. Additional MC samples fixed to decay via common background modes may also be used for better modelling, as was done in the 2020 Belle search described in Section 1.4.3.

Event generation is performed using the Belle II analysis software, to be described in Section 2.4.3. The software acts as an interface with physics generators, which are independent packages that encode the relevant particle physics theory. Different packages specialise in different decay types, though a large number of them are featured in `EvtGen` [80]. `EvtGen` was originally developed for BaBar and the CLEO experiments, specifically for B and D decays. For continuum processes and their fragmentation to hadrons, both `EvtGen` and `PYTHIA` are used [81]. Belle II maintains a '.dec file' that is used in `EvtGen`, a list of all particle decays theorised to occur in its e^+e^- collision events and their branching fractions. This also lists all the possible sub-decays that can take place. Where possible, the decay is assigned its physical branching fraction according to its listing in the PDG. For the case where the explicit branching fraction is not listed in the PDG, the value in the '.dec file' can be estimated via symmetry arguments (lepton universality, isospin symmetry etc.) from another listed mode, calculated from SM, or inferred to be a fraction of relevant PDG listed inclusive branching fractions. Once generated, a full decay chain is created for however many events are required, with a four-momentum for each particle.

As mentioned earlier, MC methods are used not only to generate physics events, but also also to simulate the detector response. The principle is to take the particles created during event generation and model the physics of how they would interact with the materials of the Belle II subdetector. This includes phenomena such as the ionisation of the gas in the CDC, the production of Cherenkov radiation in the TOP, the scintillation and electromagnetic shower evolution in the ECL crystals, and so on. The Belle II software also interfaces with external packages to simulate the physical detector response, namely `Geant4` [82]. Particles created in the material interactions of `Geant4` are referred as secondary particles, as opposed to the 'physics' or 'primary' particles from event generators. Generally, requiring that particles have origins near the IP is sufficient to avoid secondary particles being used for analysis reconstruction. Custom Belle II software is also written to turn the secondary particles in the subdetectors into digital signal.

Once event generation is complete, the created MC file is identical to that of raw collision data after the trigger stage, save for the additional truth information. They are thus subject to the same data processing flow described in Section 2.4.1, including calibration and skimming. MC events for use by Belle II analysts are created centrally by the Belle II Data Production group.

Having explained the production of data and MC, it is now relevant to explain the Belle II-specific software used to perform such tasks.

2.4.3. `basf2`: Belle II Analysis Software Framework

`basf2` is the software written to analyse data and MC events at Belle II [83]. The package is written in C++ and contains modules which can perform most steps of a data analysis: MC generation and detector simulation as described in the previous section, event skimming, event reconstruction, calculation and saving of particle attributes, and basic histogramming. These modules are part of the `basf2` library, `modularAnalysis` and can be added by calling the relevant convenience functions. To perform these tasks, the user writes a Python script to 'steer' the taskflow, in which the required `basf2` modules required are added to an analysis path object. When the path is executed, the modules are run in order on a given data sample.

Belle II data and MC are stored in the '.root' format, a filetype provided by the `ROOT` package [84]. `ROOT` is a C++-based analysis package used across high-energy physic. It is optimised for processing and storage of large amounts of data and is thus well suited for use in a high-luminosity experiment like Belle II. The package provides dataframes for processing column-based data, such as the data tables of particles in an event and its variables used in this analysis. Particle data tables are also referred to as *ntuples*. In addition to basic operations on the columns of dataframes, more sophisticated particle physics functions such as Lorentz boosting quantities are provided by `ROOT`. `ROOT` also provides statistical methods such as histogramming and parameter fitting, as well as data presentation methods. `basf2` calls `ROOT` for a number of operations, including data input/output and data analysis of ntuples and histograms.

As the Belle II Data Production group provides the skimmed data and MC samples, the implementations of `basf2` relevant to this work are in the final stages of event reconstruction and the creation of particle data tables - histogramming will be performed using external software.

`basf2`, as used to reconstruct the dataset of this analysis

To create the ntuples for this analysis, a variety of `basf2` modules are necessary. Up-to-date experimental conditions must be incorporated into the taskflow, such as the magnetic field map within the detector, further calibration and correction factors, as well as training weights for machine learning algorithms. Such data is provided on the online conditions database maintained by Belle II. The advantage of this is that corrections can be added at any time, rather than waiting for `basf2` software releases. Thus, a function to query this database is added to steering script. The path object is created, after which the function to read in the skimmed sample is added. This also initialises the particle list of pre-reconstructed B_{tag}'s, for the sample used in this analysis.

To use variables such related to the missing momentum of event, the function `modularAnalysis.buildEventKinematics` is added to the path. The missing momentum of an event, which is expected to be large for $B^+ \to \mu^+\nu_\mu$, can be derived as initial energy of the electron-positron collision is fixed. The corresponding module sums all tracks and clusters of each event in the sample, including those used in the B_{tag}, evaluates the contributing momentum and energy and subtracts this from the known initial conditions. Such tracks and clusters have minimum quality requirements to avoid fake tracks or beam-background based clusters, The `modularAnalysis.buildEventShape` function operates in a similar way, using the same tracks and clusters but instead evaluates the previously discussed event shape variables, such as calculating the event-based Fox-Wolfram moments. To evaluate the thrust variables between the B_{tag} object and tracks and clusters not yet used in reconstruction, a common selection to reject continuum in previous $B^+ \to \mu^+\nu_\mu$ searches described in Section 1.4, the function `modularAnalysis.buildContinuumSuppression` is used.

The list of muon candidates in the event can now be built from the `stdCharged` library of basf2, so as to perform $B^+ \to \mu^+\nu_\mu$ reconstruction. The function `stdCharged.stdLep` creates the muon list from available tracks, with a `muonID` probability of 50% or higher of being a true muon. As will be decided in Section 4.1.1, the best signal muon candidate to be used in $B^+ \to \mu^+\nu_\mu$ reconstruction will be decided by choosing the highest $\Upsilon(4S)$-frame momentum, via `modularAnalysis.rankByHighest`. Lower ranked muons remain in the event, but not as the signal muon candidate.

The full $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B_{\text{tag}}$ process is then reconstructed sequentially. Using `modularAnaysis.reconstructDecay`, a B meson with matching charge is reconstructed from the signal muon. With no neutrino as in $B^+ \to \mu^+\nu_\mu$, this effectively ensures the four momentum of the muon is lower than or consistent with the B invariant mass, and then relabels the object as a B meson. The function `modularAnalysis.reconstructDecay` is called again to then perform the $\Upsilon(4S)$ reconstruction of the new B meson and the B_{tag}. With two candidates, the reconstruction becomes more complex. Within the `modularAnalysis.reconstructDecay` function, the four vector of the single signal B candidate is summed uniquely with every possible B_{tag} candidate four vector, for each event. Still within the function, each sum is verified against the $\Upsilon(4S)$ invariant mass and omitted if they disagree. This completes the analysis reconstruction. To ensure only a single $\Upsilon(4S)$ candidate is allowed per event, each $\Upsilon(4S)$ is ranked using `modularAnalysis.rankByHighest` on the basis of the highest value of the B-tagging algorithm's classifier variable assigned to the underlying B_{tag} candidates. The properties of this classifier is to be presented in Section 3.4.2. Additionally, with the particles of the decay of interest full decided, the total energy and momentum of the unused tracks and clusters can be calculated using the function `modularAnalysis.buildRestOfEvent`, where this information can be saved to the $\Upsilon(4S)$ candidate of each event. This is an especially powerful constraint given the guarantee that no additional particles will exist if the $\Upsilon(4S)$ is reconstructed correctly. Further selections on the basis of such particle variables can be applied using `modularAnalysis.cutAndCopyList`.

It was discussed in Section 2.4.2 that an advantage of using MC to study particle decays is that the truth conditions of the event generation can be consulted. Within `basf2`, the truth information such as the generated momentum, particle identity and whether or not the reconstruction was performed correctly can be accessed. To be able to access the truth information in the taskflow, the function `ma.matchMCTruth` to be called on the $\Upsilon(4S)$ list is added to the analysis path. Such a function only works with MC simulated data and not collision data. The nuances of the truth-matching algorithm are explained in the next section.

This taskflow is altered slightly for the B_{tag}-only investigation of Chapter 3, where no $B^+ \to \mu^+\nu_\mu$ reconstruction takes place. Here B_{tag} ranking and `modularAnalysis.build-RestOfEvent` is applied directly to the B_{tag} list. The same applies in Section 4.1.2, when analysing the muons that remain after B-tagging but before reconstruction. In this case, the muon list is created but reconstruction does not occur. Moreover, for the muon only analysis in Chapter 4, truth-only tuples are saved via `modularAnalysis.fillParticleListFromMC`.

Additional `basf2` functions are added if collision data is analysed, to add correction factors defined in the corrections database to particle variables. These are created by the Belle II Performance group. They will be described in Section 5.2.

Finally, the variables pertaining to each events $\Upsilon(4S)$ candidate can be saved as an ntuple. Such variables include those built using the various functions from `moduleAnalysis`, including extra variables such as the kinematics of the daughter muon (i.e. the signal muon) and the daughter B_{tag}, as well as the truth variables. This is performed using `modularAnalysis.variablesToNtuple`. When the `basf2` path object is processed, each equipped module is executed, leading to a final '.root' file being saved.

The wide range of tasks that can be achieved using `basf2` for analysis of Belle II data and MC has been presented for a typical reconstruction steering script. Throughout this taskflow, there has been no necessity for the analyst to perform calculations on four-vectors directly. This emphasises the simplicity of the `basf2` package, that perfect knowledge of decay kinematic is not required and that complex combinatorial tasks such as creating candidates and ranking can be performed efficiently. All that is required of the analyst is that they decide on the logic of the taskflow, in order to create ntuples and immediately begin a statistical analysis of particle variables.

In this analysis, the final `ROOT` ntuples are read using `uproot` and implemented via an `awkward-array` dataframe rather than `ROOT`, to allow for fully Pythonic computation [85] [86]. Histogram creation and data presentation is performed using `numpy` and `matplotlib` [87] [88].

Truth-matching in MC simulations

It was mentioned in the previous section that truth information of a reconstructed particle in `basf2` can be accessed. This is performed within `basf2` by performing a truth-matching algorithm [89]. For a given reconstructed particle, the algorithm performs three key steps.

1. The reconstructed particle is associated with their MC generated counterparts for all particles in the reconstructed particle's decay chain, recursively in a top-down manner. An FSP is matched with a truth counterpart if the underlying FSP

tracks hits and cluster energy can be associated with a single truth particle. A composite particle is matched with a truth counterpart if the counterpart is the closest common parent (or grandparent, great-grand parent etc.) to the truth-matches of the particle's children. If no matches can be made, the truth-matching fails.

2. The reconstructed particle's identity is verified against the matching truth particle, as well as the identities between the child particles and their matches.

3. For each descendant in the truth particle decay tree, it is determined if they are matched to a descendent of the reconstructed particles decay tree. If there are unmatched truth particles, this means the reconstructed decay tree is incomplete.

The association of the truth match is performed in Step 1, whereas the validation is performed in Steps 2 and 3. Thus, a reconstructed particle can be matched to a truth particle, but is still reconstructed incorrectly if the particle identity is incorrect or if there are unmatched particles from the truth decay tree. There are also no requirements that observables such as momentum and energy match that of the truth particle, only the identity and the specification of all parent/child particle decay relationships.

The errors in reconstructed particle and its descendants can be encoded into the `mcErrors` variable and assigned to the candidate particle. If all three steps described above succeed, `mcErrors` will take a value of 0. Each error encountered in the matching, such as missing a final state particle in the reconstruction or adding the wrong particle, a power of two unique to the type of error is added to the value of `mcErrors`. If the first step fails for a reconstructed particle or any of its children, the internal error flag (512) is added to `mcErrors`. If the second or third step fails, all other relevant error flags are summed to `mcErrors`. The various errors are summarised in Table 2.4.

From `mcErrors`, the more useful `isSignal` is defined within `basf2`. `isSignal` takes the value 1 for correct MC matching (i.e. `mcErrors`=0), `NAN` for no MC match (i.e. `mcErrors`=512) and `isSignal`=0 if the MC matched particle identity is not the same as the reconstructed one.

When reconstructing $B^+ \to \mu^+\nu_\mu$, or any other decays with neutrinos, no events will be assigned `isSignal`=1 due to the missing neutrino. This will also affect the $\Upsilon(4S)$ that is reconstructed with this B meson. Thus, the alternative variable `isSignalAccept-MissingNeutrino` will be used. The variable `isSignalAcceptMissingNeutrino` yields

Table 2.4. Table of `mcErrors` flags in truth-matching.

Error Flag	Explanation
0	Correct truth-matching
1	Missing final state radiation photon from reconstruction, (usually from fragmentation processes)
2	Missing resonance state in reconstruction decay chain, (but all other decays identified)
4	Particle decays in flight, and this product is treated as the reconstructed particle.
8	Neutrino missing from reconstruction
16	Physics photon is missing from the decay
32	Massive FSP is missing from the reconstruction
64	K_L^0 is missing from reconstruction
128	Charged FSP has wrong charge
256	Particle from outside the decay has been used in reconstruction
512	No truth match found
1024	Missing photon from reconstruction (specifically from PHOTOS generator)
2048	A Bremsstrahlung photon (from Geant4) has been incorrectly added to reconstruction

a value of 1 even if the missing neutrino error flag is raised and as such, will be the variable used to assess the reconstruction.

2.4.4. The dataset for this analysis

This Belle II search for the $B^+ \to \mu^+\nu_\mu$ process will be performed in $362\,\mathrm{fb}^{-1}$ of e^+e^- collisions recorded at the $\Upsilon(4S)$ resonance. Internally, this is referred to as the 'Long Shutdown 1' (LS1) dataset, as it is the first set of data recorded prior to the 2022-2023 detector maintenance period.

Data-taking began in April 2019 and finished in June 2022. Each year of data-taking is roughly broken into four periods. The 'a' period, extends from February/March/April until May/June, then 'b' from May/June until June/July. The third is a break, to save operations costs due to cooling during Japanese summer. The final period is 'c' commences in October until mid-December. This analysis is restricted to those runs with collision energy equivalent to $\Upsilon(4S)$ production.

The LS1 dataset contains 7409 runs, between periods 2019a and 2022b. 6425 runs were recorded at the $\Upsilon(4S)$ resonance and will thus be relevant to this $B^+ \to \mu^+\nu_\mu$ search. The processing of the dataset is the 13th reprocessing (proc13), with prompt buckets 26-36. This excludes buckets 27 and 34, which contain non-$\Upsilon(4S)$ runs. The relationship between the relevant experiment number, data-taking period, processing number and bucket number is presented in Table 5.17.

Table 2.5. Summary of the LS1 data-taking period, including integrated luminosity of $\Upsilon(4S)$ decays per experiment. Experiments containing test-beam runs (i.e. electron or positron beams operating individually without collision) or cosmic runs (i.e. no beam, recording cosmic rays) have been omitted.

Exp.	Period	Duration	Energies	Processing	Bucket	$\mathcal{L}_{int,\Upsilon(4S)}$ [fb^{-1}]
7	2019a	26/04 - 8/05	$\Upsilon(4S)$	proc13	6	0.506
8	2019b	09/05 - 3/06	$\Upsilon(4S)$ (+scan, offres)	proc13	7	1.663
10	2019c	01/11 - 10/12	$\Upsilon(4S)$	proc13	8	3.665
12	2020a, 2020b	04/03 - 30/06	$\Upsilon(4S)$ (+offres.)	proc13	9 - 15	54.573
14	2020c	29/10 - 17/12	$\Upsilon(4S)$	proc13	16, 16b	16.500
16	2021a,2021b	26/02 - 22/03	$\Upsilon(4S)$	proc13	17, 18	10.294
17	2021a,2021b	24/03 - 06/04	$\Upsilon(4S)$	proc13	17, 18	10.715
18	2021b	07/04 - 04/07	$\Upsilon(4S)$ (+offres.)	proc13	19 - 25	89.900
20	2021c	29/10 - 09/11	$\Upsilon(4S)$	prompt	26	3.788
21	2021c	11/11 - 28/11	$\Upsilon(5S)$ scan	prompt	27	-
22	2021c	29/11 - 22/12	$\Upsilon(4S)$, $\Upsilon(5S)$ scan	prompt	28,29	32.060
24	2022a	02/03 -26/04	$\Upsilon(4S)$	prompt	30 - 33	85.642
25	2022a, 2022b	27/04 - 10/05	$\Upsilon(4S)$ offres.	prompt	34	-
26	2022b	11/05 - 21/06	$\Upsilon(4S)$	prompt	35, 36	54.795

In order to optimise the reconstruction of $B^+ \to \mu^+\nu_\mu$, an MC dataset will be investigated first. Once the reconstruction method has been finalised, including selection criteria to reject $B^+ \to \mu^+\nu_\mu$-mimicking events, then the dataset described above will be utilised.

The MC samples implemented in this analysis originate from the 15th run-independent MC production campaign, specifically `MC15ri_b`. This MC is based on the `release-06-00-08` version of `basf2`. Run-independent MC uses the same detector conditions across all events, differing from run-dependent MC where the detector conditions vary, such that each event matches that of a real run in data. Now that a large amount of data is available, the Belle II Collaboration is looking to move towards only using run-dependent

MC. However, this analysis will continue to use run-independent MC, where the extension to run-dependent MC is left for a future work.

The specific samples used in this analysis is listed in Table 2.6, created by Belle II Data Production group. A sample of 8 million $B^+ \to \mu^+\nu_\mu$ MC decays, that proceed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$, will be used as the signal MC, where charge conjugation is implied. The alternate B is free to decay generically to any charged B mode dictated by its charge. MC samples of $\Upsilon(4S) \to B^+B^-$ and $\Upsilon(4S) \to B^0\overline{B}^0$ equivalent to 2.8 ab^{-1} will form part of the *background* modelling. Inside the Collaboration, B^+B^- events are referred to as *charged* MC, while $B^0\overline{B}^0$ are referred to as *mixed* rather than 'neutral'. This is because the sample also allows encodes flavour oscillations as $B^0 \leftrightarrow \overline{B}^0$ for individual B's. For modelling the background continuum contribution, MC samples equivalent to 1000 fb^{-1} each of $e^+e^- \to q\overline{q}$ for $u\overline{u}$, $d\overline{d}$, $s\overline{s}$ and $c\overline{c}$ are used. Each MC event will be assigned a weighting according to the scale factor in the table, so that the sample can be used to predict the yield of the mode in the 362 fb^{-1} dataset.

It is noted that the related $B^+ \to \tau^+\nu_\tau$ decay, whose τ^+ could proceed as $\tau^+ \to \mu^+\bar{\nu}_\mu\nu_\tau$ and produce background events to the $B^+ \to \mu^+\nu_\mu$ search in this thesis, is treated within the charged $B\overline{B}$ background sample. Thus, there is no additional MC sample used to model this background. It would be of interest to understand the retention of this particular background mode after the B-tagging and signal-optimising data selections in Chapter 3 and Chapter 4, though quantifying such a contribution has been left for future work. Additionally, a background muon from such an tau event would be expected to have momentum much smaller than that of $B^+ \to \mu^+\nu_\mu$, given that it would originates from a three-body decay. The eventual selection criteria applied to the signal muon momentum decided in Chapter 4 would put such a large number of events (though not zero) outside the region of consideration.

While e^+e^- collisions can produce other physics processes, such as $e^+e^- \to \tau^+\,\tau^-$, they are not implemented in this analysis. In the application of the B-tagging algorithm in this analysis, a set of minimum restrictions are applied to remove events most likely to fail, thus saving computation time. Such restrictions include requiring that there are more than 3 tracks in the event. This immediately culls a number of the non-hadronising $e^+e^- \to \ell^+\,\ell^-$ for $\ell^+ = e^+$ and μ^+ modes and $e^+e^- \to \gamma\gamma$. While $\tau^+\,\tau^-$ events can exceed the three track requirement, their dynamics are very different to B^+B^- events and thus less likely to be selected by the algorithm. As such samples admits a negligible number of events, sources other than those in Table 2.6 are omitted.

Table 2.6. Statistics of Belle II MC used in this analysis. MC is from `MC15ri_b` campaign used. Cross sections taken from Ref. [67].

Sample ($e^+e^- \to X$)	Cross-section [nb]	Sample lumi. [fb^{-1}]	No. of events (in sample)	Scale factor	No. of events (in 362 fb^{-1})
B^- [$B^+ \to \mu^+\nu_\mu$]	-	-	8.00×10^6	2.17×10^{-5}	1.74×10^2
B^+B^-	0.5654	2800	1.58×10^9	1.29×10^{-1}	2.05×10^8
$B^0\overline{B}^0$	0.5346	2800	1.50×10^9	1.29×10^{-1}	1.94×10^8
$c\overline{c}$	1.329	1000	1.33×10^9	3.62×10^{-1}	4.81×10^8
$u\overline{u}$	1.605	1000	1.61×10^9	3.62×10^{-1}	5.81×10^8
$d\overline{d}$	0.401	1000	4.01×10^8	3.62×10^{-1}	1.45×10^8
$s\overline{s}$	0.383	1000	3.83×10^8	3.62×10^{-1}	1.39×10^8

2.5. Summary

After having learned about the phenomena within the $B^+ \to \mu^+\nu_\mu$ decay in Chapter 1, in this Chapter it was discussed how $B^+ \to \mu^+\nu_\mu$ could be identified using B mesons produced at the Belle II experiment. The diversity of the Belle II physics program was presented, as a multi-flavour physics experiment operating at the luminosity frontier to precisely measure both SM and BSM processes. The production of well-constrained B mesons from electron-positron collisions within the SuperKEKB collider was examined. This was done so as to motivate why Belle II is the best experiment suited to identifying the missing energy decay of $B^+ \to \mu^+\nu_\mu$, with three key advantages over such an analysis at the energy frontier experiments. These advantages are the fragmentation-free $B\overline{B}$ pair production from the $\Upsilon(4S)$ resonance, the fixed total collision energy to allow energy and momentum conservation to be applied to missing energy analyses, as well as the hermeticity of the closed cylindrical detector leading to almost all particles produced are detected.

The cases when electron-positron collisions do not create a $\Upsilon(4S) \to B\overline{B}$ process were also considered, so as to be familiar with the background contributions to the $B^+ \to \mu^+\nu_\mu$ search. How the multiple subdetectors of Belle are used to identify the decay products of the $B^+ \to \mu^+\nu_\mu$ decay was addressed, to better understand how the observables such as the momentum of the signal muon will be measured and how the identity of the particle is determined. Additionally, the sophisticated workflow to turn the electrical signal induced in subdetectors to tuples of data for statistical analysis was presented. In order to undertake this analysis, the role of the Belle II Analysis Software Framework or `basf2` was explained, which was used to perform the $B^+ \to \mu^+\nu_\mu$ reconstruction in the collision data and the MC-simulated data for this work.

In the next Chapter, the B-tagging algorithm used to skim the events used in $B^+ \to \mu^+\nu_\mu$ search will be introduced. It will be applied to the MC simulated electron-positron collisions so that the efficiency and purity of the technique can be evaluated, before the $B^+ \to \mu^+\nu_\mu$ signal reconstruction is performed, and before B-tagging is applied to collision data from the Belle II experiment.

$\bowtie \bowtie \bowtie$

Chapter 3.

The Full Event Interpretation

In Chapter 3, the technique of B-tagging used in this search for $B^+ \to \mu^+\nu_\mu$ is described, namely the Full Event Interpretation (FEI).
The role of B-tagging in electron-positron colliders generally will be discussed, as a method of constraining B-physics searches to a dataset most likely to contain true B processes. A general overview of the use of the FEI within Belle II is given.
This is followed by an in-depth discussion of the stages of the FEI B-tagging algorithm. The two main operating modes of the FEI, B-tagging via hadronic decays or semileptonic decays, are illustrated in an MC sample of charged B decays, including evaluation of the FEI via key performance variables and a best FEI B candidate selection procedure.
A brief discussion will take place on current Belle II efforts to improve the FEI algorithm. Having compared FEI B-tagging in its available modes, the mode most suited to the $B^+ \to \mu^+\nu_\mu$ search will be applied to both signal MC and background MC samples that mimic B decays. This is done so that the distributions of key variables can be explored in direct scale to what will be seen in the Belle II collision dataset.
The final section will summarise the FEI B-tagging efficiency for the chosen operation mode, in both $B^+ \to \mu^+\nu_\mu$ MC and background MC contributions. This tagged sample will subsequently be used in signal reconstruction in the following Chapter.

$\bowtie \bowtie \bowtie$

3.1. B-tagging in e^+e^- colliders at the $\Upsilon(4S)$ resonance

B-tagging involves identifying collision products which pass criteria that suggest they originate from B mesons. Such tagging can be performed to search for the B decay of interest directly (i.e. the B_{sig}) or to constrain the properties of B_{sig} using a B meson that decays alongside it (i.e. the B_{tag}). This is namely implemented at e^+e^- experiments at the $\Upsilon(4S)$ resonance, where B mesons are guaranteed to be produced in pairs.

What it means to 'tag' a B meson differs depending on the B-physics search. In CP violation studies, one searches for asymmetries in the number $B_{\mathrm{sig}} = B^0$ decays versus $B_{\mathrm{sig}} = \overline{B}^0$ decays to CP eigenstates, such as $J/\psi\, K^0_S$ or $\pi^+\pi^-$. As these eigenstates can be produced by both B^0 and $\overline{B}^0$ decays, to tag a B is to ascertain the flavour (i.e. decide if the B_{tag} is B^0 or $\overline{B}^0$) of the partnered B decay event and thus know the B_{sig} to have had the opposite flavour when produced. For missing energy signal modes, such as $B^+ \to \mu^+\nu_\mu$, B_{sig} reconstruction is made complicated by the neutrino, as was discussed in Section 1.2.3. The principle of B-tagging in missing energy B searches is not to determine the unknown flavour of the B_{sig}, but to constrain the difficult B_{sig} reconstruction by determining what particles in the event belong to the partner B i.e. the B_{tag}. Once the particles of a collision event are assigned to the B_{tag} products, all remaining products can be assumed to originate from B_{sig}, or vice versa.

In the latter case, reconstruction of a B_{tag} can create a powerful additional feature to use in the selection of a rare, missing energy B decay. From energy and momentum conservation, subtracting the identified B_{tag}'s four momentum with the known initial energy of the $e^+e^- \to B\bar{B}$ collision will estimate the four-momentum of the remaining B meson. Unlike when B mesons are sourced from $\mathcal{O}(\mathrm{TeV})$ energy collisions, $B\overline{B}$ pairs originating from the $\Upsilon(4S)$ resonance are created with no additional particles.

Both inclusive and exclusive B-tagging were used in the previous $B^+ \to \mu^+\nu_\mu$ searches summarised in Section 1.4. Here, it was seen that exclusive B-tagging methods explicitly reconstruct the B_{tag} via its decay mode, while inclusive B-tagging assigns all energies and momentum detected to be from a single B meson and thus the B_{tag} itself. The selection of inclusive over exclusive tags and vice versa is driven by conditions of the analysis conducted. The explicit decay mode reconstruction of exclusive B_{tag} provides additional tag-side decay-specific information. These can be used as additional selections, thus ensuring a higher ratio of correct B_{tag}'s (i.e. higher *purity*). While the lack of decay mode information in inclusive B-tagging can result in lower purity, using an inclusive B_{tag} means B mesons decaying alongside the signal B are not restricted to a given decay. As each B meson decays independently of one another, B decays of interest are not limited to be produced along side the decay modes implemented by an exclusive B_{tag}. With a larger number of B_{tag} candidates identified, inclusive decays have higher *efficiency* in B-tagging than exclusive ones. The choice of a higher efficiency tagging technique may be useful for a signal decay mode with a lower branching fraction. In such a case, the number of expected events will be quite low, and restricting the search based on B_{tag} decay may further decrease the yield. High-efficiency tagging is also advantageous when the size of the available dataset is also small, such as in the preliminary physics runs of Belle II. If, however, the dataset is larger, as will be the case as Belle II approaches its target with the $50\,\mathrm{ab}^{-1}$ of e^+e^- collisions, having the highest possible efficiency can be sacrificed in the pursuit of reconstruction quality. Implementing the exclusive B_{tag} limits the analysis to only the most accurately identified (i.e. highest purity) events, and thus allowing for highest precision measurements of signal modes.

In order to balance the purity and efficiency of either B-tagging approach, additional techniques can be implemented. The purity of the inclusive tag can be improved using specific tag-side improvement techniques, as was implemented in the 2020 Belle search discussed in Section 1.4.3. Recall that in this case, the energy of the B_{tag} was replaced with that of the nominal B energy. Additionally a calibration factor was derived to account for the cases of missing particles from the detector. This resulted in improved estimate of the B_{tag} and thus the B_{sig}-frame muon momentum resolution.

In Section 1.4.2, the hadronic B-tagged $B^+ \to \mu^+\nu_\mu$ search performed by BaBar was presented, where a 'semi-exclusive' tagging was used. By reconstructing the D meson produced from the B, the analysis can use more decay-specific selection variables, like the D invariant mass. This would not be an option for fully inclusive B-tagging, as the underlying decay mode is not known. Additionally, by not fully specifying the sub-decays that could have produced the added pions and kaons alongside the D meson, the analysis takes advantage of having lesser restrictions on the B decays that can be used for B-tagging.

To increase the efficiency of exclusive B-tagging, an analysis can implement more than one tag mode. The choice of these tag modes requires careful consideration, balancing how commonly the decay mode occurs to contribute to efficiency, as well as how easy to establish it is, contributing to purity. Moreover, a procedure needs to be developed to which of the chosen decay modes are chosen if each mode admits a suitable candidate for

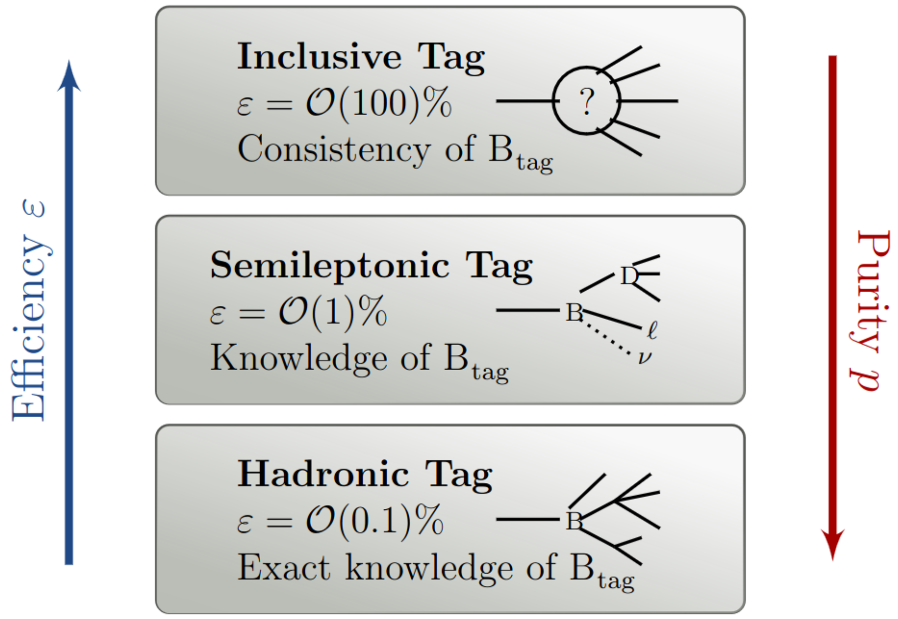


Figure 3.1. A comparison of B-tagging categories. Efficiencies here refer to the proportion of the B decays potentially covered by the tagging mode. Image taken from [90].

a given event. This is called the *best candidate selection*, and will be developed for this analysis in Section 3.4.2.

The FEI, as will be the tagging choice of this analysis, is one such approach to improving exclusive B-tagging via a large number of decay modes, with a model that provides a natural choice for B_{tag} best candidate selection.

3.2. An Overview of the FEI

The *Full Event Interpretation* is an exclusive B-tagging machine learning algorithm developed by Belle II [47]. The algorithm reconstructs B mesons into 1 of up to 34 channels depending on the operating mode, or $\mathcal{O}(10^5)$ channels when allowing for different intermediate decay modes. The FEI is written as a module for the Belle II analysis software, `basf2`.

Reconstruction takes place in a hierarchical manner. All final state particles are trialled (hence using the *full event*), but only the 'best' go on to create the intermediate states. This is repeated on the intermediate states, until a B_{tag} is reconstructed. The FEI is used in analyses of $e^+e^- \to B\overline{B}$ events that rely on the companion B being identified to learn about the desired B decay process, such as in signal modes involving missing energy.

The FEI determines the 'best' candidates of a B_{tag} decay tree by training a multivariate classifier for each particle in each decay channel. In the training phase, the algorithm learns the kinematics of correctly truth-matched candidates. From this, the algorithm derives general weights that could be used to classify a previously unknown event based on its kinematics alone. In the application phase, these weights are used to calculate the multivariate classifier, *signal probability*. Given on a scale of zero to one, the variable functions as a prediction of correctness for a given reconstruction. Thusly, only the particles with the highest signal probability are used in the reconstruction of the B_{tag}.

The FEI is broadly modelled on its Belle predecessor, the *Full Reconstruction* [45]. The FEI improves on the Full Reconstruction algorithm by introducing new decay modes, as well as switching from the `NeuroBayes` neural network package to a boosted-decision tree (BDT) method as a multivariate classification technique. The FEI is also able to be run on the original Belle dataset, allowing for comparison between the two methods, as well as re-analysis of previous data.

Within the first 5 years of collisions at Belle II, 6 out of the 9 missing energy B-decays have relied on the FEI in their analysis [50, 51, 91–94]

A more in-depth look at the stages of the FEI algorithm is discussed in the following section.

3.3. Structure of the FEI Algorithm

B_{tag} candidates are reconstructed in stages, with each stage corresponding to reconstruction of a class of particles in a given level of the B_{tag} decay tree. The particles in each stage of reconstruction is represented in the graph in Figure 3.2. For each decay channel of each particle candidate considered in a stage, two main phases are carried out: reconstruction prior to the multivariate analysis ('pre-reconstruction') and the further reconstruction that follows ('post-reconstruction').

During the pre-reconstruction, the particle lists pertaining to the stage are initialised. For the first stage, this consists of loading in the final state particles present in the

Table 3.1. Table of particles used in FEI B_{tag} reconstruction and their training variables. Δm denotes the difference between the nominal particle mass and the measured mass. Δq denotes the difference between the nominal energy released and the measured energy released.

Particle	No. of Channels	Best Cand. Var.
Charged FSPs	1: tracks	`particleID`
γ	2: unmatched ECL clusters, V^0's	E_{gamma}
π^0	1: $\pi^0 \to \gamma\gamma$	$\lvert\Delta m\rvert$
K_S^0	3: pion decays and V^0's	$\lvert\Delta m\rvert$
Λ^0	2: $\Lambda^0 \to p^+\pi^+$ and V^0's	$\lvert\Delta m\rvert$
K_L^0	1: KLM clusters	$\lvert\Delta m\rvert$
Σ^+	1: $\Sigma^+ \to \pi^+\pi^0$	$\lvert\Delta m\rvert$
Λ_c^+	18: $pK\pi$ modes, $\Lambda^0\pi\gamma$ modes, $\Sigma^+\pi$ modes	$\lvert\Delta m\rvert$
D^0	15: $K\pi$ modes	$\lvert\Delta m\rvert$
D^+	11: $K\pi$ modes	$\lvert\Delta m\rvert$
J/ψ	2: $\ell^+\ell^-$ modes	$\lvert\Delta m\rvert$
D^{*+}	3: $D\pi$ modes, $D\gamma$	$\lvert\Delta Q\rvert$
D^{*0}	3: $D\pi$ modes, $D\gamma$	$\lvert\Delta Q\rvert$
D_s^+	10: $K\pi$ modes	$\lvert\Delta m\rvert$
D_s^{*+}	2: $D_s^+\gamma$, $D_s^+\pi^0$	$\lvert\Delta Q\rvert$
B_{had}^+	36: See Table 3.2	Product of daughter `sigProbs`
B_{SL}^+	8: See Table 3.2	Product of daughter `sigProbs`
B_{had}^0	32: consisting of $D^{(*)}$, π, J/ψ, K, $D_s^{+(*)}$,Λ,Σ	Product of daughter `sigProbs`
B_{SL}^0	8: $D^{(*)}\ell\nu_\ell$, $D^{(*)}\pi\ell\nu_\ell$	Product of daughter `sigProbs`

event from tracks, clusters and V^0's. For all other stages, this consists of reconstructing the required candidates via predetermined decay modes, using the particles loaded (or reconstructed themselves) in previous stages. Depending on the channel, there might be multiple ways to combine the loaded particles to create the reconstructed particle candidate required. Thus, once all possible combinations have been reconstructed for one decay channel, they are ranked according to a 'best candidate' variable. Usually this variable is correlated with accurate reconstruction. This can include particle identity likelihoods from subdetectors for charged final state particles, or the unsigned difference between the measured and nominal masses, as used for a number of intermediate particles. The number of decay modes used per particle and the corresponding best candidate variable for all particles in the FEI are listed in Table 3.1. Only the best 20 candidates

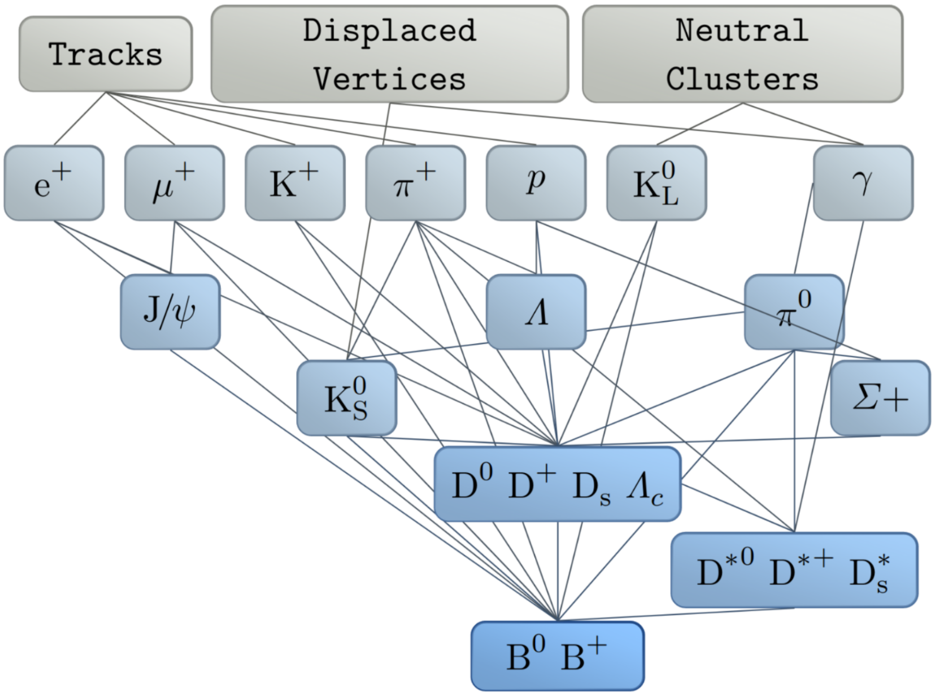


Figure 3.2. A schematic of the hierarchial structure of particles using in each stage of reconstruction, as performed by the FEI. Image taken from [95].

of the channel (40 for the case where photons are the final state particle) are kept to be included in the next stage's reconstruction.

The next step of pre-reconstruction is to perform a *vertex fit* of the particle. This derives the location that the particle was created. Better reconstructed particles will originate closer to the IP, as opposed to those occurring in particle-material interactions which will be displaced from the IP. For composite particles, the fit traces back the tracks of all charged decay products to a single point, the vertex. Vertexing information is an important variable used in the multivariate analysis (MVA) application that follows.

If the FEI is in training phase, the training weights of each particle in each decay channel for the particles of a given stage are derived at this stage, prior to post-reconstruction. The weights will indicate how probable a particle is to belong to a B decay chain, on the basis of a set of key variables. These weights are obtained via the `FastBDT` implementation of boosted decision tree regression. Using a set of kinematic variables and observables of the given particle, up to 400 decision trees are 'grown'. Each tree consists of 10 levels of successive random variables selections, with each selection dividing the data into two groups. The goal is to identify the series of selections that best splits the training data into correctly reconstructed and incorrectly reconstructed. To 'boost' the decision tree learning, each new tree is created to improve misclassification in prior trees, thus strengthening the outputted classifier.

The training variables used to train each particle's BDT vary per particle identity. FSPs use any relevant particle ID variables, energy and momentum components, fitted track parameters and cluster variables. Composite particles use fitted vertex parameters, energy and momentum components, masses, distance from IP, as well as daughter variables such as combinations of invariant masses and opening angles. Creation of the training weights of the FEI is performed centrally on behalf of analysts at Belle II, as it requires 100 millions of MC events for proper training. For the FEI training relevant to the MC campaign used in this analysis, 200 fb^{-1} of $e^+e^- \to B^+B^-$ MC simulated events (around 108 million) have been used to train charged FEI B-tagging and an additional $200\,\text{fb}^{-1}$ of of $e^+e^- \to B^0\overline{B}^0$ events (around 102 million) to train the neutral FEI B-tagging. For each decay channel, there must be at least 500 events that correctly pass truth-matching for a particle and 500 that do not, so that the training has sufficient statistics. This can prove challenging for the high purity - low efficiency modes, such as $B^+ \to J/\psi\ K^+$, a standard tag mode. In this case, additional statistics of this mode alone are implemented. Once produced, the FEI weights are uploaded to the Belle II conditions database, so that they can accessed by `basf2` applications of the FEI algorithm.

During the post-reconstruction phase, the MVA classifier is evaluated for each candidate. This is the signal probability, calculated from the training weights according to the values of training variables for a reconstructed candidate. If the particle has multiple decay channels, they are now merged into a single group for consideration. A minimum signal probability is implemented to remove the worst candidates, after which they are ranked on this variable. The minimum signal probability is stricter for FSPs than intermediate ones, owing to their higher multiplicity in an event. There is no minimum signal probability implemented for B mesons during the post-reconstruction stage, as it is expected that analysts will decide on their own minimum selection, according to their signal B mode. Generally, up to 10 candidates per particle per event are kept for following stages, increasing to 20 candidates in the final stage, when B mesons are created.

In addition to creating the FEI training weights, the application of the FEI to relevant data and MC samples is performed by the Belle II Data Production group. What is provided to analysts are data samples with a pre-reconstructed B_{tag} particle list, with all tracks and clusters of the event preserved. Analysts then can combine the B_{tag} with tracks and clusters outside the B_{tag} to select their signal B decay mode. Any event that has not produced an appropriate B_{tag} is removed from the sample. To limit the disk size the data consumes, additional selections (*post-selections*) are applied to remove candidates well outside of the range of correct B reconstruction. These post-selections are specific to which class of decays the B meson is reconstructed in. Thus, the application of the FEI algorithm to data and MC constitutes a skim, the role of which was discussed in Section 2.4.1.

The FEI and its role in reconstructing B_{tag}'s by systematically choosing particles best suited to their decay has been discussed. The available classes of decays that B_{tag} reconstruction takes place in now can be discussed.

3.4. FEI operating modes

When one implements the FEI in B-tagging, one has a choice of which class of B decay modes to reconstruct. These options are summarised in `default_channels.py` in `basf2` [96].

Firstly, one chooses the B type to reconstruct, as required by the charge and collision energy of the analysis: B^+, B^0 or B_s. As discussed in Section 2.2.4, it only makes sense to use a B_s mode in searches above $\sqrt{s}$ =10.58 GeV, as B_s production is only possible for analyses at the $\Upsilon(5S)$ resonance. With this analysis focusing on the $B^+ \to \mu^+\nu_\mu$ decay, a charged decay with no neutral B counterpart, the strange or mixed B cases are not discussed in any more depth.

Additionally, the FEI can be used to analyse data from the original Belle experiment. This allows for re-analysis of $B\overline{B}$ samples with an improved B-tagging method, as opposed to analyses that may have previously implemented the Full Reconstruction method [97]. It also allows for simultaneous measurements in Belle and Belle II datasets with the same tagging strategy, taking advantage of extra 711 fb^{-1} of $B\overline{B}$ events in Belle data. There are currently no published combined Belle and Belle II searches using the FEI, though there are studies in preparation [98].

The choice that affects the B-tagging efficiency most strongly is that of the of specific B decay mode class: *hadronic* modes, *semileptonic* modes or the lesser used K_L^0 mode. The K_L^0 modes constitute those with an explicit K_L^0 in the B decay mode (such as $B^+ \to \overline{D}^0\ D^+\ K_L^0$) or contain an intermediate state reconstructed using a K_L^0 (such as B^+

$\rightarrow [\overline{D}^0 \rightarrow K_L^0\ \pi^0\ ]\ \pi^+$). While K_L^0 mesons can be identified in the dedicated K_L^0 and muon subdetector at Belle II, their accurate identification can be challenging. Other neutral hadrons, like neutrons from beam background interactions, can also create KLM clusters. With limited other subdetector information to reject such a case, the K_L^0 modes are rarely used. They thus will not be discussed further.

3.4.1. The theory of hadronic and semileptonic *B*-decays

Similar to the choice of inclusive versus exclusive tagged decays within a particle search, implementation of hadronic or semileptonic B-tagging also requires considered thought.

It is generally considered that semileptonic B-tagging (i.e. a B_{tag} that decays according to $B \rightarrow X\ \ell\nu_\ell$, where the X system is at least one hadron) is a higher efficiency tagging method. This is due to their overall higher branching fractions and thus higher likelihood of occuring alongside a decay mode of interest. Within the 1157 B^+ Standard Model decays that are simulated in Belle II MC (i.e from the Belle II '.dec file' defined for the `EvtGen` physics generator), the modes with the highest estimated branching fractions are the $B^+ \rightarrow \overline{D}^{*0}\ \ell^+\ \nu_\ell$ for $\ell \in \{e^+, \mu^+\}$ at $\approx 5.5\%$ each. Implementing these as B_{tag} reconstruction modes, however, can diminish the purity of the analysis, due to the unreconstructable neutrino. Thus, the B meson four-vector cannot be completely determined and attributes usually available to assess accuracy of the B-tagging, such as B meson mass, are unavailable.

By definition, hadronic decay modes do not contain neutrinos. Thus a B_{tag} with a suitable hadronic decay mode can be fully reconstructed, allowing for stricter tag-based selection criteria that can ensure more pure reconstructed events. However, hadronic decays generally are lower in branching fraction, thus less probable to tag a signal B decay of interest. The highest branching fraction that is also a hadronic branching fraction in Belle II charged B MC is shared between $B^+ \rightarrow \overline{D}^{*0}\ a_1^+$ (a_1^+ is a spin-1 $u\overline{d}$ meson with mass $1230\,\text{MeV}/c^2$) and $B^+ \rightarrow D^{*0}\ \pi^-\ \pi^+\ \pi^+\ \pi^0$ at $\approx 1.8\%$ each. This is less than half of the semileptonic case.

In designing the FEI's exclusive tagging algorithm, the charged B decays utilised were a balance of both their branching fraction as well as ease of identification and reconstruction procedure within Belle II. Decay modes were limited to those with minimal successive decays, as well as to those that contain more common particles, such as those in Table 3.1 rather than ρ or a mesons. Thanks to V_{cb}, $b \rightarrow c$ transitions represent the large proportion of charged B decays. Thus, both the hadronic and semileptonic FEI B-tagging modes involve this transition, with most being specifically $B \to D^{(*)}$. The specific hadronic and semileptonic FEI modes are listed alongside their branching fractions in Table 3.2.

In the following sections, the efficacy of both the hadronic and semileptonic modes will be analysed in a short demonstration of FEI B-tagging capability in both decay mode classes. The goal of this demonstration is to investigate key observables in service of deciding the best FEI tagging mode for the $B^+ \to \mu^+\nu_\mu$ search.

The FEI will be applied to a sample of $200\,\mathrm{fb}^{-1}$ $e^+e^- \to \Upsilon(4S) \to B^+B^-$ MC-simulated decays. Utilising MC allows inspection of the 'truth-status' of the B-tagging, allowing assessment of how the FEI performs in correctly and incorrectly reconstructed B_{tag} decay modes. Such a check is not possible when applied to real collision data and so is an important part in examining trends in FEI performance before looking at data. It is ensured that this $200\,\mathrm{fb}^{-1}$ sample is different to the sample the FEI is trained with. In the MC simulated events, the B decays are 'generic' i.e. are free to decay via any allowable mode, save for the eventual signal mode, $B^+ \to \mu^+\nu_\mu$. As the current goal is to see how FEI B-tagging occurs generally, (i.e. for any signal B-decays), no signal-side reconstruction or selection has been specified. Moreover, no further restrictions have been applied other than the in-built post-selection criteria as well as *best candidate selection* for the B_{tag}.

Before investigating distributions of key observables of FEI B-tagging in generic B^+B^- decays, it is important to discuss the development of the best candidate selection method that has been applied in producing the distributions.

3.4.2. Best Candidate Selection of FEI B_{tag} candidates

After its application, the FEI algorithm returns a set of 'best' B_{tag}'s for an analyst to choose from per event. This can be up to 20 candidates for a given event. As there is ultimately only one true B that partners a B decay of interest, a prescription must be decided on for choosing the reconstructed candidate that fits the best, from the range of FEI B_{tag}'s provided. Furthermore, choosing a single candidate that is most likely to be the true B decay prevents the distributions of observables being biased towards events where multiple B reconstructions are likely, such as the high combinatoric events that will be seen within the hadronic FEI. Such a choice also addresses situations where the independently decaying B pairs both occur via taggable decay paths - which one acts as the B_{tag} is decided as which one performs better in best candidate selection. The aim of best candidate selection criteria is different to the selection criteria that will be explored in Chapter 4, which are implemented to completely remove events that mimic the decay of interest, $B^+ \to \mu^+\nu_\mu$. The variables investigated here may also be useful in this background rejection, but are not used to remove events entirely. By definition, the number of events remains the same, but the multiplicity of candidates will be reduced to one per event.

The common prescription is to use a variable correlated with better reconstruction to decide the B_{tag} candidate per event that is most likely to represent the underlying B decay. A natural variable to use for this case is the signal probability. This is the classifier produced as the result of the FEI's MVA for each B_{tag}. Before applying this method, i.e. choosing the B_{tag} candidate with the highest signal probability out of all B_{tag} identified for the event, it is important to test that such a candidate is indeed more likely to be correctly reconstructed. Moreover, it is of interest to determine if using signal probability is the best variable to use for candidate selection.

What follows is an analysis to determine which variable, inclusive of signal probability, will yield the highest proportion of correctly reconstructed FEI B_{tag} with a B^+B^- sample, when implemented as a best candidate selection variable. For all events in the sample, B_{tag} of the same event will be ranked according to a distinguishing variable. This will be applied to both the hadronic and semileptonic modes of the FEI. The FEI B_{tag} candidates will be truth-matched using `basf2`, to determine correctness. This will allow investigation into how the proportion of correct B_{tag} (i.e. purity) changes with rank. The best candidate selection itself will be to choose the rank-1 B_{tag} on the basis of this variable.

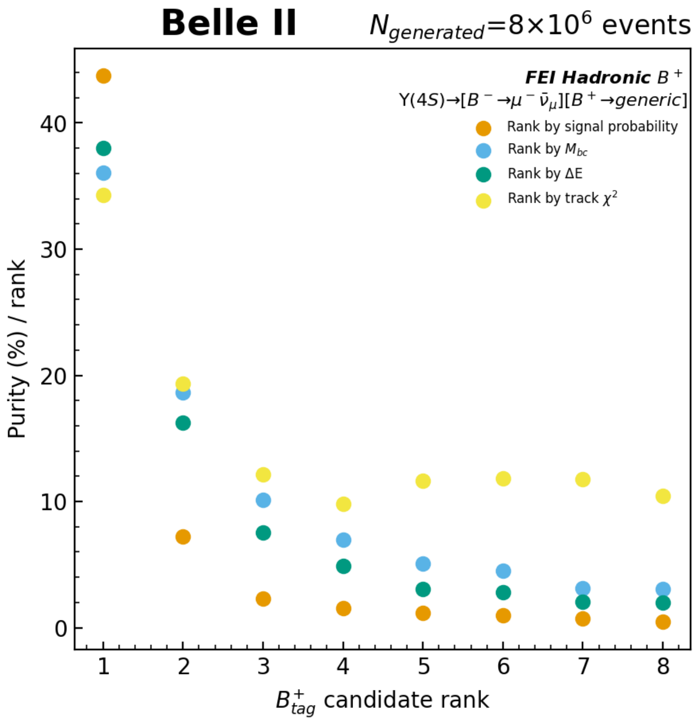
The variables considered against signal probability will vary depending on the FEI reconstruction mode, hadronic or semileptonic. Thus B_{tag} candidates for these variables will be ranked by lowest absolute differences between the variables and their nominal values, i.e. $|M_{\text{bc}}$- 5.285$|$ GeV/c^2 and $|\Delta E\,|$ GeV. For the semileptonic FEI, the two other variables used is the angle between the B_{tag} and their non-neutrino products, denoted as $\cos\theta_{BY}$, and the momentum of the charged lepton B_{tag} in the $\Upsilon(4S)$ frame, denoted as p^*_ℓ. In Section 1.4.2, the $\cos\theta_{BY}$ variable was seen to be a key discriminating variable in the BaBar 2010 semileptonic-tagged $B^+ \to \mu^+\nu_\mu$ search. The momentum of the charged lepton in the B_{tag} had played a lesser role in the BaBar search, though it will be investigated as a key variable of this analysis. Distributions of these variables will be presented in Section 3.4.3 for hadronic B_{tag} variables and Section 3.4.4 for semileptonic

B_{tag} variables. In this study, semileptonic FEI B_{tag} will be ranked according to lowest p^*_ℓ (less likely to be a lepton from a non-semileptonic B event) and the lowest absolute difference $|\cos\theta_{BY}$- 1$|$ (indicating lesser missing energy from neutrinos, as well as more likely to be a true single neutrino B_{tag}). Both modes will also investigate the use of the χ^2 probability of the vertex fit of the B_{tag} products (i.e. how likely it is the B_{tag} products originate from the same point in space) in best candidate selection. Vertex fitting, as a key step in reconstruction in the FEI, was introduced in Section 3.3

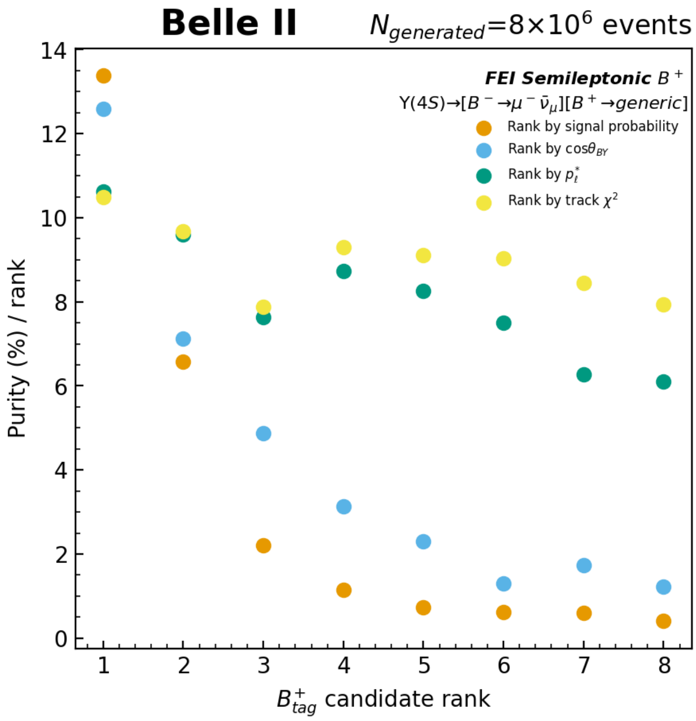
After this section, a $200\,\text{fb}^{-1}$ B^+B^- sample will be used to investigate the effect of the FEI generally, without a signal mode reconstructed. Thus it would be natural to investigate best candidate selection in the same sample. Unfortunately due to data loss, the corresponding samples and distributions are not currently available. As an alternative, the best candidate selection plots are shown for a sample of 8×10^6 $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ MC events, the signal MC of this thesis. The physics in this sample is identical to that of the $200\,\text{fb}^{-1}$ B^+B^- sample, aside from one of the charged B mesons fixed to decay according to $B^+ \to \mu^+\nu_\mu$ (or the charged conjugate). It is known from preliminary studies that the trends of the best candidate selection in the restricted $B^+ \to \mu^+\nu_\mu$ sample matches that of the generic B^+ B^- sample qualitatively, aside from the higher purity associated with the FEI in the $B^+ \to \mu^+\nu_\mu$ MC. There will be further comparison between the tagged-B mesons derived from the generic charged B MC and the signal MC in Section 3.4.5 and Section 3.5.1.

Figure 3.3a demonstrates the purity versus B_{tag} ranking within the same event for the hadronic FEI in $B^+ \to \mu^+\nu_\mu$ MC. The variable which admits the highest purity for its rank-1 hadronic B_{tag}'s is the signal probability. Moreover, looking at the purity of B_{tag} ranking up to 8th rank, the purity drops considerably more than other variables after rank 1. From first to second rank, the signal probability ranking falls from approximately 44% correct to 8% correct, below that of the equivalent purity in the next best ranking variable, M_{bc}. This reveals that there are fewer correct B_{tag} in the later ranks when using signal probability as the best candidate criterion. Thus, it is less likely that a B_{tag} that would be correctly truth-matched falls below ranking first. Choosing the highest signal probability B_{tag} per event is therefore better correlated with correct B_{tag} reconstruction, than any of the other variables trialled.

Figure 3.3b illustrates a similar trend between purity and B_{tag} ranking per event for the semileptonic FEI. The signal probability also performs better than the other FEI B_{tag} quantities, though with a lesser drop between rank 1 and rank 2 B_{tag}'s compared to the hadronic case. With a purity of around 14%, the purity of choosing the rank 1



(a) Truth-matching is via `isSignal=1` on hadronic B_{tag}'s. Ranking is according to the highest signal probability/ lowest $|M_{\mathrm{bc}}$- $m_B|$/lowest $|\Delta E\ |$/highest χ^2 probability of vertex fit, amongst all B_{tag}'s in the event.



(b) Truth-matching is via `isSignalAcceptMissingNeutrino=1` on semileptonic B_{tag}'s. Ranking is according to highest signal probability/ lowest $|\cos\theta_{BY}$- 1$|$/lowest p^*_ℓ/highest χ^2 probability of vertex fit, amongst all other B_{tag}'s in the event.

Figure 3.3. Purity of candidates (i.e. percentage correctly truth-matched) according to their variable ranking, in FEI B_{tag}'s identified from 8×10^6 $\Upsilon(4S) \rightarrow [B^+ \rightarrow \mu^+\nu_\mu$ $]B^+$ MC events.

(highest) signal probability B_{tag} of the event is larger than choosing B_{tag} with $\cos\theta_{BY}$ closest to 1, highest χ^2 of of the vertex fit or lowest p^*_ℓ. Using the χ^2 of the vertex fit for semileptonic B_{tag}'s as a ranking variable also proved difficult due to the large number of B_{tag} that failed the fit, which can be owed to the missing neutrino. In addition to this, the ranking as performed within `basf2` has taken place without allowing for duplicate values. B_{tag} that share the same value of a variable are arbitrarily assigned a rank above one another. Traditionally this would require an additional criterion to properly break the tie. Fortunately, it is found that there are no repeated values of signal probability within the whole sample, let alone within the same event. This was not the case for p^*_ℓ and $\cos\theta_{BY}$, further exhibiting signal probability as the simplest and most appropriate choice for best candidate selection.

It is important to note that the best candidate selection will need to be considered further when incorporating a signal side B decay. For the case of $B^+ \to \mu^+\nu_\mu$ in this search, this is not only to ensure the 'best' muon is adopted, but also to ensure that the best candidate procedure rejects B_{tag} reconstructed from signal side particles. This is ultimately the reason why the FEI algorithm is not fixed to yield a single 'best' B_{tag} according to signal probability, as it may be important to first reject candidates more likely to contain products of the desired signal B before ranking. Careful consideration will be given to avoid this in Chapter 4, when the full reconstruction of $\Upsilon(4S) \to B^-_{\text{tag}}[B^+ \to \mu^+\nu_\mu\ ]$ (and charge conjugate) takes place. In this case, the B_{tag} should be chosen such that it is less likely to contain the signal side μ.

For now, the best candidate procedure for FEI B_{tag}'s verified. Using only the highest signal probability candidates per event will now be applied to the assessment of the B-tagging in Belle II MC of B^+B^- events in the following section.

Table 3.2. Comparison of decay modes implemented in hadronic FEI (including the baryonic modes) and semileptonic FEI, with their branching fractions. Neutrinos are missing particles in the semileptonic FEI modes. Decay mode ID numbering begin from zero within the basf2 framework. Branching fractions are taken from the collaboration-managed Belle II decay file, which are either directly sourced from the PDG, or approximated from PDG inclusive branching fractions when necessary. The effective branching fraction of each FEI mode will be lower in application, owing to the finite set of sub-decay modes chosen for intermediate particles.

HADRONIC FEI	
0. $B^+ \to \overline{D}^0\ \pi^+$	4.68×10^{-3}
1. $B^+ \to \overline{D}^0\ \pi^+\ \pi^0$	5.00×10^{-4}
2. $B^+ \to \overline{D}^0\ \pi^+\ \pi^0\ \pi^0$	9.89×10^{-3}
3. $B^+ \to \overline{D}^0\ \pi^+\ \pi^+\ \pi^-$	5.10×10^{-3}
4. $B^+ \to \overline{D}^0\ \pi^+\ \pi^+\ \pi^-\ \pi^0$	1.60×10^{-3}
5. $B^+ \to \overline{D}^0\ D^+$	1.92×10^{-4}
6. $B^+ \to \overline{D}^0\ D^+\ K_S^0$	7.75×10^{-4}
7. $B^+ \to \overline{D}^{*0}\ D^+\ K_S^0$	1.03×10^{-3}
8. $B^+ \to \overline{D}^0\ D^{*+}\ K_S^0$	1.91×10^{-3}
9. $B^+ \to \overline{D}^{*0}\ D^{*+}\ K_S^0$	4.59×10^{-3}
10. $B^+ \to \overline{D}^0\ D^0\ K^+$	1.45×10^{-3}
11. $B^+ \to \overline{D}^{*0}\ D^0\ K^+$	2.26×10^{-3}
12. $B^+ \to \overline{D}^0\ D^{*0}\ K^+$	6.32×10^{-3}
13. $B^+ \to \overline{D}^{*0}\ D^{*0}\ K^+$	1.12×10^{-2}
14. $B^+ \to D_s^+\ \overline{D}^0$	9.01×10^{-3}
15. $B^+ \to \overline{D}^{*0}\ \pi^+$	4.90×10^{-3}
16. $B^+ \to \overline{D}^{*0}\ \pi^+\ \pi^0$	5.00×10^{-4}
17. $B^+ \to \overline{D}^{*0}\ \pi^+\ \pi^0\ \pi^0$	5.00×10^{-4}
18. $B^+ \to \overline{D}^{*0}\ \pi^+\ \pi^+\ \pi^-$	1.03×10^{-2}
19. $B^+ \to \overline{D}^{*0}\ \pi^+\ \pi^+\ \pi^-\ \pi^0$	1.80×10^{-2}
20. $B^+ \to D_s^{*+}\ \overline{D}^0$	7.60×10^{-3}
21. $B^+ \to D_s^+\ \overline{D}^{*0}$	8.22×10^{-3}
22. $B^+ \to \overline{D}^0\ K^+$	3.63×10^{-4}
23. $B^+ \to D^-\ \pi^+\ \pi^+$	1.07×10^{-3}
24. $B^+ \to D^-\ \pi^+\ \pi^+\ \pi^0$	2.00×10^{-3}
25. $B^+ \to J/\psi\ K^+$	1.01×10^{-3}
26. $B^+ \to J/\psi\ K^+\ \pi^+\ \pi^-$	5.51×10^{-5}
27. $B^+ \to J/\psi\ K^+\ \pi^0$	9.33×10^{-5}
28. $B^+ \to J/\psi\ K_S^0\ \pi^+$	9.33×10^{-5}
continues to baryonic modes	
29. $B^+ \to \overline{\Lambda}_c$- $p\ \pi^+\ \pi^0$	1.81×10^{-3}
30. $B^+ \to \overline{\Lambda}_c$- $p\ \pi^+\ \pi^-\ \pi^+$	2.25×10^{-3}
31. $B^+ \to \overline{D}^0\ p\ \overline{p}$- π^+	3.72×10^{-4}
32. $B^+ \to \overline{D}^{*0}\ p\ \overline{p}$- π^+	3.72×10^{-4}
33. $B^+ \to D^+\ p\ \overline{p}$- $\pi^+\ \pi^-$	1.66×10^{-4}
34. $B^+ \to D^{*+}\ p\ \overline{p}$- $\pi^+\ \pi^-$	1.86×10^{-4}
35. $B^+ \to \overline{\Lambda}_c$- $p\ \pi^+$	2.25×10^{-4}
TOTAL	$\mathbf{1.20 \times 10^{-1}}$

SEMILEPTONIC FEI	
0. $B^+ \to \overline{D}^0\ e^+\ (\nu_e)$	2.31×10^{-2}
1. $B^+ \to \overline{D}^0\ \mu^+\ (\nu_\mu)$	2.31×10^{-2}
2. $B^+ \to \overline{D}^{*0}\ e^+\ (\nu_e)$	5.49×10^{-2}
3. $B^+ \to \overline{D}^{*0}\ \mu^+\ (\nu_\mu)$	5.49×10^{-2}
4. $B^+ \to D^-\ \pi^+\ e^+\ (\nu_e)$	1.00×10^{-3}
5. $B^+ \to D^-\ \pi^+\ \mu^+\ (\nu_\mu)$	1.00×10^{-3}
6. $B^+ \to D^{*-}\ \pi^+\ e^+\ (\nu_e)$	1.00×10^{-3}
7. $B^+ \to D^{*-}\ \pi^+\ \mu^+\ (\nu_\mu)$	1.00×10^{-3}
TOTAL	$\mathbf{1.60 \times 10^{-1}}$

Table 3.3. Final FEI B_{tag} selections (post-selections).

Hadronic	Semileptonic
$M_{\text{bc}} > 5.2$ GeV/c^2	$p^*_\ell > 1$ GeV/c
$\lvert\Delta E\rvert < 0.3$ GeV	-4 $< \cos\theta_{BY} < 3$
$\log_{10}(\texttt{sigProb}) > -3$	$\log_{10}(\texttt{sigProb}) > -2.4$

3.4.3. Hadronic FEI and its B-tagging performance in B^+B^- MC

The hadronic FEI is the most popular FEI mode, where all six of the published FEI analyses have used the hadronic mode. Table 3.2 lists the 36 charged B decays implemented in the FEI. The larger number of tagging modes compensates for the lower branching fraction of the individual hadronic modes. The B^+ decay modes are restricted to contain ground-state, sub- GeV/c^2 mass hadrons and at most two charm hadrons. Such modes are easier to reconstruct, due to their lower number of intermediate reconstructions and well-known mass constraints. Decays with other hadrons that may even have larger branching fractions , like $B^+ \to \overline{D}^{*0}\ a_1^+$ or $B^+ \to \overline{D}^0\ \rho^+$, are not used. While they are not explicitly included, they may still be identifiable and thus taggable by the FEI, if the higher order states decay to an equivalent mode. For example, $B^+ \to \overline{D}^{*0}\ [a_1^+ \to \pi^+\ \pi^+\ \pi^-\ ]$ being tagged by $B^+ \to \overline{D}^{*0}\ \pi^+\ \pi^+\ \pi^-$.

Not only does the hadronic FEI decay modes only contain hadrons, but the intermediate particles also are reconstructed from mostly hadrons. No semileptonic subdecays are considered. Thus, the only place where leptons are involved in the final state of a B_{tag} decay mode is the charged dilepton decay of J/ψ, the properties of which will be soon discussed. Similarly, no neutrinos are present in the chosen modes, nor their subdecays.

The efficacy of the hadronic FEI applied to 200 fb^{-1} of $\Upsilon(4S) \to B^+\ B^-$ MC decays will now be investigated. The number of hadronic FEI B_{tag} candidates per decay mode is plotted in Figure 3.4. This histogram demonstrates that the most commonly identified modes are that of the $B^+ \to \overline{D}^0 n\pi$ modes, despite Table 3.2 listing their mid-range branching fraction. Their high tag rate in this sample could then be attributed to their simple reconstruction. This is owed to the single charmed intermediate reconstruction of $\overline{D}^0$and the track-based charged pions and cluster-based neutral pions. A trend that is shared between these modes as well as the excited $B^+ \to D^{*0}\ n\pi$, is that a higher

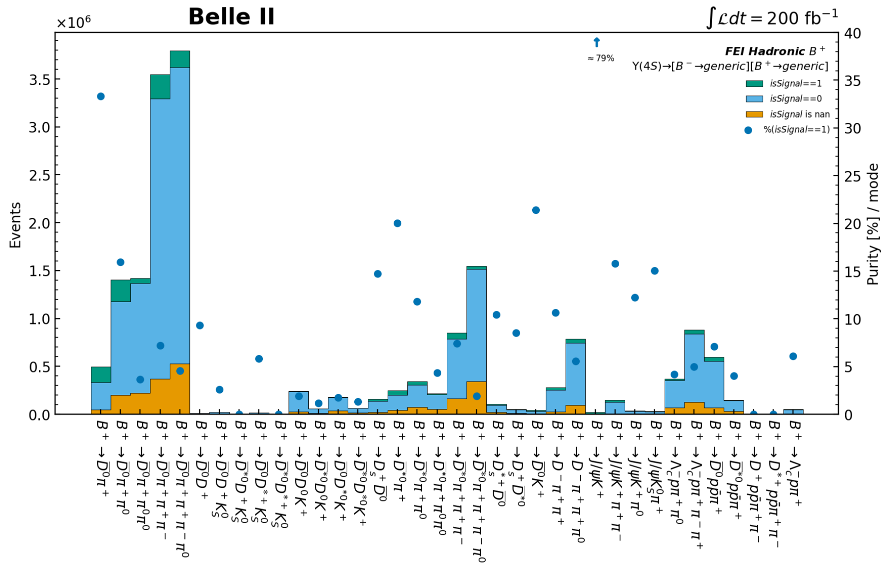


Figure 3.4. Reconstructed B_{tag} decay modes identified via hadronic FEI in $200\,\text{fb}^{-1}$ $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. Overlaid scatter points are the purity (percentage correctly truth matched) of each decay mode.

number of B_{tag} child particles is correlated with a larger number of events identified in this mode. This effect is expected, as large N-body decays are susceptible to *combinatoric background*. The dynamics of each pion in multi-π decay modes can vary substantially, and so permuting the candidate pions in each position within the decay can yield a unique B_{tag} candidate. This can lead to over-representation of B_{tag} candidates in these modes and higher likelihood that such a B_{tag} is mistaken by the FEI algorithm as the best candidate of the event. The high yields as the result of combinatoric background in the multi-π modes can be evidenced by the histogram stacking components. The histogram events of each mode are grouped according to the `basf2` truth-matching status variable suitable for fully-reconstructed final states, `isSignal`. As discussed in Section 2.4.3. B_{tag} candidates whose reconstruction fails to be matched yield `isSignal` is `NaN`, those that are matched but are misreconstructed are `isSignal`=0 and those that match and succeed are `isSignal`=1. With this legend, it is the single pion B_{tag} mode, $B^+ \to \overline{D}^0$ π^+, that has the overall largest number of truth-matched candidates, not the multi-pion ones. If instead the proportion of correctly truth-matched B_{tag} events per decay mode is considered, as per the overlaid plot of purity, the $B^+ \to J/\psi\ K(n)\pi$ modes perform the best. Of note is the purity spike specifically present in the $B^+ \to J/\psi\ K^+$ modes, the only mode with more correctly truth-matched events than incorrect ones. While these modes have the smallest B_{tag} branching fractions in Table 3.2, J/ψ are simply reconstructed via light lepton pairs. Moreover, the invariant mass distributions of the dilepton system is generally very narrow, and thus easily identifiable by the FEI with respect to other decays. In contrast, the D meson masses are also well known, but not as narrowly due to their varied decay paths with larger combinations of kaons and pions.

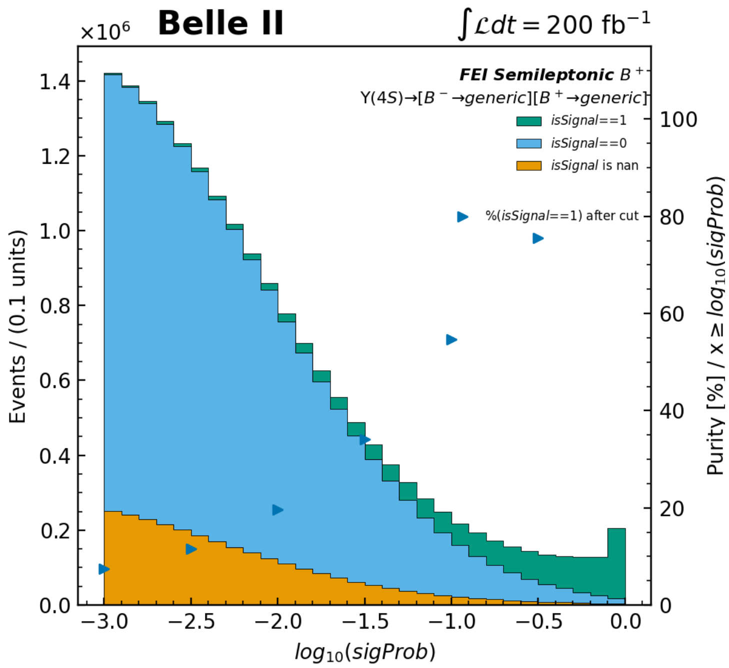


Figure 3.5. log10(signalProbability) B_{tag} decay modes identified via semileptonic FEI in 200 fb^{-1} $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. Overlaid scatter points are the purity (percentage correctly truth matched) obtained when increasing the minimum allowed $\log_{10}$(signalProbability).

The ability of the FEI classifier `sigProb` to identify hadronic charged-B processes in $\Upsilon(4S) \to B^+B^-$ decays is examined in Figure 3.5. `sigProb` is trained to be a continuous analogue to `isSignal`, with `sigProb` values closer to 1 suggesting the reconstructed decay has a larger probability of being correct and closer to 0 being incorrect. Signal probability will be analysed as $\log_{10}$(`sigProb`), as a log-transformed quantity where more negative values indicate poor reconstruction and closer to zero indicates better reconstruction. Use of the log-scale shows better demarcation of events favoured and disfavoured by the FEI. As listed in Table 3.3, B_{tag} candidates in the hadronic FEI are only accepted if they achieve `sigProb` > 0.001, or `log`$_{10}$`(sigProb)`$>$ -3.

The distribution demonstrates a larger proportion of correctly reconstructed B_{tag} for higher values of assigned signal probability. Overlaid on this distribution is a plot of how the purity of the B-tagged events changes if the $\log_{10}$(`sigProb`) lower bound is tightened. It is demonstrated that the proportion of correct B_{tag} samples is indeed correlated with tightening the signal probability lower bound, the behaviour expected as a multivariate classifier. Thus, `sigProb` does indeed behave as a reasonable proxy for correctness, as intended by its training. The purity bounds are chosen for illustration purposes; it is generally not recommended to limit the analysis of B_{tag} candidates with a lower bound as tight as $\log_{10}$(`sigProb`) $>$ -1. Such a harsh selection could lead to data 'shaping' in other key B_{tag} variables, where the only background events remaining are those that exist in the signal-like regions of the dataset, leaving no additional variable features to perform further signal to background demarcation. The affect of data shaping from the FEI process is already seen in the distributions of some later B_{tag} variables in this section, where misreconstructed B events are in the same range as correctly reconstructed ones.

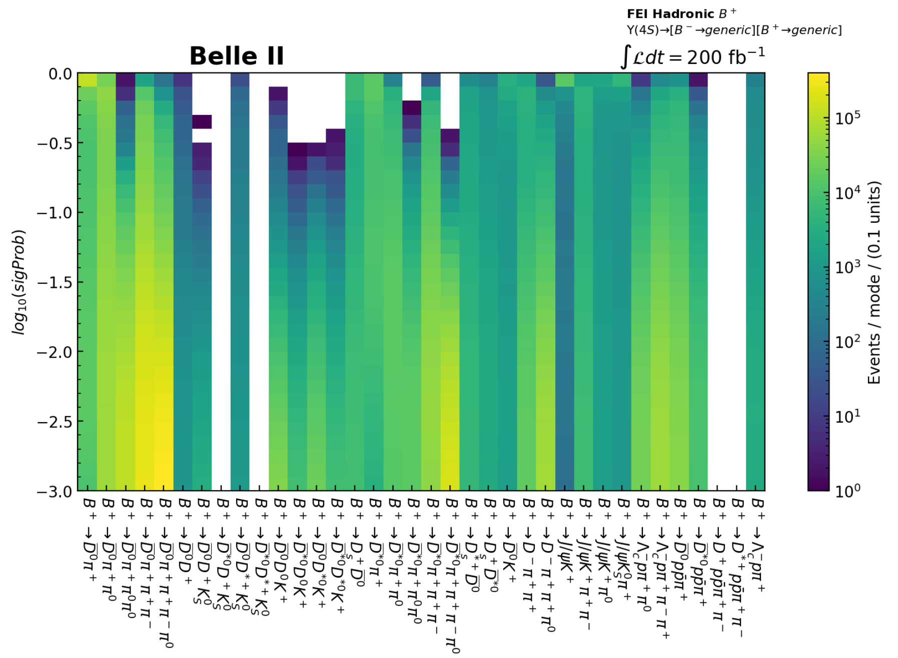


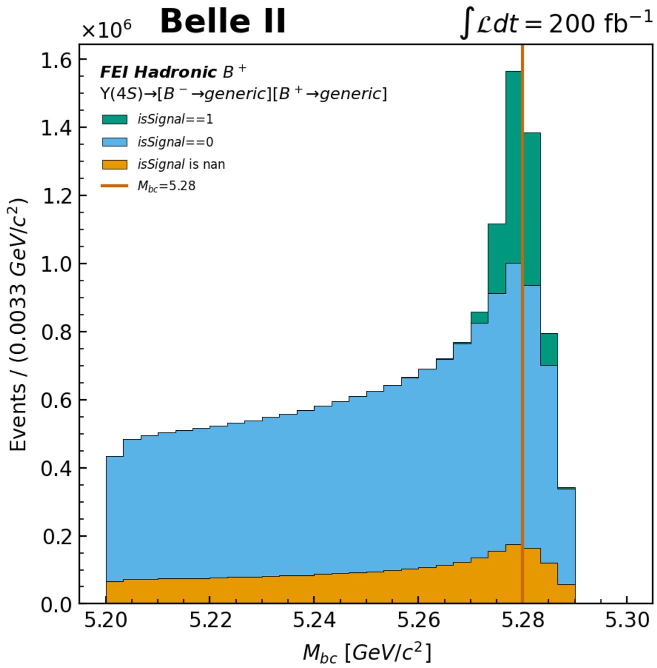
Figure 3.6. A 2D histogram of $\log_{10}$(`sigProb`) according to identified hadronic FEI B_{tag} decay modes in $200\,\text{fb}^{-1}$ $\Upsilon(4S) \to B^+B^-$ generic MC events.

A tagmap is a 2D distribution that can be used to examine how quantities may vary between decay modes. The specific tagmap in Figure 3.6 is a 2D histogram of the signal probability in hadronic FEI B^+-tagging per decay mode. Here, it is more obvious which decay modes are favoured by hadronic FEI. While the number of events tagged as $B^+ \to \overline{D}^0\pi^+\ \pi^0\ \pi^0$ decrease by many orders of magnitude for higher `sigProb`, the $B^+ \to J/\psi\ K^+\ (n\pi)$ modes are some of the few that continually increase towards larger signal probability. Such a tagmap could therefore be used to determine mode-specific selections to enhance purity, such as optimising the minimum allowed `sigProb` for these well-identified decays. Given the early stage of the Belle II dataset, where analyses are prioritising event-tagging retention over absolute purity, there have been no studies into decay mode specific selection criteria beyond what has been performed by the FEI. Once the integrated luminosity of Belle II reaches the attobarn scale, there will be sufficient data so that such a technique could be considered. There would also be potential to limit analyses to only using B_{tag} decays with the highest purity of the 36 available modes.

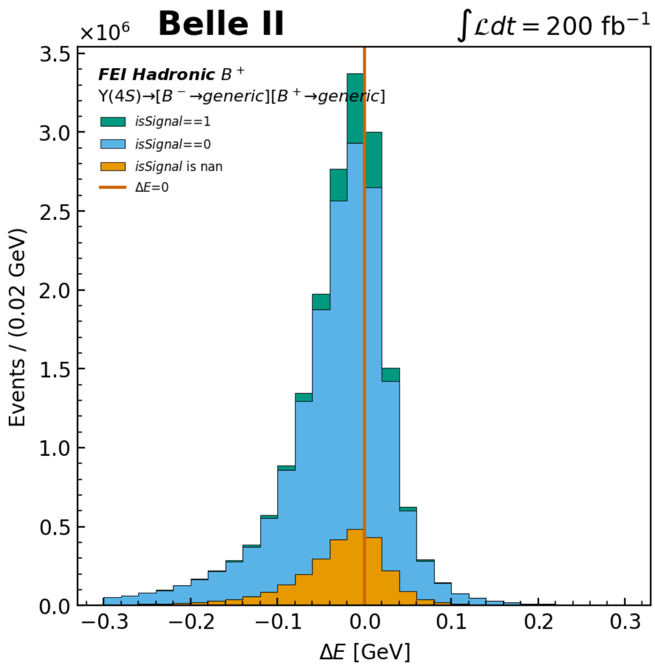
The tagmap of Figure 3.6 is also a useful way of determining which modes have not been identified in B-tagging and which are just low yield, like in $B^+ \to J/\psi\ K^+$ As the minimum threshold of the histogram is one event, it is obvious that the $B^+ \to \bar{D}^{*0} D^{(*)+} K_S^0$ and the $B^+ \to D^{(*)+} p\bar{p}\pi^+\ \pi^-$ modes have not been identified for any value of $\log_{10}($`sigProb`$)$ in this sample, despite similar modes having reasonable yields. It may also be the case that they have been reconstructed, but not survived the $\log_{10}($`sigProb`$)$, M_{bc} or ΔE post-selections, specified in Table 3.3. This is despite the $B^+ \to \overline{D}^{*0}\ D^{*+}\ K_S^0$ representing one of the larger branching fractions amongst those implemented within the hadronic FEI. More investigation is needed, to investigate if this behaviour is expected or if the modes are to be omitted.

In addition to `sigProb`, the quality of hadronic FEI reconstruction can also be assessed through other variables, such as M_{bc} and ΔE. These are the beam-constrained variables, where it was shown in Section 2.2.2 that the symmetry of the $e^+e^- \to \Upsilon(4S)$ $B\overline{B}$ process led to being able to set the B meson energy to be the centre-of-mass frame beam energy (i.e. $E_B^* = \sqrt{s}/2$). Doing so allows for the definition of M_{bc} from Eq. (2.11), thus calculating B meson mass from the beam energy and reconstructed momentum to obtain M_{bc}. The accuracy of this can be investigated by comparing how closely the B meson energy matches that of the beam, i.e. testing if ΔE =0.

Examples of these distributions for M_{bc} and ΔE are given in Figure 3.7a and Figure 3.7b. The distribution has no values below M_{bc} =5.2 GeV/c^2, due to the post-FEI selection, as listed in Table 3.3, well below the nominal B mass of $m_B = 5.28$ GeV/c^2. M_{bc} values closer to this mass indicate that the total reconstructed three-momentum of the B_{tag} is consistent with a B meson-beam energy hypothesis. For M_{bc} there is a large amount of failed truth-matching and incorrectly reconstructed B_{tag} before the 5.28 GeV/c^2 spike of `isSignal`=1 events. Following from the mathematical definition of M_{bc}, the sub-5.28 GeV/c^2 events must exceed the $|p_B^*|$= 0.331 GeV/c condition to yield lower values of M_{bc}. High momentum reconstructions confused for true B events will also be a feature of FEI B_{tag} misidentification in non-$\Upsilon(4S)$ events, such as $e^+e^- \to q\bar{q}$, which



(a) M_{bc}: The PDG B^{+} mass is indicated by the vertical line.



(b) ΔE: The desired value of ΔE =0 for well-reconstructed B_{tag} is indicated by the vertical line.

Figure 3.7. Beam constrained parameters of hadronic FEI B_{tag} in $200\,\mathrm{fb}^{-1}$ $\Upsilon(4S) \to B^{+}B^{-}$ generic MC events. Stacking is by truth-matching status.

will be seen in Section 3.5.2. While the M_{bc} distribution for FEI B_{tag} identified in B^+B^- processes is near bounded above by the nominal B meson mass, the ΔE distribution is more-symmetric about the nominal value of 0. The B_{tag} candidates that have been identified by the hadronic FEI algorithm are produced with a $|\Delta E\,| < 0.3\,\mathrm{GeV}$ selection, to reduce the number of misreconstructed events considered. `isSignal`=1 events have generally less than 0.1 GeV difference from beam energy. The slight tail in the negative ΔE region, however, may mean however that misreconstructed B_{tag} are more likely to be below that of the nominal B energy.

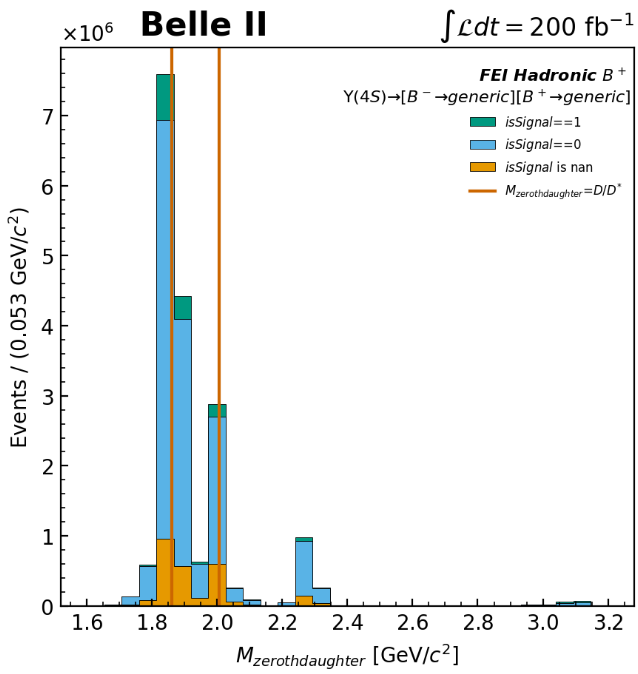


Figure 3.8. Measured mass zeroth daughter B_{tag} (i.e. D/D^*) identified via hadronic FEI in $200\,\mathrm{fb}^{-1}$ $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. The approximate PDG masses of D^+/D^0 and D^{*+}/D^{*0} are indicated by the vertical lines.

The *zeroth daughter* is the nomenclature used to refer to the 'first' child particle in a given written decay, such as the $\overline{D}^0$ in $B^+ \to \overline{D}^0\,\pi^+$. The ordering of daughter particles is merely a convention, though for all of the hadronic FEI decay modes, the zeroth daughter is a charm hadron that requires multiple levels of intermediate reconstruction.

The distribution of hadronic FEI B_{tag} zeroth daughter mass is demonstrated in Figure 3.8. Investigating distributions such as this one can be used to determine if the particles used in B_{tag} reconstruction match that of the precisely determined charm hadron masses. Also, it is evident from the distribution that the J/ψ modes (m=3.1 GeV/c^2) and newly added baryonic Λ_c^+ (m=2.3 GeV/c^2) modes are less frequently tagged compared to the standard D/D^* modes.

The 4.8 MeV/c^2 difference between D^+ and D^0 masses and the 2.6 MeV/c^2 difference between the D^{*+} and D^{*0} masses are not discernible in 200 fb^{-1} with the scale of this distribution, nor when zoomed in. With further selection criteria, the mass splitting could be made more obvious. For a sufficiently large dataset, fine enough binning can be implemented to resolve their masses. Moreover, the proximity of the D_s^+ mass of $m = 1.97$ GeV/c^2 to that of the D^* mass peak makes the D_s^+/D^* system irresolvable in this distribution. However, the general mass difference between the combined $D = D^+/D^0$ peak and $D^* = D^{*+}/D^{*0}$ peak can be discerned.

The D^{*0}-D^0 mass splitting, $\Delta m = 142$ MeV/c^2 is known to sub- MeV/c^2 precision. It is traditionally measured within a single decay chain, such as via $D^{*0} \to D^0\ X$ for $X = \pi^0, \gamma$ [99]. Thus, measuring the mass difference in the B_{tag} zeroth daughter and the zeroth daughter's zeroth daughter (which will be referred to as *zeroth granddaughter*) can provide further validation for hadronic FEI modes containing the $D^{*0} \to D^0$ transition.

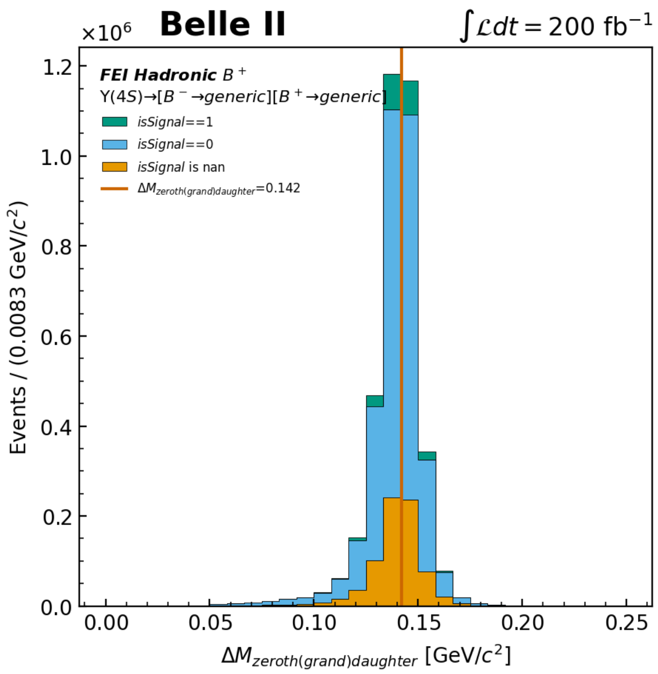


Figure 3.9. Difference in measured masses between zeroth daughter and zeroth granddaughter B_{tag} identified via hadronic FEI in 200 fb^{-1} $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. The horizontal axis has been restricted to focus on $\Delta M = m_{D^*} - m_D$. Thus, only B_{tag} decay modes of the form $B \to D^* \to D$ are considered here.

Figure 3.9 demonstrates the difference in the measured masses between the zeroth daughter and zeroth granddaughters of hadronic FEI B_{tag}. The distribution's range is limited to less than 250 MeV/c^2, thus inclusive of D^{*0}/D^0 mass splitting. For events in this range, it is seen that those closer to the nominal value of ΔM= 142 MeV/c^2 are indeed associated with a larger proportion of correctly truth-matched events. This peak demonstrates that the FEI has performed as desired for B_{tag} containing the $D^{*0} \rightarrow D^0$ transition. B_{tag} decay modes with transitions from other zeroth daughters, such as D^0, D^+, J/ψ, Λ_c^+ are still included in the overall analysis, but are omitted by the range of the distribution. They yield mass splittings that are larger in magnitude due to the mass difference between the ground state charmed hadrons and their sub- GeV/c^2 products. They are also less precisely known in literature and thus less effective as a validation of B-tagging. The given range, however, is inclusive of other excited-D transitions included in the FEI, such as $D^{*+} \rightarrow \overline{D}^0$ (145 MeV/c^2), $D^{*+} \rightarrow D^+$ (141 MeV/c^2) as well as $D_s^{*+}\rightarrow D_s^+$ (144 MeV/c^2). Any peaks due to B_{tag} of these individual transitions are not resolvable. The restriction thus includes 12 out of the 36 hadronic FEI decay modes. By definition, the distribution also doesn't include cases where the D^* is a non-zeroth daughter, such as in $B^+ \rightarrow \overline{D}^0\ D^{*0}$.

The performance of the hadronic FEI in identifying charged B_{tag} candidates in generic $\Upsilon(4S) \rightarrow B^+B^-$ events has been investigated. From the peaking correctly-truth matched tagged events around M_{bc} =5.28 GeV/c^2 as well as $\log_{10}(\texttt{sigProb})$ approaches zero, it can be seen that obtaining B_{tag}'s via the FEI's hadronic modes is well founded. Thus the hadronic FEI could be used to tag B mesons that occur alongside $B^+ \rightarrow \mu^+\nu_\mu$ decays.

3.4.4. Semileptonic FEI and its *B*-tagging performance

While the hadronic mode of the FEI is limited to reconstructing B_{tag} directly from hadrons, the semileptonic version utilises decay modes that contain both hadrons and leptons. Moreover, the charmed hadron present in the initial decay mode is fixed to be D-type mesons with isospin, i.e. not baryons, J/ψ's or D_s^+ as in the hadronic FEI. The specific semileptonic modes implemented in charged B meson reconstruction via the FEI are listed in Table 3.2. Each implemented semileptonic decay covers a much larger branching fraction of B decays, and thus less decay modes than what is implemented in the hadronic FEI are needed to obtain a suitable B_{tag} candidate. Whereas in the FEI the hadronic decay mode with the highest theoretical B^+ branching fraction, $B^+ \rightarrow \overline{D}^{*0}$ a_1^+, was not practical to implement, the semileptonic counterpart ($B^+ \rightarrow \overline{D}^{*0}\ \ell^+\ \nu_\ell$ for

$\ell^{\pm}$=$e^{\pm}, \mu^{\pm}$) are key modes in the semileptonic FEI. The chosen modes theoretically cover approximately 16% of all B decays. When accounting for the intermediate decays modes in the FEI, this reduces to 4.78%, mostly due to only considering hadronic D decays. One has the option with the FEI to allow for semileptonic D modes (mostly of the form $D \to K\ell\nu$) to increase efficiency, though the additional tag-side neutrino can complicate variables which assess B_{tag} quality, which will be discussed soon. The charged leptons considered are only electrons or muons, not taus. This is because tau leptons require reconstruction from FSPs, and create additional tag-side neutrinos.

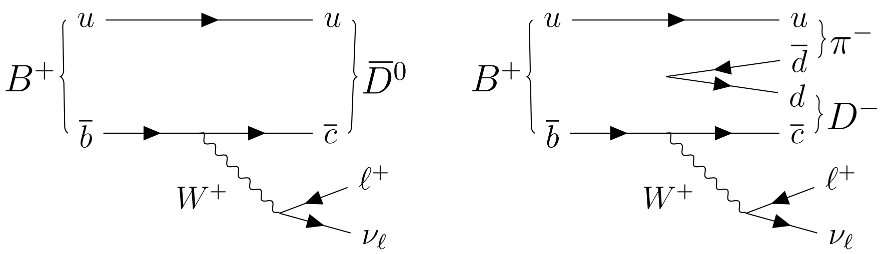


Figure 3.10. Feynman diagram of semileptonic FEI B^-_{tag} decay modes: $B^-_{\text{tag}} \to D^{(*)0}\ell^+\ \nu_\ell$ (left) and $B^-_{\text{tag}} \to D^{(*)-}\pi^+\ \ell^+\ \nu_\ell$ (right), for $\ell \in$\{e,μ\}.

The first four decay modes *resonant* decays, where the D meson is created alongside the charged-neutral lepton pair. The further decay modes are *non-resonant* decays, where the charged B meson decays directly to a four-body system, inclusive of a charged pion. As non-resonant processes, there is no known substructures within these B^+ decays i.e. the $D^{*-}\ \pi^+$ does not originate from an intermediary hadron. The Feynman diagrams of these processes are demonstrated in Figure 3.10

The key difference in using semileptonic decay modes is the undetectable neutrino. Thus the B_{tag} reconstruction is inherently incomplete, creating a four-vector that sums the hadronic system (i.e. D and π when necessary) and the charged lepton. Thus to 'tag' in this case means technically that the hadrons and charged leptons of the underlying semileptonic B_{tag} mode have been found (save for the missing neutrino), and technically not the full B_{tag} four-vector.

With knowledge of the included decay modes of the semileptonic FEI, its efficacy in identifying B_{tag} in 200 fb^{-1} of $\Upsilon(4S) \to B^+\ B^-$ MC decays at Belle II can now be investigated. Figure 3.11 demonstrates the number of B_{tag} events per semileptonic decay mode that were identified in this study. The events in these distributions are also grouped according to truth-matching status. This time, `isSignalAcceptMissingNeutrino` is now used to consider correctly tagged events, a variable which was introduced Section 2.4.3.

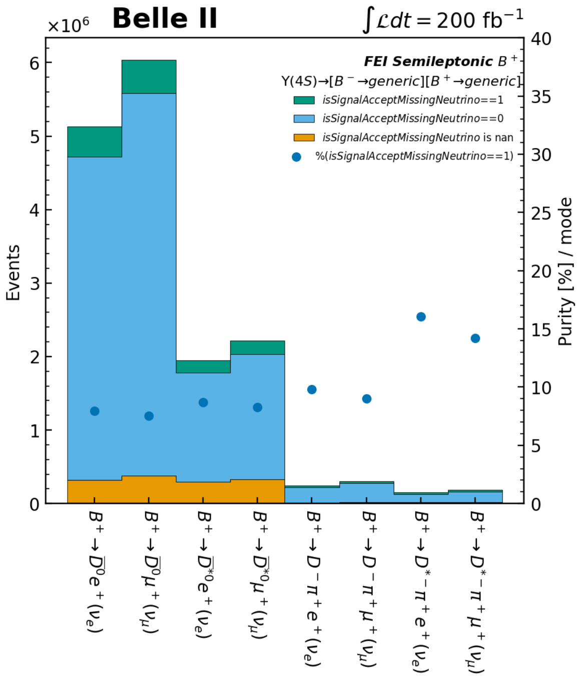


Figure 3.11. Reconstructed B_{tag} decay modes identified via semileptonic FEI in 200 fb^{-1} $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. Overlaid scatter points are the purity (percentage correctly truth matched) of each decay mode.

Using this `basf2` variable to check truth-matching of FEI B_{tag} candidates allows events to be deemed correctly reconstructed even in absence of the unrecontructable neutrino.

The decay mode most commonly tagged by the semileptonic FEI in this sample is one of the resonant modes, as expected from their higher branching fractions in Table 3.2. Specifically, it is the $B^+ \to \overline{D}^0\ \mu^+\ \nu_\mu$ mode that is identified the most, with approximately six million events containing the B_{tag} decay. By comparison, the most frequently tagged hadronic FEI mode, $B^+ \to \bar{D}^0\pi^+\pi^+\pi^-\pi^0$, occurs in two-thirds the number of events as its semileptonic counter part. This illustrates the higher event efficiency of the semileptonic decay modes. Moreover, the muon modes of the semileptonic FEI demonstrate a higher number of identified events than the elctron ones in Figure 3.11. This difference is not due to the underlying physics; lepton universality means the branching fractions of the decays with muons are identical to those with electrons. The difference in number of events may be attributed to difference in detector efficiency in muons and electrons, or a preference for tagging muon over electron modes within the training of the FEI. If instead the number of electron vs muon modes are compared in the truth-matched components of the histogram, the difference is less prominent, as is expected from lepton universality. Any potential preference of tagging semimuonic over semielectronic events will need to be considered if it is used with a signal B decay that contains muons (i.e. like $B^+ \to \mu^+\nu_\mu$), to avoid the signal-side muons being used in B_{tag} reconstruction.

Overlaid in this distribution is the percentage purity of each semileptonic FEI B-tagged decay mode, where the proportion of correctly identified B_{tag}'s also does not seem to change between electronic and muonic modes. It is of interest that the multibody non-resonant modes demonstrate higher purity than the resonant ones. The opposite effect was seen in the multibody decays of the hadronic FEI in Figure 3.4, though in this case this could be a potential lower branching fraction and thus low statistical effect. The reliability of the non-resonant modes will also be challenged in Chapter 4 as it will be seen there that they rarely pass basic signal-enhancing selection criteria. Overall, the included modes yield 5-15% and it is clear when compared to the hadronic version in Figure 3.4 that the semileptonic FEI is sacrificing purity of B_{tag} selection for increased initial yield.

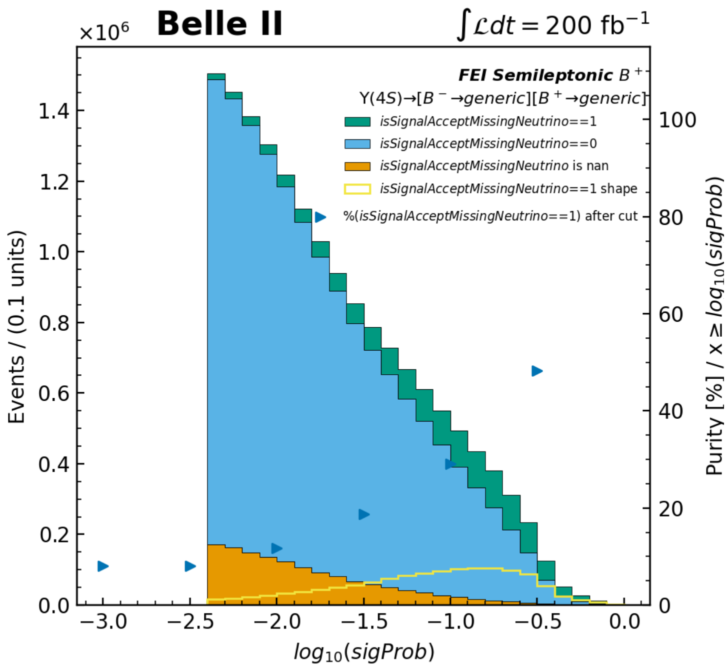


Figure 3.12. log10(signalProbability) B_{tag} decay modes identified via semileptonic FEI in $200\,\text{fb}^{-1}$ $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. Overlaid scatter points are the purity (percentage correctly truth matched) obtained when increasing the minimum allowed $\log_{10}$(signalProbability).

Figure 3.12 is a histogram of $\log_{10}$(`sigProb`) for the events with identified semileptonic B_{tag}. The distribution begins at $\log_{10}$(`sigProb`) > -2.4 due to the FEI post-selection criteria outlined in Table 3.3. The `isSignalAcceptMissingNeutrino` =1 component shows a much smaller increase in the number of B_{tag} in larger `sigProb`, when compared to the hadronic version in Figure 3.4. Moreover, unlike the hadronic FEI, the number of correctly truth-matched events does not continually increase and significantly peak as $\log_{10}$(`sigProb`) the approaches 0 (i.e. `sigProb` approaching 1). Instead, the number of correctly truth-matched B_{tag} per binned $\log_{10}$(`sigProb`) achieves its maximum between $\log_{10}$(`sigProb`) =-1.5 and $\log_{10}$(`sigProb`) =-0.5 (`sigProb` ≈0.03 and `sigProb` ≈0.3 respectively), before decreasing to a negligible number of candidates.

When comparing the overlaid purity for increasing $\log_{10}$(`sigProb`) lower bound plot, the semileptonic version generally demonstrates lower purity for the same lower bound than their hadronic counterpart, as well as the overall smaller number of events assigned to the same value of signal probability. However, it is seen that the purity does still increase for $\log_{10}$(`sigProb`) lower bound increasing up to -0.5, indicating `sigProb` is still a reasonable proxy for correctness in this limited region.

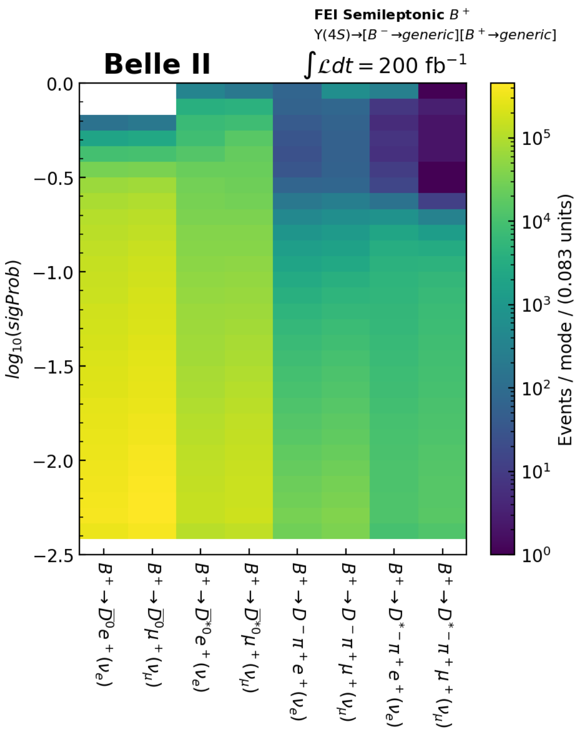


Figure 3.13. A 2D histogram of $\log_{10}$(`sigProb`) according to identified semileptonic FEI B_{tag} decay modes in 200 fb^{-1} $\Upsilon(4S) \to B^+B^-$ generic MC events.

The tagmap in Figure 3.13 demonstrates the signal probability distribution for the 8 semileptonic FEI decay modes. It gives further insight into if the overall trend of $\log_{10}$(`sigProb`) varies between decay modes, i.e. that there is a peak well before zero, leaving the final bin ($\log_{10}$(`sigProb`) >= -0.1, `sigProb` ≈0.79) with a negligible number of events. The log-scale of events indicates any increase in the number of events for up to $\log_{10}$(`sigProb`) =-1 is within the same magnitude i.e. a relatively small increase. While near-all modes do not increase again after this point, the $B^+ \to D^- \pi^+ \mu^+ (\nu_\mu)$ and $B^+ \to D^{*-} \pi^+ \mu^+ (\nu_e)$ peak in the zero bin, with around 1000 events in each bin. Given the sudden increase from around 100 events before this bin, this could be an indication of overtraining in these decay modes.

The most obvious drawback of the semileptonic implementation of the FEI compared to the hadronic one is the presence of the undetectable neutrino within the Belle II detector. With this missing momentum contribution in B_{tag} decay products, the B_{tag} four-vector is therefore incompletely reconstructed. Thus, the B_{tag} energy is now underestimated, no longer narrowly distributed about $E^*_B = 5.28\,\text{GeV}/c^2$.Additionally, the equivalence of the reconstructed B and beam energies no longer holds. The beam-constrained parameters, M_{bc} and ΔE, are therefore not be appropriate for assessing the semileptonic B_{tag} correctness. This is demonstrated in the distributions of Figure 3.14, where `isSignalAcceptMissingNeutrino` =1 semileptonic B_{tag} peak in the regions of M_{bc} and ΔE which would indicate misreconstruction, if compared against hadronic FEI B_{tag}. This is also where other backgrounds, such as events identified in $e^+e^- \to c\bar{c}$ will peak, making demarcating between B^+B^- and other processes difficult.

Instead, semileptonic FEI efficacy can be probed by using a variable tailored to identifying single missing neutrinos in B decays. This alternate observable for assessing B_{tag} is the angular variable $\cos\theta_{BY}$, which was initially presented when discussing the BaBar 2010 $B^+ \to \mu^+\nu_\mu$ search in Section 1.4.2. While the B four-vector and thus B direction not directly measurable for the semileptonic B_{tag}, the angle between the $Y = D^{(*)}(\pi)\ell$ and B momentum, θ_{BY}, can be derived. It is derived in the following, again due to the constrained kinematics in the collision frame of $e^+e^- \to \Upsilon(4S) \to B\overline{B}$.

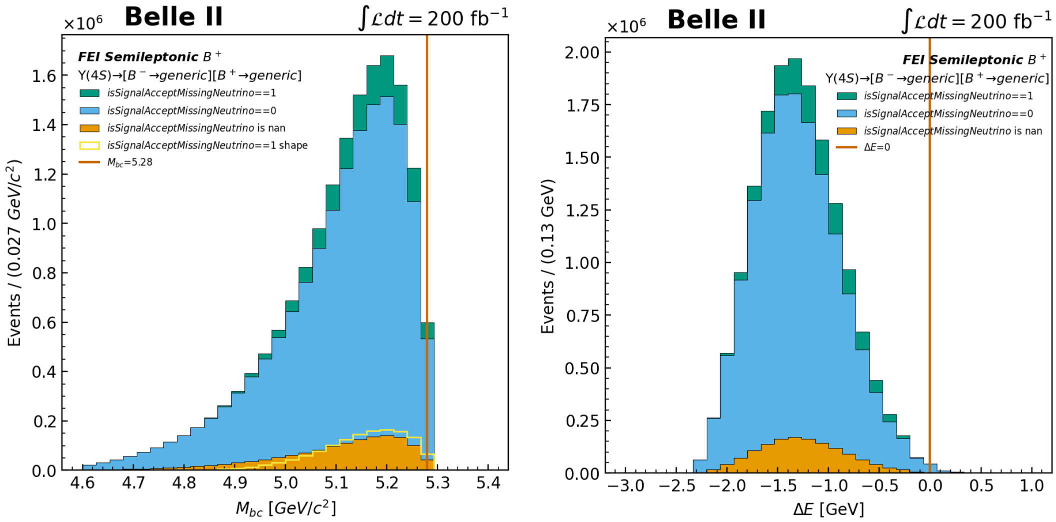


Figure 3.14. Beam constrained mass and ΔE measurements of semileptonic FEI B_{tag} in $200\,\text{fb}^{-1}$ $\Upsilon(4S) \to B^+ B^-$ generic MC events. Stacking is by truth-matching status. M_{bc} and ΔE are not appropriate variables for assessing B_{tag} quality in semileptonic FEI, as they assume no missing B_{tag} daughters. The desired values of B^+ mass and ΔE for well-reconstructed B_{tag} is indicated by the vertical line. Contrast the distribution with hadronic case in Figure 3.7a and Figure 3.7b.

$$
\begin{aligned}
p_B^* &= p_Y^* + p_\nu^* \\
\implies p_B - p_Y &= p_\nu \\
(p_B^* - p_Y^*)^2 &= m_\nu^2 = 0 \\
m_B^2 + m_Y^2 - 2p_B^* \cdot p_Y^* &= 0 \\
m_B^2 + m_Y^2 &= 2(E_B^* E_Y^* - \vec{p}_B^* \cdot \vec{p}_Y^*) \\
m_B^2 + m_Y^2 &= 2(E_B^* E_Y^* - |\vec{p}_B^*||\vec{p}_Y^*|\cos\theta_{BY}) \\
\cos\theta_{BY} &= \frac{2E_B^* E_Y^* - m_B^2 - m_Y^2}{2|\vec{p}_B^*||\vec{p}_Y^*|}
\end{aligned}
$$

Thus the angle between the Y system and the constrained B momentum expression as $\cos\theta_{BY}$ depends only on kinematically constrained or directly measurable quantities, not on the undetectable neutrino parameters. As with M_{bc} and ΔE, E_B^* is again implemented as E_{beam}^* (where $E_{\text{beam}}^* = \sqrt{s}/2$), thus also fixing $|p_B^*|$. The only unknowns in calculating

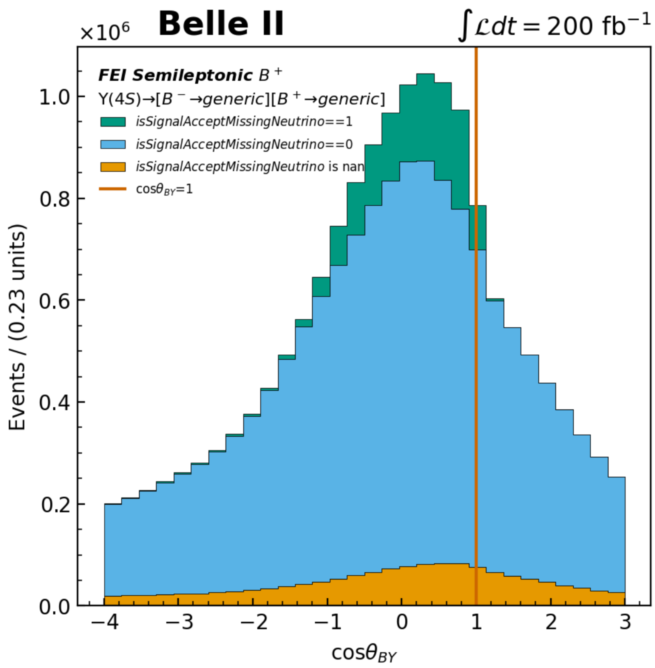


Figure 3.15. $\cos\theta_{BY}$ measurements of semileptonic FEI B_{tag} in 200 fb^{-1} $\Upsilon(4S) \to B^+ B^-$ generic MC events. Stacking is by truth-matching status. The theoretical value of $\cos\theta_{BY}$ =1 for B_{tag} with no missing energy is indicated by the vertical line.

$\cos\theta_{BY}$ have been reduced to that which is detectable in a semileptonic B_{tag}, the visible Y system.

The power of using $\cos\theta_{BY}$ to probe the correctness of semileptonic FEI B_{tag} reconstruction lies in the mathematical boundaries of the cos function. $\cos\theta_{BY}$ measured to be outside the interval [-1,1] indicate the Y system of the B_{tag} is inconsistent with a true B event. Such values of $\cos\theta_{BY}$ correspond to non-physical angles between the true B and required Y system, thus indicating a misidentified Y in the B_{tag} reconstruction. In practice, momentum smearing of detected particles can also lead to experimental $\cos\theta_{BY}$ slightly outside the [-1,1], indicating there needs to be some allowance in the $\cos\theta_{BY}$ bounds adopted. This can be seen in the $\cos\theta_{BY}$ distribution of FEI semileptonic B_{tag} of $e^+e^- \to B^+B^-$ processes in Figure 3.15, where the correctly identified B_{tag} extend between [-1.5,1.5]. Near-all B_{tag} outside of this range are truth-matched incorrectly, thus demonstrating the use of this variable in rejecting misreconstructed decays.

The centrally-produced sample is limited with a post-selection criterion $\cos\theta_{BY}$ ∈[-4,3], rejecting the worst FEI B_{tag} candidates. Notably, both the `isSignalAcceptMissingNeutrino` =1 and `isSignalAcceptMissingNeutrino` =0 distributions are not flat, instead peaking within the theoretical [-1,1] bounds. This demonstrates that the FEI is prioritising events with B_{tag} within the physical region of $\cos\theta_{BY}$, a desired effect for correct reconstruction but less so for misreconstruction. With this peak within the correctly reconstructed zone, this indicates further selection criteria will need to be considered alongside $\cos\theta_{BY}$ to determine correct semileptonic FEI B-tagging.

The derivation in Eq. (3.4.4) also indicates that assumption of $\cos\theta_{BY} \in$[-1,1] for correctly identified semileptonic B decays relies heavily on the invisible system being a single neutrino. The inclusion of the semileptonic intermediate D decays, as discussed in the beginning of this section, to the semileptonic B_{tag} would introduce a second invisible four-vector. Thus the requirement that the total invisible system's invariant scalar amounts to zero no longer holds and a dependence on the energy of the produced neutrinos enters the equation. As such, the semileptonic D modes will not be used for this analysis.

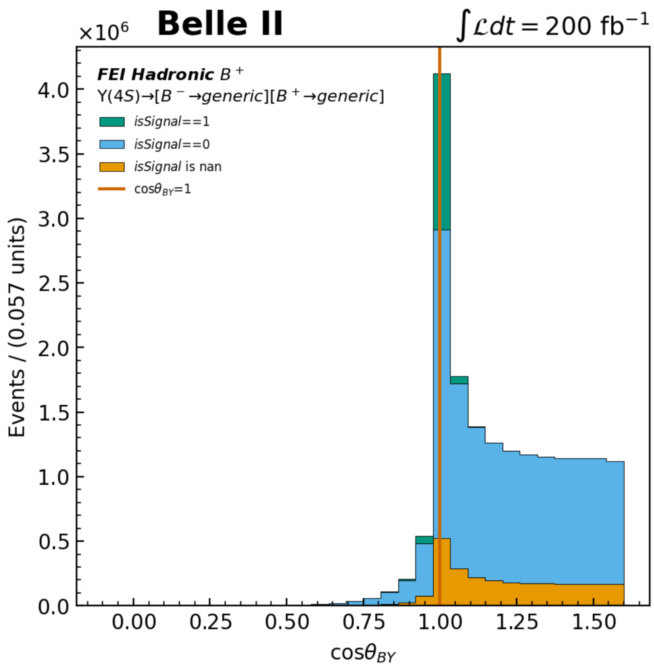


Figure 3.16. Distribution of $\cos\theta_{BY}$ measurements for hadronic FEI B_{tag} in 200 fb^{-1} $\Upsilon(4S)$ $\to B^+ \, B^-$ generic MC events. Stacking is by truth-matching status. $\cos\theta_{BY}$ is not as powerful in determining misreconstructed hadronic FEI B_{tag} and the distribution varies significantly to that of the semileptonic FEI (Figure 3.15).

For comparison's sake, Figure 3.16 demonstrates what the distribution of $\cos\theta_{BY}$ looks like when the B_{tag} is created from a full reconstructable hadronic FEI mode. In this case, $p_B = p_Y$ with no invisible system and it follows from Eq. (3.4.4) $\cos\theta_{BY}$ =1, save for the effects of any momentum smearing. There are many values above $\cos\theta_{BY}$ =1.6, though the distribution is truncated to match the semileptonic case. Most of the power that using $\cos\theta_{BY}$ has in identifying correctly identified hadronic B mesons is replicated by a selection on more direct variable like $M_{\rm bc}$. This is because $M_{\rm bc}$ are $\cos\theta_{BY}$ are directly correlated, given that both variables depend on E^*_{beam} and the reconstructed $|p^*_B| = |p^*_Y|$ for hadronically-tagged B mesons.

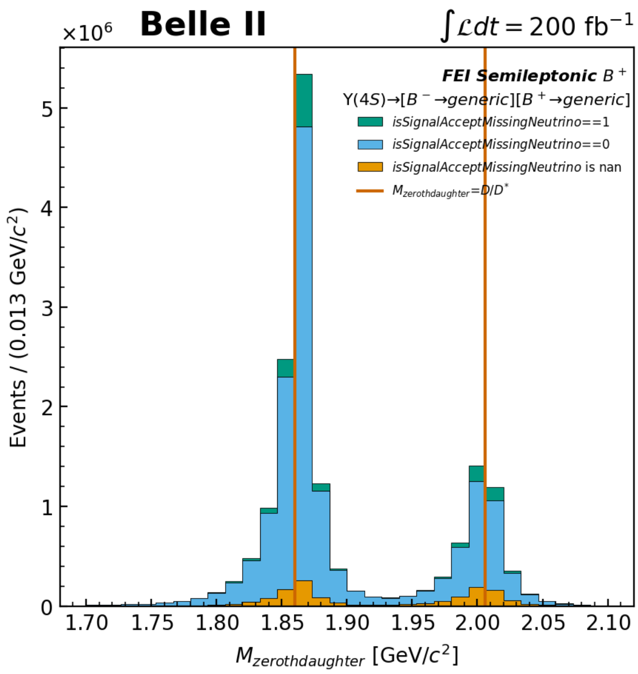


Figure 3.17. Measured mass zeroth daughter $B_{\rm tag}$ (i.e. D/D^*) identified via semileptonic FEI in $200\,{\rm fb}^{-1}$ $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. The approximate PDG masses of D^+/D^0 and D^{*+}/D^{*0} are indicated by the vertical lines.

As the semileptonic FEI $B_{\rm tag}$ reconstruction all proceed via modes of the form $B^+ \to \bar{D}^{(*)}(\pi)\ell\nu_\ell$, correctly identified $B_{\rm tag}$ will have zeroth daughter masses around $1.8\,{\rm GeV}/c^2$ or $2.0\,{\rm GeV}/c^2$. Figure 3.17 demonstrates this, where `isSignalAcceptMissingNeutrino`=1 the zeroth daughter $B_{\rm tag}$ are within $0.3\,{\rm GeV}/c^2$ of the nominal D and D^* masses. Similar to the hadronic $B_{\rm tag}$ zeroth daughter case in Figure 3.8, the well-known $4.8\,{\rm MeV}/c^2$ mass splitting between D^+ and D^0 and the $B_{\rm tag}$ mode is not discernible in this sample. The analogue of this between D^{*+} and D^{*0} is not as precisely constrained in experiment, though an absolute difference in nominal values, $2.6\ \ {\rm MeV}/c^2$, is also discernible in this scale. The intermediate decay modes between D mesons of the hadronic FEI and the semileptonic FEI are identical and so similarly, the narrow zeroth daughter peaks indicate suitable identification of $B_{\rm tag}$'s in the semileptonic FEI.

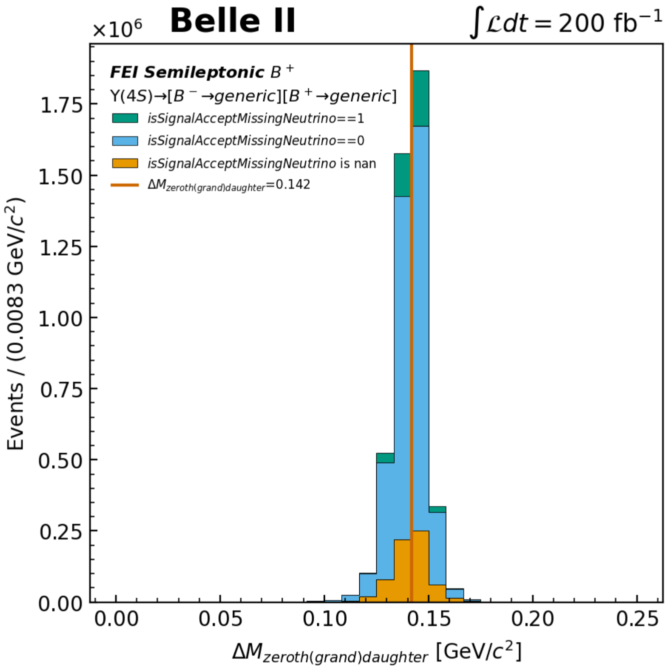


Figure 3.18. Difference in measured masses between zeroth daughter and zeroth granddaughter B_{tag} identified via semileptonic FEI in $200\,\text{fb}^{-1}$ $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status. The horizontal axis has been restricted to focus on $\Delta M = m_{D^*} - m_D$. Thus, only B_{tag} decay modes of the form $B \to D^* \to D$ are considered here.

Figure 3.18 demonstrates the difference in mass the zeroth daughter and the zeroth granddaughter, with a focus on the decay modes with the $D^* \to D$ transition. This corresponds to half of the implemented semileptonic B-tagging modes. This distribution shows little difference with the hadronic FEI analogue in Figure 3.9, aside from less misreconstruction tail towards $m_{D^*} - m_D = 0$ region suggesting better D meson identification in semileptonic B-tagging.

The final observable of the semileptonic FEI has no counterpart within the hadronic FEI. As all semileptonic decays contain a lepton by definition, one can also investigate the B_{tag} child lepton's momentum, in centre-of-mass frame. A lower bound placed at $p^*_\ell = 1$ GeV/c is in-built into the FEI sample considered, from the post-selection criteria defined in Table 3.3. This omitted region is usually dominated by $e^+e^- \to c\bar{c}$ leptons and not those from true B^+B^- processes. The distribution of correctly identified B_{tag} admits no significant peak between $1\,\text{GeV}/c$ and $2.25\,\text{GeV}/c$. In contrast, the misidentified B_{tag} peak strongly towards $1.65\,\text{GeV}/c$. This variable will be useful in rejecting misreconstructed B_{tag} candidates in B^+B^- processes, especially those containing high momentum leptons

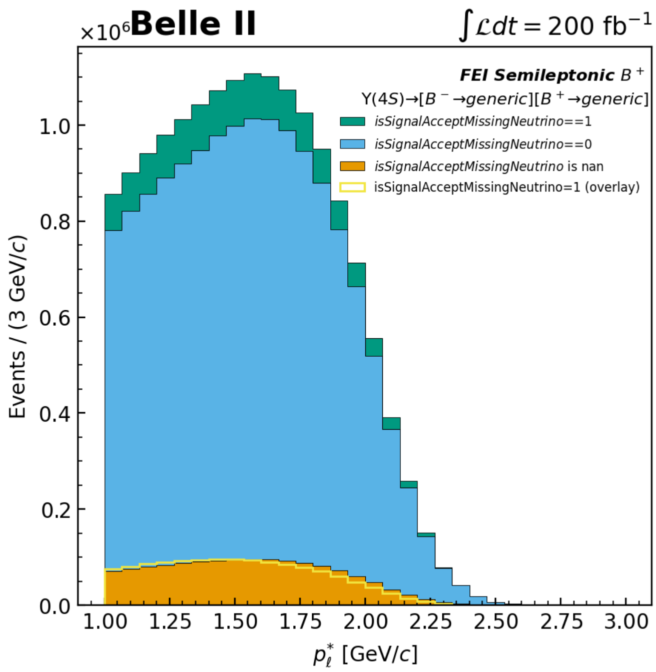


Figure 3.19. Lepton centre-of-mass frame momentum of B_{tag} identified via semileptonic FEI in 200 fb^{-1} $\Upsilon(4S) \to B^+B^-$ generic MC events. Stacking is by truth-matching status.

from other background physics processes. This will be seen in Section 3.5.2, when FEI tagged-B mesons are compared across $B^+ \to \mu^+\nu_\mu$ and background MC events. The existence of a tag-side lepton in the semileptonic FEI decay modes can prove difficult for signal-side B decays that contain a lepton themself, as is the case with the $B^+ \to \mu^+\nu_\mu$ decay. It will be important to implement selection criteria in such a case to minimise signal-side leptons from being mistaken for tag-side ones.

Using the semileptonic mode of FEI has thus been validated as a B-tagging method in generically decaying $e^+e^- \to \Upsilon(4S)$ B^+B^- processes. It has generally been seen that purity of a semileptonic B_{tag} sample is correlated with increasing the lower bound of the multivariate classifier produced by the FEI process, $\log_{10}(\texttt{sigProb})$. Additionally, there is a higher density of correctly truth-matched B candidates with $\cos\theta_{BY}$ between -1 and +1. Thus a combination of the previous variables could be used to improve the performance of a FEI B-tagged in a more specific charged B decay search, such as $B^+ \to \mu^+\nu_\mu$.

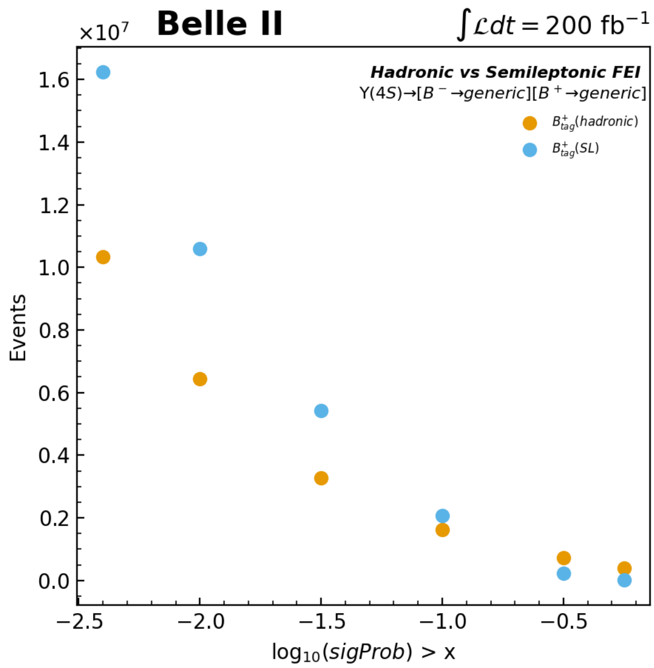
3.4.5. Choosing an appropriate FEI tagging mode for this analysis

The properties of the hadronic and semileptonic FEI B-tagging in $e^+e^- \to \Upsilon(4S) \to B^+B^-$, where the signal B is left unconstrained, has been discussed. SM calculations of the B^+ branching fractions implemented in the two modes show that the semileptonic FEI would be capable of tagging more events, which would benefit the search for a rare signal B-decay, such as $B^+ \to \mu^+\nu_\mu$. In inspecting the number of events measured per decay mode, it was one of the semileptonic decays that yielded the highest number of B_{tag}. However, the lower purity in the range 5-15% for the semileptonic decays indicated that there was difficulty in correctly reconstructing the modes. Better performance in this regard is seen in the hadronic FEI, especially with the rare but 79% correct $B^+ \to J/\psi\ K^+$ decays.

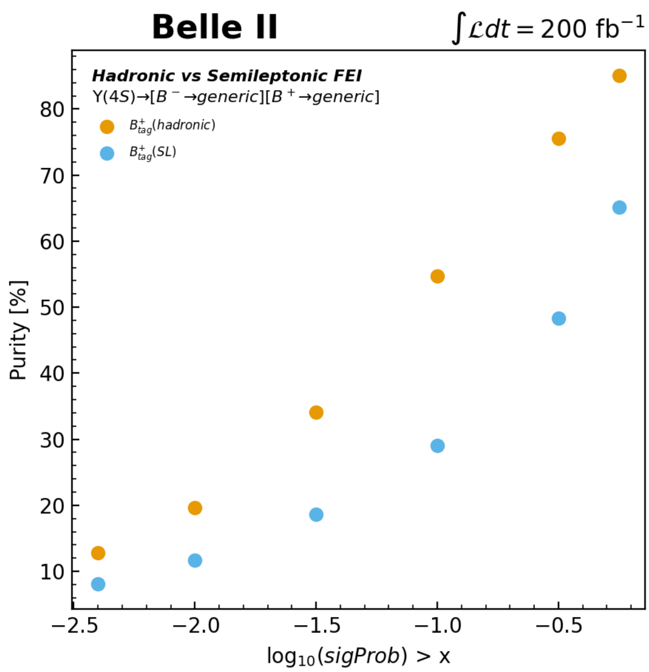
Both the hadronic and semileptonic FEI modes of the algorithm applied to B^+B^- decays was indicated to be relatively well-trained, given that increasing the minimum allowed values of the outputted multivariate classifier, $\log_{10}(\texttt{sigProb})$, resulted in a higher proportion of correctly reconstructed B_{tag}. Furthermore, both FEI modes demonstrated suitable D/D^* reconstruction within their nominal mass values, and nominal mass splittings. It was also illustrated that both modes have a unique set of variables that could be used to further reject misreconstructed B_{tag} and enhance the purity of the yield, with the beam constrained quantities of M_{bc} and ΔE used in the hadronic FEI, and the p^*_ℓ and $\cos\theta_{BY}$ used in the semileptonic FEI.

The purpose of this investigation is ultimately to inform the decision of which tagging mode will be implemented in the search for the rare process, $B^+ \to \mu^+\nu_\mu$. To use both FEI tagging modes would be to conduct two separate searches that require their own selection criteria and optimisation, which is beyond the scope of this thesis. As a rare process, the tagging mode that is ultimately chosen should be one that maximises retention of $B^+ \to \mu^+\nu_\mu$ events, and where further purity can be obtained by implementing tailored selection criteria to reject misreconstruction. While the semileptonic FEI seems to achieve this end, it is important to numerically verify this, as well as to understand whether the hadronic FEI has any feasibility for a future study.

To more directly compare the performance of hadronic and semileptonic FEI in generic B^+B^- processes, the number of events for stricter $\log_{10}(\texttt{sigProb})$ criteria are directly compared in Figure 3.20a. This is done so as to infer how tagging efficiency



(a) Number of events versus minimum allowed $\log_{10}(\texttt{sigProb})$.



(b) Purity versus minimum allowed $\log_{10}(\texttt{sigProb})$.

Figure 3.20. Scatter plots of yields and purities in B_{tag} samples identified by the hadronic or semileptonic FEI in generic $B\overline{B}$ MC, and how this changes when increasing the minimum allowed $\log_{10}(\texttt{sigProb})$. Comparison of the hadronic FEI begins from $\log_{10}(\texttt{sigProb}) > -2.4$, so as to match the semileptonic FEI post-selections.

and purity may change, with increasing $\log_{10}$(`sigProb`) standing in for future selection criteria. In an attempt to show each mode on an equal footing, comparisons begin from $\log_{10}$(`sigProb`) >-2.4, given the fixed post-selection criteria in the semileptonic FEI. The semileptonic FEI admits more B_{tag} than the hadronic FEI for $\log_{10}$(`sigProb`) lower bounds between -2.4 and -1. Beyond this, the hadronic has higher efficiency, which can be expected given the lack of peak past -1 in $\log_{10}$(`sigProb`) for semileptonic FEI in Figure 3.12 versus the hadronic one. For very tight selection, the yields of the two modes are more comparable. Consulting Figure 3.20b, the sacrifice of the purity from the missing neutrino in the semileptonic FEI is illustrated. While around $\log_{10}$(`sigProb`) = -2.4 the purity of hadronic FEI is only marginally larger, this difference becomes more pronounced for harsher selections. This becomes as much as a 20% higher purity in the hadronic modes for $\log_{10}$(`sigProb`) = -0.5.

Table 3.4. Number of candidates / events after truth-matching and best candidate selection in hadronic and semileptonic FEI B^+-tagging in B^+B^- MC. Efficiency is the percentage of FEI B^+-tagged events in the 108 million generated events. Purity is the percentage of correctly truth-matched events in all events identified events.

FEI mode	Hadronic (no. cands. / no. events)	Semileptonic (no. cands / no. events)
All candidates / events	5.70×10^7 / 1.82×10^7	4.16×10^7/ 1.62×10^7
Correctly truth matched only	1.70×10^6/ 1.64×10^6	1.98×10^6/ 1.96×10^6
Best candidate selection only	1.82×10^7/ 1.82×10^7	1.62×10^7/ 1.62×10^7
Truth-matched after best candidate selections	1.36×10^6/ 1.36×10^6	1.32×10^6/ 1.32×10^6
EFFICIENCY[%]	- / 16.8	- / 15.0
PURITY[%]	- / 7.5	- / 8.1

Finally, the overall number of B_{tag} can be compared, as truth-matching and best candidate selection are applied to both the semileptonic and hadronic FEI. This is outlined in Table 3.4. It is noted that in this table that $\log_{10}$(`sigProb`) is restored to its maximal value in the hadronic FEI. Both tagging modes begin with an average of around 3 candidates per event, with the hadronic FEI having identified B_{tag} in more events. Of the smaller number semileptonic FEI identified events, there are a larger number of events containing a correct B_{tag}. There is no guarantee, however, that this correct B_{tag} also is deemed 'best' performing B_{tag} amongst all others in the event. After the best candidate selection of choosing the B_{tag} with the highest signal probability, it is now the hadronic FEI that has found the larger number of correct B_{tag}. Overall, the hadronic FEI identifies B_{tag} in 16.8% of the 108 million events, with 7.5% of them correct. This is compared with the semileptonic FEI's 15.0% efficiency within the same number of events, at 8.1% purity. It initially seems that the assumption of semileptonic decays having high efficiency/lower purity and hadronic decays having high purity/lower efficiency in B^+ decays is contradicted.

For this study, only the B_{tag} were reconstructed, with no limitation to the opposing B. A better method to compare the hadronic and semileptonic FEI would be to reconstruct the same signal B mode in each FEI tag mode and then measure the efficiencies and purities. With both a signal B and a tag B, the products can be controlled for to prevent the opposing B being misreconstructed into the B_{tag}, and thus ensures true semileptonic and hadronic decays are assessed. This is actually the preferred way of conducting studies into the FEI within Belle II, especially in comparing its performance between data and MC collision [100]. It was avoided in this case as this would also be beyond the scope of this chapter, as a simple illustrative exercise of the features of the FEI.

As ultimately, the B-tagging mode desired would be one that specifically favours the identification of B_{tag} alongside $B^+ \to \mu^+\nu_\mu$ decays, it is best to instead consider the yields of the hadronic FEI and semileptonic FEI algorithm when the signal B-decay is fixed to $B^+ \to \mu^+\nu_\mu$. Using $B^+ \to \mu^+\nu_\mu$ MC has an advantage in comparison of FEI modes over generic B^+B^- processes, in that there is only a single visible particle on the signal side that can be misreconstructed into FEI B_{tag}.

Table 3.5 compares the efficiencies and purities of semileptonic and hadronic FEI B-tagging in a sample of 8 million $e^+e^- \to \Upsilon(4S)$ $[B^- \to \text{generic}][B^+ \to \mu^+\nu_\mu\,]$ MC events. While the signal B is fixed to occur as $B^+ \to \mu^+\nu_\mu$, it remains unreconstructed for now. The statistics of the signal B reconstruction and the requirement that the B_{tag} is compatibly combined to create a $\Upsilon(4S)$ as per $\Upsilon(4S) \to B^-$ $[B^+ \to \mu^+\nu_\mu\,]$ is left to be the focus of Chapter 4.

In the table of B_{tag} yields found in $B^+ \to \mu^+\nu_\mu$ MC, the assumption has been restored: the semileptonic FEI has identified near 3 times the number of events than the hadronic FEI, with 6.1% of the 8 million events containing a suitable FEI B_{tag}. While that hadronic FEI mode does have more events after best candidate selection that are correct, it is still comparable in performance to the semileptonic mode.

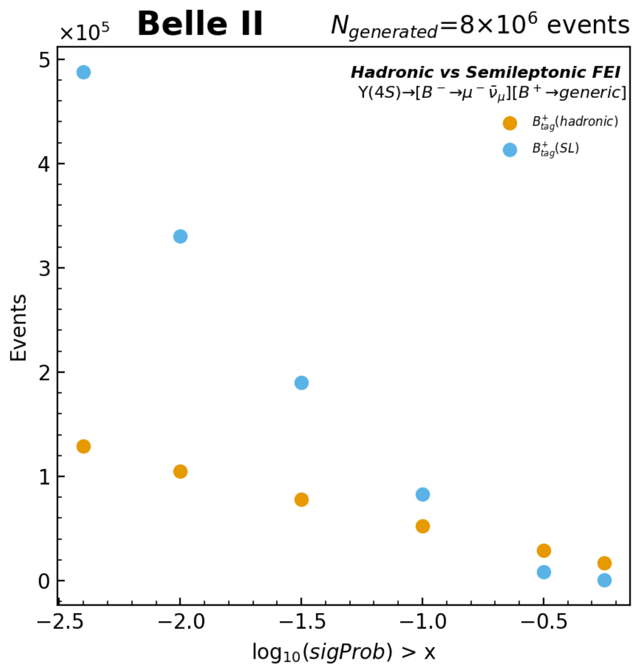
Table 3.5. Number of candidates / events after truth-matching and best candidate selection in hadronic and semileptonic FEI B-tagging in $B^+ \to \mu^+\nu_\mu$ MC.

FEI mode	Hadronic (no. cands. / no. events)	Semileptonic (no. cands / no. events)
All candidates / events	$3.07{\times}10^5/1.65{\times}10^5$	$8.14{\times}10^5/4.88{\times}10^5$
Correctly truth matched only	$7.87{\times}10^4/7.56{\times}10^4$	$8.12{\times}10^4/8.09{\times}10^4$
Best candidate selection only	$1.65{\times}10^5/1.65{\times}10^5$	$4.88{\times}10^5/4.88{\times}10^5$
Truth-matched after best candidate selections	$7.20{\times}10^4/7.20{\times}10^4$	$6.53{\times}10^4/6.53{\times}10^4$
EFFICIENCY [%]	- / 2.1	- / 6.1
PURITY[%]	- / 43.8	- / 13.4

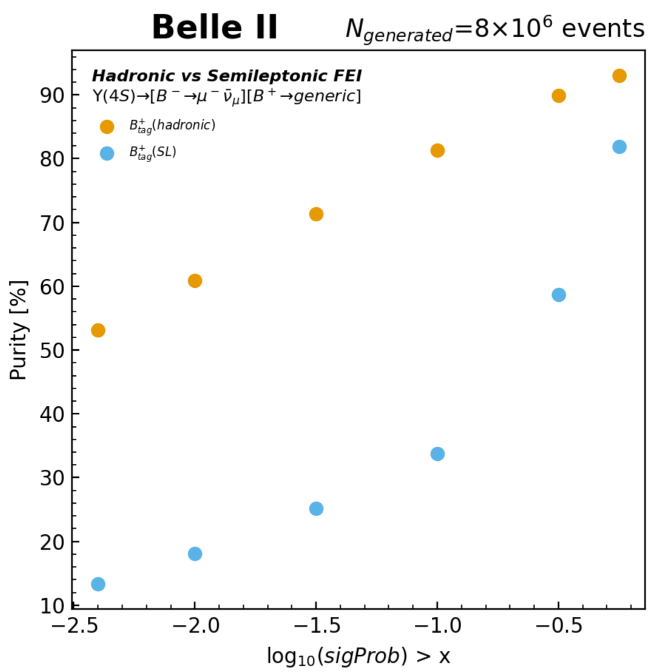
For completeness, the plots of number of events and the purity of FEI B_{tag} against increasing $\log_{10}$(`sigProb`) are provided in Figure 3.21a and Figure 3.21b, respectively. They showcase similar trends as the B^+B^- versions in Figure 3.20a and Figure 3.20b. The semileptonic FEI identifies more B_{tag} than the hadronic FEI, with this difference decreasing until around $\log_{10}$(`sigProb`) =-0.5, where the hadronic FEI now has more events. The difference in purity is consistently higher by about 40% for the hadronic FEI until $\log_{10}$(`sigProb`) =-1, where the two modes begin to converge in purity.

It is important to note that the comparison above of the hadronic and semileptonic FEI in $B^+ \to \mu^+\nu_\mu$ MC, a selection was made to consider the modes more equally. The trend in the semileptonic FEI showing a preference for muonic modes, as was demonstrated in the decay mode histogram of Figure 3.11, is enhanced in the $B^+ \to \mu^+\nu_\mu$ MC sample due to the signal muon being mistaken as a tag-side lepton. This effect is made clear in Figure 3.22. Figure 3.22a demonstrates tag-side lepton momentum in the centre-of-mass frame, a large peak exists after 2.25 GeV/c. The grouping of the tagged events by truth-matching status demonstrates that almost no B_{tag} in this region are `isSignalAcceptMissingNeutrino` correct. This same peaking structure is not seen in Figure 3.19, for generic B^+B^- decays and where notably the $B^+ \to \mu^+\nu_\mu$ process is not present. As will be seen when the dynamics of the $B^+ \to \mu^+\nu_\mu$ decay are studied in Chapter 4, the range of this peak in the tag-side momentum matches that of correctly identified signal-side muon.

To demonstrate further that this is a muonic effect, the secondary distribution of tag-side momentum grouped by semileptonic decay mode is presented in Figure 3.22b



(a) Number of events versus minimum $\log_{10}$(`sigProb`).



(b) Purity (no. of events with correctly truth-matched B_{tag} candidate) versus minimum $\log_{10}$(`sigProb`).

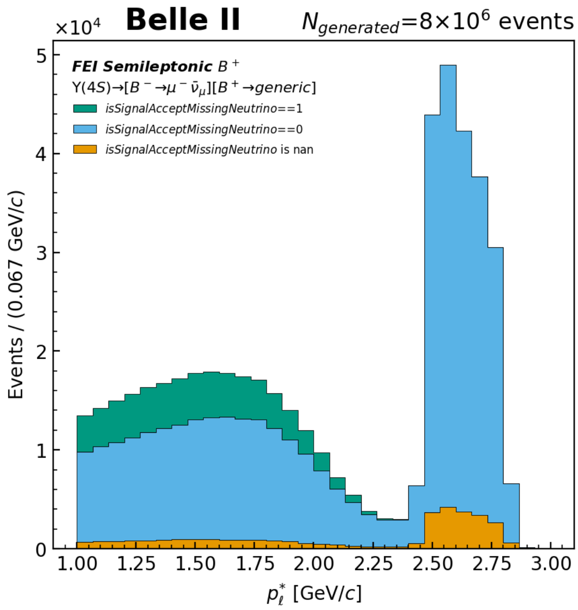
Figure 3.21. Scatter plots comparing yields and purity of events identified by the FEI in $\Upsilon(4S) \to B^- \, [B^+ \to \mu^+\nu_\mu\,]$ (and charge conjugate) MC and how this changes when increasing the minimum allowed $\log_{10}$(`sigProb`).

The electron component (i.e. all events with $B_{\rm tag} \to D(\pi)e^+\nu_e)$) does not have events with lepton momentum beyond 2.25 GeV/c. The semimuonic tagged events exist both in the sub-2.25 GeV/c and post-2.25 GeV/c region, with the peak consisting entirely of semimuonic $B_{\rm tag}$. To prevent the effect of $B_{\rm sig}$ contamination in $B_{\rm tag}$ reconstruction, all events with $p^*_\ell < 2.25\,{\rm GeV}/c$ were removed from the analysis, including in the results of the purity and efficiency calculations of Table 3.5, as well as the $B^+ \to \mu^+\nu_\mu$ $B_{\rm tag}$ yield and purity plots in Figure 3.21a and Figure 3.21b. It was not necessary to implement similar selections in the hadronic FEI applied to $B^+ \to \mu^+\nu_\mu$, as there is no tag-side lepton and as no such effects were observed. Muons are only used in J/ψ reconstruction in the hadronic FEI, and the efficiency of these modes are not large enough to give the same effect as in the semileptonic FEI.

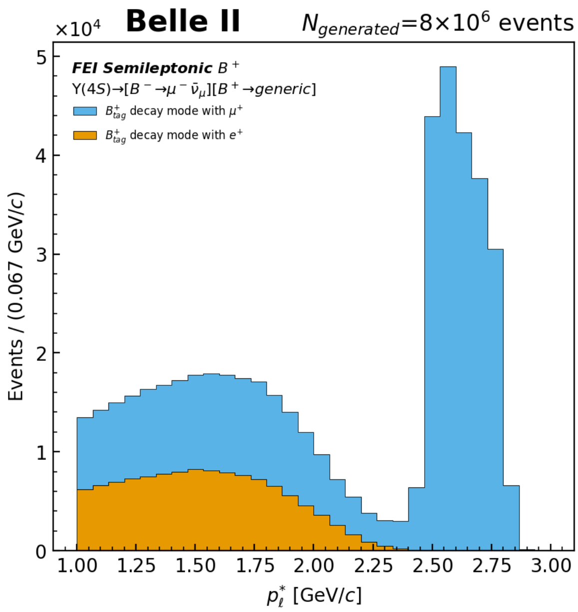
After investigating the tagging of semileptonic and hadronic decays, the total purity and efficiency of both modes of the FEI, its performance in both generic $\Upsilon(4S) \to B^+B^-$ decays as well as via the signal mode, $\Upsilon(4S) \to B^-\ [B^+ \to \mu^+\nu_\mu\ ]$, it is finally justified to select the FEI mode that will be used in this analysis. This $B^+ \to \mu^+\nu_\mu$ search is to be performed in a real collision dataset of 362 fb^{-1}, corresponding to 195 million $e^+e^- \to B^+B^-$ collisions. Of these decays, theoretically 83.3 of them will include $B^+ \to \mu^+\nu_\mu$, if the decay occurs according to the SM branching fraction. Combined with the FEI efficiencies derived in Table 3.5 this suggests approximate yields of 1.7 and 5.1 events for hadronic and semileptonic FEI respectively. Thus, to maximise efficiency, the analysis will proceed using semileptonic FEI B-tagging. While semileptonic FEI has demonstrated lower purity in $B^+ \to \mu^+\nu_\mu$ MC, this can be improved via signal-enhancing selection criteria of events ultimately used in calculating the experimental $B^+ \to \mu^+\nu_\mu$ branching fraction. As more $e^+e^- \to B\overline{B}$ collisions are created at the Belle II experiment in the future, a large enough dataset will eventually be available, such that applying the lower efficiency hadronic FEI will still yield a significant number of events with high purity. In the future, the techniques that take advantage of having a complete $B_{\rm tag}$ can be explored, such as estimating the signal-B momentum from the tag-B as was done in Section 1.4.3. With the B-tagging mode decided, the hadronic FEI will not be explored any further.

3.4.6. Improvements to the FEI

Having seen the application of the FEI, a brief discussion on improving the FEI method more widely will take place.



(a) Stacking by truth-matching status.



(b) Stacking by lepton in decay-mode. Note new legend.

Figure 3.22. Lepton centre-of-mass frame momentum of B_{tag} identified via semileptonic FEI in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+ \nu_\mu\,]B^-$ MC events.

At the end of the previous section, the effects of B_{sig}-B_{tag} contamination was discussed. There are options within the FEI to run in *specific* mode, where a signal-side is identified prior to B-tagging. The intention here is to tailor the FEI training phase to signal-side specific backgrounds, rather than generic misreconstructable B_{tag}. Such a method could be useful in future iterations of this analysis, to avoid semimuonic B_{tag} being constructed from signal muon.

However, implementation of the specific FEI is uncommon within Belle II analyses, owing to the need for large amounts of specific signal MC to train the signal-specific backgrounds, as well as the computational knowledge required to train. The generic FEI is trained by the Belle II Analysis Tools group and so it is generally easier for analyses to use this than create their own specific training. Previously, the specific FEI has been successfully applied to improve a Belle search for hadronically-tagged $B^+ \to \mu^+ \nu_\mu \gamma$ in Ref. [101]. In contrast to this success, Ref. [102] found that the specific FEI performed worse than the generic FEI in identifying B_{tag} in $\Upsilon(4S) \to B^- [B^+ \to \rho^+ \mu^+ \nu_\mu]$ decays. Further investigations into assessing performance of the specific FEI is not currently a priority of the Belle II Analysis Tools group.

Another aspect of the FEI that requires consideration is accounting for differences in B-tagging capability in MC and data. FEI training is performed in MC $B\overline{B}$ events, where truth-matching is used to verify the attempts to create B_{tag}'s throughout the learning phase. MC events can never completely match that of collision data; differences can come about for a variety of reasons including differences in particle-identification, or track reconstruction efficiency between the simulated MC and the real performance of decay products in the detector. As the training has taken place in MC, it is expected that FEI B-tagging is better suited to MC events than events in data. This can then lead to discrepancies in the expected yields of the FEI in data, as well as mismatches in the yields in data and MC. Analyses implementing the FEI require calibration, such that the MC better represents the trends in data, which is done in one of two ways. The first is to derive correction factors for each FEI decay mode, which will result in better agreement between distributions of B-tagged data and MC. These correction factors can be calculated by using the B_{tag} to reconstruct an alternate signal-side, referred to as a control mode. In this case, the factors will be the ratio of yields of the control mode reconstruction between data and MC. These correction factors can then be applied to an individual analysis to better estimate the signal yield and tagging-capability of the FEI. For the FEI, these are calculated by the Analysis Tools group. The control mode implemented for the semileptonic FEI is $B \to D^*\ell\nu$ modes [103]. As the signal decay of

$B^+ \to \mu^+\nu_\mu$ does not match the topology of this three-body control mode, an alternate method will be adopted to account for data/MC differences. This will be discussed in Chapter 4.

The need for calibration could be remedied by training the FEI using Belle II collision data. Such a method requires careful consideration, as truth-matching is not possible in collision data. There is functionality within the FEI to perform a data-based training using the *sPlot* tool, where the properties of discriminating variables are used to determine relationships between lesser discriminating variables and event type, in unknown data [104]. This technique may be more feasible when Belle II has a larger dataset, such that there will be sufficient collision data remaining after a portion of it is used for training.

An alternate formulation of the FEI has also become available to analysts, called the *Graph-based Full Event Interpretation* or GraFEI [105] [106]. In this version, graph neural network learning is implemented. Here, a neural network learns what combinations of particles are likely in an event, by representing a B decay tree as a directed graph. For each possible pair of final state particles in a given event, the lowest common ancestor is determined (i.e. the closest relative common to both particles). Thus the neural network learns to replicate common decay trees based on the attributes of FSPs and their likely lowest common ancestors. This then allows for full B decay trees to be built in a variety of unknown events, without hard-coding identifiable decays like in the standard FEI. As it has only recently been incorporated into `basf2`, there are not yet any published particle searches using this tool, though there are preliminary analyses underway.

3.5. Performance of the FEI in a Belle II $B^+ \to \mu^+\nu_\mu$ search

In the previous sections, the two main FEI operating modes relevant to charged B decays were discussed, where it was decided that the semileptonic FEI is better suited to tagging $B^+ \to \mu^+\nu_\mu$ events. The preliminary statistics of the semileptonic FEI applied to $B^+ \to \mu^+\nu_\mu$ MC has already been shown. In this section, the FEI performance variables in $B^+ \to \mu^+\nu_\mu$ events will be explored graphically, similar to the histograms seen earlier for charged MC.

In Section 3.5.1, the distributions of the performance variables will be presented for the 8 million events generated as $\Upsilon(4S) \rightarrow B^- \, [B^+ \rightarrow \mu^+\nu_\mu \,]$ (and charge conjugate) MC only, where the events will be grouped according to truth-matching status of the FEI B_{tag}. They will be compared to the corresponding distributions with generic B^+B^- MC in Section 3.4.4 After this, Section 3.5.2 will present the same distributions, but inclusive of background MC. This will namely include the MC samples outlined in Table 2.6, with charged and mixed B decays, as well as $u\overline{u}$, $d\overline{d}$, $s\overline{s}$ and $c\overline{c}$ continuum processes at $\sqrt{s}$ =10.58 GeV. Thus instead of grouping events according to truth match status, events will be grouped according to which physics process they were generated as. Both the background and signal MC, will be scaled to their appropriate branching fractions so that their yields will be equivalent to what is expected within the 362 fb^{-1} real Belle II dataset. Finally, a numerical summary of the B-tagging selection for the semileptonic FEI in both signal and background MC will be presented. The large number of background MC will further motivate the need for background rejection criteria to be developed in Chapter 4

In both the signal MC only and the background added analyses, all reconstructed B_{tag} containing a muon will have selection of $p^*_\ell < 2.25\,\text{GeV}/c$ to avoid μ_{sig} contamination, as was motivated in Section 3.4.5. Moreover, best candidate selection on B_{tag}'s in all samples is performed by choosing the one with the largest signal probability, as in previous sections.

3.5.1. B_{tag} in Belle II MC: $B^+ \rightarrow \mu^+\nu_\mu$ only

To begin, the histogram of $\log_{10}(\texttt{sigProb})$ for B_{tag} identified via semileptonic FEI in $B^+ \rightarrow \mu^+\nu_\mu$ is presented in Figure 3.24. It must be first acknowledged that there are fewer events in the 8 million signal event sample versus the 108 million event B^+B^- sample. Regardless, comparing the generic B^+B^- version in Figure 3.11, the distribution shape has changed considerably, with a more substantial proportion of correctly truth-matched events. This supports what was discussed in Section 3.4.5, that the semileptonic FEI is better able to identify high-signal probability B_{tag} with a fixed signal B^+ meson final state. Such truth-matched events are also occurring for larger values of log-transformed signal probability, suggesting that signal probability again is a reasonable proxy for correctness in the $B^+ \rightarrow \mu^+\nu_\mu$ MC. The overlaid purity of the B_{tag} for increasing lower bounds on $\log_{10}(\texttt{sigProb})$ also supports this idea. A maximum purity of around 65% is

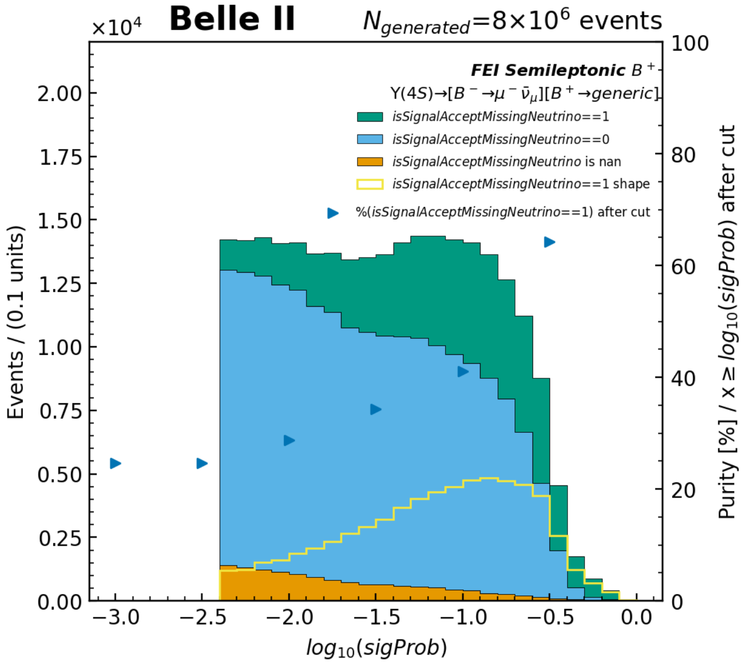


Figure 3.23. Log-transformed signal probability of B_{tag} decay modes identified via semileptonic FEI in 8×10^6 MC generated $B^+ \to \mu^+\nu_\mu$ events. Stacking is by truth-matching status.

achieved for $\log_{10}(\texttt{sigProb}) \geq$ -0.5, which is also higher than its B^+B^- counterpart at approximately 50%.

The distribution in Figure 3.24 demonstrates the number of events tagged by the semileptonic FEI in each decay mode, when applied to the $B^+ \to \mu^+\nu_\mu$ MC. The yields of the first four decay modes demonstrates the same preference for the 3-body semileptonic tagging modes, as was seen in the generic B decay version presented in Figure 3.11. Each mode is also made up of a larger proportion of correctly-truth matched modes. The superimposed purity of each B_{tag} decay mode identified alongside the $B^+ \to \mu^+\nu_\mu$ is at most 4 times larger than its counterparts in the generic B^+ sample. The tagging purity in excited-D and $D^{(*)}\pi$ modes benefit the most. These are modes which require more particles (for D π) as well as more sub-reconstructions (for the $D^* \to D$). Thus, with only a single signal side particle in this MC, there are lesser particles which can be confused by the semileptonic FEI into creating incorrect B_{tag}, relative to the $\overline{D}^0\ell^+$ ν_ℓ modes.

Figure 3.25 contains the tagmap of $\log_{10}(\texttt{sigProb})$ distributed across each semileptonic B_{tag} decay mode. This tagmap shows little relative difference in the relationship between the signal probability and the different B_{tag} modes when contrasted against the generic

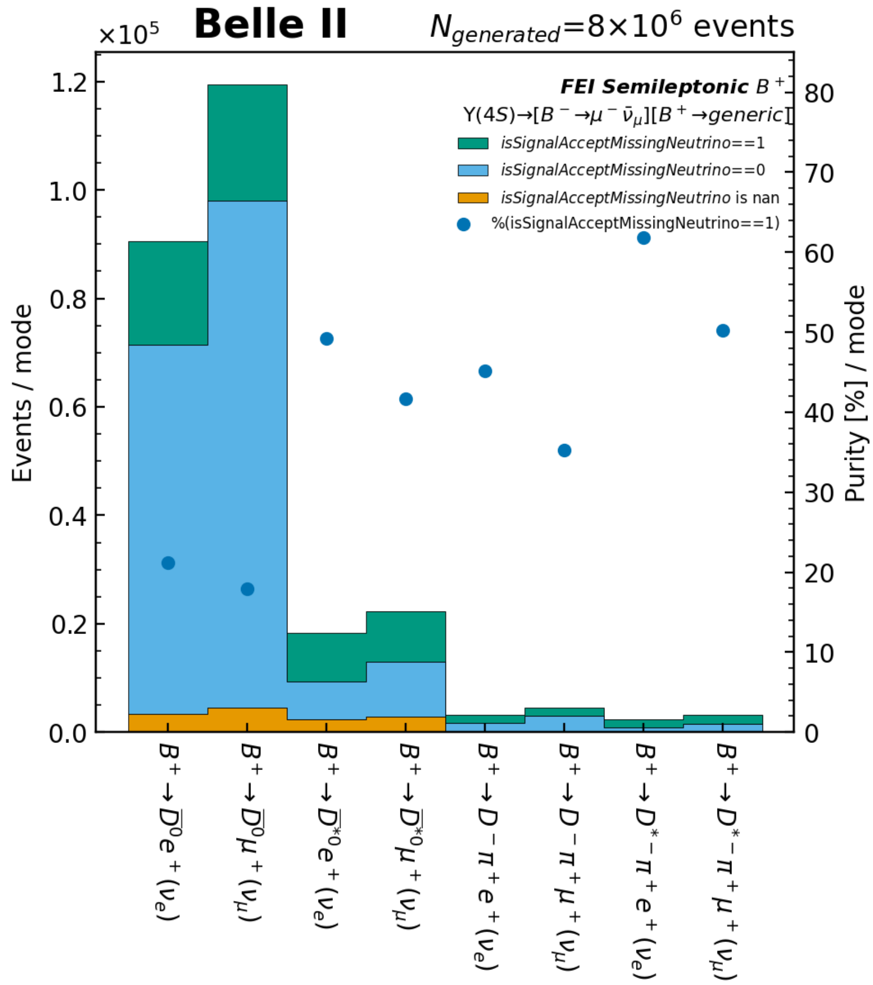


Figure 3.24. Reconstructed B_{tag} decay modes identified via semileptonic FEI in 8×10^6 MC generated $B^+ \to \mu^+\nu_\mu$ events. Stacking is by truth-matching status.

B^+ version in Figure 3.13. The sole difference between the distributions of applying the semileptonic FEI between the B^+B^- and the signal MC is the lesser number of events in the $D\ \pi$ modes in the $\log_{10}$(`sigProb`) approaches 0 region. It was suggested in the discussion of the generic B^+ version that the previous peaking in the $\log_{10}$(`sigProb`) =-0.1 bin could indicate some overtraining in these modes. The events that do remain in this region of the signal MC are of lesser relative magnitude, indicating that potential overtraining will in turn have a smaller effect in the signal MC.

In Figure 3.26, the $\cos\theta_{BY}$ angular variable of semileptonic B_{tag} in $B^+ \to \mu^+\nu_\mu$ MC is depicted. When compared to the B^+B^- counterpart in Figure 3.15, both demonstrate the same 'tail' below the physical $\cos\theta_{BY}$ ∈[-1,1] region that is dominated by misreconstructed B_{tag}. Where the distribution shape varies the in region near $\cos\theta_{BY}$ =1. In the signal

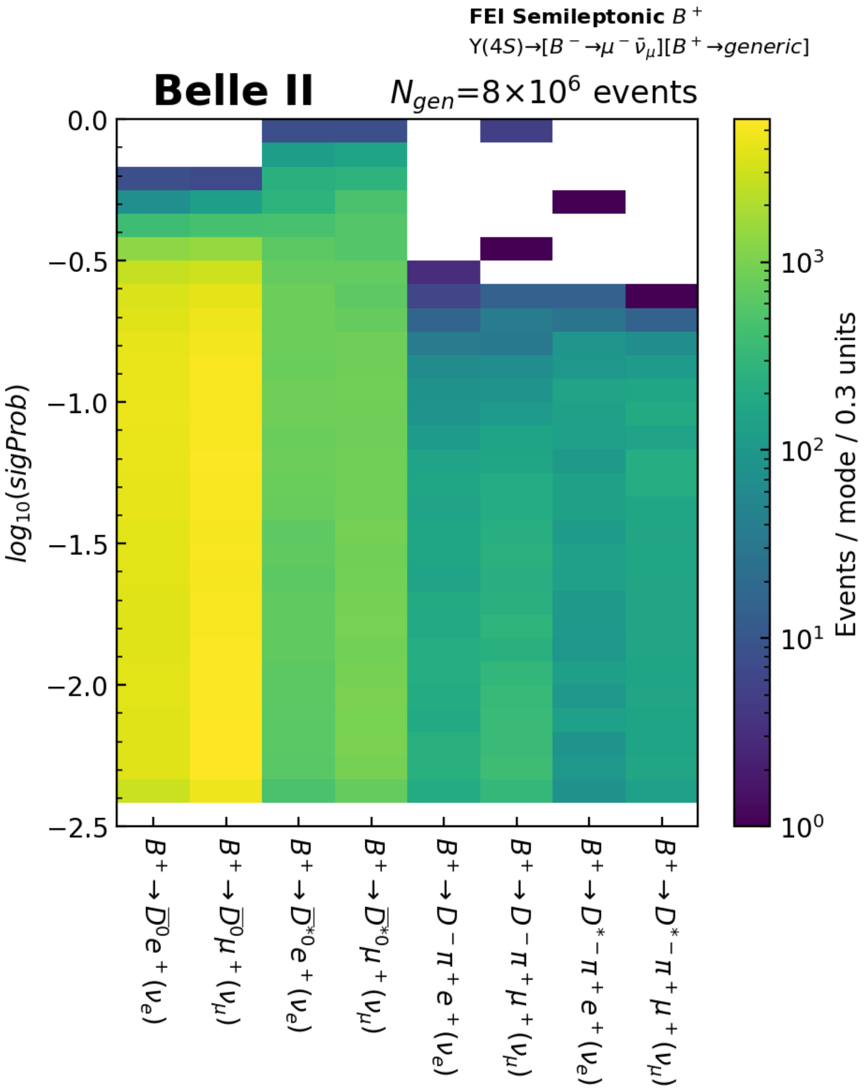


Figure 3.25. A 2D histogram of $\log_{10}(\texttt{sigProb})$ according to identified semileptonic FEI B_{tag} decay modes in $B^+\to\mu^+\nu_\mu$ MC.

MC, this is close to a local minimum in the number of events, whereas in the B^+B^- sample there is a surplus of incorrectly truth-matched B_{tag}. This change is due to both less misreconstructed events in the physical $\cos\theta_{BY}$ region of [-1,1], as well as less events overall in the $\cos\theta_{BY} > 1$ region. The decrease in the number of incorrect events in the [-1,1] interval will be advantageous, as such a restriction to the allowed values of $\cos\theta_{BY}$ for considered events could be used when calculating the signal yield in data. The overall decrease in events can be attributed both to the fixed signal B decay discussed previously as well as the selection of $p^*_\ell < 2.25$ GeV/c. Originally, this was made to avoid the signal muon being mistaken as the tag-side one by the FEI. From Eq. (3.4.4), it was seen that

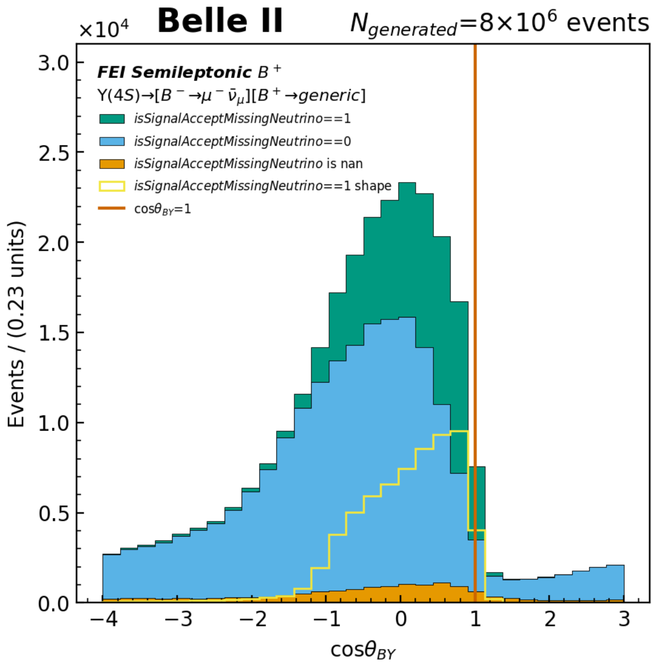
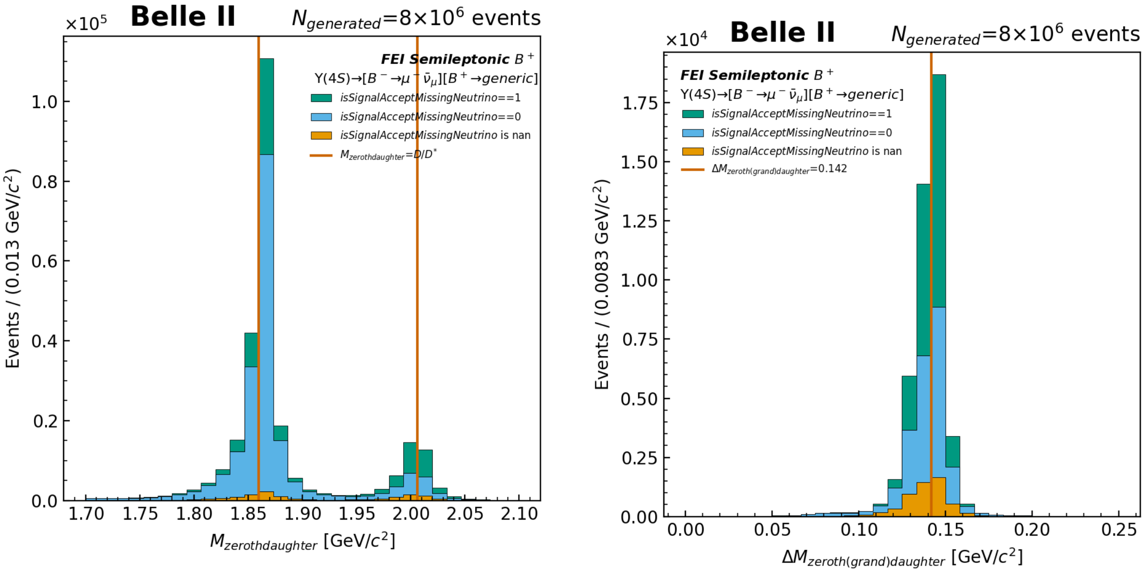


Figure 3.26. $\cos\theta_{BY}$ measurements of semileptonic FEI B_{tag} in 8×10^6 MC generated $B^+ \to \mu^+\nu_\mu$ events. Stacking is by truth-matching status. The theoretical value of $\cos\theta_{BY}$ =1 for B_{tag} with no missing energy is indicated by the vertical line.

$\cos\theta_{BY}$ is dependent on the three-momentum of the Y system, which is inclusive of the tag-side lepton. With the removal of the fake B_{tag} via the $p^*_\ell < 2.25$ GeV/c selection, it is therefore not surprising that the $\cos\theta_{BY} > 1$ region of misreconstruction has been affected. Thus, the values of $\cos\theta_{BY}$ in semileptonic B_{tag} created in signal MC are generally better reconstructed, as they result in a lesser proportion of events above the physical $\cos\theta_{BY}$ region.

Figure 3.27a depicts the histogram of B_{tag} zeroth daughters identified in $B^+ \to \mu^+\nu_\mu$ MC. There are no major differences between the B_{tag} zeroth daughter masses in signal MC and those in the generically decaying B version of Figure 3.17, aside from a larger proportion of correctly truth-matched events near the nominal D/D^* masses. The lack of shape difference and enhanced purity near nominal mass values is also seen in distributions of the related D^*/D mass splitting, when comparing semileptonic B-tagging in signal MC (Figure 3.27b) to semileptonic B-tagging in generic B^+B^- events (Figure 3.18).

In summary, semileptonic FEI B-tagging has performed better in the $e^+e^- \to \Upsilon(4S)$ $\to [B^+ \to \mu^+\nu_\mu\,]B^-$ (and charge conjugate) sample compared to when both B mesons were previously allowed to decay generically. This has been attributed to both the lower



(a) Zeroth daughter mass (i.e. D/D^*), with labelled PDG masses of D^+/D^0 and D^{*+}/D^{*0}.

(b) $\Delta M = m_{D^*} - m_D$ for $B \to D^* \to D$ FEI modes.

Figure 3.27. Daughter mass and (grand-)daughter mass splitting in B_{tag} identified via semileptonic FEI in 8×10^6 MC generated $B^+ \to \mu^+\nu_\mu$ events. Stacking is by truth-matching status.

number of $B^+ \to \mu^+\nu_\mu$ signal particles that can be mistaken in tag-side reconstruction, as well as the role of the $p^*_\ell < 2.25$ GeV/c selection. Performing better has corresponded to higher proportions of correct B_{tag} in regions that indicate good reconstruction, such as being close to nominal mass peaks, higher signal probability and in physically allowable regions of variables like $\cos\theta_{BY}$. Having understood the differences between distributions of reconstructed and misreconstructed events, the distributions of background samples can now be investigated.

3.5.2. B_{tag} in Belle II MC: $B^+ \to \mu^+\nu_\mu$ with background

The ranges of key semileptonic FEI variables as applied to both MC where two B's decay generically and when one B is fixed to decay via $B^+ \to \mu^+\nu_\mu$ has been presented. In both the studies of the $200\,\text{fb}^{-1}$ generic B^+B^- MC sample and the 8 million event $B^+ \to \mu^+\nu_\mu$ MC sample, no normalisations were applied. Eventually, the $B^+ \to \mu^+\nu_\mu$ reconstruction strategy will be applied to 362 fb^{-1} of real Belle II e^+e^- collisions. It is essential that in using MC to model the data, that the yields reflect what is possible within the number of events in 362 fb^{-1}. Thus, scaling factors must be applied to understand what $B^+ \to \mu^+\nu_\mu$ events could look like in the search. To this end, $B^+ \to \mu^+\nu_\mu$ MC are scaled according to their predicted SM branching fraction. If significantly more $B^+ \to \mu^+\nu_\mu$ events are identified in data than what is predicted using the SM branching fraction, this could indicate contributions of BSM mechanisms in creating $B^+ \to \mu^+\nu_\mu$ events.

Having investigated the charged B-tagging in two kinds of charged B MC, the susceptibility of the B-tagging procedure to misidentifying other physics processes as B_{tag}'s must be examined. An example where a background event can mimic a semileptonic B decay is in the $e^+e^- \to c\bar{c}$ process. The Feynman diagram in Figure 3.28 illustrates such a case. Just as $e^+e^- \to b\bar{b}$ (i.e. $\Upsilon(4S)$) events can coalesce as $B^0\overline{B}^0$, so can e^+e^- $\to c\bar{c}$ (i.e. J/ψ) as $D^0\bar{D}^0$. Here, true leptons and true D mesons are created, while a real B meson is not. The semileptonic FEI could combine the $\overline{D}^0$ with the charged lepton from the other D^0 and be mistaken as the products of a B meson parent. Other background processes may also mimic semileptonic B-decay modes, with potential for fake D mesons created from real kaons and real pions in $e^+e^- \to u\overline{u}/d\overline{d}/s\overline{s}$ processes. This effect is compounded with potential for final state particle misidentification, such as if the tracks of real charged kaons are mistakenly determined as pions, and reconstructed into a hadronic D decay.

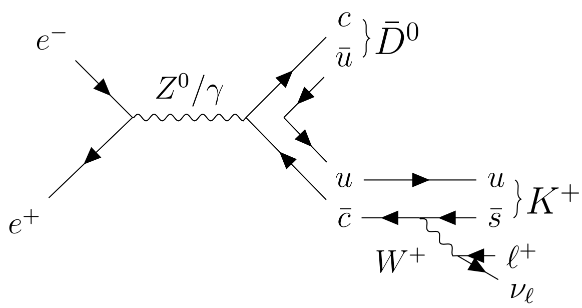


Figure 3.28. Feynman diagram of a potential $e^+e^- \to c\bar{c}$ process, containing a $D^0 \to K^+$ ℓ^+ ν_ℓ decay. This is an example of a process which could be mistaken by the semileptonic FEI as $B^+ \to \overline{D}^0\ell^+$ ν_ℓ.

Other backgrounds which could be retained by the B^+ semileptonic FEI include those from other real $e^+e^- \to B\overline{B}$ events. This could be the FEI mistaking a non-semileptonic FEI mode within real charged B's in $e^+e^- \to B^+B^-$ events or attempting to reconstruct a charged B_{tag} by mixing particles products from non-semileptonic decays of both B's. Moreover, the FEI is also susceptible to reconstructing a charged semileptonic B_{tag} from products from a neutral $e^+e^- \to B^0\overline{B}^0$ event containing a neutral semileptonic B decay. This can occur, if for example, the π^- in a real $B^0 \to \overline{D}^0\pi^-$ e^+ ν_e is 'lost' due to incorrect detection, potentially being reconstructed via the mode $B^+ \to \overline{D}^0e^+$ ν_e. The numerous ways in which both $e^+e^- \to q\overline{q}$ and non-semileptonic $e^+e^- \to B\overline{B}$ processes can be mistaken as semileptonic FEI modes motivate how important it is to investigate how the performance variables may differ between fake and real B_{tag}'s, such that these fake can be removed from future calculations of the signal yield.

The following distributions will be constructed from semileptonic FEI-skimmed signal and background MC. When investigating the key semileptonic FEI performance variables, all MC samples are combined with the events now grouped according to their underlying physics process. Datasets equivalent to $2.8\,\text{ab}^{-1}$ of charged and mixed events are used, with $1\,\text{ab}^{-1}$ of each continuum sample and the 8×10^6 $B^+ \to \mu^+\nu_\mu$ MC. While it is a charged B decay, the $B^+ \to \mu^+\nu_\mu$ process has not been generated in the $2.8\,\text{ab}^{-1}$ B^+B^- sample. The yield of each MC event type component is scaled to what is expected to be produced by 362 fb^{-1} of e^+e^- collisions. These scalings were summarised in Table 2.6. It is best practice to use a larger luminosity of MC than collected in data, so that even if a very small area of the parameter space is studied, there would still be sufficient number of underlying simulated events to support the analysis.

As will be seen, the high cross-section of background physics processes overwhelms the $B^+ \to \mu^+\nu_\mu$ component which is plotted on top of the following histograms. To help, the signal MC has been overlaid with significantly higher scaling to get an idea of the signal shape prior to any future data selections. This yet again highlights the importance of developing selection criteria that prioritises continuum and generic $B\overline{B}$ event rejection, so that the yields of the $B^+ \to \mu^+\nu_\mu$ component in data can be easily identified. To be able to decide on future selection criteria, it will be explored what FEI observables best differentiate between signal and background events. Such distributions are now presented, for signal and background MC samples scaled to 362 fb^{-1}, with pre-selection criteria of $p^*_\ell < 2.25$ GeV/c made prior to best candidate selection of accepting the highest signal probability B_{tag} per event .

To begin, the histogram of $\log_{10}(\texttt{sigProb})$ for B_{tag} identified via semileptonic FEI in $B^+ \to \mu^+\nu_\mu$ and background events scaled to 362 fb^{-1} is presented in Figure 3.30. For the first time, the magnitudes of the signal $B^+ \to \mu^+\nu_\mu$ process when compared to background ones is demonstrated. While the signal events are plotted in the stacked histogram, they are not visible with only eleven $B^+ \to \mu^+\nu_\mu$ events taggable in 362 fb^{-1} of e^+e^- collisions. The overlaid signal shape is plotted with 5 million times the number of events expected, in order to be visible. With this emphasis, the same signal shape that was plotted in Figure 3.23 is seen in this histogram, where it would be expected that correct reconstruction to dominated between $\log_{10}(\texttt{sigProb})$ = -1.5 and $\log_{10}(\texttt{sigProb})$ =-0.5 . The background components vary between $\mathcal{O}$(1 million-10 million) events each.

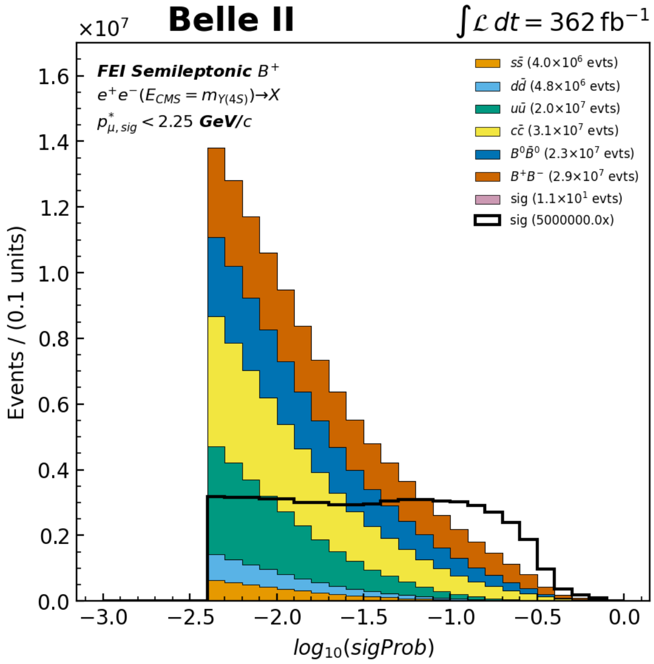


Figure 3.29. log10(signalProbability) B_{tag} decay modes identified via semileptonic FEI applied to MC simulated e^+e^- collisions at the $\Upsilon(4S)$ resonance. Stacking is according each type of process, each scaled to 362 fb^{-1}.

The $q\bar{q}$ components peak towards the more negative values of $\log_{10}$(`sigProb`), indicating they are being rated poorly by the semileptonic FEI, despite the large number of events that do survive. The $B\bar{B}$ components do not increase in yield as strongly towards $\log_{10}$(`sigProb`) =-2.5, which as true B events, the FEI struggles to reject with higher values of $\log_{10}$(`sigProb`). The overall shape difference is significant, indicating that background events are assigned different signal probabilities of being real B's by the FEI, and thus could be used in background rejection.

Figure 3.30 illustrates the semileptonic FEI decay modes reconstructed in signal and background MC, as expected within 362 fb^{-1} of e^+e^- collisions. Any differences between signal shape and background components are not obvious. For all MC types, the FEI tags more B mesons in the $B^+ \to \overline{D}^0\ \ell^+\ \nu_\ell$ modes than in the excited ones, and even more than the $B^+ \to D^{(*)}\pi^+\ \ell^+\nu_\ell$, a trend that was also seen in the B^+B^- samples. This suggests that all B_{tag} decay modes are more or less equally susceptible to mistaking background events for taggable ones.

Figure 3.31 demonstrates the 2D distributions of $\log_{10}$(`sigProb`) against semileptonic FEI charged B decay mode, where each of the distributions pertain to a difference MC type: $[B^+ \to \mu^+\nu_\mu\ ]B^-$, B^+B^-, $B^0\overline{B}^0$ and the four combined $q\bar{q}$ samples. The signal

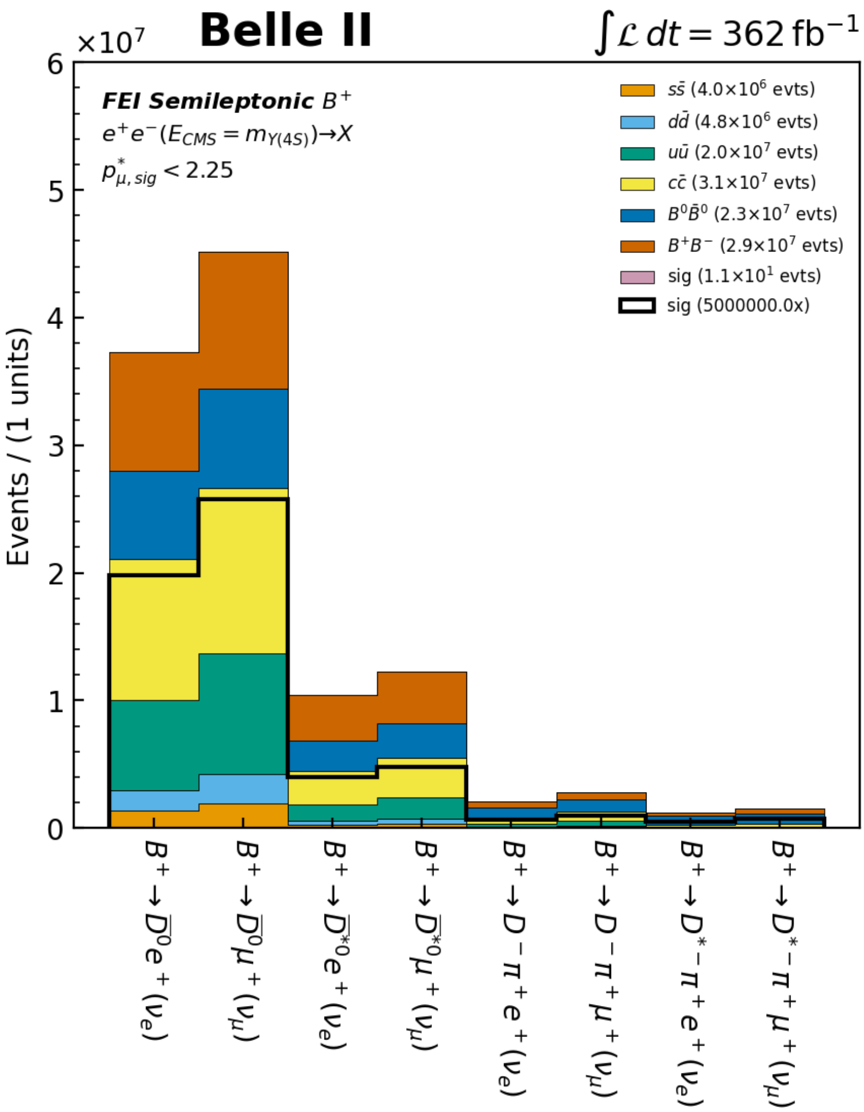


Figure 3.30. Reconstructed B_{tag} decay modes identified via semileptonic FEI applied to MC simulated e^+e^- collisions at the $\Upsilon(4S)$ resonance. Stacking is according to each type of process, each scaled to 362 fb^{-1}.

distribution in the top left is a reproduction was previously shown in Figure 3.25, while the charged version varies from what was shown in Figure 3.13 by the addition of the p^*_ℓ < 2.25 GeV/c selection as well as being produced with a larger MC dataset. The tagmaps do not provide more information than what has already been seen in the histograms of $\log_{10}$(`sigProb`) and decay mode, that B_{tag} from the signal MC tend do have more events between $\log_{10}$(`sigProb`) between -1.5 and 0.5 while other peak towards -2.5, and that the decay modes are uniform between MC sources. Though one feature of the tagmaps present in the background MC that is not seen in the signal sample is the

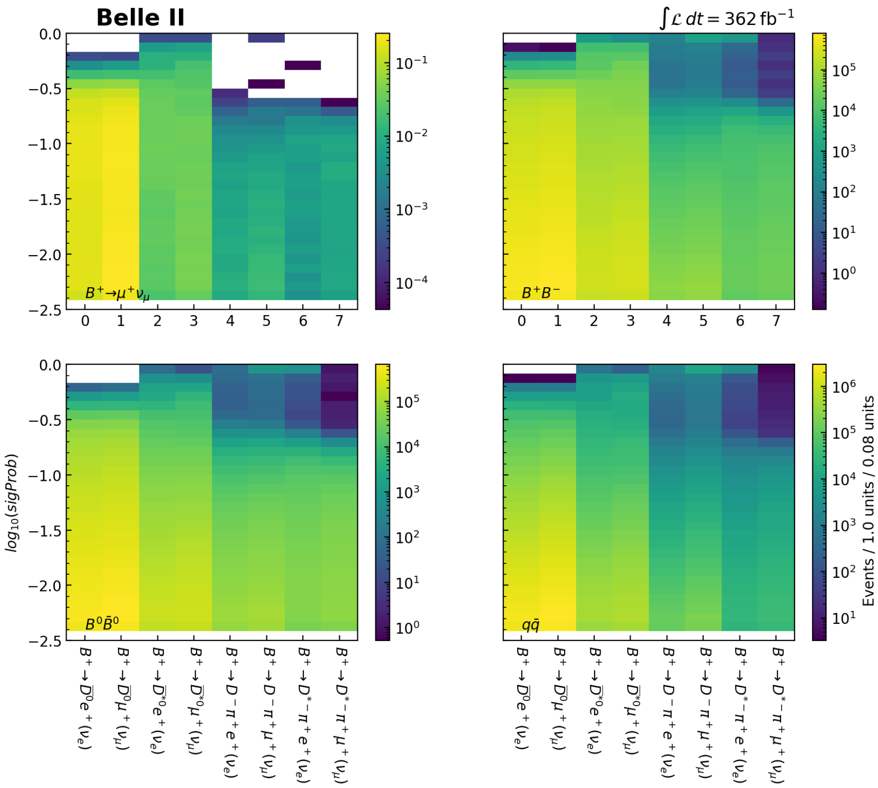


Figure 3.31. 2D Histogram of $\log_{10}(\texttt{sigProb})$ according to decay modes identified via semileptonic FEI, applied to MC simulated e^+e^- collisions at the $\Upsilon(4S)$ resonance. Top left: signal MC, top right: B^+B^- MC, bottom left: $B^0\overline{B}^0$ MC, bottom right: Combined $q\overline{q}$ MC.

relative increase in the number of events in the final $\log_{10}(\texttt{sigProb})$ bin. It was previously discussed that there should be very few events closer to zero in $\log_{10}(\texttt{sigProb})$ (i.e. closer to $\texttt{sigProb}$=1), especially for background events and that this could be due to some overtraining in these less prevalent 4-body modes. The effect of this will be monitored as the analysis continues.

The distribution of $\cos\theta_{BY}$ in background MC in Figure 3.32 varies substantially from signal MC. The signal shape is reproduced from what was seen in the signal-only $\cos\theta_{BY}$ histogram in Figure 3.26, which itself peaks between [-1,1] and has few events passed $\cos\theta_{BY} > 1$. As both the mixed and charged $B\overline{B}$ components of the stacked background contain real B decays, they also peak between [-1,1]. This was also seen in the 200 fb^{-1} charged MC B_{tag} analysis of the $\cos\theta_{BY}$ distribution in Figure 3.15. However, as the semileptonic FEI is specifically reconstructing charged B's, the B_{tag} candidates in the $B^0\overline{B}^0$ MC will always be misreconstructed. This results in the larger number of events in the misreconstructed $\cos\theta_{BY} > 1$ region than in the charged component. On the

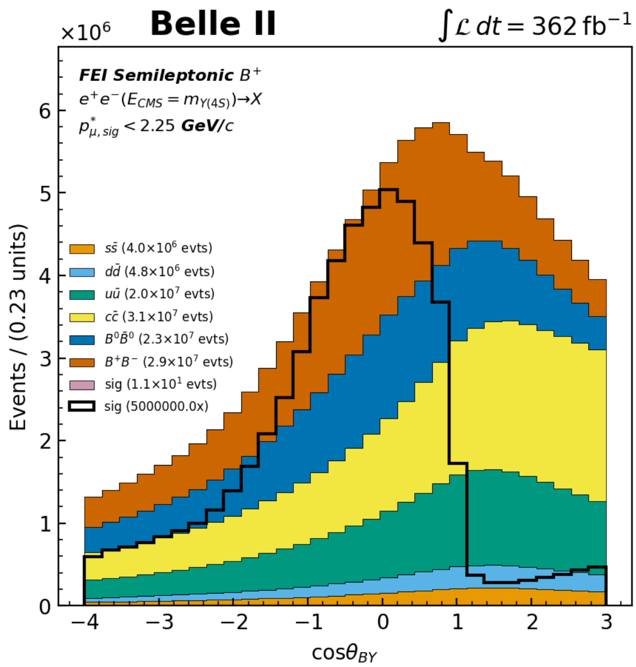
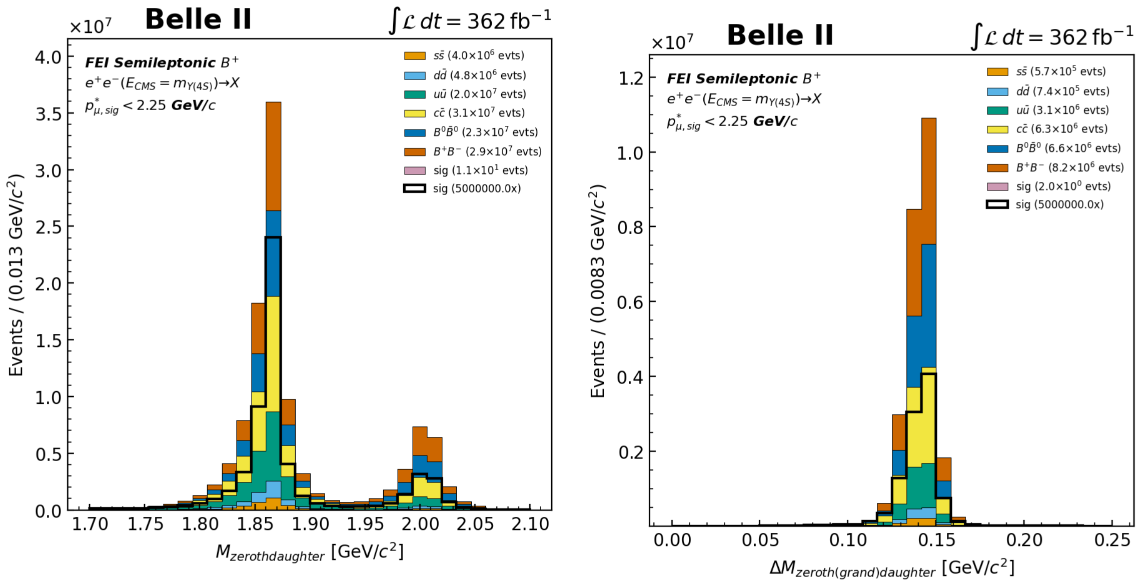


Figure 3.32. $\cos\theta_{BY}$ measurements of B_{tag} identified via semileptonic FEI applied to MC simulated e^+e^- collisions at the $\Upsilon(4S)$ resonance. Stacking is according to each type of process, scaled to 362 fb^{-1}.

other hand, the charged MC sample can contain true semileptonic FEI charged B decays, leading to more events in [-1,1] than the mixed sample. The $B\overline{B}$ signal and background components contrast strongly with the $q\overline{q}$ components, which instead peak on the interval [1,2], well into the misreconstructed region of $\cos\theta_{BY} > 1$.

In the discussion of the signal only $\cos\theta_{BY}$ histogram, it was decided that the $p^*_\ell <$ 2.25 GeV/c selection had rejected the misreconstructed events in the $\cos\theta_{BY} > 1$ region, by excluding the signal muon from tag-side selection. With no real signal muon to exclude in the background components, it is clear why the background components continue to admit events in the positive misreconstructed region of $\cos\theta_{BY}$. This continued shape difference provides a worthwhile selection to limit event analysis to $\cos\theta_{BY} <$ 1, thus removing a great deal $q\overline{q}$ events without considerable loss of signal events. All components, signal or background, have long 'tails' in the $\cos\theta_{BY} <$ -1 region. This corresponds to B_{tag} with more missing particles than just the expected neutrino in the semileptonic decay mode [107]. For signal and charged MC, this commonly occurs when excited $B \to D^*$ modes are mistagged as B decays containing lower state D. An example of this is if π^0 is not detected in $B^+ \to [\ \overline{D}^{*0} \to \overline{D}^0\ \pi^0\ ]\ \ell^+\ \nu_\ell$, leaving potential to be identified instead as $B^+ \to \overline{D}^0\ \ell^+\ (\nu_\ell)$. While the B^+ decay is mistagged, it is still true semileptonic B^+ decay, so it is of lower priority to reject. Implementing a lower bound on $\cos\theta_{BY}$ will be made primarily to reject the large amount of continuum and $B^0\overline{B}^0$ events than mistagged signal, which will be seen in Section 4.2.3.



(a) Zeroth daughter mass (i.e. D/D^*), with labelled PDG masses of D^+/D^0 and D^{*+}/D^{*0}.

(b) $\Delta M = m_{D^*} - m_D$ for $B \to D^* \to D$ FEI modes.

Figure 3.33. Daughter and (grand-)daughter mass splitting in B_{tag} identified via semileptonic FEI applied to MC simulated e^+e^- collisions at the $\Upsilon(4S)$ resonance. Stacking is according to each type of process, scaled to 362 fb^{-1}.

The masses of the zeroth daughter (i.e. D/D^*) of the semileptonic B_{tag} in signal and background MC, as well as the corresponding zeroth-daughter-zeroth-granddaughter mass splitting (i.e. $m_{D^*} - m_D$) restricted to $B^+ \to D^*$ modes are presented in Figure 3.33a and Figure 3.33b, respectively. There are no discernible differences in the shape between each of the background and signal components in both the daughter mass and mass splitting distributions. For the B_{tag} daughter masses, this implies that there is no MC sample that prefers D against D^* modes or vice versa. This corroborates what was seen in the relative yields of B_{tag} decay modes per MC type in Figure 3.30. It is clear that in the construction of the zeroth daughters of the B_{tag} candidates that there is biasing or 'shaping' of background events towards the D/D^* mass distribution, even in $e^+e^- \to q\bar{q}$ processes that cannot produce charmed hadrons. This is a result of the best-candidate variable for D reconstruction in the FEI to be those with lowest mass difference to the nominal D mass. This was previously mentioned in the table of best candidate variables for FEI particles in Table 3.1. Therefore, further selections on D/D^* mass variables will not be useful in rejecting background events over signal ones.

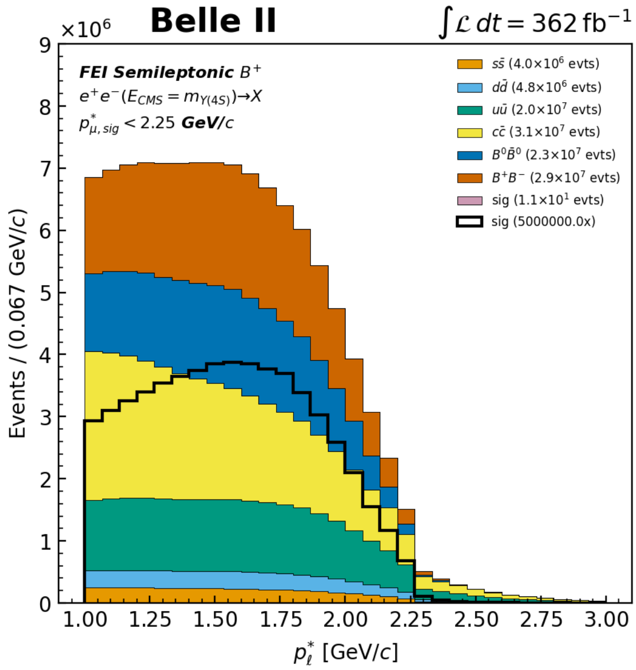


Figure 3.34. Lepton centre-of-mass frame momentum of B_{tag} identified via semileptonic FEI in signal and background MC scaled to 362 fb^{-1}.

The final distribution of semileptonic B_{tag} determined in signal and background MC scaled to 362 fb^{-1} is the tag-side leptonic momentum in centre-of-momentum frame. The majority of events for all components have $p^*_\ell < 2.25$ GeV/c. The $p^*_\ell < 2.25$ GeV/c was applied only to the semimuonic B_{tag} decay modes, as it was only in these modes that the μ_{sig} would be mistaken as the tag-side lepton, as was shown in Figure 3.22b. Given the small number of semielectronic events for which $p^*_\ell \geq 2.25$ GeV/c, this selection could be implemented across all decay modes without major impact on the yields of each MC component.

For the light quark continuum components, the tag-side lepton momentum spectrum peaks broadly between 1 and 1.75 GeV/c. The leptons identified in the $c\bar{c}$ component accumulate towards 1 GeV/c, emphasising the utility of the $p^*_\ell \geq 1$ GeV/c pre-selections hard-coded into the application of the semileptonic FEI. This is in contrast to the signal and background $B\overline{B}$ modes, which all peak between 1.5 and 2 GeV/c. Increasing the p^*_ℓ lower bound in analysed events could therefore be useful in rejecting $e^+e^- \to c\bar{c}$ events.

Table 3.6. Yields and Efficiencies of FEI Semileptonic B_{tag} in Belle II MC, with $p^{*}_{\mu,\text{tag}} <$ 2.25 GeV/c.

Sample ($e^+e^- \to X$)	No. of events (in sample)	No. of events (in 362 fb^{-1})	No. of events (B_{tag} in sample)	No. of events (B_{tag} in 362 fb^{-1})	Retention [%]
$[B^+ \to \mu^+\nu_\mu\ ]B^-$	8.00×10^6	1.74×10^2	5.28×10^5	1.15×10^1	6.60
B^+B^-	1.58×10^9	2.05×10^8	2.26×10^8	2.92×10^7	14.25
$B^0\overline{B}^0$	1.50×10^9	1.94×10^8	1.78×10^8	2.30×10^7	11.91
$c\overline{c}$	1.33×10^9	4.81×10^8	8.68×10^7	3.14×10^7	6.53
$u\overline{u}$	1.61×10^9	5.81×10^8	5.59×10^7	2.03×10^7	3.49
$d\overline{d}$	4.01×10^8	1.45×10^8	1.32×10^7	4.79×10^6	3.30
$s\overline{s}$	3.83×10^8	1.39×10^8	1.10×10^7	3.97×10^6	2.86

3.6. Summary

In this chapter, the role that the Full Event Interpretation plays to identify rare, missing energy decays at Belle II has been discussed. The general principles of B-tagging in e^+e^- colliders was presented and how they leverage the lack of background produced in $e^+e^- \to B\overline{B}$ processes, the hermetic detector and the precisely known collision energy to constrain $B\overline{B}$ decays. The structure of the FEI, as a hierarchial exclusive reconstruction tagging algorithm, was introduced, as well as the phenomenology of hadronic versus semileptonic B decays. These two charged B meson reconstruction modes of the FEI were compared in an MC sample of $e^+e^- \to B^+B^-$ decays, where the different key performance variables were assessed. As only one could be chosen for the $B^+ \to \mu^+\nu_\mu$ search, the analysis proceeded with the semileptonic FEI modes, as their higher efficiency in $B^+ \to \mu^+\nu_\mu$ MC would would assist in providing more events of this rare B decay. The distributions of B_{tag} produced by the semileptonic FEI in $B^+ \to \mu^+\nu_\mu$ MC was investigated, to understand the properties of misreconstructed and reconstructed B_{tag} that decay alongside $B^+ \to \mu^+\nu_\mu$. This investigation was then extended to distributions including full set of background processes expected in the 362 fb^{-1} collision dataset created at Belle II. Here, it became obvious that the magnitude of background to this rare B decay was substantial, as well as the observables that were better correlated with a B_{tag} that decays alongside $B^+ \to \mu^+\nu_\mu$.

In the next chapter, the B_{tag}'s identified with the semileptonic FEI in signal and background MC will be combined with muons most likely to originate from the $B^+ \to$

$\mu^+\nu_\mu$ process. From this, the $\Upsilon(4S)$ that produces the $B\overline{B}$ pairs will be reconstructed. The large magnitude of background seen in this Chapter will also be addressed by restricting the analysed dataset to events more likely to contain $B^+ \to \mu^+\nu_\mu$ than background events. From this, the best estimate of the number of signal events in 362 fb^{-1} of real Belle II collision data can be constructed.

$\bowtie\bowtie\bowtie$

Chapter 4.

Deriving selection criteria for identifying $B^+ \to \mu^+\nu_\mu$ in $362\,\mathrm{fb}^{-1}$ of Belle II data

The candidate events potentially containing the rare $B^+ \to \mu^+\nu_\mu$ decay have been partially reconstructed in MC via B-tagging. Chapter 4 will present the method of completing reconstruction via the addition of the signal muon.
The properties of the signal muon will be explored before and after its combination with the tagged-B candidate. This is done by consulting distributions of key observables that discriminate between true $B^+ \to \mu^+\nu_\mu$ MC events and background ones.
The process of further restricting the analysed dataset to events that more closely behave like $B^+ \to \mu^+\nu_\mu$ will be optimised via Figures-of-Merit procedures. Potential selection criteria are explored by varying observables and the limits on these observables, where each selected criterion will maximise a significance quotient of $B^+ \to \mu^+\nu_\mu$ MC events against $B^+ \to \mu^+\nu_\mu$-mimicking ones.
Once finalised, the same selections are applied to real Belle II collision data. In this way, the performance of MC in representing collision data can be evaluated, with further minor selections if necessary to regions of better agreement. Signal and background yields within MC will be used to estimate corresponding yields in data.
These yields will be summarised in the final section, to be used in the following Chapter for evaluating the ultimate $B^+ \to \mu^+\nu_\mu$ branching fraction.

$\bowtie \bowtie \bowtie$

4.1. Signal reconstruction

$B^+ \to \mu^+\nu_\mu$ events occur via $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ (and charge conjugated) processes at Belle II. The FEI algorithm has been applied to Belle II MC simulated events to restrict the search for $B^+ \to \mu^+\nu_\mu$ to those decaying alongside a charged, semileptonic decaying B meson. While this has limited the number of events that can be analysed, using fixed-decay B_{tag} provides additional observables that can be used to reject $B^+ \to \mu^+\nu_\mu$-mimicking events, thus supporting the goal of this chapter. A large amount of $B^+ \to \mu^+\nu_\mu$-mimicking events are not created via $\Upsilon(4S)$ decays. To limit these background events, identified B_{tag} will undergo $\Upsilon(4S)$ reconstruction with the signal B meson candidates.

With the B_{tag} reconstruction complete, focus is now given to reconstruction of the signal B meson. In the absence of neutrino four momentum information, B_{sig} reconstruction consists only of choosing a signal muon. If the neutrino were reconstructable, the choice of muon could simply be the candidate that yields the most accurate B mass when combined with the neutrino. Instead, signal muons will be selected as those most likely to originate from a B and decay alongside a single missing particle. Namely, this will exploit the properties of the two-body B decay, by enforcing a large muon momentum as well as limiting the number of detectable particles that remain after the muon is chosen. Moreover, signal muon choice will require consideration of those that can be assigned to a B_{sig}, which can then go on to be reconstructed into $\Upsilon(4S)$. This reconstruction can be expected to reject a number of background processes from consideration, including those that do not contain muons in them after the FEI B_{tag} has been identified.

Using the signal MC sample introduced in Chapter 3, the origins of the muons that can occur in the $e^+e^- \to \Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ (and charge conjugated) process will be studied. This will be used to inform best signal muon candidate selection prior to $\Upsilon(4S)$ reconstruction with the semileptonic FEI B_{tag}.

4.1.1. The signal muon best candidate selection

As final-state particles, muons are some of the most commonly occurring particles in B decay trees at Belle II. This makes selecting the best muon candidate (i.e. the muon most likely to originate from a $B^+ \to \mu^+\nu_\mu$ event) non-trivial.

In this subsection, the muons within the 8 million event $B^+ \to \mu^+\nu_\mu$ MC will be analysed, so as to determine features that distinguish signal muons from others in the $e^+e^- \to B^+B^-$ event. The sample of muons is of all charged final-state particles that can be detected within the simulation and can be truth-matched to a muon. This is done to include the cases of real muons that might be detected as other final-state particles, such as kaons or pions. This does not extend to cases where real kaons or pions mimic muons. Within `basf2`, the subset of charged final-state particles that are true muons is created by requiring `abs(mcPDG)`=13, where -13/13 is the number assigned to μ^+/μ^- in the Particle Data Group Monte Carlo Particle Numbering Scheme [4]. Moreover, this study will take place on the MC sample prior to the application of the semileptonic FEI, as the effect that tagging has on the underlying muon sample will be quantified later.

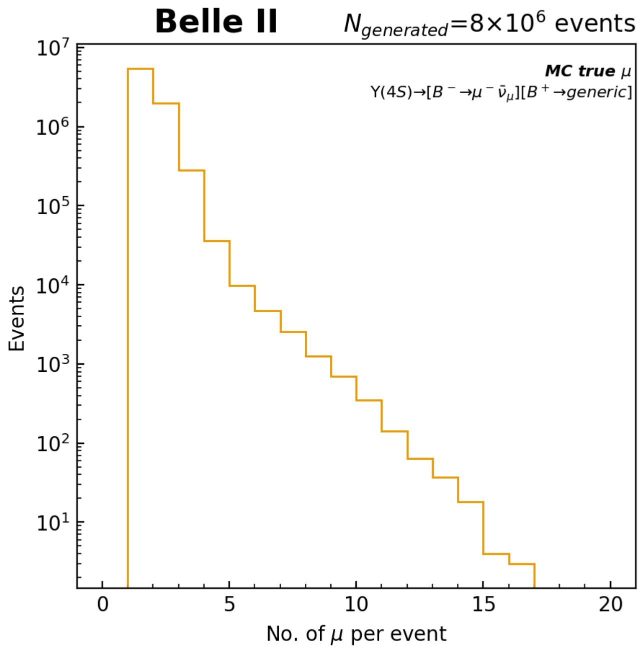


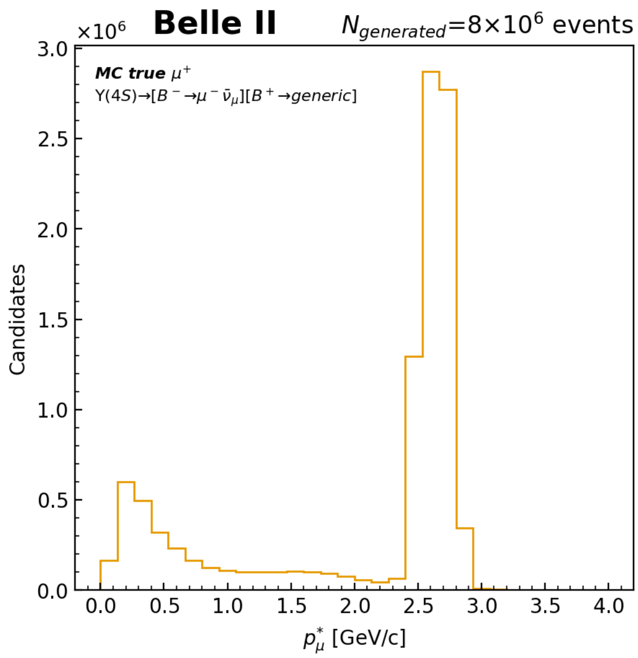
Figure 4.1. Number of muons per event, for all MC-true muons in $\Upsilon(4S) \to B^-$ $[B^+ \to \mu^+\nu_\mu$].

Figure 4.1 summarises the number of true muons per simulated $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu$ $]B^-$ event. This is derived from the muon sample derived above. Of the 8 million signal MC events utilised, 7.7 million events contained charged tracks which were simulated within the detector acceptances and thus could be truth-matched to a muon. While not indicated in the zero bin due to the requirement that the sample contained at least one muon, this means 0.3 million events did not contain muons generated within the detector acceptances. Across these 7.7 million events, 10.4 million muons are detected. As the sample is fixed to contain exactly one muon from the signal B meson, it can be inferred that any extra muons originate from the partnered B meson, or from material interactions. The origins of each muon will be seen later in Figure 4.3. With an average muon event multiplicity larger than 1 and thus the potential for multiple muons per event, this result emphasises the need for best signal muon selection criteria. Over 5 million events contain one muon in the event. 99% of events contain at most three.

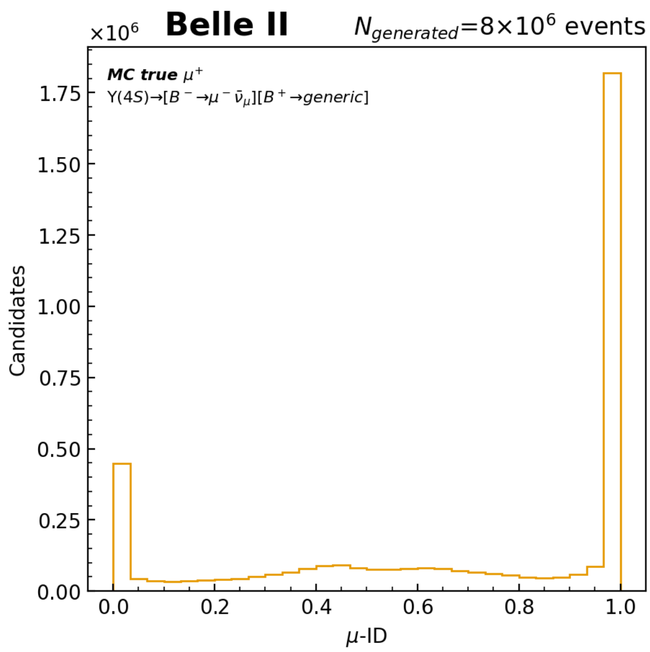
The distribution of p^*_μ (i.e. muon momentum magnitude in the $\Upsilon(4S)$ frame) for all muons across all events is presented in Figure 4.2a. As has been discussed in Section 1.2, a key feature of the muon in $B^+ \to \mu^+\nu_\mu$ is its large-but-narrow momentum distribution, originating from its fixed momentum in the B_{sig} frame. With the missing neutrino making the B_{sig} four-vector and thus frame inaccessible, the muon momentum is instead analysed in the $\Upsilon(4S)$ frame. The dominant feature between 2.25 GeV/c and 3 GeV/c can be taken to be the signal signal muon, by comparison with Figure 1.5 The origin of muons with $p^*_\mu < 2.25$ GeV/c as well as the small number of events with $p^*_\mu > 3$ GeV/c will be revisited soon.

An additional observable that can be used to appraise the sample of track-based final state particles is the muon identification probability, (referred to as `muonID` within `basf2`), as was described in Section 2.3.2. This variable will give an indication of how likely the charged particle chosen is in fact a muon and not a different track-based particle with kinematics that mimic muons, such as a charged pion or kaon. Figure 4.2b gives the distribution of this variable for all muons in the signal MC sample. Values of 1 indicate the charged track is muon-like, while 0 indicate the charged track is more likely to be one of the other charged FSPs (i.e. electron, charged pion, charged kaon, proton or deuteron). The distribution is bimodal in nature, indicating that the muon identification method can easily classify the tracks as least-fitting or best-fitting the muon hypothesis, while perhaps struggling in less obvious cases.

The two key variables for assessing muons in the signal MC have been demonstrated. In order to limit the analysis to one signal muon per event, which variables are most associated with the "correct" (i.e. signal) muon will be studied. In the same style of the best candidate selection of the B_{tag} in Section 3.4.2, the best candidate variable for signal muons will be determined by ranking methods. One option is to leverage the high-momentum property of the muon from $B^+ \to \mu^+\nu_\mu$ and rank the muons of each event according to their momentum in the $\Upsilon(4S)$ frame. Picking the signal muon to be the one with highest p^*_μ of the event will also be a useful way of rejecting the cases of B_{tag} candidates that mistake the signal muon with the muon in a semileptonic tag, as was seen in Figure 3.22. Best candidate selection via momentum ranking will be compared to



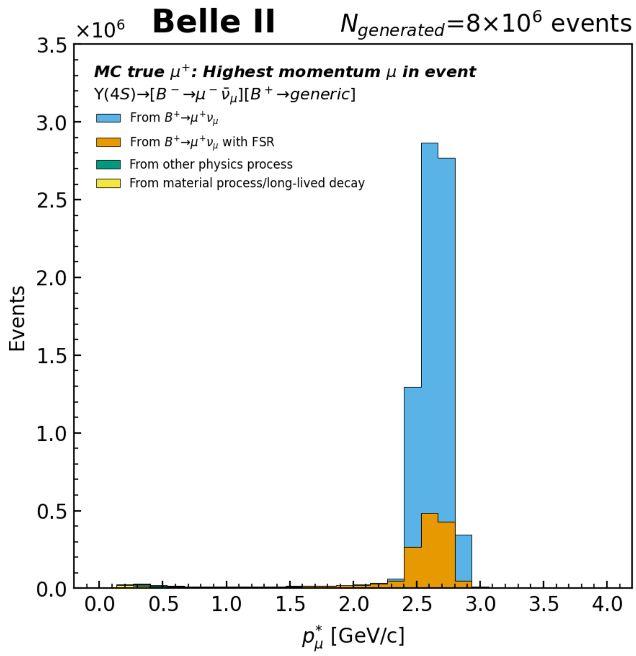
(a) muon momentum in $\Upsilon(4S)$ centre-of-momentum frame.



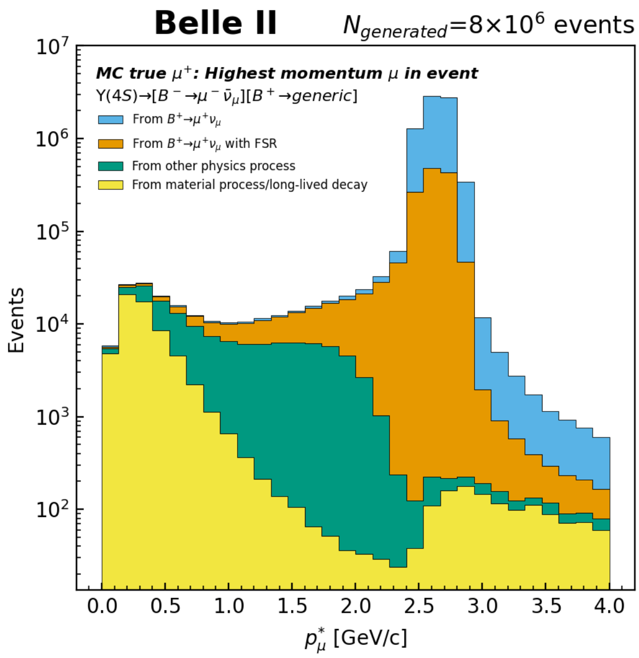
(b) Muon identification probability.

Figure 4.2. p^*_μ and μ-ID for all MC-true muons in $\Upsilon(4S) \to B^-$ $[B^+ \to \mu^+\nu_\mu$]. Note that the histogram counts number of muon candidates, not events containing muon.

ranking the muons of the event according to `muonID`. Ranking by this variable will avoid cases of other particles mimicking the signal muon, which can lead to other non-muonic charged B meson decays being mistaken for $B^+ \to \mu^+\nu_\mu$.

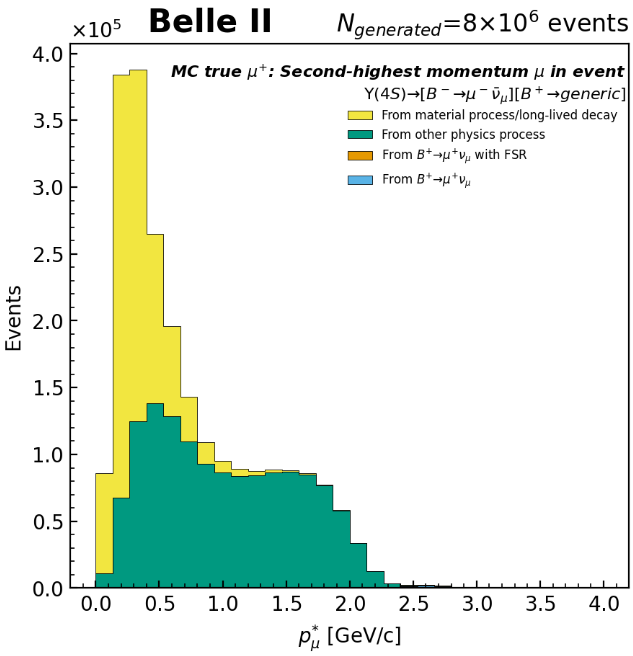


(a) Linear event scale.

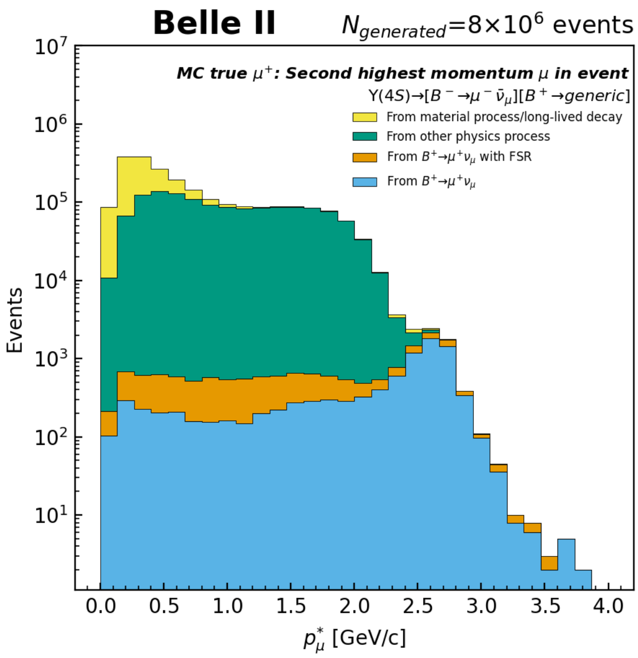


(b) Log event scale.

Figure 4.3. Muon momentum in $\Upsilon(4S)$ centre-of-mass frame (p^*_μ), for the muon with the largest p^*_μ in each event. Stacking is according to generated muon origin.



(a) Linear event scale.



(b) Log event scale.

Figure 4.4. Muon momentum in $\Upsilon(4S)$ centre-of-mass frame (p^*_μ), for the muon with the second-largest p^*_μ in each event. Stacking is according to generated muon origin.

The plots of Figure 4.3 demonstrate the range of p^*_μ for the highest ranked (i.e. largest momentum in the $\Upsilon(4S)$ frame) muons of each event. The muons are grouped according to how they were generated in MC, appearing as stacked components in the histogram.

Figure 4.3a demonstrates that the largest grouping of muons are those from $B^+ \to \mu^+\nu_\mu$ decays, largely residing in the $p^*_\mu > 2.4$ GeV/c peak. There is some momentum smearing for $p^*_\mu > 3$ GeV/c, though this proportion is small enough that it is only visible on the log-scaled of Figure 4.3b

The secondary grouping of muons are those from $B^+ \to \mu^+\nu_\mu$ generated to undergo final-state radiation (FSR). FSR refers to the emission of a photon from the muon (i.e. Bremstrahlung photons). Momentum is lost in the creation of the photon resulting in a lower recorded momentum than what it was created with. In the distribution, this corresponds to muons which can have p^*_μ below the peak. This can be partially remedied via Bremstrahlung recovery techniques, where suitable photons are combined with muons to obtain their original momentum and thus enhance the high momentum peak. This strategy was not implemented as preliminary studies indicated a drop in efficiency. Moreover, only 2% of real muon candidates undergo FSR that results in a sub-peak p^*_μ. As no Bremstrahlung recovery method is being implemented, it is natural to restrict the $B^+ \to \mu^+\nu_\mu$ search to high-momentum, non-FSR muons, as will be implemented when the signal region is decided on in Section 4.2.2.

The third largest group of muons in the $B^+ \to \mu^+\nu_\mu$ MC are other physics processes. Within the context of the MC utilised, this will be muons from the background B meson decays. The majority of such muons have momenta between 0 and 2.25 GeV/c. This same range was seen in the analysis of all candidate muons, suggesting the excess in Figure 4.2a is due to other B meson decays.

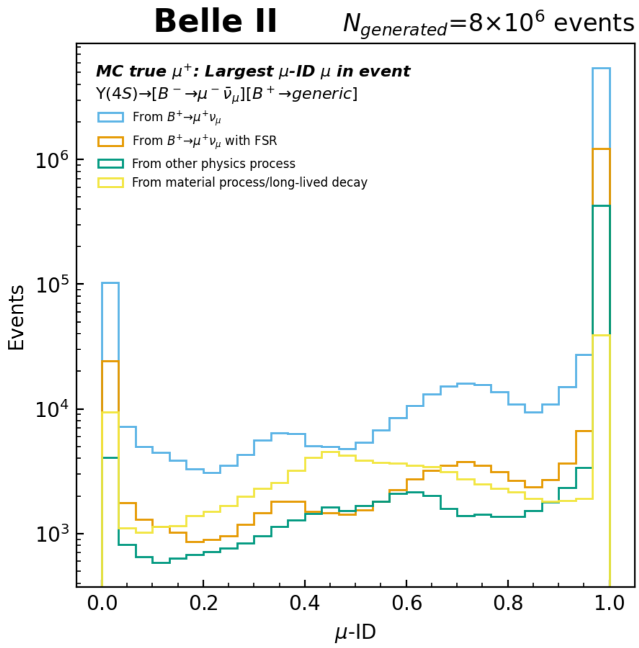
The final component is muons from 'secondary' physics processes. These differ from 'primary' final-state particles i.e. particles belonging to either B decay tree. Secondary muons can be created when primary particles interact with the materials of the detector or in the rarer instances of final state particles decaying within the detector medium (e.g. $\pi^+ \to \mu^+ \nu_\mu$). From the log-scaled distribution, the majority of secondary muons recorded are below 1 GeV/c. Secondary muons can be rejected by requiring that they are closer to the e^+e^- interaction point, thus more likely to originate from $B^+ \to \mu^+\nu_\mu$ and not outer material interactions. From the relative magnitudes of each component of the distribution, it can be seen that by only accepting the highest momentum muons per event, the majority of muons chosen originate from the $B^+ \to \mu^+\nu_\mu$ process.

To verify that choosing rank-1 muons is best correlated with the muon of $B^+ \to \mu^+\nu_\mu$, the rank-2 candidates are examined. Figure 4.4 demonstrates the range of p^*_μ for muons with the second largest p^*_μ in each event. From the linear scale of Figure 4.4a, is is clear that rank-2 muons are dominated by secondary muons and those from other decays within the B decay chain. On the log-scaled of Figure 4.4b, there is a very small proportion of both FSR and non-FSR muons from $B^+ \to \mu^+\nu_\mu$, much smaller than what was seen in the rank-1 version. It can therefore be said that choosing the muon candidates with the largest p^*_μ per event results in large retention of true $B^+ \to \mu^+\nu_\mu$ muons. The distinct differences in the shape of the two ranked p^*_μ distributions, as well as the differences in the sources of muons indicate a large rejection of non-$B^+ \to \mu^+\nu_\mu$ muons at rank-1, with only a few true signal muons lost at rank-2.

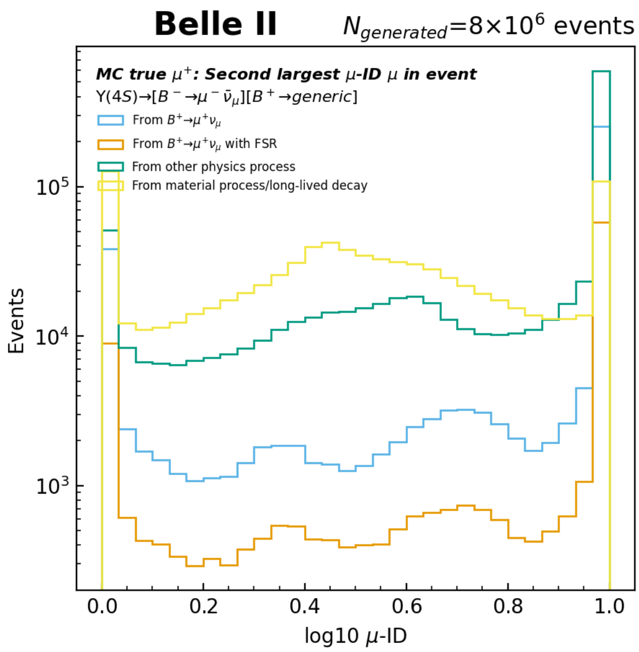
Comparison of p^*_μ as a ranking variable can be made with the alternate muon quality variable, `muonID`. Figure 4.5 presents the muons with the highest and second highest `muonID` per event in the $B^+ \to \mu^+\nu_\mu$ MC sample. The muons are grouped as before, though this time overlaid rather than stacked, to emphasise shapes of the individual log-scaled distributions. Unlike the p^*_μ rank-1 muons, Figure 4.5a shows lesser shape discrimination between the muons of true $B^+ \to \mu^+\nu_\mu$ events and others. The plot contains muons of all origins demonstrating large peaks at `muonID`=1 and slightly lesser ones at `muonID`=0. For `muonID` to be a well-performing ranking variable for identifying the muon in $B^+ \to \mu^+\nu_\mu$, we'd expect rank-1 muons to consist mostly of signal muons closer to 1 and for the rank-2 muons of Figure 4.5b more background muons than $B^+ \to \mu^+\nu_\mu$ ones towards `muonID`=0. Thus choosing muons of higher `muonID` isn't necessarily correlated with choosing the muon of $B^+ \to \mu^+\nu_\mu$, only that the charged track is muon-like.

To numerically verify the performance of p^*_μ over `muonID` as a muon ranking variable, the purity of each rank (i.e. fraction of muon originating from $B^+ \to \mu^+\nu_\mu$ out of all muons at this rank) for each variable is examined. Figure 4.6 demonstrates that choosing p^*_μ rank-1 muon results in more than 95% of charged tracks being the muon in $B^+ \to \mu^+\nu_\mu$. Less than 5% of p^*_μ rank-2 muons originate from $B^+ \to \mu^+\nu_\mu$. As for `muonID` ranking, 90% of rank-1 muons are from $B^+ \to \mu^+\nu_\mu$, with as many as 20% of rank-2 muons originating in $B^+ \to \mu^+\nu_\mu$. There is thus a fair proportion of real signal muons with lower `muonID` amongst other muons in the event, that will be lost if choosing the rank-1 `muonID` muon candidates.

By analysing two potential ranking variables, the best-candidate variable for identifying signal muons from any number of muons in the event is decided to be p^*_μ. This study



(a) muon with largest μ-ID in the event.



(b) Muon with second largest μ-ID in the event.

Figure 4.5. Muon identification probability for any muon in $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ events. Components are according to muon origin. They are overlaid rather than stacked for visualisation.

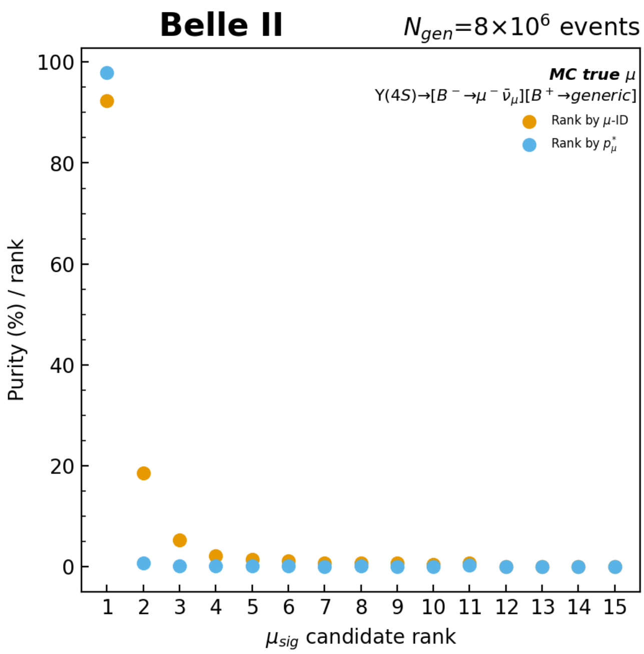


Figure 4.6. Purity of $\mu_{\rm sig}$ candidates (i.e. percentage of muons that are generated from $B^+ \to \mu^+\nu_\mu$) according to their variable ranking, in `abs(mcPDG)`=13 charged tracks from 8 million $e^+e^- \to \Upsilon(4S)$ $[B^+ \to \mu^+\nu_\mu\ ]B^-$ processes.

has been conducted using all muon in the event without any reconstruction. Now the semileptonic $B_{\rm tag}$ reconstruction can be incorporated to see the effect this has on muon selection.

4.1.2. The signal muon after B-tagging

In the previous section, the dynamics of true muon in $B^+ \to \mu^+\nu_\mu$ MC was investigated, to learn how to choose the signal muon from generic ones. Now, how muon selection takes place when the particle identity isn't guaranteed is examined, as well its compatibility with $B_{\rm tag}$ in being reconstructed into $\Upsilon(4S)$. In Figure 3.22b, it was seen that the semileptonic FEI could confuse muons from the signal side in $B^+ \to \mu^+\nu_\mu$ MC as tag-side muons. To avoid such cases of misreconstruction, a restriction to tagged-B mesons with $p^*_\ell < 2.25$ GeV/c was implemented, as this would exclude signal muons with $p^*_\mu > 2.25$ GeV/c. With the identification strategy of the signal muon now determined, this selection can now be revisited, where p^*_ℓ can be compared to the tag-side lepton directly.

The goal is to determine which selection is better in identifying signal events. It will also be investigated how to further constrain the dataset in other variables. Due to the rarity of the $B^+ \to \mu^+\nu_\mu$ mode compared to the larger cross-sections of other e^+e^- processes, utilising an unrestricted dataset of background MC can take up considerable disk size. Selection criteria applied prior to reconstruction or *pre-selections*, can reduce the size of the background dataset by removing the events least likely to survive reconstruction. For the same reason, it is also best to remove events well outside the region where signal events can occur.

Table 4.1. Selections implemented in determination of μ_{sig}. These are recommendations by Belle II's LeptonID performance group.

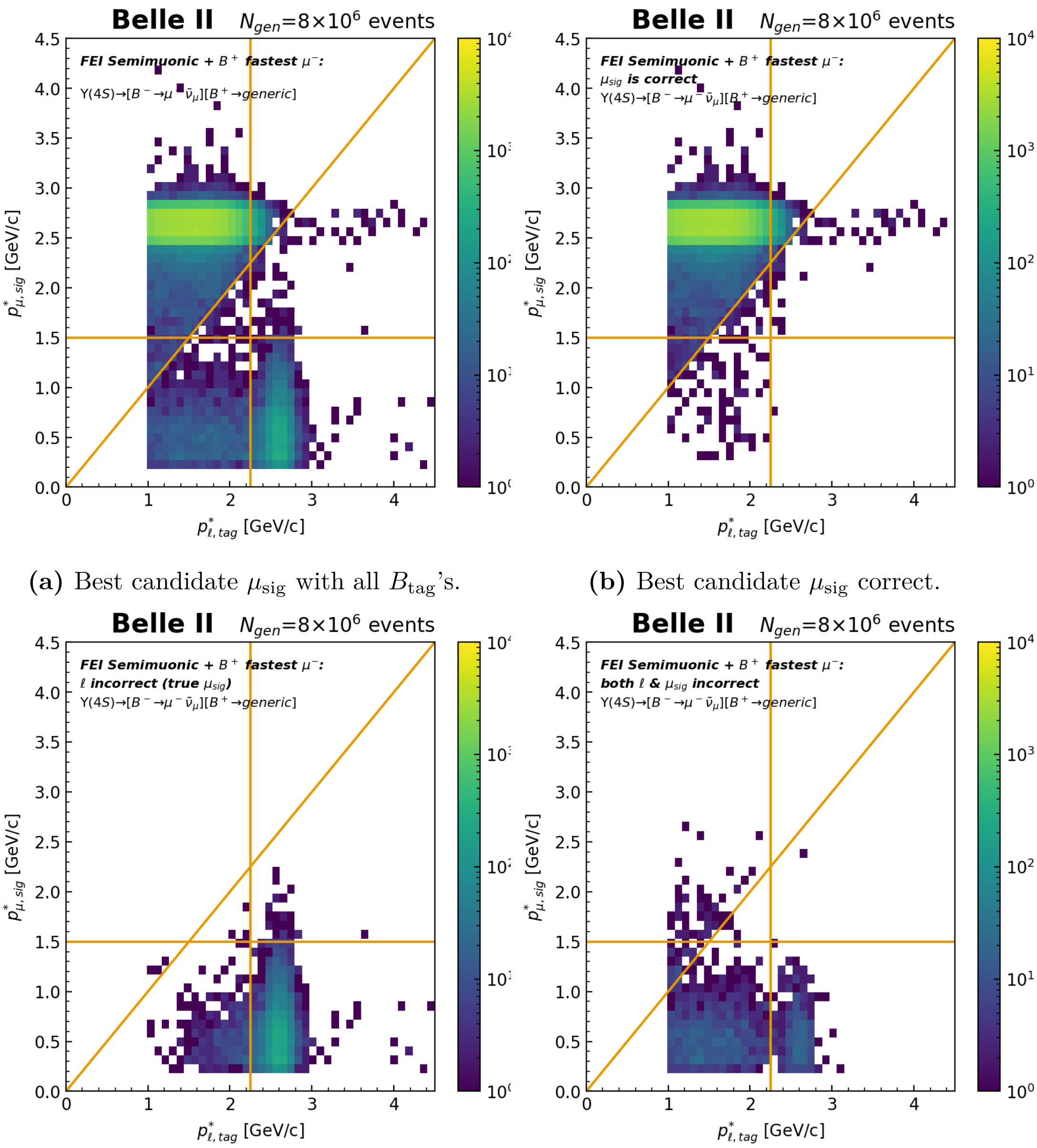
Muon selection (from track recommendations)	**Explanation**
`dr` < 2 and `abs(dz)` < 4 (cm)	Maximum transverse and z-distance of track origin from interaction point, ensures muon originates from $\Upsilon(4S)$ event.
`pt` > 0.2 (GeV/c)	Minimum transverse momentum, ensures μ_{sig} has enough outwards momentum to reach later subdetectors like the KLM.
`thetaInCDCAcceptance` = True	Only accept signal muons produced at polar angle θ that can reach CDC subdetector, $\theta \in (17°, 150°)$.
Muon selection (imposed by `stdCharged.stdLep` function)	**Explanation**
`muonID` > 0.5	Minimum threshold on tracks implemented in muon list, must have 0.50 likelihood of being a muon over other charged particles.
$0.2 \leq$ `p` ≤ 6.5 and $0.4 \leq$ `theta` ≤ 2.6 (GeV/c and rad)	Magnitude and angular ranges for which data/MC corrections exist.

To compare the efficacy of the pre-selection $p^*_\ell < 2.25$ GeV/c against $p^*_\ell < p^*_\mu$, the required leptons are compared within `basf2`. A list of candidate muons, from the 8 million event $e^+e^- \to \Upsilon(4S)$ $[B^+ \to \mu^+\nu_\mu\ ]B^-$ signal sample. The requirement of `abs(mcPDG)`=13 is dropped, so as to incorporate potential muon backgrounds. The recommendations of the Belle II Lepton ID group to fill a list of all muons using the `basf2` function `stdCharged.stdLep`. This includes implementing the minimum `muonID`, and the allowable ranges of momentum magnitude and momentum polar angle for which data/MC correction factors exist. These selections are described in Table 4.1. The necessity of such correction factors will be discussed in Chapter 5. Within the list, muons of the same event are ranked by p^*_μ to obtain the muon most likely to be the muon from $B^+ \to \mu^+\nu_\mu$. Finally, cases where the best candidate muons are used in $B_{\rm tag}$ reconstruction are rejected, so that the signal muon and the tag-side muon are distinct particles. This can functionally act as event selection, as there may be cases where only a single $B_{\rm tag}$ candidate is created within the FEI, and that $B_{\rm tag}$ fails as it has been reconstructed using the true signal muon.

Figure 4.7 contains 2D histograms that compare the tag-side lepton from tagged B mesons before signal probability ranking, against the corresponding best muon candidate. With the plots created with the intention of rejecting signal-side/tag-side muon confusion, the tagged B mesons plotted are only the semimuonic modes. Overlaid are a number of prospective selection criteria boundaries to investigate: $p^*_\ell < 2.25$ GeV/c (i.e. tag-side lepton momentum in $\Upsilon(4S)$ frame larger than 2.25 GeV/c, as was implemented in Section 3.5.1), $p^*_\ell < p^*_\mu$ (newly proposed) and $p^*_\mu > 1.5$ GeV (alternate choice). $p^*_\mu > 1.5$ GeV is an example loose bound, as it was seen in Figure 4.3b that such outside regions are dominated by secondary and non-$B^+ \to \mu^+\nu_\mu$ muons.

Figure 4.7a contains all $B_{\rm tag}$ candidates and their corresponding event's signal muon. The greatest density of candidates correspond to p^*_ℓ with between 1 and 2.5 GeV/c and p^*_μ between 2.4 and 2.9 GeV/c. These are the expected regions of correctly assigned tag-side muon and signal muon. This is corroborated by Figure 4.7b, where the same event density region is seen when p^*_μ is limited to be from correctly identified signal muon. This distribution of tag-side muon momentum when the correct signal muon is identified does not have the secondary mirrored structure on the opposite axis, which is present in Figure 4.7a. The high momentum tag-side muon and sub-1.5 GeV/c signal muon structure can be said to originate from tag-side muon/signal muon misreconstruction, as the same structure is seen when the events are limited to this condition in Figure 4.7c. For completeness, Figure 4.7d demonstrates the distribution when the true signal muon is skipped entirely, both in tagging and signal muon choice. This further demonstrates that the peaking high momentum tag-side lepton must be a result of signal muon-tag-side muon confusion, as the peak is not present in this distribution when the signal muon is not tagged.

From Figure 4.7a, the selection of $p^*_\ell < p^*_\mu$ provides a natural boundary between high momentum and correct signal muon against high momentum and misreconstructed tag-side muon. The previous selection of $p^*_\ell < 2.25$ GeV/c omits the majority of the misreconstructed tag-side muon's, though leaves behind the smeared signal muon, likely from final state radiation. Moreover, the vertical selection removes a small number



(a) Best candidate $\mu_{\rm sig}$ with all $B_{\rm tag}$'s.

(b) Best candidate $\mu_{\rm sig}$ correct.

(c) $\ell_{\rm tag}$ is misreconstructed from μ^+ in $B^+ \to \mu^+\nu_\mu$.

(d) Neither best candidate $\mu_{\rm sig}$ nor $\ell_{\rm tag}$ are reconstructed from μ^+ in $B^+ \to \mu^+\nu_\mu$.

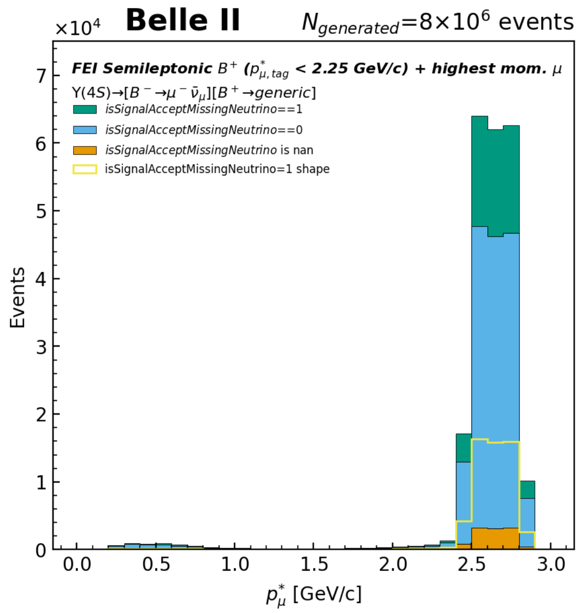
Figure 4.7. A 2D histogram of p^*_ℓ from potential semimuonic $B_{\rm tag}$ against p^*_μ for best candidate $\mu_{\rm sig}$ in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu]B^-$ signal MC events, grouped by how successful the identification is. Potential selection boundaries prior to best $B_{\rm tag}$ candidate selection are overlaid: $p^*_\ell < p^*_\mu$, $p^*_\ell < 2.25$ GeV/c (previously used in Section 3.5.1) and $p^*_\mu > 1.5$ GeV/c.

of correct signal muon candidates that occur alongside tag-side muon with $p^*_{\mu,\mathrm{tag}} \geq 2.25\,\mathrm{GeV}/c$, events that will be retained by the diagonal selection. The $p^*_\ell < p^*_\mu$ selection also has the added benefit of removing the signal muon reconstruction with $p^*_\mu < 1\ \mathrm{GeV}/c$, where a secondary particle peak was seen in Figure 4.3b. The selection $p^*_\mu > 1.5\ \mathrm{GeV}/c$ will be revisited in determining a signal counting region in Section 4.2.3.

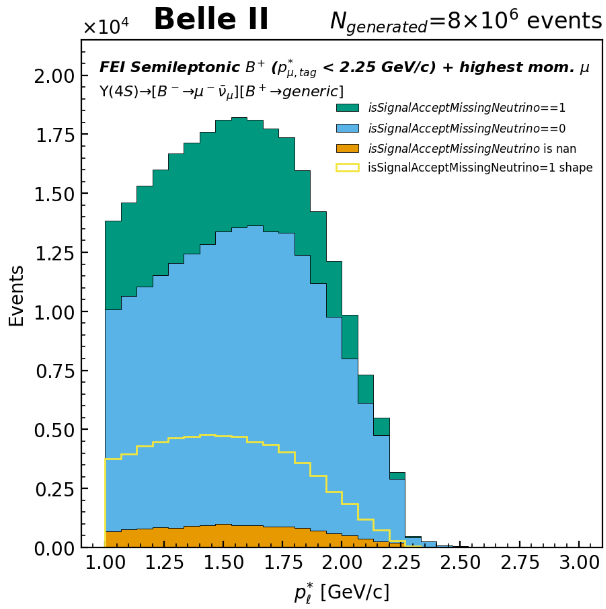
Thus, moving forward, a pre-selection of $p^*_\ell < p^*_\mu$ will be implemented, prior to $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ (and charge conjugated) reconstruction. For uniformity, this will be applied to both semielectron and semimuonic events. The effect of this on the p^*_μ and p^*_ℓ distributions are demonstrated in Figure 4.8, where the two selections are applied before B_{tag} best candidate selection. As desired, this selection removes a negligible number of events from the signal peak with the $p^*_\ell < p^*_\mu$ in Figure 4.8c. What is removed is the low momentum, secondary and other physics structure, which is present in in the $p^*_\ell < 2.25\ \mathrm{GeV}/c$ Figure 4.8a. The $p^*_\ell < p^*_\mu$ selection has also resulted in some shape reduction in the `isSignalAcceptMissingNeutrino` =0 component in the p^*_ℓ distribution Figure 4.8d. Specifically, there are less events at high p^*_ℓ which is expected and desired from the $p^*_\ell < p^*_\mu$ selection when compared to Figure 4.8b.

The impact of the new $p^*_\ell < p^*_\mu$ selection is also investigated in the signal muon `muonID`, in Figure 4.9. Here, `muonID` is limited to `muonID` ≥ 0.5, as per the limitation described in Table 4.1. Moreover, the events are on log-scale, due to the large difference between the number of events in the `muonID` ≥ 0.9 bin and those outside of this. The distributions are similar between the two p^*_ℓ selections, with only minor reduction in the number of events towards `muonID`=0.5. This suggests that the $p^*_\ell < p^*_\mu$ selection has resulted in muon candidates with a higher probability of being muons and not other charged particles.

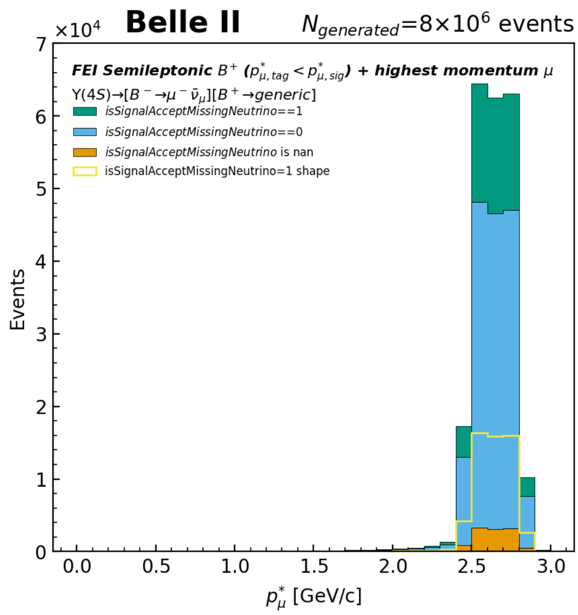
The $p^*_\ell < p^*_\mu$ selection is now adopted prior to best candidate signal muon and B_{tag} selection. An additional quality variable can now investigated, in a comparison between the two leptons of the $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ event. Figure 4.10 contains distributions of the opening angle between the tag-side lepton and signal muon in signal MC, both as cosine and in radians. Specifically, the opening angle is the angle between p^*_ℓ and p^*_μ at the point of their creation in the $\Upsilon(4S)$ reference frame. For `isSignalAcceptMissingNeutrino` =1 B_{tag} in Figure 4.10a i.e. the events with correctly identified B_{tag}, the $\cos\theta^*_{\mu_{\mathrm{sig}},\ell_{\mathrm{tag}}}$ is uniformly distributed. The corresponding $\theta^*_{\mu_{\mathrm{sig}},\ell_{\mathrm{tag}}}$ distribution between 0 and π in Figure 4.10b is not uniform, instead sinusoidal with maximum at $\theta^*_{\mu_{\mathrm{sig}},\ell_{\mathrm{tag}}} = \frac{\pi}{2}$. This indicates that for correct B_{tag}, the leptons will be preferentially emitted orthogonal to one another in the $\Upsilon(4S)$ frame, with lesser



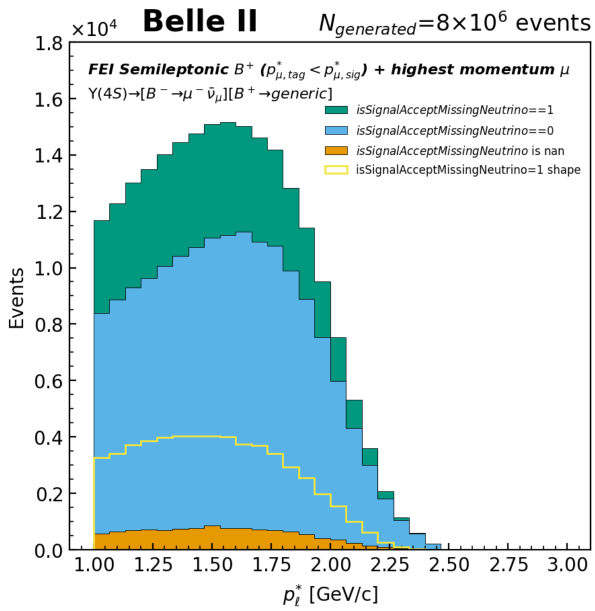
(a) μ_{sig} after $p^*_\ell < 2.25$ GeV/c.



(b) ℓ_{tag} after $p^*_\ell < 2.25$ GeV/c.

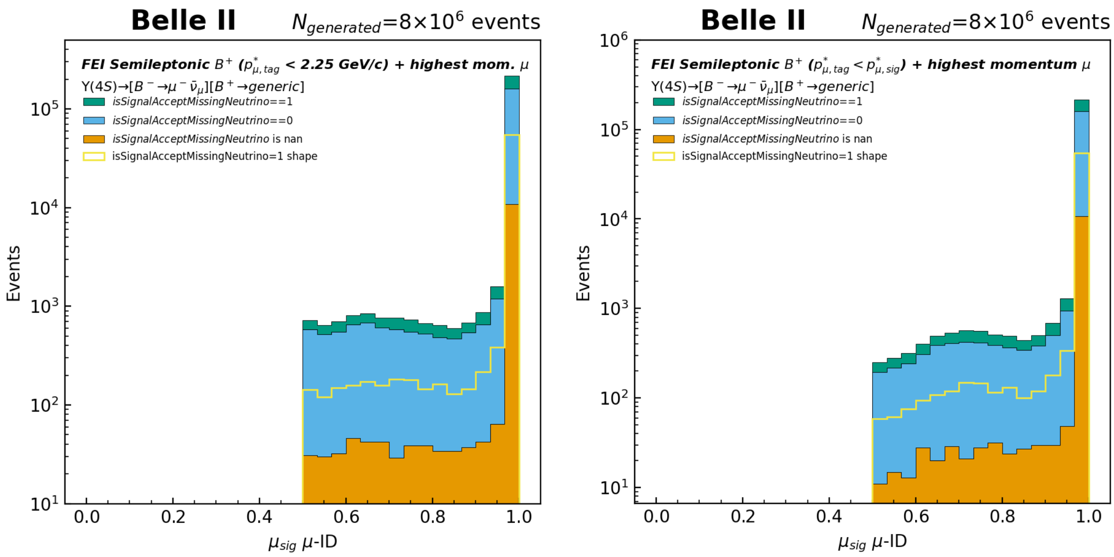


(c) μ_{sig} after $p^*_\ell < p^*_\mu$.



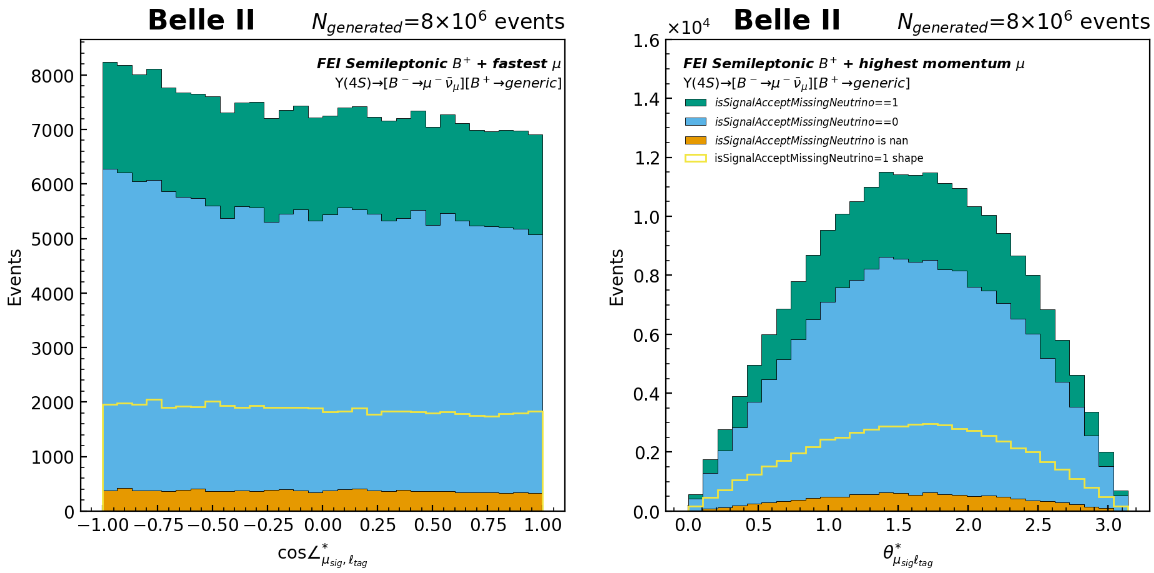
(d) ℓ_{tag} after $p^*_\ell < p^*_\mu$.

Figure 4.8. Semileptonic FEI tag-side lepton and signal muon $\Upsilon(4S)$ frame momentum with two different pre-selection criteria, in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ MC events.



(a) $p^*_\ell < 2.25$. **(b)** $p^*_\ell < p^*_\mu$.

Figure 4.9. Signal muon identification probability with two different pre-selection criteria, in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ MC events.



(a) Cosine of opening angle. **(b)** Opening angle.

Figure 4.10. The cosine opening angle and opening angle in radians between semileptonic FEI tag-side lepton and signal muon in $\Upsilon(4S)$ frame momentum, in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ (and charge conjugate) MC event.

likelihoods of being produced near parallel/anti-parallel. The cosine distribution of the `isSignalAcceptMissingNeutrino` =0 component of B_{tag} demonstrates slight peaking towards $\cos\theta^*_{\mu_{\text{sig}},\ell_{\text{tag}}}$, suggesting that misreconstructed and background physics events prefer extreme cosine values i.e. prefer being produced near parallel/anti-parallel. Such angular behaviour will provide a key method of signal-background shape demarcation which will be seen in Section 4.1.4.

4.1.3. The Rest of Event for B_{tag} and signal muon

The majority of $B^+ \to \mu^+\nu_\mu$ searches were performed using an inclusive B_{tag}. The strategy of the inclusively tagged searches was that after identifying the tracks and clusters attributed to the signal B decay, all others could be treated as particles of the B_{tag} decay chain. This assumption held because no additional particles would be created in the $\Upsilon(4S) \to B\overline{B}$ process. Additionally, the detectors were treated as hermetic, such that all particles of the event could be measured and thus a full B_{tag} four-momentum could be obtained. The set of unused tracks and clusters remaining after reconstruction has taken place is called the *Rest of Event*, or (ROE).

If instead the B_{tag} has been exclusively reconstructed, the ROE can be used to assess the quality of the reconstruction. For this analysis, if both the B_{tag} decay chain and the signal muon have been chosen correctly in a $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-$ (and charge conjugated) event, there should be no further primary particles measured by the detector. This of course excludes the tag-side and signal-side neutrinos, as these are incapable of leaving charged tracks or creating electromagnetic clusters in the detector. If after reconstruction there are particles remaining in the event, this could indicate that either the B_{tag} has insufficiently reconstructed the event, or that the event was not an $e^+e^- \to \Upsilon(4S)\ [B^+ \to \mu^+\nu_\mu\,]B^-$ process.

Specifically, the ROE object will be created for the FEI semileptonic B_{tag}, where only a single muon should be present if reconstruction has been performed correctly. Variables concerning the ROE object, such as the number of charged tracks that remain in the ROE, can reject $e^+e^- \to q\overline{q}$ events, as it is unlikely that all tracks and clusters will be used in creating a fake B candidate of non-$b\overline{b}$ events. Additionally, the ROE variables will provide powerful pre-selection criteria to minimise the disk size of the resulting dataset. In Section 4.1.5, when the B_{tag} and signal muon are combined, a second ROE will be created to reject events with particles remaining after a $\Upsilon(4S)$ candidate is reconstructed.

In Section 2.4.3, it was described that `basf2` has an in-built function to create the ROE for a given particle. It will now be described how the ROE is created in `basf2` for the semileptonic FEI B_{tag}. The goal currently is to measure the ROE variables prior to $\Upsilon(4S)$ reconstruction. This will be used as an indication of the ROE variables after reconstruction, so that selections can be made to reduce the size of the dataset in regions where correct $\Upsilon(4S)$ candidates are not expected. Thus, the signal muon is also explicitly removed from the rest of event, as it will be used in $\Upsilon(4S)$ reconstruction. This step won't be necessary once they are combined in the ultimate $\Upsilon(4S)$ ROE.

To prepare the ROE in `basf2`, a list of all the photons and charged tracks measured in the event are created using the `modularAnalysis.fillParticleList` function. Using these lists and the particle list around which the ROE is built (in this case the B_{tag} from the semileptonic FEI), the `basf2` function `modularAnalysis.buildRestOfEvent` filters out which are used in the B_{tag} reconstruction. The signal muon determined via the procedure of Section 4.1.2 is also removed from the ROE. A minimum threshold (referred to as a *mask*) is applied to the remaining tracks and clusters, so as avoid inclusion of particles outside the B-decay chains, like the secondary muons described in Section 4.1.1. The mask used for this analysis is described in Table 4.2.

With the ROE object, minus the signal muon, created around the B_{tag}, two related variables in the signal MC sample can be consulted. The first of such variables is referred to as `nROE_Tracks` within `basf2`. This is the number of charged tracks in the ROE, displayed in Figure 4.11. Here, there are up to three unused tracks, though there are a comparatively small number of events that contain exactly 3. The vast majority of events contain no additional tracks. This zeroth bin also contains near-all the `isSignalAcceptMissingNeutrino` =0 truth-matched B_{tag}. Thus, it is motivated that a selection of `nROE_Tracks`=0 could be effective in rejecting misreconstruction via background contributions, where excess tracks are expected.

In addition to the number of tracks that remain in the ROE, the electromagnetic clusters that remain in the ROE can also be investigated. Unused clusters can be energy deposits from unreconstructed photon or associated with the ROE tracks. Specifically, the total energy of the unused ECL clusters is examined, referred to as the ECL energy (E_{ECL}, also E_{extra}). In a correctly reconstructed event, there should be zero unused ECL clusters, as all clusters used to create physics photons or the clusters created by charged tracks will have been utilised. The energy selections in the ROE mask will minimise the effect of secondary photons. Limiting the ECL energy will reject background physics events as well

Table 4.2. Selections implemented in ROE mask. These are combined recommendations by Belle II's Neutrals performance group and Analysis Tools group.

Track selection (model as pions)	**Explanation**
`dr` < 0.5 *and* `abs(dz)` < 3 (cm)	Minimum transverse and z-distance of track origin from interaction point, ensures tracks originate from $\Upsilon(4S)$ event.
`pt` > 0.2 (GeV/c)	Minimum transverse momentum, ensures tracks has enough directional momentum to reach ECL subdetector
`thetaInCDCAcceptance` = True	Only accept tracks produced at polar angle θ that can reach CDC subdetector, $\theta \in (17°, 150°)$.
Cluster selection (model as photons)	**Explanation**
`clusterNHits` > 1.5	Number of crystals in ECL cluster, cluster must consist of more than one.
`abs(clusterTiming)` < 200 and `abs(clusterTiming/clusterErrorTiming)` < 2	Duration after e^+e^- collision event that the ECL cluster is created, and its proportion fo uncertainty. Ensures ECL cluster timing corresponds to being created from $\Upsilon(4S)$ decay chain and not secondary interactions.
`thetaInCDCAcceptance` = True	Only accept photons produced at polar angle θ that can reach CDC subdetector, $\theta \in (17°, 150°)$.
0.2967 < `clusterTheta` < 2.6180 (rad)	Allowable angles polar angle of the cluster centroid, can be different to the polar angle of parent photon
[`clusterReg` = 1 and `E` > 0.1] or [`clusterReg` = 2 and `E` > 0.09] or [`clusterReg` = 3 and `E` > 0.16] (GeV)	Minimum energy threshold, according to region of the ECL subdetector (forward endcap, barrel, backward endcap respectively) the cluster is created in. Differing regions have different energy tolerances.

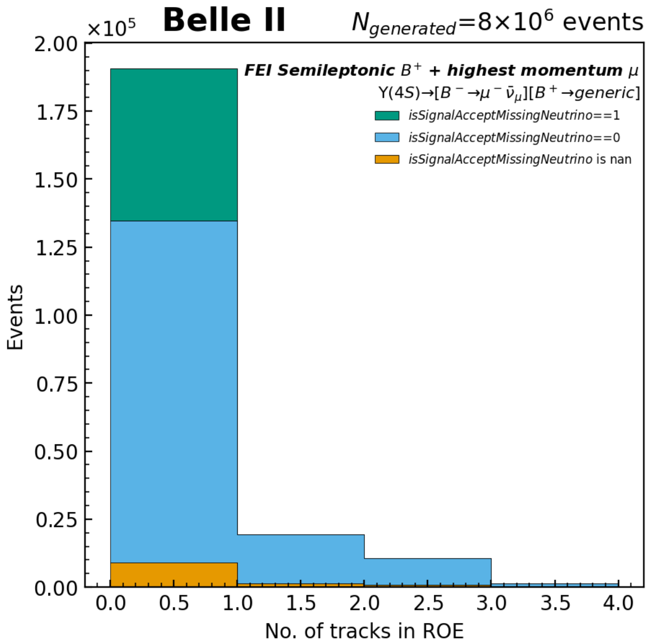


Figure 4.11. Number of charged tracks that remain in the rest-of-event after identifying the semileptonic FEI B_{tag} and μ_{sig} in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$ MC events.

as events where the muon has undergone final-state radiation. Figure 4.12 contains the distribution of ECL energy in the signal MC. The `isSignalAcceptMissingNeutrino` =1 truth-matched component peaks at 0 GeV, while the `isSignalAcceptMissingNeutrino` =0 one peaks in the first bin after the zeroth. These trends are as expected, that misreconstruction can be judged by excess unused clusters in the ECL. Excess clusters will also be a feature of background MC. Notably, all MC events are $E_{\text{ECL}} < 2.5$ GeV. As will be seen, enforcing such a restriction as a pre-selection criteria will be a powerful method of background reduction.

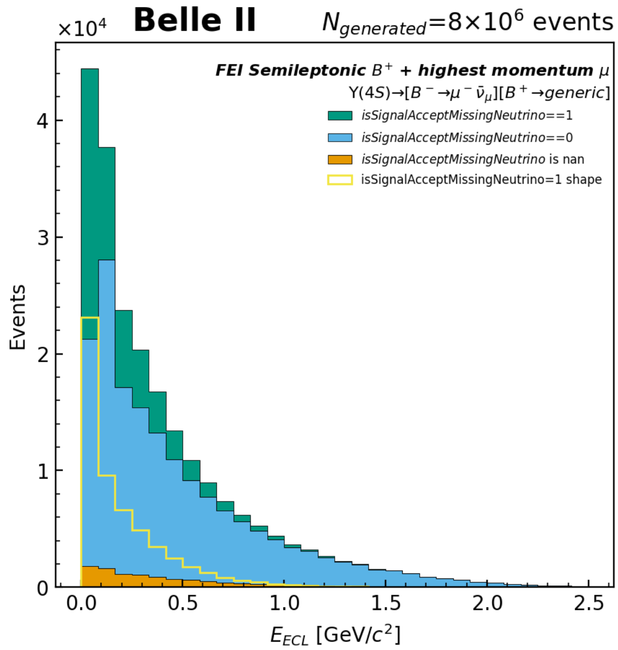


Figure 4.12. Energy in ECL subdetector due to clusters that remain in the rest-of-event after identifying the semileptonic FEI B_{tag} and μ_{sig} in 8 million $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu]B^-_{\text{SL}}$ MC events.

4.1.4. B_{tag} and signal muon in 362 fb^{-1} of signal and background MC

In the previous section, the features of ROE variables were presented for $e^+e^- \to \Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu]B^-$ MC, with $p^*_\ell < p^*_\mu$ and the best candidate selection of muons and the tagged B mesons. This examination will now be extended to include background MC samples, in a method similar to what as done in the stacked B_{tag} histograms of signal and background MC in Section 3.5.2. The MC of signal and background samples will again be scaled to what is expected in the 362 fb^{-1} data sample of the Belle II collision dataset that was described in Section 2.4.4.

While again shapes of signal and background samples are compared for future signal-enhancing selections, it will also be a comparison of the effect that muon selection has had on the overall yield and tagging performance. The following distributions will finally be used to decide on pre-selection criteria to be applied prior to the combination of the B_{tag} and signal muon into $\Upsilon(4S)$ candidates. The goal of such pre-selections are to exclude events least likely to contain signal events, whilst preserving the signal events themselves. After $\Upsilon(4S)$ reconstruction in Section 4.1.5, more substantial background

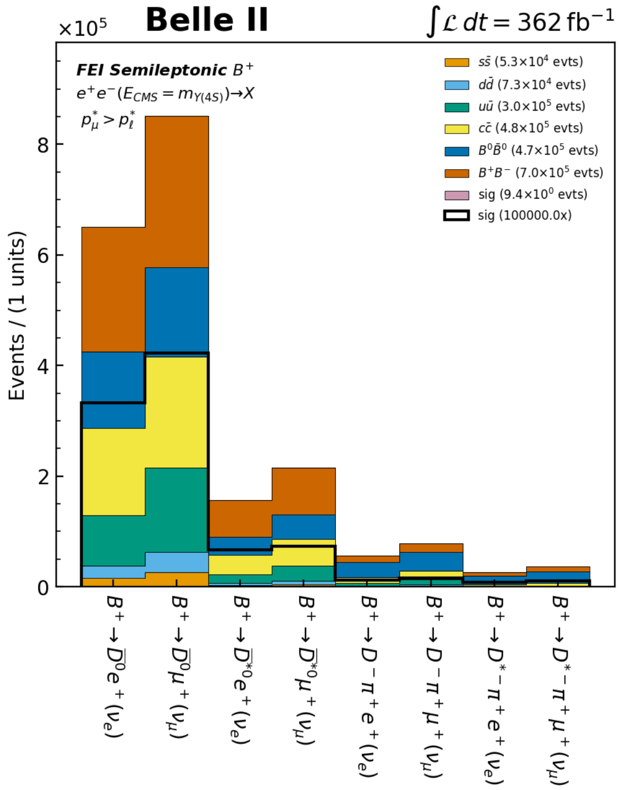


Figure 4.13. Reconstructed B_{tag} decays modes identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified signal muon. Stacking is according to event type, each scaled to 362 fb^{-1}.

rejection techniques will be implemented, informed by the study of the distributions to be shown.

The first distribution of signal and background MC events that contain a suitable semileptonic FEI B_{tag} and a signal muon that $p^*_\mu > p^*_\ell$ is that of the B_{tag} decay mode in Figure 4.13. The restriction to events with a compatible signal muon has dramatically reduced the yield of background MC by a factor of 100, in comparison with B_{tag} only version in Figure 3.30. In doing so, the signal component in the stack is not yet visible, with 9.4 events across all 8 decay modes. The overlaid signal component for all plots in this section is now 10^6 times larger than what is expected in the 362 fb^{-1} collision dataset. At this stage in the analysis, the distribution of the decay modes are beginning to show preference between the charged and mixed $B\overline{B}$ components, with more B^+B^- than $B^0\overline{B}^0$ events in the 3 body decays (i.e. the first 4 decay modes) and more $B^0\overline{B}^0$ than B^+B^- in the 4 body ones. The basis of this could be in real neutral three-body semileptonic decays, which proceeds as $B^0 \to D^{(*)-}\ell^+\,\nu_\ell$. Combining the products of this mode with an additional pion from the $B^0\overline{B}^0$ event could therefore mimic these charged B decay modes. Selecting a subset of the FEI B decay modes may be a selection made in order to avoid B^+/B^0-based misreconstruction. Other than the mixed B component and overall yields, there does not seem to be large impact of the signal muon on the creation of the FEI B_{tag}.

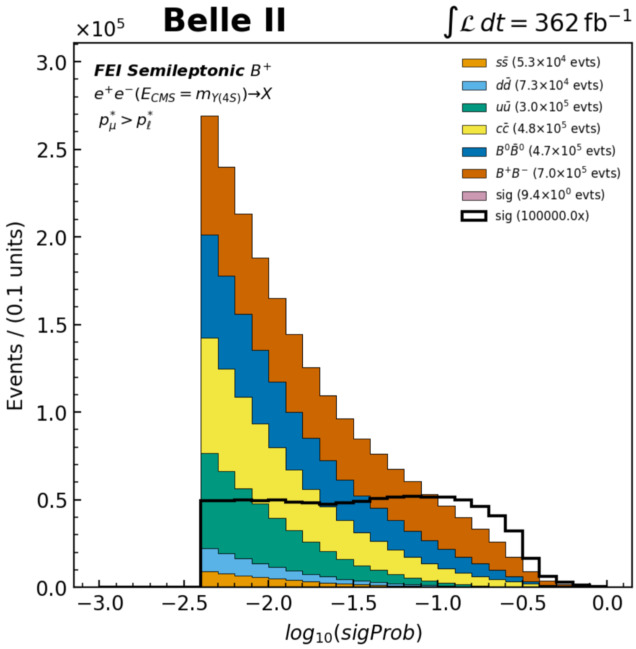


Figure 4.14. $\log_{10}(\texttt{sigProb})$ of B_{tag} identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified signal muon. Stacking is according to event type, each scaled to 362 fb^{-1}.

Figure 4.14 illustrates the log-scaled signal probability of B_{tag} identified in signal and each type of background process expected in 362 fb^{-1} of e^+e^- collisions. With the reduction in events from the signal muon selection, there is very subtle shape change in the stacked components. What was a sharp decrease in the number of background events as $\log_{10}(\texttt{sigProb})$ increased in the B_{tag} only version of Figure 3.29, is now more gradual in the charged MC component between -1.5 and -0.5. This suggests muon selection has preferentially removed B_{tag} with lower rather than larger $\log_{10}(\texttt{sigProb})$, corresponding to correctly reconstructed B_{tag}. Prior to $\Upsilon(4S)$ reconstruction and further $B^+ \to \mu^+\nu_\mu$ specific selections, it can still be expected that charged events are reconstructed correctly. At this stage, the only conditions have been a B_{tag} and a high-momentum muon in the event that is not the tag-side lepton. Thus the enhancement of the area of $\log_{10}(\texttt{sigProb})$ between -1.5 and -0.5 can consist of correctly tagged charged B mesons, as was seen in the corresponding histogram of semileptonic tags in charged MC alone in Figure 3.12.

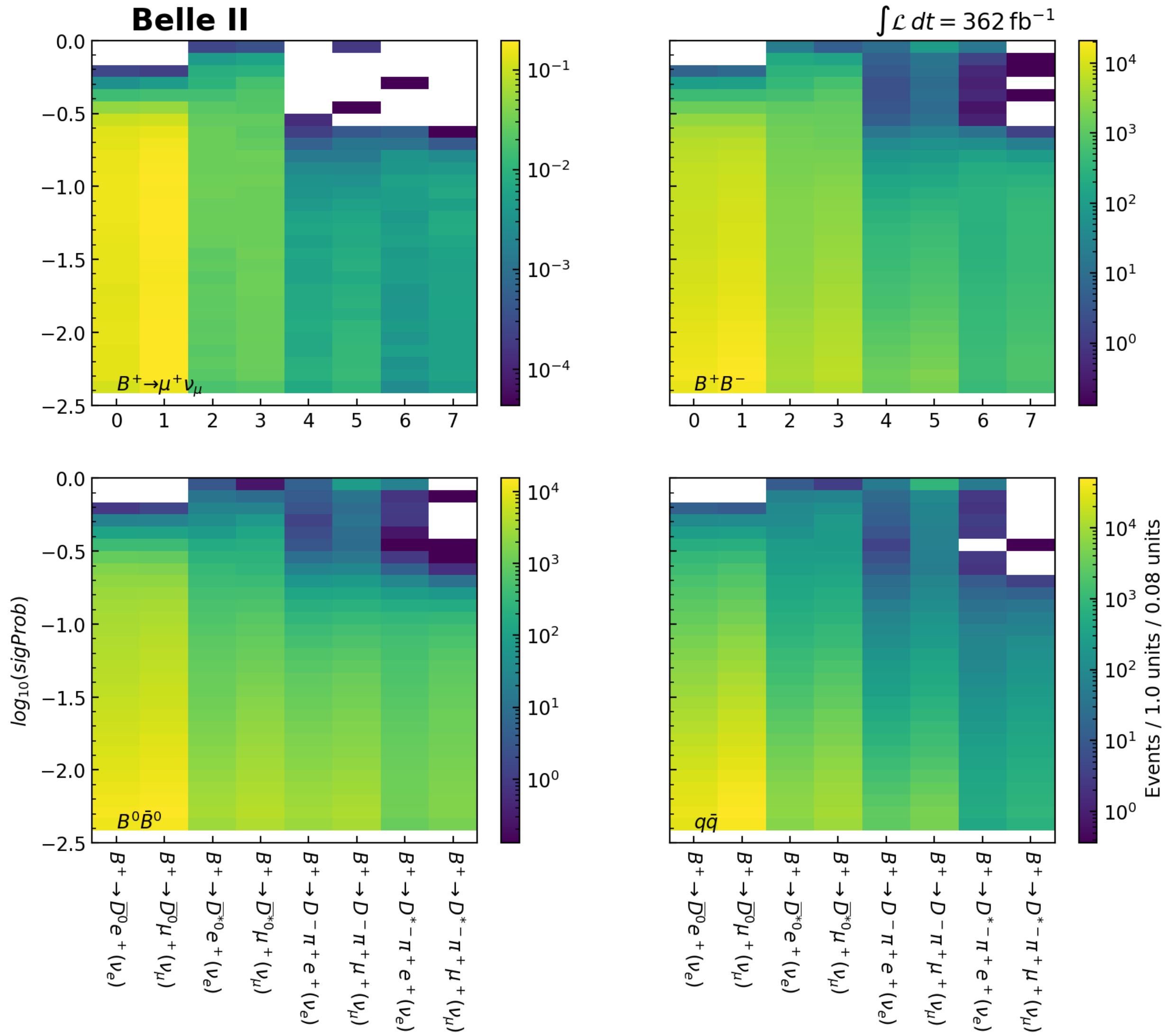


Figure 4.15. 2D histogram of $\log_{10}$(`sigProb`) according to decay modes identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

The tagmaps of Figure 4.15 provide little additional understanding of B_{tag} after signal muon selection, when contrasted with the B_{tag} only version of Figure 3.31. There are lesser events in the region of $\log_{10}$(`sigProb`) > -0.5 for the 4-body semileptonic decays, while there also continues to be a proportion of background events in the $\log_{10}$(`sigProb`) > -0.1 (i.e. misreconstructed events rated highly by the FEI).

Displayed in Figure 4.16 are the $\cos\theta_{BY}$ values of B_{tag} determined in signal and background MC after the signal muon selection is made. Compared with Figure 3.32

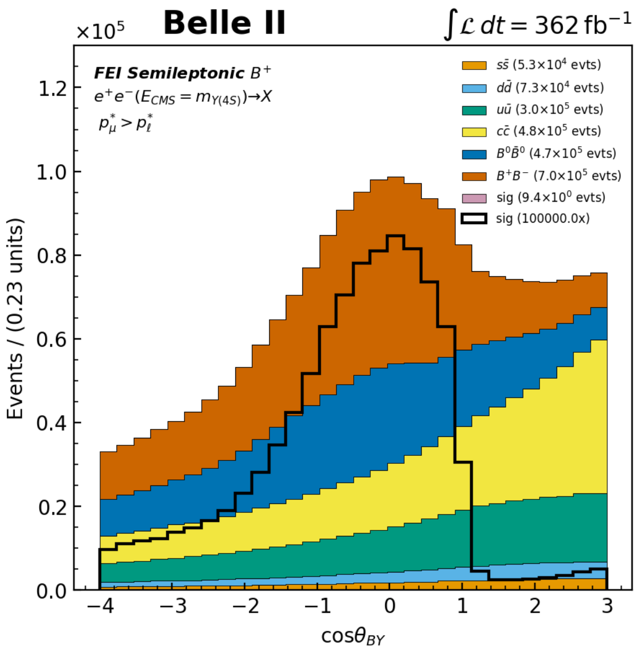
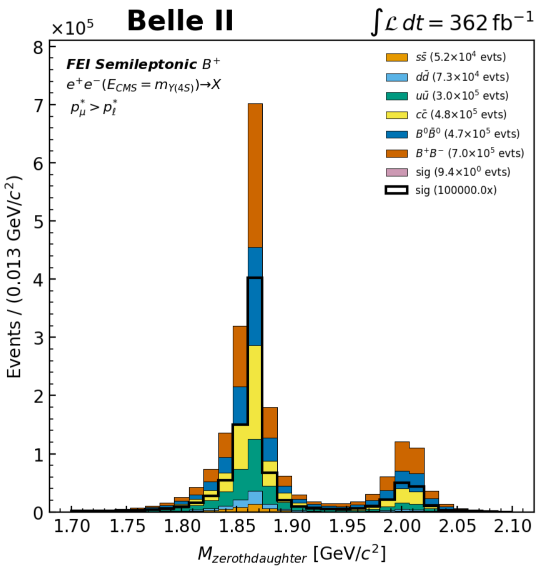


Figure 4.16. $\cos\theta_{BY}$ of B_{tag} identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

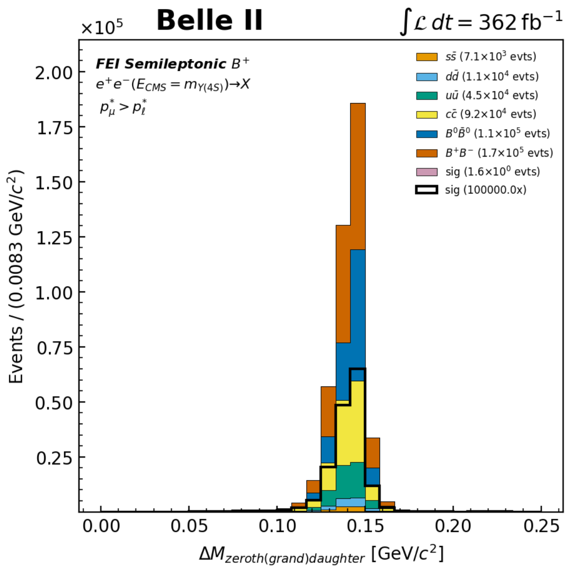
before muon selection, there has been a reduction in number of events in the light quark components between $\cos\theta_{BY}$ =1 and $\cos\theta_{BY}$ =2. As a result, the $c\bar{c}$ component stacked on top is more obviously increasing towards the boundary of $\cos\theta_{BY}$ =3. Moreover, the $B\overline{B}$ components are more noticeably peaking between -1 and 1. Both the charged and mixed components are also more symmetrical in their tails outside of the physical [-1, 1] $\cos\theta_{BY}$ region. This also leads to continued difference between the signal and charged distributions, with the $p^*_\ell < p^*_\mu$ having a similar effect to the $p^*_\ell < 2.25$ GeV/c selection, having lesser events in the positive $\cos\theta_{BY}$ tail. Thus, implementing a future upper bound on $\cos\theta_{BY}$ continues to be a viable option in rejecting background events.

Figure 4.17 contains distributions of the zeroth daughter mass of the B_{tag} and the D^*/D mass splitting for the excited D B_{tag} decays, both after muon selection has been made. There are no major shape changes when compared to distributions prior to μ_{sig} selection in Figure 3.33a and Figure 3.33b.

The requirement on the tag-side lepton momentum relative to the signal muon (in the $\Upsilon(4S)$ frame) has led to a distinct change its distribution in Figure 4.18a. Whereas the $p^*_{\mu,\text{tag}} < 2.25$ GeV/c version in Figure 4.18a has background components with broad maxima between 1 and 2 GeV/c, ensuring that the tag-side lepton is bounded



(a) $B_{\rm tag}$ zeroth daughter (i.e. D^*/D) mass.

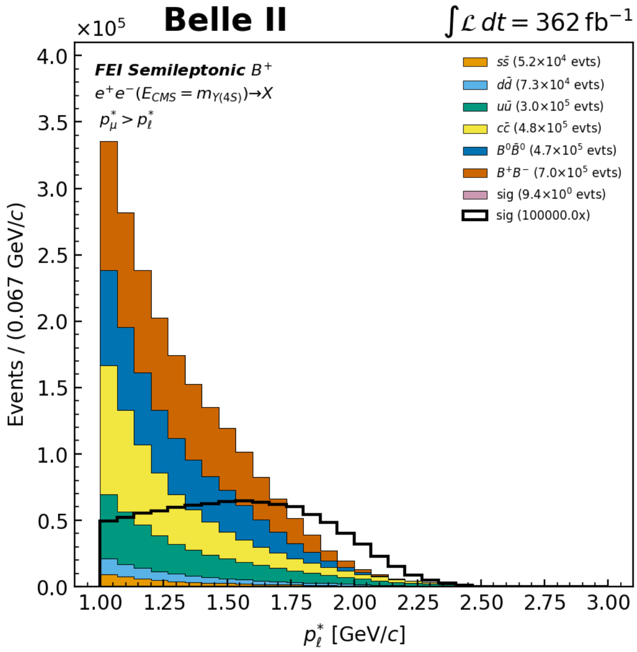


(b) ΔM=$m_{D^*} - m_D$ for $B^+ \to D^* \to D$ FEI modes.

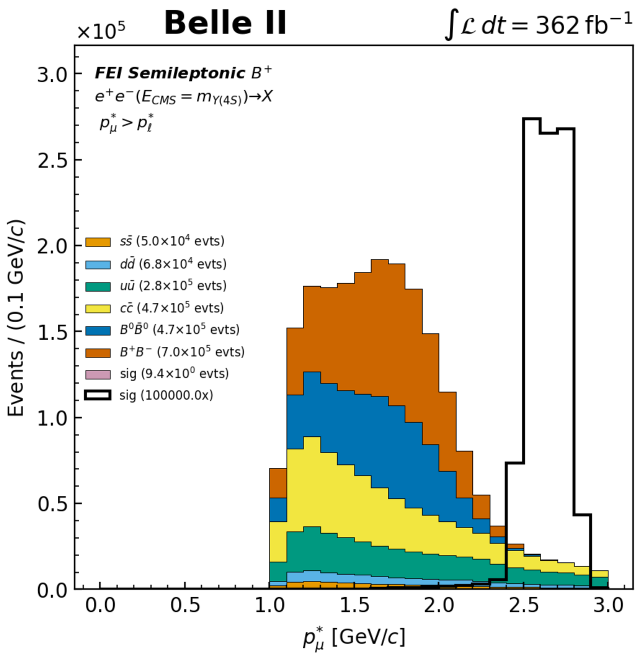
Figure 4.17. Daughter and (grand-)daughter mass splitting in $B_{\rm tag}$ identified via semileptonic FEI applied to MC simulated e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified $\mu_{\rm sig}$. Stacking is according to event type, each scaled to 362 fb^{-1}.

above by a signal muon momementum has led to a steep decline in the momentum of tag-side lepton candidates in $q\bar{q}$ events. This has also led to shape discrimination between the $B\overline{B}$ background and the signal components, where the excess of $B\overline{B}$ tag-side leptons occur strongly towards p^*_ℓ =1 GeV/c. This contrasts against the overlaid signal component, which remains broadly distributed towards 1.75 GeV/c. Placing a lower bound requirement may be useful in rejecting the peak misreconstructed $B_{\rm tag}$ background events, when eventually calculating the number of $B^+ \to \mu^+\nu_\mu$ events.

Figure 4.18b contains the first look at the momentum of signal muon candidates in the $\Upsilon(4S)$ frame, in both signal and background events that also admit $B_{\rm tag}$. The requirement of $p^*_\ell < p^*_\mu$ has resulted in no signal muons of any component below 1 GeV/c, as per the p^*_ℓ post-selection criteria in the semileptonic FEI listed in Table 3.3. This was also seen in the p^*_μ in $B^+ \to \mu^+\nu_\mu$ MC only histogram in Figure 4.8c. The distribution contains significant shape distinction of signal muon (i.e. fastest muon of the event that isn't used in the $B_{\rm tag}$) chosen in the $q\bar{q}$ background, $B\overline{B}$ background and $B^+ \to \mu^+\nu_\mu$ signal component. The $q\bar{q}$ momentum component skews towards 1 GeV/c, with considerable tail towards 3 GeV/c. The muons from non-$B\overline{B}$ are relatively symmetric momentum distributions, with most events between 1.5 and 2 GeV/c. The overlaid signal component shows very



(a) Tag-side lepton momentum.



(b) Signal-side muon momentum.

Figure 4.18. Lepton momenta in $\Upsilon(4S)$ frame, identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance and restricted to those containing μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

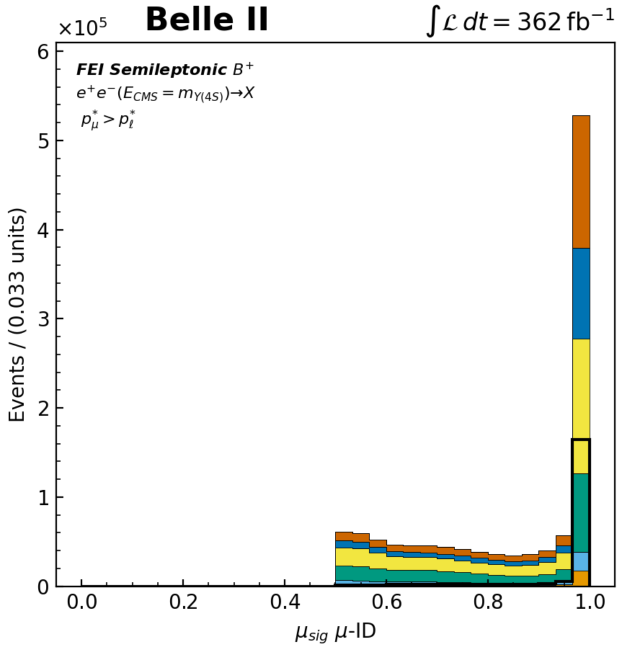


Figure 4.19. Muon identification probability (`muonID`) of signal muon in $B^+ \to \mu^+\nu_\mu$. Occurs alongside B_{tag} identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

few muon from $B^+ \to \mu^+\nu_\mu$ events where the maxima of background components are, even given the 10^5 times larger signal shape. Within the $p^*_\mu \in$[2.4,2.9) GeV/c interval where the signal component shape is, the biggest background contribution is from muons in $c\overline{c}$ and $u\overline{u}$ events. As a distinctive feature of the $B^+ \to \mu^+\nu_\mu$ decay, the p^*_μ distribution has provided one of the best visually discriminating variables between $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+\nu_\mu$ mimicking events in the distributions consulted so far, with majority background below $p^*_\mu < 2.4$ GeV/c and signal beyond this. Thus, p^*_μ could provide a suitable region in which to evaluate the yield of $B^+ \to \mu^+\nu_\mu$ events, as was done in the 2018 Belle search described in Section 1.4.3. This is provided appropriate background rejection of events in the [2.4,2.9) GeV/c region could take place, as the signal component scaled to 362 fb^{-1} remains visibly overwhelmed by other MC components in the stack. It again emphasises that the challenge to such an analysis is not rejecting the secondary, the misreconstructed or radiative muons in the signal component in this rare decay, as much as the challenge in rejecting muons from the much more probable continuum and other $B\overline{B}$ events.

In Figure 4.19, the range of `muonID` for signal muon candidates is also shown in both signal and background MC for the first time in this analysis. In principle, `muonID` is better

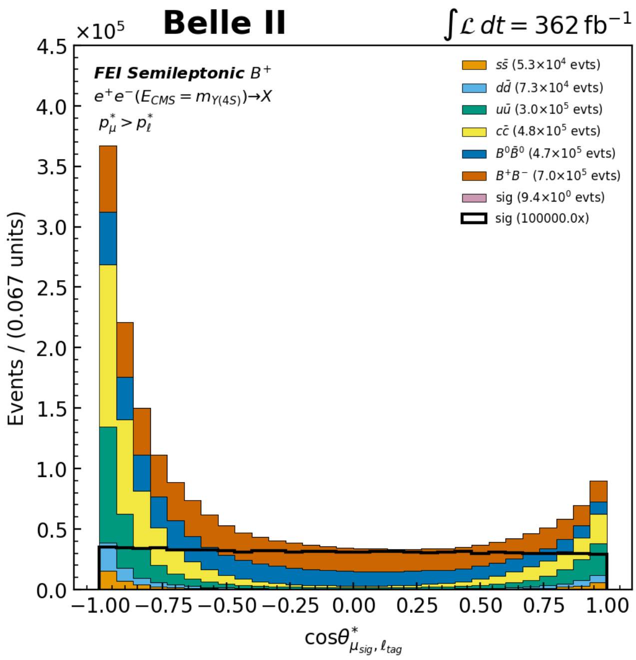


Figure 4.20. Opening angle between signal muon and tag-side lepton, in events that contain signal mode $B^+ \to \mu^+\nu_\mu$ and a B_{tag} identified via semileptonic FEI applied to e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

suited to discrimination of muons from other FSPs, than for signal from background event discrimination. Making a selection on `muonID` will increase the probability that the chosen charged particle is in fact a muon, but not that it is the high momentum muon from $B^+ \to \mu^+\nu_\mu$. While the signal sample is the only MC containing real muon from $B^+ \to \mu^+\nu_\mu$, all background components can contain real muon from non-$B^+ \to \mu^+\nu_\mu$ events. This is reflected in the histogram of `muonID`, where the muons chosen in signal and background MC peak at `muonID` ≥ 0.9. There is minor shape difference in the background components which slowly increase the number of events towards `muonID`=0.5, whereas it was seen in Figure 4.9b that the true muon in $B^+ \to \mu^+\nu_\mu$ slowly decreases. Given the rarity of the $B^+ \to \mu^+\nu_\mu$ event however and the shape of the `muonID` profile, it is most likely that the signal muon will have `muonID` ≥ 0.9, and thus could provide a suitable background rejecting selection.

Figure 4.20 also demonstrates the first look at the opening angle between the tag-side and signal side leptons in $\Upsilon(4S)$ frame, in both signal and background events. Here, there is a clear shape difference between the background events and the overlaid signal shape. As was discussed in Figure 4.10, the cosine of the opening angle was uniformly distributed between [-1, 1) in Figure 4.10a, which in turn demonstrated in Figure 4.10b that this corresponded to a slight preference near-orthogonal opening angles between the tag-side lepton and the signal muon. This contrasts with the continuum components, peaking towards -1 and 1 with very few events within these bounds. Moreover, this corresponds to a preference for angles between the leptons to be closer to parallel or anti-parallel. With more continuum events in the cosine opening angle peak corresponding to $\cos\theta^*_{\mu_{\text{sig}}\ell_{\text{tag}}}$ = -1, there is thus a preference for the anti-parallel case, where the signal muon and the tag-side muon are picked with opposing angles. This angular behaviour is consistent with the large momentum $e^+e^- \to q\bar{q}$ events, where particles are produced with non-overlapping jets of particles. This event shape was discussed as a property of continuum events in Section 2.2.3. The majority of the middle interval of the cosine opening angle consists of the $B\overline{B}$ components. B mesons are created with much smaller momentum in the $\Upsilon(4S)$ frame, resulting in a larger range of angles allowed in their more spherical, overlapping B decays. As both $B^0\overline{B}^0$ and B^+B^- are not generated to contain the $B^+ \to \mu^+\nu_\mu$ event, they are inherently misreconstructed and thus gently peak in their events towards anti-parallel opening angle. This is also consistent with the signal only histogram, where the misreconstructed tagged B mesons peak towards cosine opening angle equally -1 in Figure 4.10a.

Figure 4.21 contains the distribution of the tracks that remain in the ROE of the B_{tag}, with the contributions of the signal muon removed. From the overlaid signal component, it is seen that the vast majority of signal events have no charged tracks after reconstructing the B_{tag} and identifying the signal muon, as was also seen in Figure 4.11. Conversely, very few background events have no remaining tracks, indicating that placing requirements on the number of unreconstructed tracks would be an efficient method of reducing such events. Other analyses may undertake techniques to reassign unused to tracks to improve the B_{tag}, similar to what was done with the 'semi-exclusive' B_{tag} used in the 2008 BaBar $B^+ \to \mu^+\nu_\mu$ search described in Section 1.4.2. However, such a technique is not used, as it is assumed that the FEI has produced the best B_{tag} possible. This analysis therefore restricts the search to events with zero additional tracks. As all background components peak at two tracks in the ROE, this will provide a powerful constraint on the number background events that proceed to the next stage of the analysis. Notably, as the e^+e^- process begins with no net-charge, there should be no net charge in the final event. Thus not only does an odd number of tracks indicate misreconstruction due to unused particles in the event, but that there are potentially tracks in the event that have been missed in detection. This further motivates that only considering events with no tracks in the

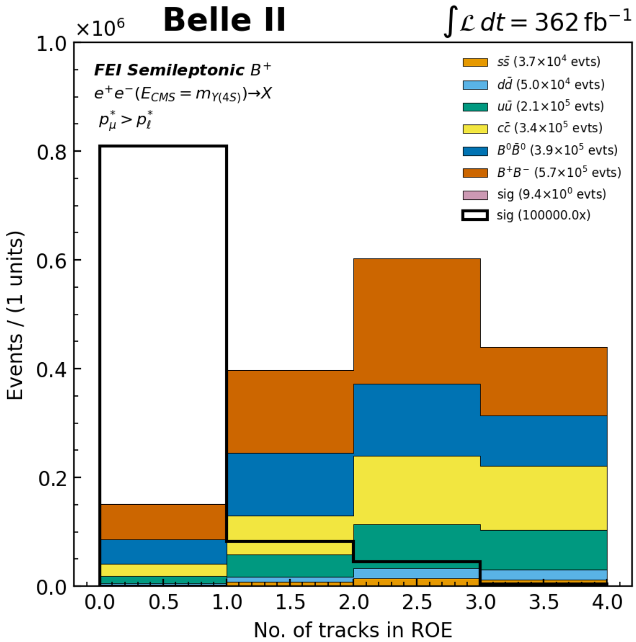


Figure 4.21. Number of charged tracks that remain in the Rest-of-Event, after identifying the semileptonic FEI B_{tag} and μ_{sig} in MC e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

ROE will reduce the number of background events when eventually estimating the yield signal events. This selection criterion will be adopted in the reconstruction procedure.

The final figure to consult before performing $\Upsilon(4S)$ reconstruction is Figure 4.22, the energy leftover in the ECL after the B_{tag} and signal muon selection. This variable also shows substantial shape discrimination between signal and background event components. The overlaid signal component peaks sharply towards zero energy in the electromagnetic calorimeter, verifying that the majority of signal events reconstructed are done so accurately, without omitting photons from the event. The same shape was seen in the $B^+ \to \mu^+\nu_\mu$ MC only version in Figure 4.12. Contrasting with this is the background contribution, where it is expected that there will be excess extra tracks and energy clusters omitted from the reconstruction of $B^+ \to \mu^+\nu_\mu$ in MC that do not contain $B^+ \to \mu^+\nu_\mu$. Specifically, the continuum components slowly increase passed 2 GeV, while the $B\overline{B}$ components peak broadly between 1 and 2 GeV of energy in unused ECL clusters. The ability of this variable to discriminate strongly between the signal and background events was presented in the discussion of the 2008 hadronic-tagged BaBar $B^+ \to \mu^+\nu_\mu$ search in Section 1.4.2. The distributions in the current figure broadly match the trends in the BaBar distribution in Figure 1.12b, where E_{ECL} was used as a

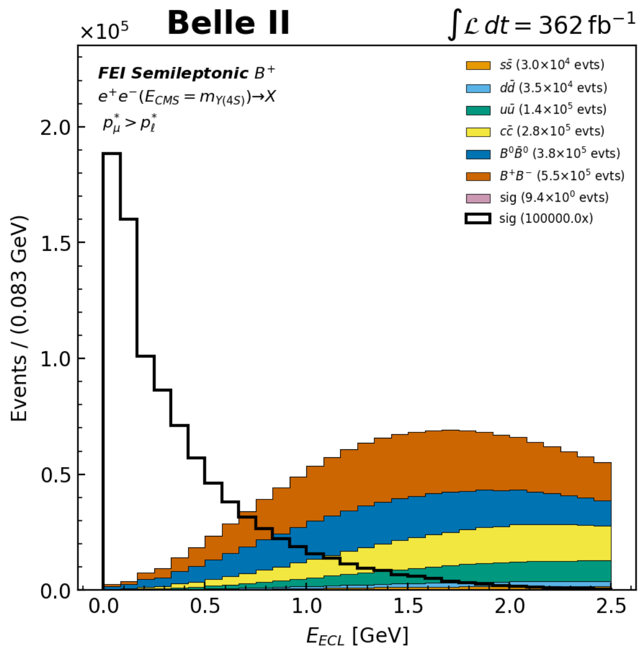


Figure 4.22. Energy in ECL subdetector due to clusters that remain int he Rest-of-Event, after identifying the semileptonic FEI B_{tag} and μ_{sig} in MC e^+e^- collisions at the $\Upsilon(4S)$ resonance. Events restricted to those with identified μ_{sig}. Stacking is according to event type, each scaled to 362 fb^{-1}.

signal variable. The choice of E_{ECL} or p^*_μ as the variable that best distinguishes between signal and background will be determined in Section 4.2.2.

This distribution has been limited to $E_{\text{ECL}} < 2.5$ GeV/c for plotting purposes. From the trends of the MC, there would be a large proportion of background events with $E_{\text{ECL}} > 2.5$ GeV/c, with a negligible number of signal events that leave this magnitude of energy in the electromagnetic calorimeter. For this reason, a pre-selection of $E_{\text{ECL}} < 2.5$ GeV/c will be adopted to both limit the size of the reconstructed dataset and remove events least likely to contain true $B^+ \to \mu^+\nu_\mu$ events.

The MC distributions of signal and background events at the $\Upsilon(4S)$ resonance have been presented, restricted to those that have both a signal muon and a semileptonic FEI B_{tag}. Table 4.3 summarises the yields of each MC type expected within 362 fb^{-1} of data, comparing between B-tagging alone and after the signal muon has been identified. The requirement of a muon that has the largest p^*_μ to be the signal muon has resulted in $\mathcal{O}(1\%)$ retention out of the events that mimicked B_{tag} in background MC, while retaining 82% of signal events. It must be noted that despite 0.12% of background events rejected with FEI B-tagging and signal muon selection, the 9.43 $B^+ \to \mu^+\nu_\mu$ events remain

Table 4.3. Yields and Efficiencies of Belle II MC of events with FEI Semileptonic B_{tag} and μ_{sig}.

Sample ($e^+e^- \to X$)	No. of events (after FEI, 362 fb^{-1})	No. of events (after μ_{sig}, 362 fb^{-1})	Eff. (selection) [%]	Eff. (total) [%]
$[B^+ \to \mu^+\nu_\mu\ ]B^-$	11.47	9.43	82.25	5.43
B^+B^-	2.92×10^7	6.96×10^5	2.39	0.34
$B^0\overline{B}^0$	2.30×10^7	4.73×10^5	2.05	0.24
$c\overline{c}$	3.14×10^7	4.78×10^5	1.52	0.10
$u\overline{u}$	2.03×10^7	3.02×10^5	1.49	0.05
$d\overline{d}$	4.79×10^6	7.28×10^4	1.52	0.05
$s\overline{s}$	3.97×10^6	5.25×10^4	1.32	0.04
tot. background	1.13×10^8	2.07×10^6	1.84	0.12

overwhelmed by the 2.07 $\times 10^6$ $q\overline{q}$ and generic $B\overline{B}$ events. The analysis will thus continue to combining the best candidate B_{tag} and signal muon into suitable $\Upsilon(4S)$ candidates, where further event rejection will take place according to those inconsistent with the $\Upsilon(4S)$ invariant mass. From the variables consulted, two pre-selections are decided on in $\Upsilon(4S)$ reconstruction: zero tracks in ROE and $E_{\text{ECL}} < 2.5$ GeV/c. This is consistent with the scope of the analysis, in that the tagged-B mesons with missing tracks are not considered, nor signal muons that emit photons via final-state radiation. Moreover, these selections fulfil the goal of a pre-selection, which in this case is to reject background, while retaining as many $B^+ \to \mu^+\nu_\mu$ signal events as possible. More stringent background rejection is explored soon in Section 4.2.3.

4.1.5. Reconstruction of $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\text{SL}}$

In the previous sections, a procedure for obtaining the best semileptonically decaying charged B mesons was determined. Furthermore, a similar procedure was implemented to find a muon most likely to originate from $B^+ \to \mu^+\nu_\mu$ decays. The sets of best candidate B mesons and signal muons can finally be combined to further constrain the analysed events to the conditions of their creation, namely as $e^+e^- \to \Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\text{SL}}$ (and charge conjugate). The reconstruction of the $\Upsilon(4S)$ event is essentially combining the B_{tag} four vector with that of the signal muon one and ensuring that the resulting invariant mass is consistent with the 10.58 GeV/c^2 $\Upsilon(4S)$ one.

While this could be done by writing code to calculate the combined B_{tag}-μ_{sig} invariant mass of each event, there is instead the `modularAnalysis.reconstructDecay` function provided by `basf2`. This function ensures that all returned combinations are physically allowable, e.g. that charge is not violated, that each particle is only used once throughout a candidate's decay chain, that the combined mass does not significantly exceed the nominal mass etc. Reconstruction also takes place for charge conjugated modes. In this study, reconstruction is performed in two stages. The first stage is to create the B^+_{sig}, using the following decay string.

$$\texttt{B-:sig -> mu-:sig} \tag{4.1}$$

The $\Upsilon(4S)$ reconstruction then takes place using the following decay string.

$$\texttt{Upsilon(4S):sig -> B-:feiSL B+:sig} \tag{4.2}$$

Notably, the neutrino contribution is not added, as these are not reconstructible in the Belle II detector. There are no physics implications in the ordering of the B mesons in reconstruction, with ordering purely for labelling purposes. Thus, any further references to the "zeroth daughter" of the $\Upsilon(4S)$ will be referring to the B_{tag}, whereas the "first" will be the B_{sig}.

Significant effort was dedicated to determining the best candidate selection of B_{tag} and signal muon in Section 3.4.2 and Section 4.1.1 respectively. $\Upsilon(4S)$ reconstruction is performed using best candidate muon per event, but before best candidate selection in B_{tag} candidates. Reconstruction of the $\Upsilon(4S)$ will thus produce multiple $\Upsilon(4S)$ candidates per event. This is done to avoid cases where the highest signal probability B_{tag} and high momentum signal muon were incompatible, therefore losing such an event because the other real candidate B_{tag} has been removed. The best candidate $\Upsilon(4S)$ selection will be the one who survives reconstruction, has the B_{tag} daughter with the highest signal probability and has a tag-side lepton such that $p^*_\ell < p^*_\mu$. After this, the Rest-of-Event is built on the best $\Upsilon(4S)$ candidate per event, such that the pre-selections of zero tracks in ROE and $E_{\text{ECL}} < 2.5$ GeV/c can be applied.

Table 4.4 contains the yields of each component after $\Upsilon(4S)$ reconstruction and after the conservative ROE pre-selections. 4.23 $B^+ \to \mu^+\nu_\mu$ events remain in 1.14 $\times 10^5$ background events. Near 50% of signal events did not survive reconstruction, while around 95% of background was rejected. The 4.23 weighted events expected in 362 fb^{-1}

Table 4.4. Yields and Efficiencies of Belle II MC of events after $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction and selections zero tracks in ROE and $E_{\mathrm{ECL}} < 2.5$ GeV/c.

Sample ($e^+e^- \to X$)	No. of events (after μ_{sig}, 362 fb^{-1})	No. of events (after reco., 362 fb^{-1})	Eff. (selection) [%]	Eff. (total) [%]
$[B^+ \to \mu^+\nu_\mu\,]B^-$	9.43	4.23	44.8	2.44
$B^+\,B^-$	$6.96{\times}10^5$	$5.79{\times}10^4$	8.31	$2.83{\times}10^{-2}$
$B^0\overline{B}^0$	$4.73{\times}10^5$	$3.88{\times}10^4$	8.21	$2.01{\times}10^{-2}$
$c\overline{c}$	$4.78{\times}10^5$	$1.03{\times}10^4$	2.16	$2.14{\times}10^{-3}$
$u\overline{u}$	$3.02{\times}10^5$	$4.32{\times}10^3$	1.43	$7.44{\times}10^{-4}$
$d\overline{d}$	$7.28{\times}10^4$	$1.21{\times}10^3$	1.66	$8.31{\times}10^{-4}$
$s\overline{s}$	$5.25{\times}10^4$	$1.11{\times}10^3$	2.12	$8.04{\times}10^{-4}$
tot. background	$2.07{\times}10^6$	$1.14{\times}10^5$	5.48	$6.52{\times}10^{-3}$

is represented by $1.9{\times}10^5$ generated MC events and so there is sufficient statistics to further investigate trends in $B^+ \to \mu^+\nu_\mu$ events.

The key variables to this analysis will be revisited, to examine how their distributions have changed after $\Upsilon(4S)$ reconstruction. The variables will be the standard B_{tag} and signal muon variables we've seen previously, as well as introducing some new variables that were used in previous $B^+ \to \mu^+\nu_\mu$ searches and perform well in demarcation between $B\overline{B}$ and $q\overline{q}$ events. Once this is completed, the strategy of determining the optimal selections in creating a signal-enhanced region will be discussed. To inform which variables will take part in the strategy, the key variables in the following distributions are qualitatively investigated. Focus will be on differences from histograms at previous stages of the analysis, rather than trends that have already been discussed.

In the following distributions, statistical uncertainty will be considered for the first time. In the remaining Chapter and for the rest of this thesis, statistical uncertainties will be considered using Poisson $\sqrt{N}$ error. Statistical uncertainties on histograms will be treated as $\sqrt{N}$ per bin. For the case of the MC generated events where a scaling factor has been applied to model yields at 362 fb^{-1}, N will be treated as the yield after scaling. This is a more conservative approach to statistical uncertainty estimation. It is generally more appropriate for scaled events to assign a statistical uncertainty based on the number of raw events used, specifically as $N/\sqrt{N_{raw}}$ for N_{raw} that has been scaled to N events. This thesis will continue, however, to use the more conservative $\sqrt{N}$, where the evaluation of the more stringent statistical uncertainty is left for future work.

It is noted that the following histograms will not contain collision data. Collision data will eventually be overlaid with the simulated data once selection criteria is fully determined, which will be in Section 4.3. For reasons explained previously, data and MC is not expected to have perfect agreement at this stage. Once selection criteria is implemented, constraining both datasets to represent a smaller set of better-modelled decays, it is then that data-MC agreement could be achieved. At this stage in the analysis, the goal is to observe how key variable distributions have changed with $\Upsilon(4S)$ reconstruction and not to worry about data-MC agreement. Thus, to not distract from this goal, data is omitted until selections are determined.

Key variables after $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$ reconstruction and ROE pre-cut selection criteria, in signal and background MC scaled to 362 fb^{-1}

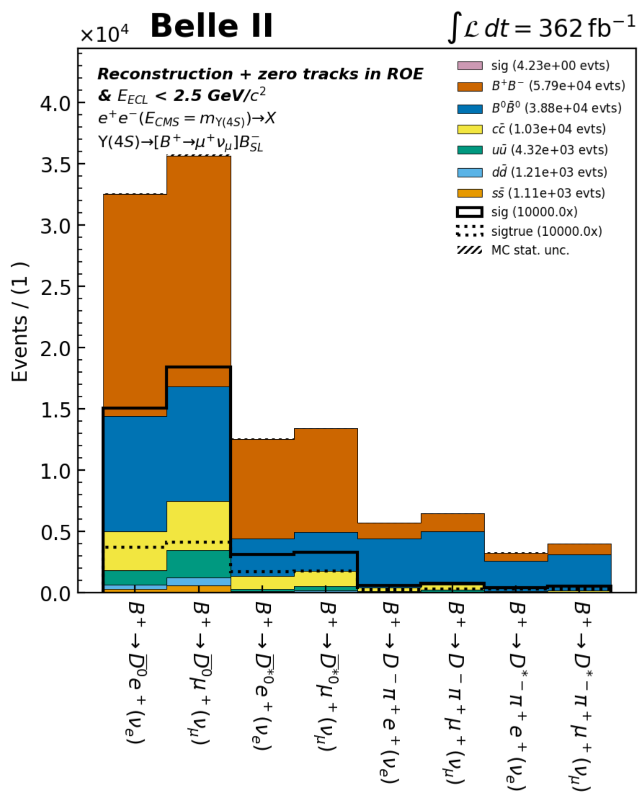


Figure 4.23. Semileptonic FEI B_{tag} decay modes used in $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

Figure 4.23 shows the distribution of B_{tag} decay modes, for the underlying daughter B_{tag} that go on to reconstruct the $\Upsilon(4S)$. The overlaid signal is now scaled 10^4 times, whereas it was 10^5 times in the plots prior to reconstruction and ROE selection. An additional truth-matched signal component has now also been overlaid, and will be used in all following histograms. Such a component is scaled the same as the overlaid regular signal histogram, and allows comparison between the general signal MC and events that are reconstructed correctly. The trend continues where the $B^0\overline{B}^0$ MC component dominates the 4-body semileptonic decay modes. Compared to the distribution prior to reconstruction in Figure 4.13, it is seen especially in $B^+ \to \overline{D}^0\ \ell\nu_\ell$ that the $\Upsilon(4S)$ reconstruction plus the ROE pre-selection criterion has resulted in a substantial reduction in the number of $q\overline{q}$ events. Notably, the stacked signal component is not yet visible, despite the number of background events rejected with the $\Upsilon(4S)$ hypothesis. Additionally, the representation of the statistical uncertainty per bin in the histogram is also not yet visible at the current number of events.

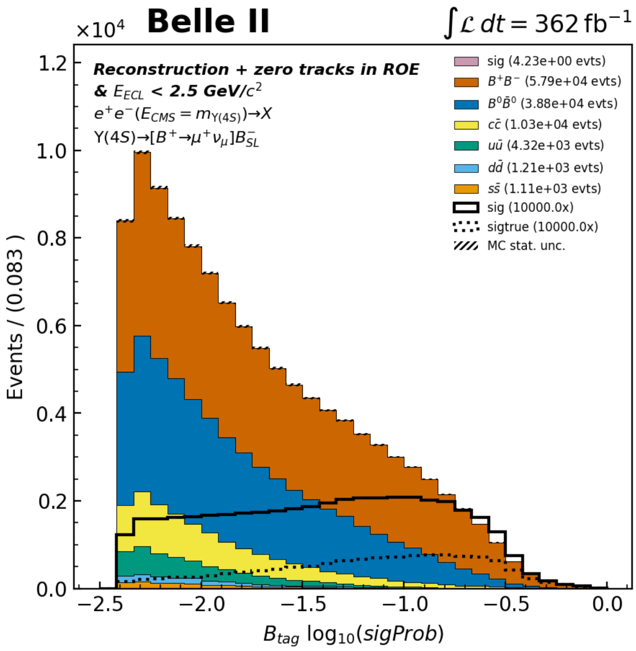


Figure 4.24. $\log_{10}$(`sigProb`) for B_{tag} used in $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\text{SL}}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

Figure 4.24 demonstrates the $\log_{10}$(`sigProb`) of the B_{tag} daughters of $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\text{SL}}$. With the reconstruction and the ROE pre-selection criteria, the contribution from $q\overline{q}$ events in the peak towards $\log_{10}$(`sigProb`) =-2.5 has diminished.

This has led to the $B\overline{B}$ components now dominating the $\log_{10}(\texttt{sigProb})$ distribution, though they still maintain a shape difference to the signal $B^+ \to \mu^+\nu_\mu$ decay.

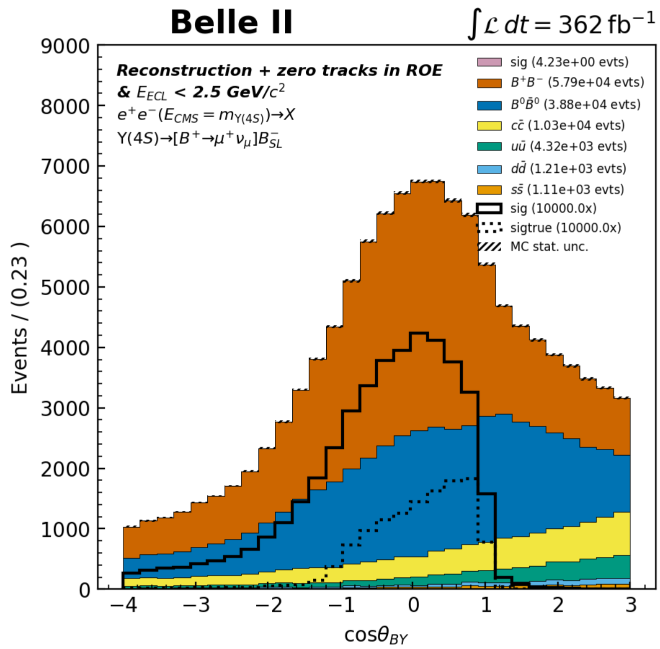
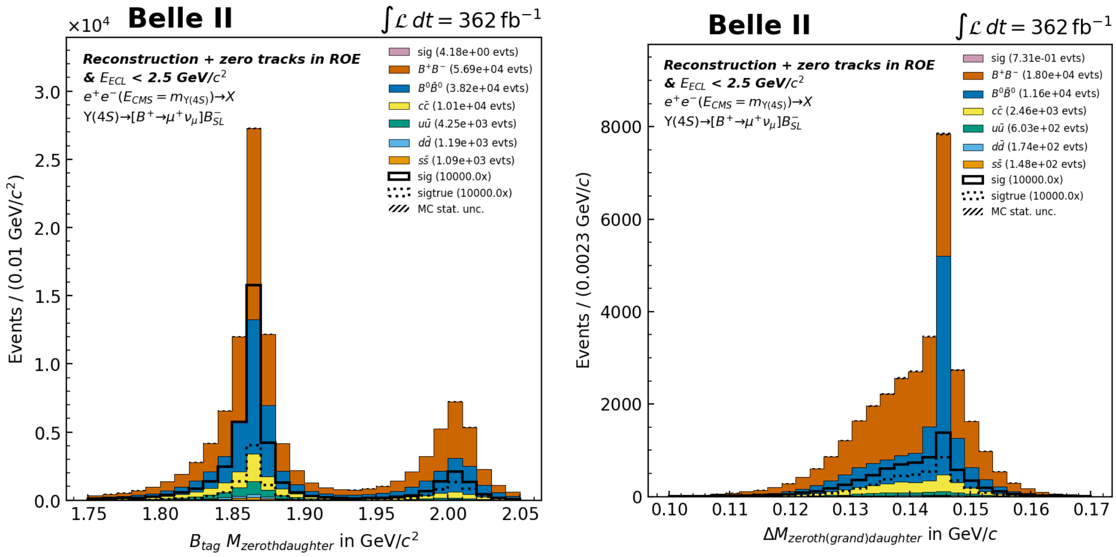


Figure 4.25. $\cos\theta_{BY}$ for B_{tag} used in $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

Figure 4.25 illustrates the $\cos\theta_{BY}$ distribution of the B_{tag} used to reconstruct $\Upsilon(4S)$. The $\cos\theta_{BY}$ region of [-1, 1], which indicates better B_{tag} reconstruction, is now dominated by the $B\overline{B}$ components. Both the charged and $B^+ \to \mu^+\nu_\mu$ components peak in this region, while the $B^0\overline{B}^0$ component peaks broadly between $\cos\theta_{BY}$ = -1 and $\cos\theta_{BY}$ = 2 The shape of the remaining $q\overline{q}$ events is distinct from the $B\overline{B}$ components, where they linearly increase towards the misreconstructed $\cos\theta_{BY}$ >1 region. With minimal signal events in the $\cos\theta_{BY} > 1$ region, even after the signal shape has been scaled up 10^5 times, it is clear that a restriction should be made to reject the background events in this region.

The distribution of zeroth daughter masses (D/D^*) of the B_{tag} used in the $\Upsilon(4S)$ reconstruction is demonstrated in Figure 4.26a. The shapes are consistent with the pre-reconstruction version in Figure 4.17a, save for the reduction in $q\overline{q}$ events. As for the D^*/D mass splitting in the $B^+ \to D^*\ell^+\,\nu_\ell$ decay modes, the distribution is placed on a smaller range for the first time. The peak mass splitting continues to be sharply peaked at the nominal value of 0.142 GeV/c^2. The majority of events in this bin are from



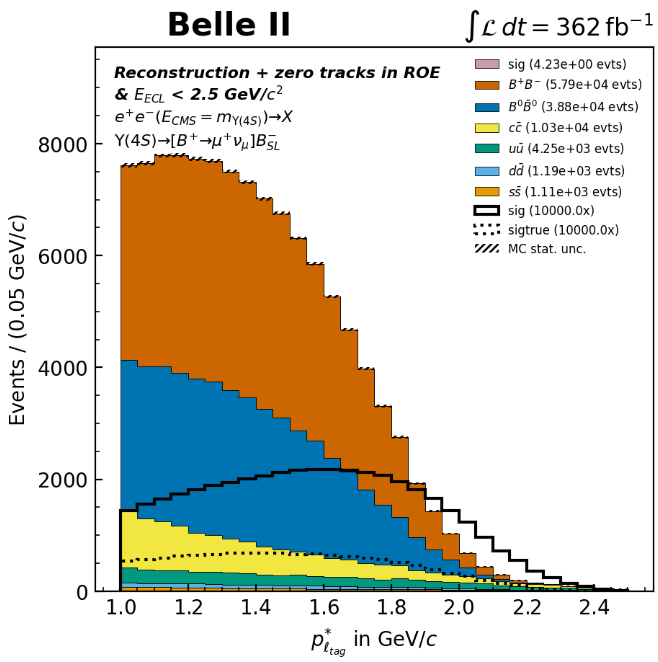
(a) Zeroth daughter mass. **(b)** Mass splitting of D^*/D.

Figure 4.26. Mass variables of $D^{(*)}$ from $B_{\rm tag}$ used in $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\rm SL}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

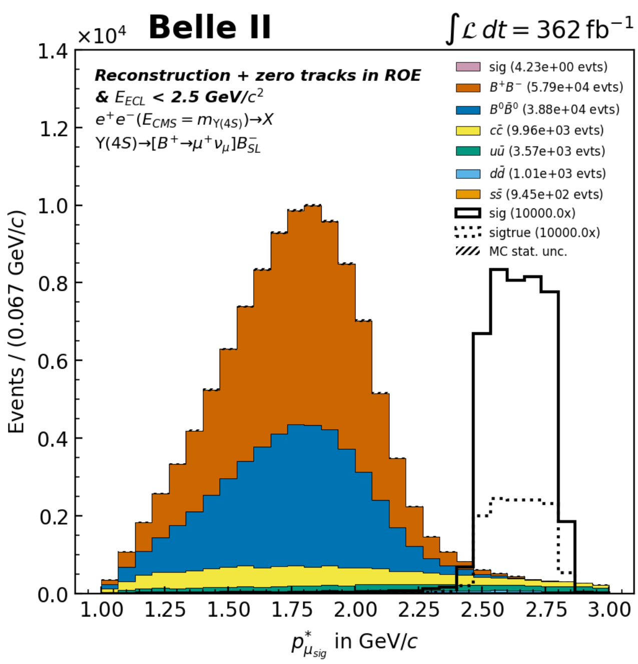
the mixed (i.e. neutral) B MC. As neutral B events cannot be accurately reconstructed from the B^+ charged modes, this is seemingly an anomaly. Instead, it suggests that there is a real $B^0 \to D^* \to D$ transition (e.g. $B^0 \to [D^{*-}\!\to D^-\ \pi^0\,]\ell^+\ \nu_\ell$) in the mixed sample, that mimic a charged B decay when combined incorrectly with other tracks. These events of the splitting peak would likely demonstrate misreconstruction in other variables.

Figure 4.27a contains the distribution of the tag-side lepton momentum in $\Upsilon(4S)$ frame, from the $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\rm SL}$ reconstruction. Prior to reconstruction in Figure 4.18a, the effect of the $p^*_\ell < p^*_\mu$ selection lead to a sharp reduction in the number of events from the $p^*_\ell > 1$ GeV/c boundary. After reconstruction and the ROE pre-selection criteria, the number of $q\bar{q}$ events have fallen considerably to reveal the broader peaking lepton momentum of the $B\overline{B}$ components.

The signal muon momentum again shows significant signal-to-background distinction in Figure 4.27b. With the reconstruction and the ROE selection, the $q\bar{q}$ components are now broadly distributed in the stack. This has lead to the $B\overline{B}$ components on top peaking symmetrically at p^*_μ =1 GeV/c, distinctly demarcated from the overlaid signal shape. Despite the continuum reduction below 2 GeV/c, the range of p^*_μ corresponding with the



(a) Tag-side lepton momentum in $\Upsilon(4S)$ frame.



(b) Signal muon momentum in $\Upsilon(4S)$ frame.

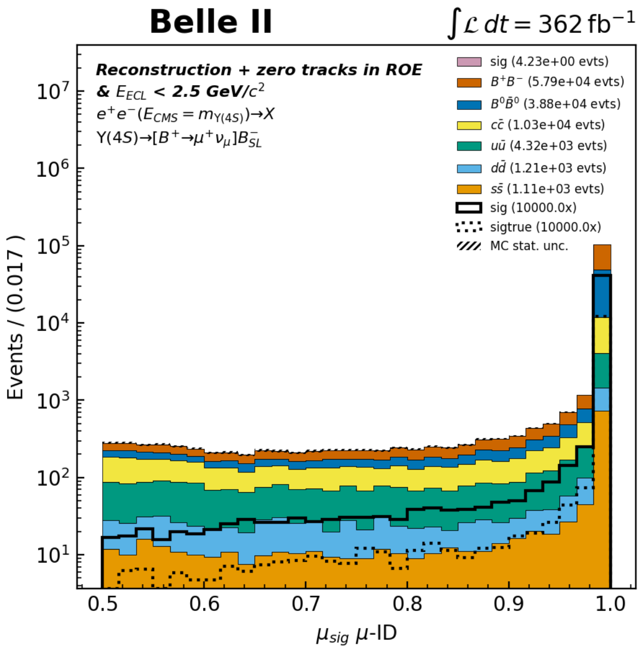
Figure 4.27. Lepton momenta from $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

signal shape continues to contain more $q\overline{q}$ events than generic $B\overline{B}$ events. Setting a lower bound on p^*_μ of events will thus remove a large amount of $B\overline{B}$ events in displayed in the Figure.

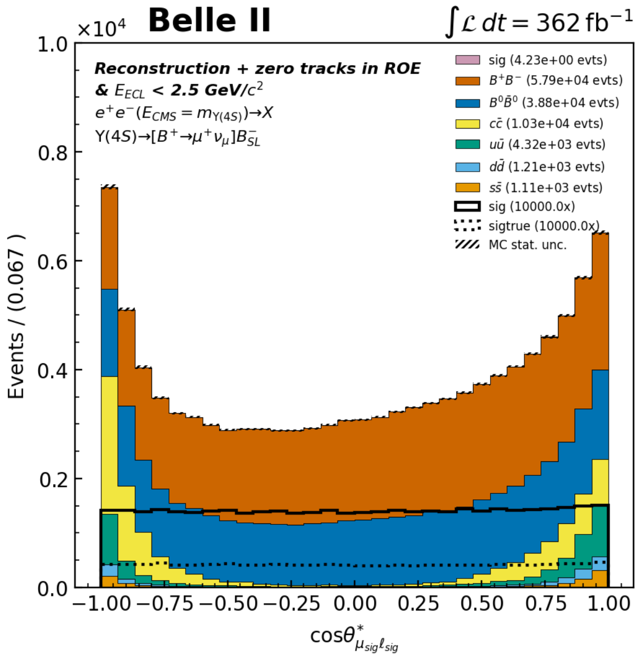
The muon identification probability of the signal muon in the $B^+ \to \mu^+\nu_\mu$ reconstruction is presented in Figure 4.28a. Events are on log-scaled due to the surplus of events in the muonID$\geq$0.99 bin, which wasn't necessary in the pre-reconstruction distribution of Figure 4.19. The muons identified in each MC type are mostly uniformly distributed in muonID below the peak. Placing a stricter lower bound on the signal muons of the event will not lead to considerable signal enhancement, as it is seen that all MC components contain high quality muons.

The cosine angle between tag-side and signal side lepton momenta in the $\Upsilon(4S)$ frame, after reconstruction and ROE pre-selection criteria, is depicted in Figure 4.28b. It was seen in the prereconstruction version in Figure 4.20 that there were three times the events in the negative cosine peak relative to the $\cos\theta^*_{\mu_{\text{sig}},\ell_{\text{tag}}}$=1 peak. This was mostly due to the domination of the $c\overline{c}$ and $u\overline{u}$ background components. With the background rejection taking place due to the reconstruction and the ROE pre-selection criteria, this preference for the anti-parallel opening angle of the leptons has diminished. While $q\overline{q}$ events continue to prefer extreme values in cosine opening angle (i.e. leptons that are near parallel or near anti-parallel), yields in the extreme bins are now more equal in magnitude. They are also in height to the yields on the wider [-1, 1] interval, where the more uniformly distributed signal and $B\overline{B}$ components are predominant. The cosine opening angle of these components are relatively flat on the interval [-1, 1], with the most outward peaking in the more misreconstructed $B^0\overline{B}^0$ MC. Misreconstructed B events favouring extreme cosine opening angles was seen in the incorrectly truth-matched components in the $B^+ \to \mu^+\nu_\mu$ MC only version of this distribution in Figure 4.10a. Thus, a future selection to prefer events with cosine opening angle of leptons closer to zero will effectively reject the peaking continuum and misreconstructed $B\overline{B}$ events.

The energy that remains in the ECL after $\Upsilon(4S)$ reconstruction in signal and background MC is presented in Figure 4.29. The charged and mixed $B\overline{B}$ components continue to dominate the shape of this variable, peaking between 1 and 2 GeV. Prior to reconstruction in Figure 4.22, there was a large number of $c\overline{c}$ and $u\overline{u}$ continuum events that could mimic B_{tag} and signal muon, leaving increasing amounts of energy towards 2.5 GeV in the ECL. These same events have not survived reconstruction or ROE preselection, leading to a more symmetrical shape in the background $B\overline{B}$ events left behind. This shape remains distinct from the overlaid signal shape. The goal of future selections, inclusive



(a) Signal muon `muonID`.



(b) The angle between the signal muon and the tag-side lepton momenta in $\Upsilon(4S)$ frame.

Figure 4.28. Further signal and tag-side lepton attributes of $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

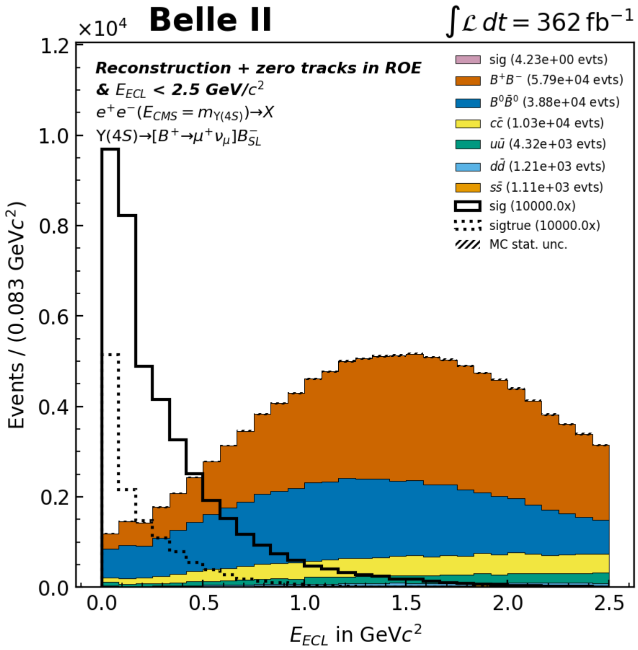


Figure 4.29. Energy remaining in the ECL after $\Upsilon(4S)$ reconstruction obtained via $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

of a tighter upper bound on E_{ECL}, will be to remove the remaining non-$B^+ \to \mu^+\nu_\mu$ $B\overline{B}$ events in the [0,0.5] GeV region, which is where a potential signal yield could be calculated.

The investigation of B_{tag} quality, signal muon quality and ROE variables that have been previously seen has been completed. Now that the full $\Upsilon(4S)$ event has been identified, three more variables that have been historically used in rejecting $q\overline{q}$ background will be consulted, also referred to as *continuum suppression*.

The first new continuum suppression variable to be presented is in Figure 4.30, $\cos\theta_{T_{B_{\mathrm{tag}}}T_{\mathrm{ROE}}}$. This corresponds to the cosine angle between the *thrust axis* of the B_{tag} and the thrust axis of the Rest-of-Event sum of tracks and clusters. Additionally, the ROE used for this calculation is not the one built on the $\Upsilon(4S)$. It is instead specially created using the unused tracks and clusters of the B_{tag}, so any potential signal muon remains in the ROE rather than extracted as it was in Section 4.1.3

The thrust is a measure of the event shape i.e. how collimated the products of the e^+e^- collision are. To construct the thrust axis of a particle, the momentum of its final state products are projected onto trial unit vectors. The thrust axis is the

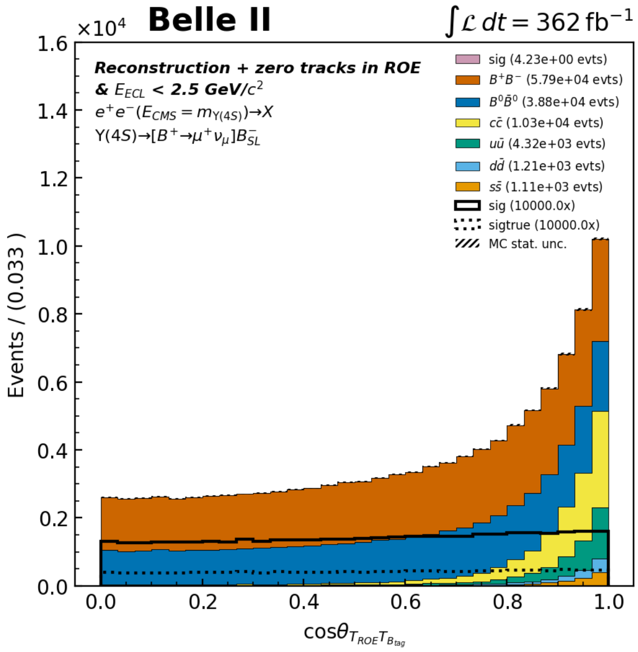


Figure 4.30. Cosine angle between the thrust axis of the B_{tag} and the thrust axis of all other particles in the event, inclusive of signal muon, from $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$ reconstruction events. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

resulting unit vector which maximises the sum of the momentum projections in the $\Upsilon(4S)$ frame. However, for $\cos\theta_{T_B T_{\text{ROE}}}$, the angle is chosen to be smallest measured at the intersection of the line extending in both directions parallel to the thrust axis unit vector i.e. $\theta_{T_{B_{\text{tag}}} T_{\text{ROE}}} \in [0,\pi/2]$ and $\cos\theta_{T_B T_{\text{ROE}}} \in [0,1]$. Thus $\cos\theta_{T_B T_{\text{ROE}}}$ gives an indication of how parallel the products of the B_{tag} and the ROE event are. From earlier discussions of the event shape in continuum versus $B\overline{B}$ events in Section 2.2.3, it is known that $B\overline{B}$ events are more likely to overlap, while $q\overline{q}$ ones will appear 'jetty' or highly collimated. In this way, continuum events are expected to have much smaller thrust axis separation within the event i.e. a larger value of $\cos\theta_{T_B T_{\text{ROE}}}$.

The Figure demonstrates this effect, where $B\overline{B}$ events are flat in $\cos\theta_{T_B T_{\text{ROE}}}$ and continuum peaks for $\cos\theta_{T_B T_{\text{ROE}}} = 1$ (i.e. a small thrust axis separation of $\theta_{T_B T_{\text{ROE}}} = 0°$) Furthermore, an upper bound in the range of 0.6 to 1 on $\cos\theta_{T_{B_{\text{tag}}} T_{\text{ROE}}}$ could effectively reject all non-$B\overline{B}$ background from consideration.

An additional powerful continuum suppression variable is the Fox-Wolfram R_2 variable illustrated in Figure 4.31, which was used frequently in the prior $B^+ \to \mu^+\nu_\mu$ searches

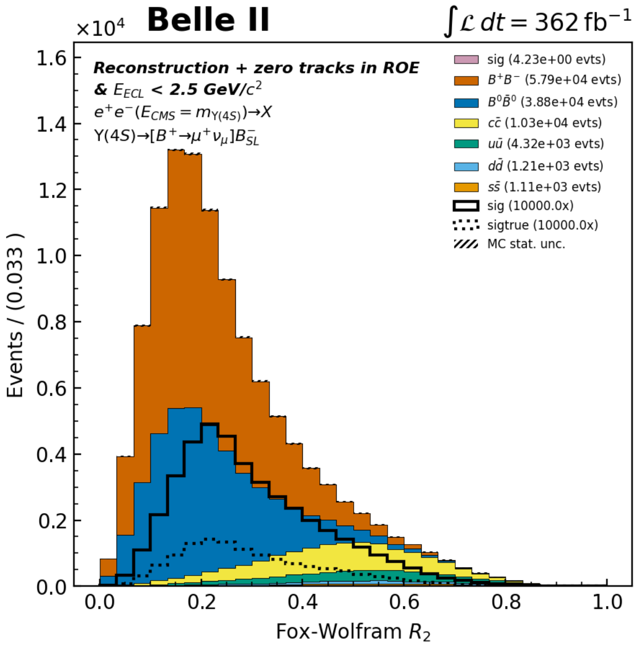


Figure 4.31. Fox-Wolfram ratio (R_2) for events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\rm SL}$. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

described in Section 1.4. The exact definition is presented here. More precisely, R_2 is the ratio of second and zeroth order Fox-Wolfram moments and are derived from spherical harmonics [68]. In practice, they are calculated from the momentum of all final state particles in the event and the angles between them. R_2 itself has a range of 0 to 1, expressing a ratio of 'spherical' to 'jetty' combinations of particles. Thus, $B\overline{B}$ events, including the signal $B^+ \to \mu^+\nu_\mu$ process, are expected to peak towards R_2=0. $q\overline{q}$ events, which are more conical in their event shapes, will peak towards R_2=1. With respect to the distribution after $\Upsilon(4S)$ reconstruction, the overlaid signal shape and the $B\overline{B}$ background components peak around 0.2. The $q\overline{q}$ components peak around 0.5, though it could be that events that peaked more towards R_2=1 have been removed with the ROE pre-selection criteria. There is some small distinction in the peak between that of the $B\overline{B}$ background and the overlaid signal shape. This could be because R_2 is calculated using all detectable final state particles in the event, not just those used in $\Upsilon(4S)$ reconstruction. As there are two missing neutrinos in the $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\rm SL}$ process, this could make the event seem less isotropic and thus register higher values of R_2.

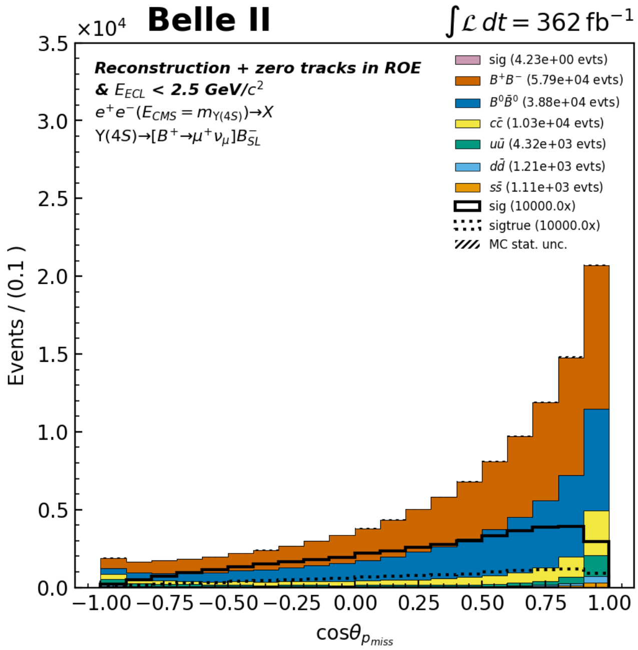


Figure 4.32. Cosine angle of missing momentum in lab frame, for events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu]B^-_{\mathrm{SL}}$. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

The final variable to be examined leverages missing neutrinos as a signifier of the $B^+ \to \mu^+\nu_\mu$ process, something that will not be present in other hadronic B decays. Figure 4.32 contains the distribution of cosine angle of the missing momentum vector, specifically in the lab frame. Provided that all measurable particles have been detected, the missing momentum will correspond to the combined momentum of the signal-side and tag-side neutrinos. Despite not being detectable, the neutrino momenta can be calculated from the known initial momentum of the e^+e^- collision and subtracting the visible momenta of the event. Specifically, the variable is cosine polar angle of the missing momentum vector relative to the forward direction of the beamline. With respect to the distribution, the number of events in the signal shape is shown to linearly increase with $\cos\theta_{\vec{p}_{\mathrm{miss}}}$ (i.e. as the missing momentum vector is rotated from anti-parallel to the forward beamline to parallel) until the final bin. This suggests that the combined neutrino momentum is less likely to be completely parallel to the beamline in the signal events. Contrasting with the overlaid signal shape, the background components all peak towards the same bin. Here, it would be expected that background distributions to match if the underlying processes also have neutrinos. With the mismatching at in the $\cos\theta$

= -1 and cosθ = 1 bin, selections that exclude these bins could be effective in rejecting non-neutrino events in background events.

A number of variables which could provide useful selection for significantly reducing the number of background events have been consulted. However, with a small number of $B^+ \to \mu^+\nu_\mu$ events left after reconstruction, it is important to consider what selections would reduce the largest amount of background while preserving signal. Choosing selections by inspection can be subject to biases. It is therefore desired to be able to defer to a method of selection optimisation. This strategy will be introduced in the next section.

4.2. Developing Selection Criteria to Prioritise $B^+ \to \mu^+\nu_\mu$

As a helicity and Cabbibo suppressed decay, $B^+ \to \mu^+\nu_\mu$ has a small Standard Model branching fraction. In 362 fb^{-1}, very few signal events are expected, compared to those from standard B decays and continuum processes. It has already been seen throughout the last Chapter and this one when plotting key variables of the $B^+ \to \mu^+\nu_\mu$ decay in MC that background events overwhelm the small number of signal events. It will therefore be important to reduce the contributions of background events from the ultimate yield in data.

For this search, a signal variable will be implemented, a variable whose distribution naturally divide data into regions of mostly signal and mostly background. If possible, this 'mostly signal' region will be where the signal yield is to be evaluated, with a low enough number of background events such that any signal ones are not overwhelmed by Poisson counting error. After this region is determined, additional selection criteria can be explored, to further enhance the significance of the signal over background events. Implementing selection criteria within high-energy physics is sometimes referred to as making 'cuts', in the sense of cutting the data into subsets. A number of selections have been implemented already: there were pre-selection criteria performed in the semileptonic FEI to limit to the best quality B_{tag}, the $p^*_{\ell,\text{tag}} < p^*_{\mu,\text{sig}}$ selection to avoid the confusion between the leptons in signal MC, as well as the ROE selections to remove the regions of the dataset where background processes dominated. Aided by the knowledge of the signal and background distributions of key variables pertaining to the $\Upsilon(4S)$

reconstruction, selections will finally be implemented to limit the analysis to the subset of events where signal events are most probable. Once selections are applied, the signal variable distribution can be compared with their collision data counter parts to thus estimate the yield of $B^+ \to \mu^+\nu_\mu$ events in the 362 fb^{-1} dataset.

There are many methods of determining selections to obtain the signal-enhanced subset of events in which yields will be evaluated. Many of the previous $B^+ \to \mu^+\nu_\mu$ searches discussed in Section 1.4 use machine learning methods to identify which events are more likely to contain a $B^+ \to \mu^+\nu_\mu$ process. Machine learning has already been implemented in this analysis, in the form of the semileptonic FEI used in B-tagging. Machine learning methods, such as boosted decision tree and neural network learning, are well suited to identifying subtle relationships between the observables of physics decays that can be used to reject backgrounds from consideration. This can generally be done faster and with less potential for biases and inconsistencies than if a human were to determine selections by hand for the considered variables. A downside of using machine learning techniques is the need for larger MC datasets, in order to train, test and then utilise the method. For this analysis, it will be seen that the amount of background MC is limited when compared to what would be required to achieve proper sensitivity to observing a $B^+ \to \mu^+\nu_\mu$ event. Machine-learning methods can also be opaque to what has taken place during training phases, making it difficult to understand and validate their performance. This does not contradict the use of the FEI; the FEI has already been subject to intense scrutiny within the Belle II collaboration.

In view of the downsides of using further machine learning methods to obtain a background suppressed dataset, a more manual approach will be adopted. Rather than deciding variable bounds by consulting the signal and background distributions of histograms from Section 4.1.5, a signal optimisation procedure will be defined in the next Section. A *significance estimator* will be implemented to ensure the chosen boundaries on the signal region and in selection variables are optimised to reject background events, while preserving a non-zero number of signal events.

4.2.1. Using significances estimation in optimising selection criteria: $\frac{S}{\sqrt{S+B}}$

Significance, within the context of statistical testing, refers to how likely a piece of evidence rejections the null case over a given hypothesis [108]. For this $B^+ \to \mu^+\nu_\mu$

search, high significance that the yield is not within statistical fluctuation of background events is desired, which instead better supports the presence of the $B^+ \to \mu^+\nu_\mu$ process.

A *significance estimator* will be used to decide on the limits of signal regions, as well as limits on background rejection selections applied to said region. Thus, the boundaries of selection criteria ultimately utilised will be those that maximise the significance estimate. As per their name, significance estimators do not yield precise values of significance. Once the selection criteria are determined, applied to data and a signal yield is derived, the true significance of this result can be evaluated. This will occur in Section 5.3, via the method of Feldman-Cousins [109].

The optimised selections will minimise the background events, (thus minimise Poisson error), whilst also maximising the number of signal events. The quantity used as an estimate on significance of the chosen signal over all events is as follows:

$$\mathcal{S}(a \leq x < b) = \frac{S}{\sqrt{S+B}}.$$

Here, the significance estimator $\mathcal{S}$ is evaluated for variable x on an interval of $[a,\, b)$. Within this interval, the number of signal events is measured to be S, while B is the number of background ones. The goal here is to implement selection criteria of the form $x \geq a$ (or $x <$ b), where x is any measurable of reconstructed $\Upsilon(4S)$ events and a (or b) is some bounding value, such that the quantity $\frac{S}{\sqrt{S+B}}$ is maximised over this region. $\frac{S}{\sqrt{S+B}}$ functions as a significance estimator, on the assumption that the measurement of the number of particle events can be modelled via a Poisson distribution. If one assumes that the measured number of events as the mean to an underlying Poisson distribution, one can obtain the standard deviation as the square root of this. Thus, if $S+B$ events are measured, the uncertainty on this can be treated as the standard deviation, $\sqrt{S+B}$. The significance can then be estimated by considering whether the signal yield S is within statistical fluctuation. Specifically, the magnitude of S is compared to the standard deviation $\sqrt{S+B}$, hence $\frac{S}{\sqrt{S+B}}$.

Cut-based selections are chosen via Figure-of-Merit methods, where selection boundaries are continuously varied and the impact on the $\frac{S}{\sqrt{S+B}}$ evaluated can be observed graphically. The selection boundary that yields a maximum $\frac{S}{\sqrt{S+B}}$ will define the region implemented in the selection. To begin, the signal variable to be used in this analysis is decided. A signal variable that separates the signal-like events and the background-like ones is desired. Thus using the Figure-of-Merit approach with $\frac{S}{\sqrt{S+B}}$ to both find a suitable variable and its bounding value is a natural path to take.

It is noted that there are alternate optimisation quantities available. Ref. [110] discusses a number of quantities, including those that prioritise a best final upper limit and those that prioritise discovery. Maximising $\frac{S}{\sqrt{S+B}}$ prioritises small statistical uncertainty, which is suitable given that this will be seen to be a limiting factor on the final branching fraction measured. Such a quantity does not, however, take into account the systematic uncertainties that will be discussed in Chapterchap:chap5.

4.2.2. Choosing a signal variable and signal region

Scope

This analysis is looking for a parameter space of collisions in which background events are minimal and where there's a non-zero amount of signal events. In such a region, the number of events in data that are due to $B^+ \to \mu^+\nu_\mu$ processes is estimated. This is what is defined as the *signal region.* Signal regions are traditionally kept 'blind', meaning that the number of events in data are not measured until the analysis strategy has been finalised. This avoids biasing selection criteria towards a desired result. An example of this would be fine-tuning calibration factors to yield a data excess in the signal region, to favour a beyond Standard Model contribution to a process.

For this analysis, the signal variable will be the parameter which defines the region. There will be multiple selections on multiple variables to optimally reduce background events whilst preserving signal ones. Instead of referring to all selection variables then as signal variables, one variable will be defined as the 'primary' signal variable. This signal variable will be taken to be the one whose distribution best divides the data into a signal region and a signal-depleted region (also called the *sideband*) To precisely determine the boundary value that separates signal and background, the $\frac{S}{\sqrt{S+B}}$ technique discussed in the previous section will be implemented. From the previous investigation into the shapes of distributions after reconstruction and ROE pre-selection criteria, a number of potential signal variables are already available. Specifically, p^*_μ and E_{ECL} have demonstrated good separation, and also benefit from having been used as signal variables in previous $B^+ \to \mu^+\nu_\mu$ measurements, with p^*_μ used in the 2018 Belle search and E_{ECL} in the 2010 BaBar search.

It will be noted that such an approach for defining a signal variable is not universal. For example, in the 2020 Belle $B^+ \to \mu^+\nu_\mu$ search, the variable on which the signal yield was calculated did not admit an interval of 'mostly signal'. Instead, the muon momentum in signal B frame was chosen as the signal variable as the shape of the signal and background distributions were distinct enough to separated in a fitting strategy, even though they overlapped in the fitted range.

Alternate signal variables will be briefly discussed. In principle p^B_μ, (i.e. momentum of the signal muon in the signal B frame) would be a fantastically suitable signal variable, where it was derived in Section 1.2.1 that this value is exactly 2.64 GeV/c, up to momentum smearing. Additionally, it was discussed in Section 1.4, p^B_μ could be derived in analyses from the B_{tag} four momentum. However, as a semileptonically-tagged analysis, complete B_{tag} four-momentum knowledge has been sacrificed for both higher efficiency (relative to hadronic tag) and knowledge of the decay mode (relative to inclusive tag). In the previous semileptonic-tagged $B^+ \to \mu^+\nu_\mu$ search at BaBar discussed in Section 1.4.2, the four momentum of the B_{tag} (and thus B_{sig}) could be estimated. This was done by averaging over trial p^*_B constructed to be angle θ_{BY} from the daughter $Y = D^{(*)}\ell$ system, with $|p^*_B|$=0.331 GeV/c. More recently, this technique has been referred to as *the Diamond-Frame method*, when it was applied to obtain the four-vector of B in $B \to D\ \ell\nu$ decays, independent of the inclusive tag in a Belle II derivation of $|V_{cb}|$ [111].

Such attempts were made to evaluate $p^B_{\mu_{\text{sig}}}$ for use as a potential signal variable in this analysis. However, the resolution of such a distribution was found to be larger than that for p^*_μ, instead of narrower as desired. This may be attributed to the dependence on deriving the angle of the B relative to Y, θ_{BY}, from the large number of signal events below the physical region of $\cos\theta_{BY}$, seen in Figure 4.25. Further investigation is required, though this is outside the scope of the current analysis.

There are also other constructed variables that could be considered as signal variables, which can leverage the presence of missing neutrinos in both B_{sig} and B_{tag}. These have been namely implemented in $B \to D^{(*)}\ell\nu$ tagged semileptonic signal decays, rather than missing energy leptonic signal. This includes the $\cos^2\phi_B$ variable derived from signal-side and tag-side $\cos\theta_{BY}$ in semileptonically tagged $B^0 \to \pi^-\ \ell^+\ \nu_\ell$ decays by BaBar [112]. The related variable x^2_B (related to the intersecting line of signal side and tag-side $\cos\theta_{BY}$ cones) was also used in semileptonically tagged $B \to (\pi/\rho)\ell\nu$ decays by Belle [113]. In the interest of simplicity, the use of such variables in discern semileptonically tagged $B^+ \to \mu^+\nu_\mu$ decays are left for future investigations.

Thus, to determine which of E_{ECL} and p^*_μ make the best signal variable, the $\frac{S}{\sqrt{S+B}}$ procedure with both variables will be performed, to investigate which yields the higher $\frac{S}{\sqrt{S+B}}$ and thus will be the signal variable. To ascertain as to if there are other potential signal variables, the $\frac{S}{\sqrt{S+B}}$ procedure will be performed on a number of other variables. This includes key variables that have been investigated, as well as other kinematic variables that may affect $B^+ \to \mu^+\nu_\mu$.

This $\frac{S}{\sqrt{S+B}}$ procedure is as follows.

1. Apply any already decided on selection criteria to the dataset and then repeat the following for each variable:
   a) For each MC sample (both signal and background), evaluate the minimum and the maximum value of the variable x in all events. If there are significant outliers in the variable, appropriate minima and maxima are substituted by hand.
   b) Store the maximum (x_{max}) of all maxima found in the MC samples. Similarly, store the minimum (x_{min}) of all minima.
   c) Construct a list of 51 values, defined as entries evenly spaced over the interval of $[x_{\text{min}}, x_{\text{max}}]$ (both inclusive).
   d) For each value x_j in the list:
      i. Make a selection of $x \geq x_j$ in all MC components.
      ii. Evaluate S_j, B_j (the number of signal and background events that survive the selection criteria, scaled to 362 fb^{-1})
      iii. Evaluate $\mathcal{S}_j = \frac{S_j}{\sqrt{S_j+B_j}}$.
      iv. Store $\mathcal{S}_j$, S_j, B_j and the selection string '$x \geq x_j$' in a list with results from other x_j.
      v. Repeat for selection $x < x_j$.
   e) From the list of $\mathcal{S}_j$ calculated across the interval, find the selection, $x < x_j$ or $x \geq x_j$, that yields the maximum $\mathcal{S}_j$
   f) Store the maximum $\mathcal{S}_j$, the corresponding S_j, B_j and the selection string '$x \geq x_j$' (or '$x < x_j$') in a list with results from other x.
2. After all x are trialled, sort the list of maximum $\mathcal{S}$ from each variable to find the variable and variable selection that yields the maximum $\mathcal{S}$ overall.

3. If the x_j of the selection is not physical (i.e. the variable only admits integers, or dividing the interval into 50 is beyond the measurement precision) repeat 1d for a more appropriate interval and accept if consistent with previous step.
4. Once the selection is chosen, apply it to the dataset and repeat 1d for the opposite bound to obtain the full region.

Results

The described $\frac{S}{\sqrt{S+B}}$ procedure to identify potential signal regions from variables concerning $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction has been performed in the sample of signal ($B^+ \to \mu^+\nu_\mu$) and background (B^+B^-, $B^0\overline{B}^0$, $c\bar{c}$, $u\bar{u}$, $d\bar{d}$, $s\bar{s}$), all scaled to their expected yields in 362 fb^{-1}. As many as 157 variables have been trialled, inclusive of a range of ROE variables, continuum suppression variables, event shape variables, event kinematic variables, as well as kinematic and identity variables of particles in the reconstructed $\Upsilon(4S)$ decay chain. The variables explicitly do not contain truth variables, as these selections will not be available for data.

Table 4.5. Potential signal variable regions, first ten ranked by $\frac{S}{\sqrt{S+B}}$.
Pre-selection values: S_i=4.2±2.1, B_i=(1.365±0.0034)×10^5, $\mathcal{S}_i$=0.013±0.0061.

Selection	$\frac{S}{\sqrt{S+B}}$	S	B
$E^*_\mu \geq 2.4658$ GeV	0.061±0.030	4.1±2.0	(4.57±0.07)×10^3
$p^*_\mu \geq 2.4626$ GeV/c	0.061±0.030	4.1±2.0	(4.58±0.07)×10^3
$E_{\mathrm{ECL}} < 0.25$ GeV	0.036±0.024	2.3±1.5	(4.08±0.06)×10^3
(p^*_μ-p^*_ℓ) ≥ 1.0234 GeV/c	0.033±0.022	2.3±1.5	(4.81±0.07)×10^3
$M_{\mathrm{ROE}} < 0.0997$ GeV/c^2	0.032±0.020	2.6±1.6	(6.71±0.08)×10^3
(no. of photons in ROE) < 1.15	0.032±0.021	2.4±1.5	(5.57±0.07)×10^3
(no. of ECL clusters in ROE) < 1.15	0.032±0.021	2.4±1.5	(5.57±0.07)×10^3
(no. of neutral ECL clusters in ROE) < 1.15	0.032±0.021	2.4±1.5	(5.57±0.07)×10^3
$p^{\mathrm{lab}}_\mu \geq 2.7215$ GeV/c	0.030±0.021	2.1±1.5	(4.92±0.07)×10^3
$p^{\mathrm{lab}}_{\mathrm{ROE}} < 0.1496$ GeV/c	0.030±0.022	1.8±1.3	(3.57±0.06)×10^3

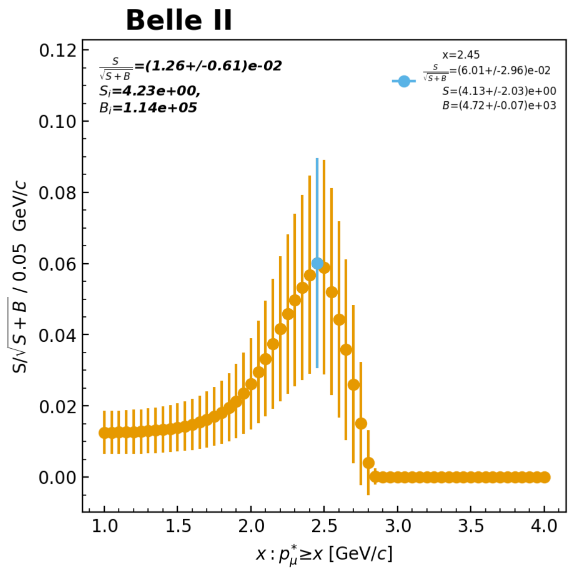
Table 4.5 contains the variables and their selections which yielded the 10 largest values of $\frac{S}{\sqrt{S+B}}$, in signal and background MC scaled to 362 fb^{-1}. The first two selections concern the kinematics of the signal muon, specifically the momentum and energy measured

in the $\Upsilon(4S)$ frame. These two variables are directly correlated, motivating why their proposed signal regions yield the same number of signal and background events. The resulting $\frac{S}{\sqrt{S+B}}$ is just less than double what is yielded by using E_{ECL} as a signal variable, suggesting that p^*_μ performs better as a primary discriminating variable. The remaining variables concern either the dynamics of the muon or the ROE, thus will function similarly to a selection in p^*_μ or E_{ECL}.

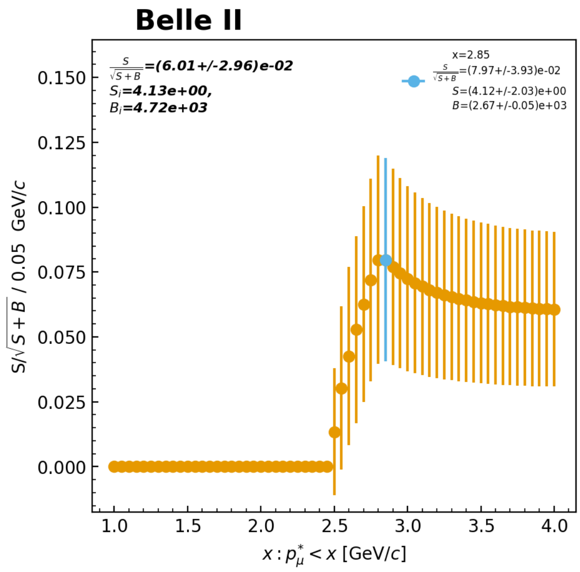
Figure 4.33 compares the Figures-of-Merit for potential signal regions in p^*_μ and E_{ECL}. Here, the intervals have been changed to 50 MeV and 50 MeV/c increments, respectively. The figures in the first column show the boundaries proposed by the $\frac{S}{\sqrt{S+B}}$ analysis, while the second column shows the corresponding Figures-of-Merit for the secondary boundary found after the first selection is applied. The $\frac{S}{\sqrt{S+B}}$ values according to $p^*_\mu \geq 2.45$ GeV/c and $E_{\text{ECL}} < 0.25$ GeV are consistent with those in the Table 4.5. The p^*_μ version admits a more distinct peak at $p^*_\mu = 2.45$ GeV/c than that of the $E_{\text{ECL}} = 0.25$ GeV one, suggesting stability in slight variations in E_{ECL}, thus less need for a fine-tuned bound. Despite this, it is decided to continue with the signal region that yields that largest $\frac{S}{\sqrt{S+B}}$, that is $p^*_\mu \in [2.45, 2.85)$ GeV/c. It will be in this region that future selections will be made that maximise $\frac{S}{\sqrt{S+B}}$. It has thus been shown the power of using p^*_μ as a signal variable over E_{ECL}. As it will soon be seen, there is still potential for a selection on E_{ECL} to reject additional background events in the newly defined signal region, despite not being implemented as the signal variable.

In addition to a signal region, the analysis design requires a sideband region. While the blinding method means that there will be no data investigated in the signal region until the analysis is finalised, data and data-MC agreement can be inspected in a sideband region. This is because the sideband is designed to not contain any $B^+ \to \mu^+\nu_\mu$ events, thus will not be used in evaluating an overall signal yield. Such a region will act as a control, so that data-MC agreement can be improved, without biasing towards a positive result. At this stage in the analysis, the sideband will be defined to the left of the signal region, covering an equivalent width of 400 MeV/c i.e. $p^*_\mu \in [2.05, 2.45)$ GeV/c.

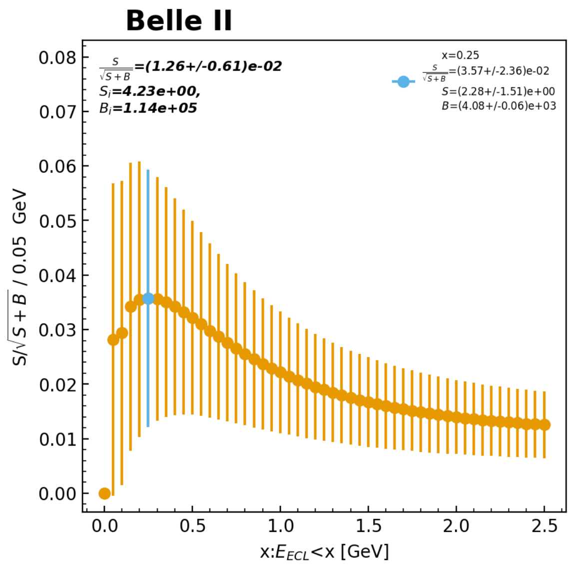
Figure 4.34 contains the distribution of p^*_μ in both the sideband and signal region, on both linear and log event scale. The continuum events are uniform in both the sideband and signal region, with the vast majority of $B\overline{B}$ background in the sideband. Despite the comparative lack of $B\overline{B}$ background in the signal region, there are still 2.67×10^3 background events, such that the signal component stacked on top of background is still overwhelmed in this region. Further selection criteria must be implemented to reveal the contribution to p^*_μ from $B^+ \to \mu^+\nu_\mu$. Table 4.6 documents the number of events



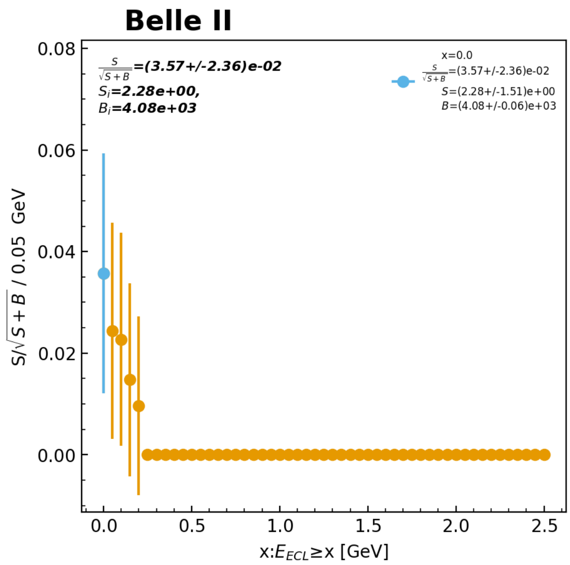
(a) Proposed: $p^*_\mu \geq 2.45$ GeV/c.



(b) Opposite bound: $p^*_\mu < 2.85$ GeV/c.



(c) Proposed: $E_{\mathrm{ECL}} < 0.25$ GeV.



(d) Opposite bound:$E_{\mathrm{ECL}} \geq 0$.

Figure 4.33. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for potential signal regions. p^*_μ is varied 50 MeV/c increments in [1,4] GeV/c, while E_{ECL} is varied in 50 MeV increments in [0, 2.5] GeV. Error bars are propagated from Poisson statistical error in the MC components.

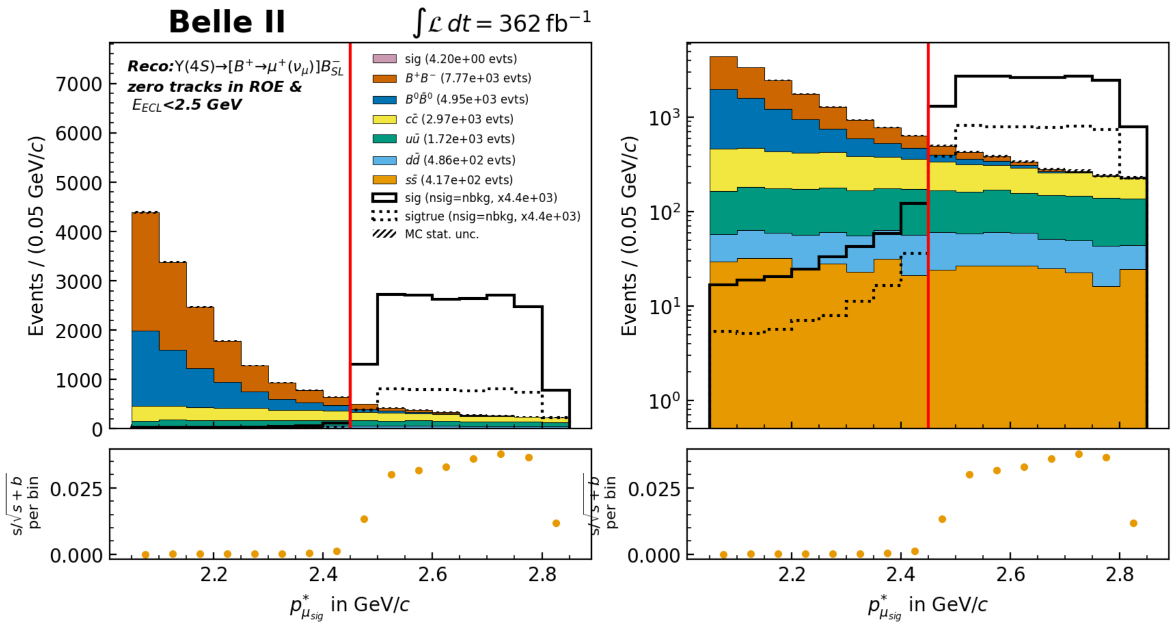


Figure 4.34. Linear and log-scale distributions of $\Upsilon(4S)$ frame muon moment, from events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\rm SL}$. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates the proposed boundary between the sideband (left) and signal region (right).

Table 4.6. Yields and Efficiencies in $p^*_\mu \in$[2.45, 2.85) GeV/c.

Sample	Yield	Selection eff [%]	Cuml. eff [%]
$B^+ \to \mu^+\nu_\mu$	4±2	97.40	2.37
B^+B^-	254±16	0.44	$1.24{\times}10^{-4}$
$B^0\overline{B}^0$	168±13	0.43	$8.66{\times}10^{-5}$
$c\overline{c}$	1016±32	9.85	$2.11{\times}10^{-4}$
$u\overline{u}$	801±28	18.52	$1.38{\times}10^{-4}$
$d\overline{d}$	236±15	19.56	$1.63{\times}10^{-4}$
$s\overline{s}$	193±14	17.34	$1.39{\times}10^{-4}$
tot. bkg.	2667±52	2.35	$1.53{\times}10^{-4}$
$\frac{S}{\sqrt{S+B}}$	0.08	-	-

per MC component in the signal region. Restricting to this region preserves 97.40% of reconstructed $B^+ \to \mu^+\nu_\mu$ events, while only 2.35% of the reconstructed background remains.

4.2.3. Choosing selection criteria

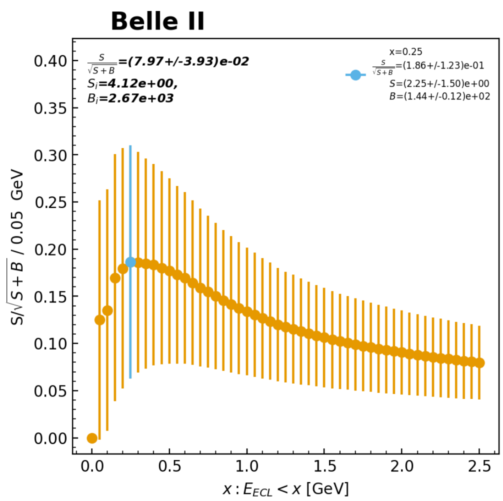
In the previous section, the signal region of $p^*_\mu \in$[2.45, 2.85) GeV/c was decided. This naturally defined a sideband region of equal width as $p^*_\mu \in$[2.05, 2.45) GeV/c, which will be used to probe data-MC agreement. Now that the signal variable and region is chosen, the set of selections that will prioritise $B^+ \to \mu^+\nu_\mu$ events in the signal region can be explored. This will again implement the $\frac{S}{\sqrt{S+B}}$ method. The steps outlined in the previous section will be used to identify successive selections until a small enough number of background events has been achieved. This includes defining both upper and lower bounds for each selection, as well as applying all previous selections to the signal region in pursuing a new one. To keep the selections independent, no selection criteria on any variables that are correlated with previously selected variables will be accepted, especially if correlated with p^*_μ. The analysis requires a non-zero number of events in the sideband to define data-MC ratios used in Chapter 5. If any selections were made on variables correlated with p^*_μ, this could entirely remove events in this required region.

First Selection: $E_{\mathbf{ECL}} \in$ [0 , 0.25) GeV

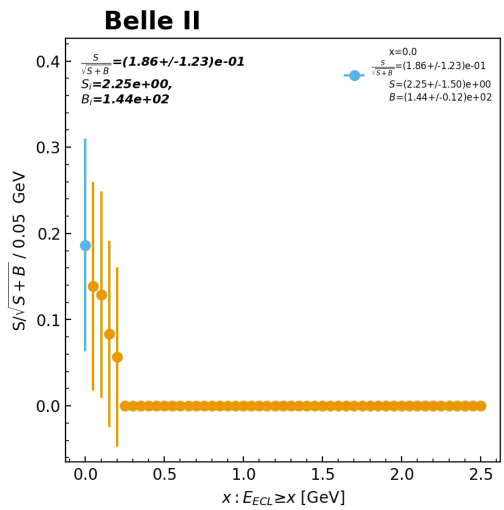
Table 4.7. Potential first selections in signal region $p^*_\mu \in$[2.45,2.85) GeV/c, ranked by $\frac{S}{\sqrt{S+B}}$. Pre-selection values S_i=4.1±2.0, B_i=(2.67±0.051)×10^3, $\mathcal{S}_i$=0.080±0.039.

Selection	$\frac{S}{\sqrt{S+B}}$	S	B
$E_{\text{ECL}} < 0.25$ GeV	0.19±0.12	2.3±1.5	144±12
$p^{\text{lab}}_{\text{ROE}} < 0.199$ GeV/c	0.16±0.11	2.2±1.5	183±14
$p^*_{\text{miss}} \geq 2.4804$ GeV/c	0.16±0.09	2.9±1.7	319±18
$M_{\text{ROE}} < 0.0982$ GeV/c^2	0.16±0.10	2.6±1.6	254±16
$p^*_{\text{ROE}} < 0.1505$ GeV/c	0.16±0.11	1.9±1.4	147±12
(no. of ECL Clusters in ROE) < 1.05	0.16±0.10	2.3±1.5	218±15
(no. of neutral ECL Clusters) < 1.05	0.16±0.10	2.3±1.5	218±15
(no. of photons in ROE) < 1.05	0.16±0.10	2.3±1.5	218±15
$\cos\theta_{T_B T_{\text{ROE}}} < 0.72$	0.15±0.09	2.8±1.7	381±20
$p^*_{T,\text{ROE}} < 0.1681$ GeV/c	0.14±0.09	2.4±1.6	292 ±17

The first potential selections are outlined in Table 4.7. A number of ROE based regions have been identified by the procedure, a large number of which appeared as potential signal regions in Table 4.5. This suggests that despite the p^*_μ signal region performing



(a) Proposed: $E_{\mathrm{ECL}} < 0.25$ GeV.



(b) Opposite bound:≥ 0.25 GeV.

Figure 4.35. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for selections in ECL energy remaining after $\Upsilon(4S)$ reconstruction. E_{ECL} is varied in 50 MeV increments in $[0,\ 2.5]$ GeV. Error bars are propagated from Poisson statistical error in the MC components.

well in separating lower momentum muons from background processes, such events still leave products in the ROE, unlike correctly identified $B^+ \to \mu^+\nu_\mu$ events. New high performing variables include p^*_{miss} the missing momentum in the centre-of-momentum frame, which for signal can be attributed to the two missing neutrinos and could be use to remove background events without neutrinos. Moreover, the variable $\cos\theta_{T_B T_{\mathrm{ROE}}}$ has performed well owing to its role as a continuum suppressing variable, which was seen in Figure 4.30.

The Figures-of-Merit for the highest performing selection criteria in the p^*_μ signal region, $E_{\mathrm{ECL}} < 0.25$ GeV, is presented in Figure 4.35a. The shapes are near identical to those in Figure 4.33, suggesting the exact same selection bounds of $E_{\mathrm{ECL}} \in [0,\ 0.25)$ GeV as those put forward for defining a E_{ECL} signal region.

As discussed in the beginning of Section 4.2.3, selection variables are required to be independent of one another, so as to preserve a non-zero number of events in the sideband. Generally, the design of the $\frac{S}{\sqrt{S+B}}$ method would mean that once a selection has been made, correlated variables will no longer improve the significance estimate, as both the admitted and correlated variable would select the same events. Regardless, for each selection chosen, its correlation with p^*_μ and all previous selection variables will be examined through 2D histograms of each MC sample in the search. The 2D correlation histograms from here onwards will only have the pre-selections on E_{ECL} and

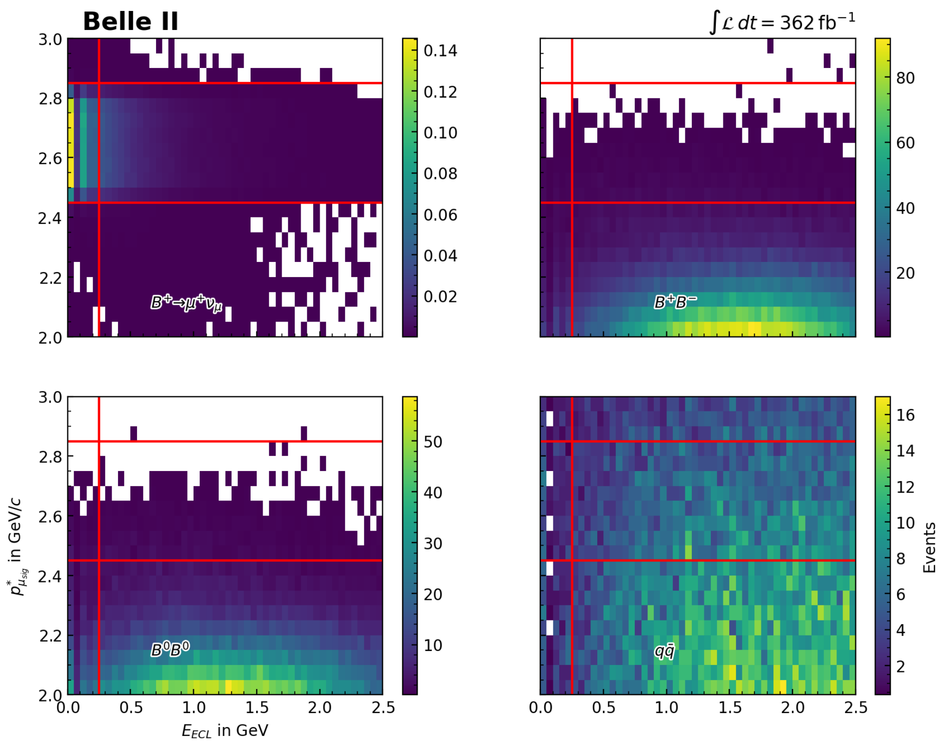
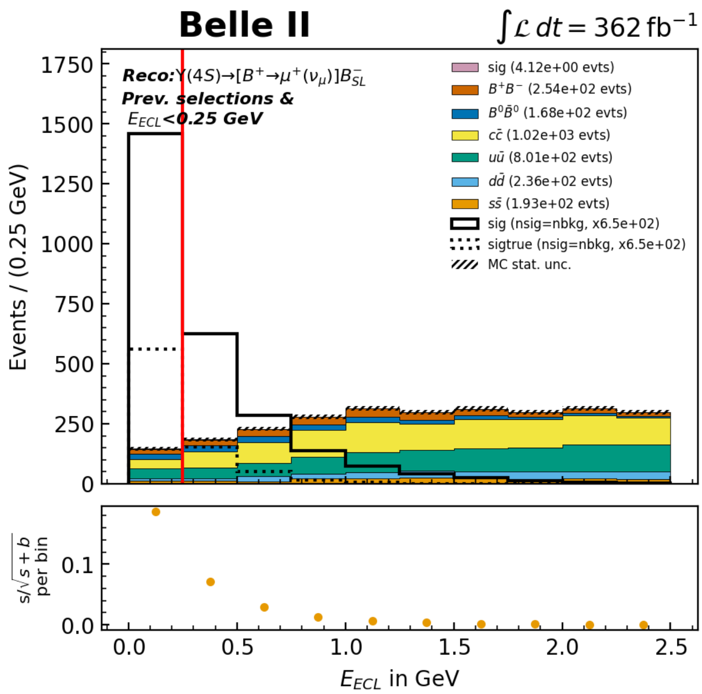


Figure 4.36. 2D histograms of E_{ECL} against signal variable p^*_μ, in different MC types. *Top left:* signal MC, *top right:* B^+B^- MC, *bottom left*: $B^0\overline{B}^0$, *bottom right*: $q\overline{q}$ for $q \in = \{u, d, s, c\}$. Proposed selections indicated by red lines. Only pre-selections applied, so that variable correlation can be visualised.

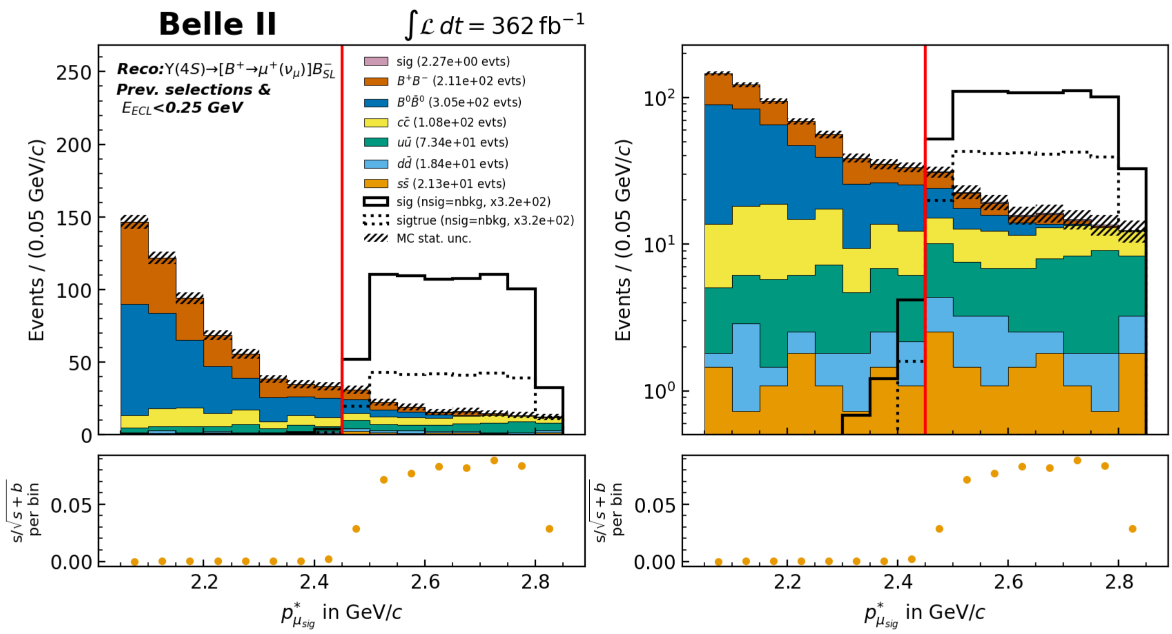
the number of tracks in the ROE. This is because future selections will reduce the dataset considerably, thus making it difficult to see potential correlations.

The 2D correlation histogram for E_{ECL} and p^*_μ is demonstrated in Figure 4.36. The red lines indicate the proposed E_{ECL} selection (vertical) and the p^*_μ signal region (horizontal). Within the signal MC, B^+B^-, $B^0\overline{B}^0$ and $q\overline{q}$ combined distributions, there is no obvious relationship seen between E_{ECL}. However, it can be seen that the p^*_μ signal region removes the peak of $B\overline{B}$ events around 1.75 GeV/c, with $E_{\mathrm{ECL}} < 0.25$ GeV continuing to reject these backgrounds while preserving the hotspot of events in the signal histogram. Unlike $B\overline{B}$, the $q\overline{q}$ events are uniformly distributed in in p^*_μ-E_{ECL}, indicating a large quantity of such events will remain.

The potential of rejecting $B\overline{B}$ from the signal region while retaining $q\overline{q}$ events via the selection $E_{\mathrm{ECL}} < 0.25$ GeV is now seen in Figure 4.37b. Figure 4.37a demonstrates the distribution of E_{ECL} in the p^*_μ signal region, where the vertical line indicates the E_{ECL} boundary yielded by the $\frac{S}{\sqrt{S+B}}$ procedure. It can be seen that the selection of $E_{\mathrm{ECL}}<$



(a) Proposed E_{ECL} selection.



(b) Linear and log-scale distributions of $\Upsilon(4S)$ frame muon momentum.

Figure 4.37. Histograms of events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$, after E_{ECL} selection. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates proposed selections.

0.25 GeV coincides with the bin of highest $B^+ \to \mu^+\nu_\mu$ yield in the overlaid histogram and lowest background in the stacked one. In the subplot below the histogram, the same bin also has the highest $\frac{S}{\sqrt{S+B}}$, as designed by the $\frac{S}{\sqrt{S+B}}$ selection procedure. The result of signal variable p^*_μ after the selection is illustrated in Figure 4.37b. This region continues to be dominated by $q\bar{q}$ events, demonstrating the need for selections focussed on continuum suppression.

Table 4.8. Yields and Efficiencies after $E_{\rm ECL}\in$[0,0.25) GeV.

Sample	Yield	Sel. Eff. [%]	Cuml. Eff [%]
$B^+ \to \mu^+\nu_\mu$	2.3±1.5	54.65	1.30
B^+B^-	19.4±4.4	76.63	9.47×10^{-6}
$B^0\overline{B}^0$	20.9±4.6	12.49	1.08± × 10^{-5}
$c\bar{c}$	38.3±6.2	3.77	7.98±10^{-6}
$u\bar{u}$	42.3±6.5	5.28	7.29±10^{-6}
$d\bar{d}$	10.8±3.3	4.60	7.48±10^{-6}
$s\bar{s}$	11.9±3.5	6.18	8.62±10^{-6}
tot. bkg.	143.7±12.0	5.34	8.25×10^{-6}
$\frac{S}{\sqrt{S+B}}$	0.18	-	

Table 4.8 contains the resulting yields and efficiencies of the selection $E_{\rm ECL} < 0.25$ GeV. All background samples are now $\mathcal{O}(10)$, rather than $\mathcal{O}(10^2)$ as they were previously. The $B^0\bar{B}^0$ decays have been most affected of the backgrounds with 12% reduction. The selection has also removed more than 50% the signal events. The majority of such events correspond to those in the secondary $E_{\rm ECL}\in$[0.25,0.5) GeV bin. With all events that leave charged tracks in the ROE removed, such remaining energy in signal MC would be the result of radiative events in the tag-side B, or B decays involving photons that have been misidentified as semileptonic. Due to the lower bound on p^*_μ signal region, it is unlikely that such photons are associated with the muon.

Second Selection: $B_{\rm tag}$ cos$\theta_{T_{B_{\rm tag}}T_{\rm ROE}}$ $\in$[0,0.75)

With the $E_{\rm ECL}$ selection added to the signal region where background events continue to dominate, additional selections are explored. The $\frac{S}{\sqrt{S+B}}$ method is applied again, yielding the potential selections in Table 4.9. In this round of $\frac{S}{\sqrt{S+B}}$ evaluation, continuum

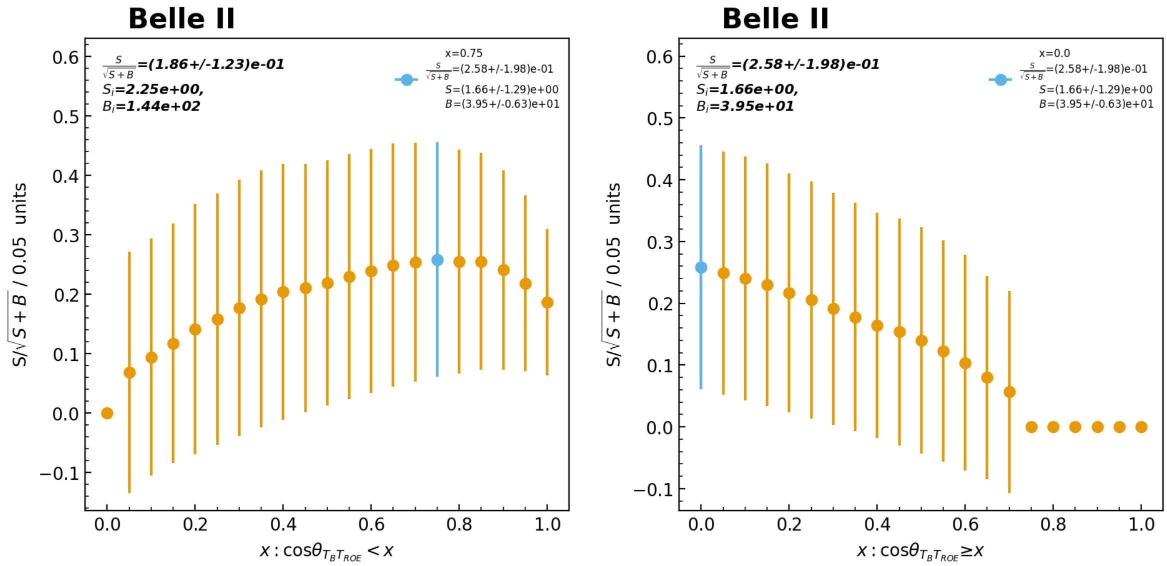
Table 4.9. Potential second selection in signal region $p^*_\mu \in$[2.45,2.85) GeV/c, with previous selection $E_{\mathrm{ECL}}\in$[0,0.25) GeV, ranked by $\frac{S}{\sqrt{S+B}}$. Pre-selection values: S_i=2.3±1.5, B_i=144±12, $\mathcal{S}_i$=0.19±0.12.

Selection	$\frac{S}{\sqrt{S+B}}$	S	B
$\cos\theta_{T_B T_{\mathrm{ROE}}}$ <0.76	0.26±0.20	1.7±1.3	40±6
B_{tag} $\cos\theta_{BY}$ <1.0383	0.26±0.17	2.2±1.5	73±9
Fox-Wolfram R_2<0.424	0.25±0.19	1.8±1.3	49±7
$\log_{10}$(`sigProb`) ≥-1.2779	0.24±0.23	1.1±1.1	20±4
`sigProb` ≥0.0507	0.24±0.22	1.1±1.1	21±5
p^*_{miss}≥1.8318	0.23±0.16	2.1±1.4	83±9
$E^*_{D^{(*)}(\pi)}$<2.9418	0.23±0.16	2.1±1.4	81±9
$E^*_{D^{(*)}}$<2.9866	0.21±0.14	2.2±1.5	102±10
E^*_Y<4.4297	0.21±0.16	1.8±1.4	70±8
B_{sig} $\cos\theta_{BY}$ ≥-0.8	0.21±0.15	2.0±1.4	87±9

suppression selection criteria in $\cos\theta_{T_B T_{\mathrm{ROE}}}$ and Fox-Wolfram R_2 rank highly. Alternatives include variables concerning the B_{tag}, such as $\cos\theta_{BY}$, both linear and log-transformed signal probability, as well as the energies of the hadronic, visible and zeroth daughter systems of the B_{tag}. The signal-side $\cos\theta_{BY}$, (i.e. the cosine angle between $Y = \mu_{\mathrm{sig}}$ and the nominal B direction) has also admitted a higher $\frac{S}{\sqrt{S+B}}$. As was seen in Eq. (3.4.4), the momentum of the Y system is an input to $\cos\theta_{BY}$. It is therefore directly correlated to p^*_μ and inappropriate as a future selection.

With $\cos\theta_{T_B T_{\mathrm{ROE}}}$ providing a selection with the largest significance estimate, the range of selections trialled for $\cos\theta_{T_B T_{\mathrm{ROE}}}$ are plotted in Figure 4.38. With a step-size of 0.05, the selection becomes $\cos\theta_{T_B T_{\mathrm{ROE}}} < 0.75$, instead of 0.76 from the default 50 bin search. There is no minimum bound applied, as suggested by Figure 4.38b. The maximum on the Figure-of-Merit is not as obvious as in the previous versions, nor is it significant when considering the propagated Poisson error in $\frac{S}{\sqrt{S+B}}$ per selection made. However, the selection is still admitted based on its highest central value of $\frac{S}{\sqrt{S+B}}$ and its reduction in the number of background from 144 events to 40 events.

The 2D histograms of $\cos\theta_{T_B T_{\mathrm{ROE}}}$ against signal variable p^*_μ and first selection variable E_{ECL} are provided in Figure 4.39. They also do not show underlying correlations between the proposed selection and the previously identified variables. It can be seen that changing $\cos\theta_{T_B T_{\mathrm{ROE}}}$ (i.e. moving along x-axis) does not affected the density of events when the



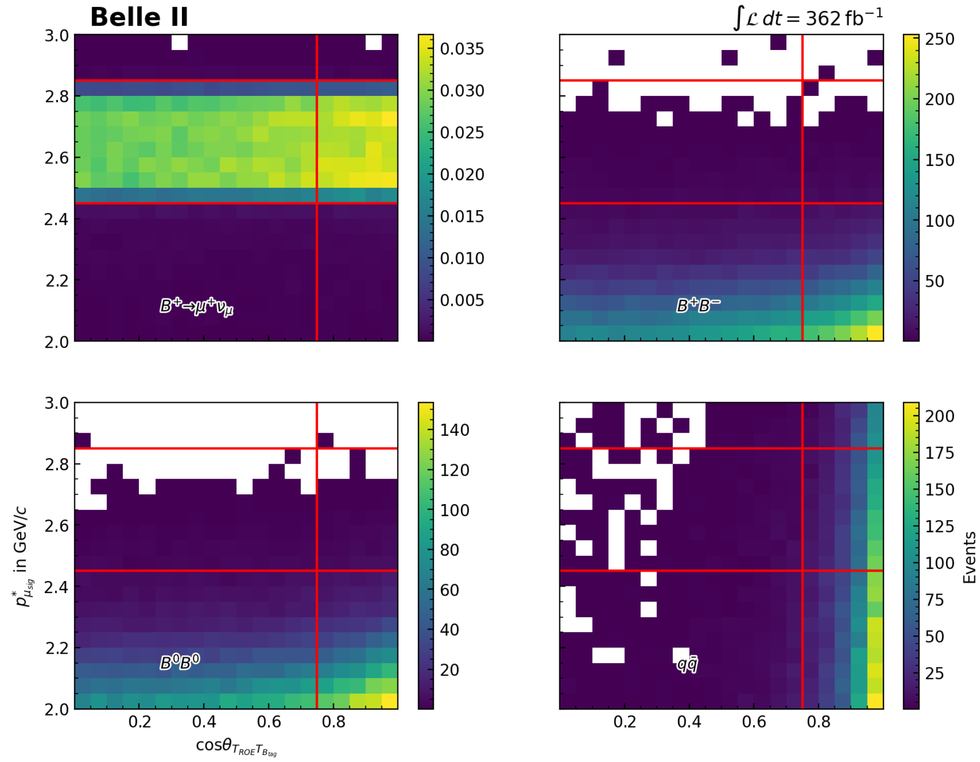
(a) Proposed: $\cos\theta_{T_B T_{\mathrm{ROE}}} < 0.75$. **(b)** Opposed bounds: $\cos\theta_{T_B T_{\mathrm{ROE}}} \geq 0$.

Figure 4.38. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for selections in cosine angle between the thrust axis of the B_{tag} and the ROE. $\cos\theta_{T_B T_{\mathrm{ROE}}}$ is varied in 0.05 increments in [0, 1] interval. Error bars are propagated from Poisson statistical error in the MC components.

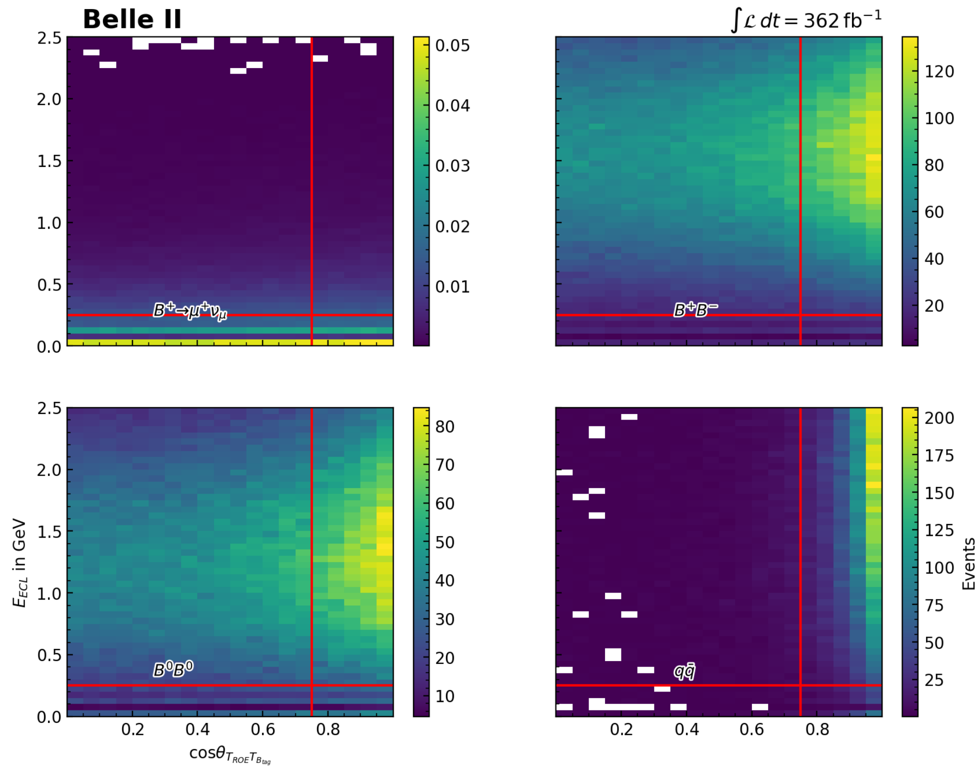
p^*_μ and E_{ECL} bin is fixed. That is to say, $\cos\theta_{T_B T_{\mathrm{ROE}}}$ is completely independent from p^*_μ and E_{ECL} in signal MC.

Figure 4.40a contains the distribution of $\cos\theta_{T_B T_{\mathrm{ROE}}}$ for tagged-B mesons used in $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction, with the proposed $\cos\theta_{T_B T_{\mathrm{ROE}}}$ selection indicated. In the discussion of Figure 4.30, $\cos\theta_{T_B T_{\mathrm{ROE}}}$ was first introduced as characterising events $q\bar{q}$ events closer to 1, whereas both signal and background $B\overline{B}$ events will have a uniform spread of $\cos\theta_{T_B T_{\mathrm{ROE}}}$. The selection $\cos\theta_{T_B T_{\mathrm{ROE}}} < 0.75$ will thus remove the large number of the $q\bar{q}$ events that was seen in the signal region of p^*_μ in Figure 4.37b. The variable $\cos\theta_{T_B T_{\mathrm{ROE}}}$ has also sequestered all $d\bar{d}$ events that remained in the signal region into its most extreme bin, with $e^+e^- \to d\bar{d}$ being 'jetty', small-mass events. In the subplot to this distribution, $\frac{S}{\sqrt{S+B}}$ per bin is plotted. This same bin also corresponds to the lowest relative $\frac{S}{\sqrt{S+B}}$, motivating its omission.

This continuum suppression of the selection is demonstrated in Figure 4.37b, where mostly $B\overline{B}$ events remain in the p^*_μ signal region. Despite $\frac{S}{\sqrt{S+B}}$ optimisation (and thus $q\bar{q}$ rejection) takes place in the signal region of p^*_μ, there is also a reduction in the remaining $c\bar{c}$ and $u\bar{u}$ events in the $p^*_\mu \in$[2.05, 2.45) GeV/c sideband. With this reduction in $q\bar{q}$ events, fractional events of the signal component can be seen stacked on top of

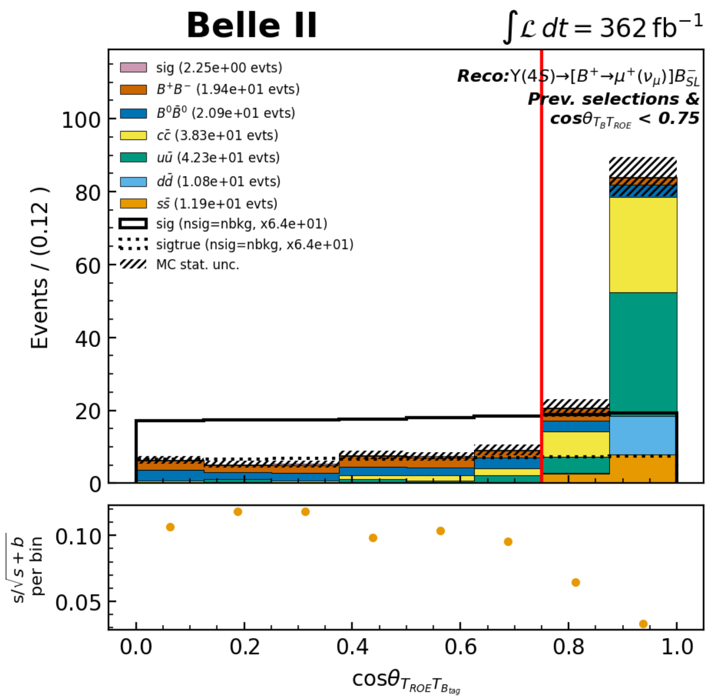


(a) $\cos\theta_{T_B T_{\mathrm{ROE}}}$ vs p^*_μ.

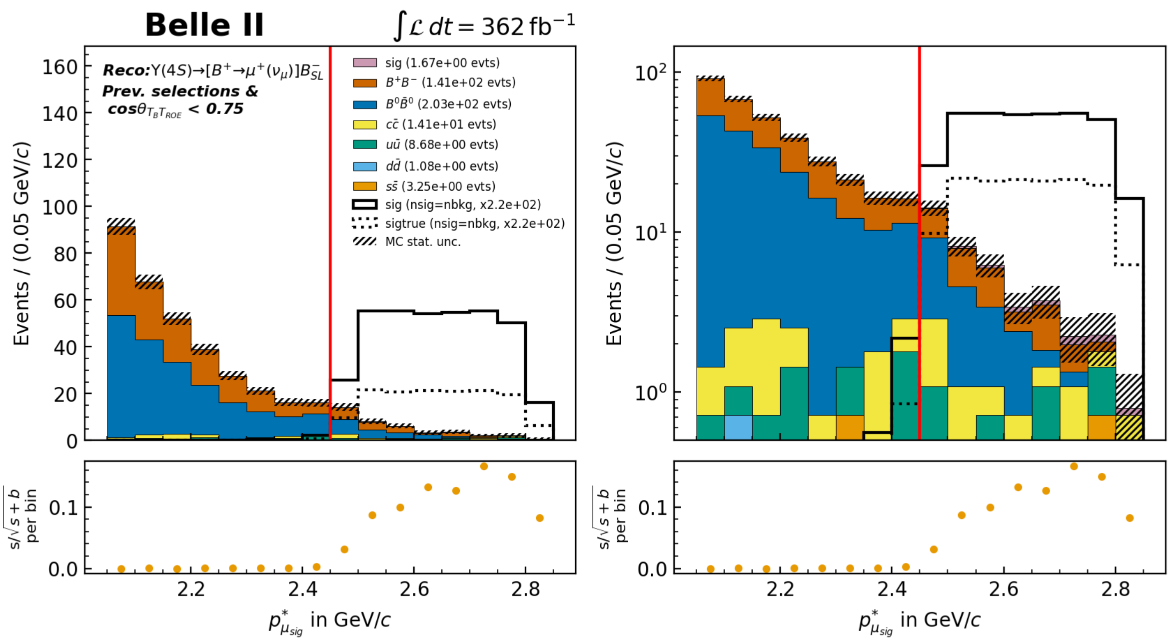


(b) $\cos\theta_{T_B T_{\mathrm{ROE}}}$ vs E_{ECL}.

Figure 4.39. 2D histograms of $\cos\theta_{T_B T_{\mathrm{ROE}}}$ against signal variable p^*_μ and first selection variable E_{ECL}, in different MC types. *Top left:* signal MC, *top right:* B^+B^- MC, *bottom left*: $B^0\overline{B}^0$, *bottom right*: $q\overline{q}$ for $q \in = \{u,\ d,\ s,\ c\}$. Proposed selections indicated by red lines. Only pre-selections applied, so that there are sufficient events to visualise correlations.



(a) Proposed $\cos\theta_{T_B T_{\mathrm{ROE}}}$ selection.



(b) Linear and log-scale distributions of $\Upsilon(4S)$ frame muon momentum.

Figure 4.40. Histograms of events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$, after $\cos\theta_{T_B T_{\mathrm{ROE}}}$ selection. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates proposed selections.

background in the p^*_μ signal region of the log-event scaled distribution. It is noted that unfortunately, their magnitude is much smaller than the statistical error in backgrounds in this region. Further reduction of the background can improve the relative magnitude between signal events and statistical error in the total bin.

Table 4.10. Yields and Efficiencies after $\cos\theta_{T_B T_{\mathrm{ROE}}} \in$[0, 0.75) GeV.

Sample	Yield	Sel. Eff. [%]	Cuml. Eff. [%]
$B^+ \to \mu^+\nu_\mu$	1.7±1.3	73.35	0.95
B^+B^-	14.2±4	73.33	6.95×10^{-6}
$B^0\overline{B}^0$	14.5±4	69.14	7.48×10^{-6}
$c\overline{c}$	5.4±2.3	14.15	1.13×10^{-6}
$u\overline{u}$	4.0±2.0	9.40	6.85×10^{-7}
$d\overline{d}$	0	0	0
$s\overline{s}$	1.4±1.2	12.12	1.04×10^{-6}
tot. bkg.	39.5±6.3	27.50	2.27×10^{-6}
$\frac{S}{\sqrt{S+B}}$	0.26	-	-

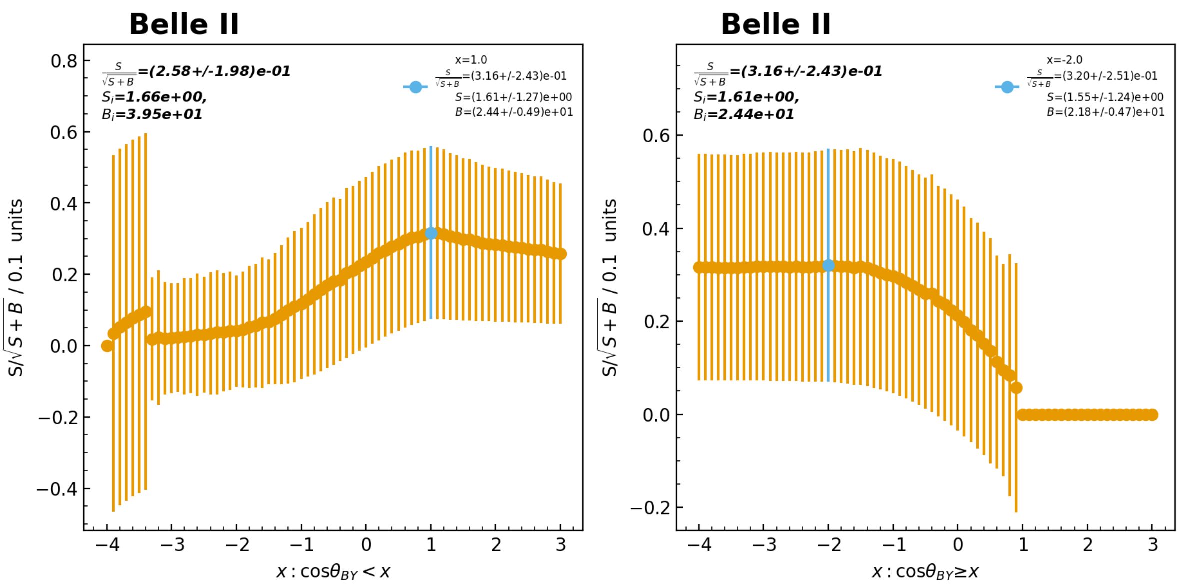
Table 4.10 summarises the yields and efficiencies in each MC component after adding the $\cos\theta_{T_B T_{\mathrm{ROE}}}$ selection. 1% of signal events remain, out of the number expected at 362 fb^{-1}. With the continuum suppressing selection criteria applied, the number of $q\overline{q}$ events in the signal region have dropped to less than half that of the $B\overline{B}$ background events left.

Third Selection: B_{tag} $\cos\theta_{BY}$ $\in$[-2, 1)

The results of trialling a third selection criterion decided by optimising $\frac{S}{\sqrt{S+B}}$ is presented in Table 4.11. With $\cos\theta_{T_B T_{\mathrm{ROE}}}$ selection implemented, there are no longer any continuum suppression variables ranking high in $\frac{S}{\sqrt{S+B}}$. The variables for this round either concern the B_{tag}, or are directly correlated to p^*_μ, like $p^*_{T,\mu}$, the transverse muon momentum in the $\Upsilon(4S)$ frame. The signal variable itself shares second largest $\frac{S}{\sqrt{S+B}}$, with the signal-side $\cos\theta_{BY}$ and muon energy, proposing a selection of the same subset of events. A tag-side B_{tag} pre-selection was proposed when investigating pre-selection criteria to reconstruction in the discussion of Figure 4.25. It is chosen to add an upper bound on the $\cos\theta_{BY}$ of the tagged-B meson, as it also yields the highest $\frac{S}{\sqrt{S+B}}$.

Table 4.11. Potential third selection in signal region $p^*_\mu \in$[2.45, 2.85) GeV/c, with previous selections $E_{\mathrm{ECL}}\in$[0, 0.25) GeV and cos$\theta_{T_{B_{\mathrm{tag}}}T_{\mathrm{ROE}}} \in$[0, 0.75), ranked by $\frac{S}{\sqrt{S+B}}$. Pre-selection values: S_i=1.7±1.3, B_i=40±6, $\mathcal{S}_i$=0.26±0.20.

Selection	$\frac{S}{\sqrt{S+B}}$	S	B
B_{tag} $\cos\theta_{BY}$ <1.0726	0.32±0.24	1.6±1.3	25±5
B_{sig} $\cos\theta_{BY}$ ≥-0.6148	0.30±0.25	1.4±1.2	19±4
p^*_μ ≥2.538	0.30±0.25	1.4±1.2	19±4
E^*_μ≥2.5402	0.30±0.25	1.4±1.2	19±4
$E^*_{D^{(*)}(\pi)}$<2.9865	0.29±0.23	1.5±1.2	27±5
$p_{T,\mu}$≥0.8739	0.28±0.23	1.4±1.2	23±5
B_{tag} Decay Mode ID<3.08	0.28±0.22	1.6±1.3	30±5
$p^*_{T,\mu}$≥0.8766	0.28±0.23	1.4±1.2	23±5
p^*_Y<1.7865	0.28±0.22	1.6±1.3	31±6
`sigProb` ≥0.0156	0.27±0.24	1.3±1.1	20±4
$\log_{10}$(`sigProb`) ≥-1.7922	0.27±0.24	1.3±1.1	20±4



(a) Proposed: $\cos\theta_{BY}$ < 1. **(b)** Opposite bound: $\cos\theta_{BY}$ ≥ -2.

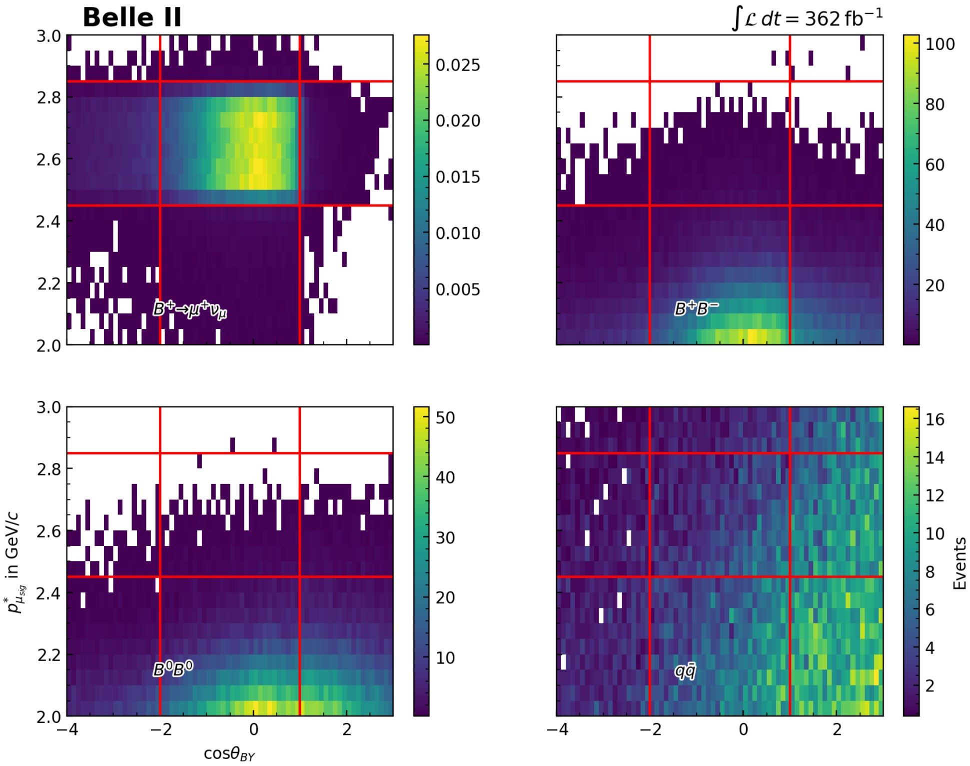
Figure 4.41. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for selections in the cosine angle between the visible B_{tag} products and the nominal B direction. $\cos\theta_{BY}$ is varied 0.01 increments between [-4, 3]. Error bars are propagated from Poisson statistical error in the MC components.

The Figures-of-Merit for B_{tag} $\cos\theta_{BY}$ are presented in Figure 4.41. Here, $\cos\theta_{BY}$ upper bounds are trialled in steps of 0.1 between -4 and 3. This yields a maximum for the selection $\cos\theta_{BY} < 1$, differing from the default binning selection with little change in the significance estimate. The trial for a consequent lower bound suggests $\cos\theta_{BY} \geq$ -2. The associated Figure-of-Merit for the lower bound shows no discernible peak at this value, suggesting that implementing a lower bound in this region will neither improve nor worsen $\frac{S}{\sqrt{S+B}}$. Despite this, the lower bound is added to $\cos\theta_{BY}$ as the region is known to be consistent with misreconstructed B_{tag}.

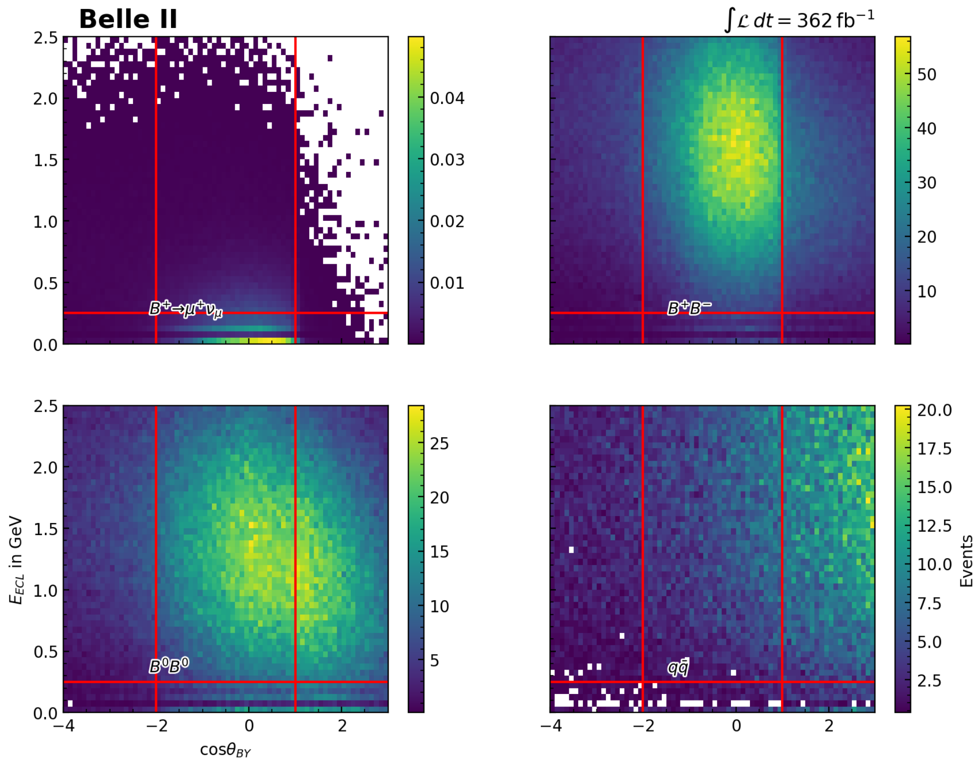
The 2D histograms in Figure 4.42 continue to show no correlation between $\cos\theta_{BY}$ previous signal and selection variables. The histograms do however provide a deeper understanding of the regions of $\cos\theta_{BY}$. It can be seen most clearly Figure 4.42a that charged B_{tag} occur mostly within the [-1, 1] region, with the hotspot of $B^0\overline{B}^0$ events spilling passed $\cos\theta_{BY}$ =1 and the highest density of $q\overline{q}$ events at $\cos\theta_{BY}$ =3.

Figure 4.43a presents the proposed $\cos\theta_{BY}$ event selection, of the $\Upsilon(4S)$ reconstructed events in the signal region of p^*_μ with all selection criteria decided so far. The upper tail of $\cos\theta_{BY}$ is removed by this selection, consisting of $q\overline{q}$ events and more $B^0\overline{B}^0$ than B^+B^-. There are lesser events lost in the lower tail of $\cos\theta_{BY}$, previously mentioned to be where B_{tag} with missing particles (such as from losing the π in $D^* \to D\pi$ reconstruction in B_{tag} processes). This potentially explains why the $\cos\theta_{BY}$ $\in$[-2, -1] region of the tail has been retained by the $\frac{S}{\sqrt{S+B}}$ procedure, that events with real D^*, such as the signal, $B\overline{B}$ and $c\overline{c}$ components reside here, as well as it being the third largest bin for $B^+ \to \mu^+\nu_\mu$ events. After this selection, the signal shape closely matches that of the background, suggesting this variable is now fully 'sculpted' i.e. background events in this region are very similar to true $B^+ \to \mu^+\nu_\mu$ events. The resultant histogram of p^*_μ for these events in $\cos\theta_{BY} \in$ [-2, 1) is presented in Figure 4.43b. For p^*_μ $\in$[2.45, 2.85) GeV/c, there are only fractional $c\overline{c}$ and $u\overline{u}$ events visible on the log event scale, leaving mostly B^+B^- with some $B^0\overline{B}^0$ backgrounds. The p^*_μ bin at p^*_μ =2.8 GeV/c is now has less than 0.5 events. The $\frac{S}{\sqrt{S+B}}$ per bin graph below indicates this to be the bin with highest estimated significance, suggesting there are fractional signal events present here. With higher luminosity, this could be a bin where sufficient signal occurs to yield a statistically significant result.

Table 4.12 contains the yields in p^*_μ signal region after previous selections, inclusive of the selection criteria on $\cos\theta_{BY}$. The light $q\overline{q}$ contributions have been reduced to fractional contributions. Meanwhile both the signal and background events with charged

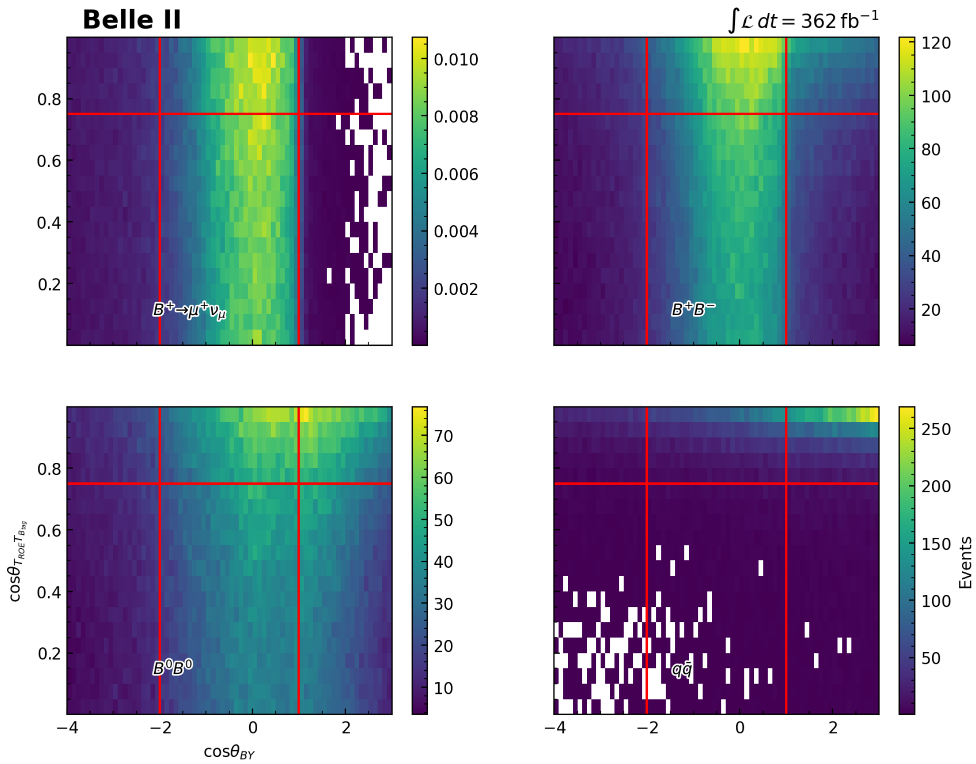


(a) $\cos\theta_{BY}$ vs p^*_μ.



(b) $\cos\theta_{BY}$ vs E_{ECL}.

Figure 4.42. 2D histograms of $\cos\theta_{BY}$ against signal variable p^*_μ and prior determined selection variables, in different MC types. *Top left:* signal MC, *top right:* B^+B^- MC, *bottom left*: $B^0\overline{B}^0$, *bottom right*: $q\overline{q}$ for $q \in=\{u, d, s, c\}$. Proposed selections indicated by red lines. Only pre-selections applied, so that there are sufficient events to visualise correlations.

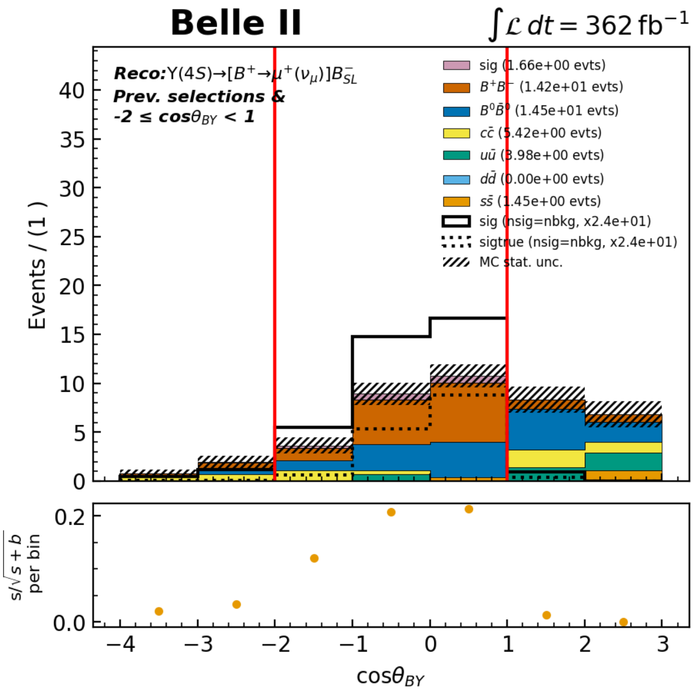


(c) $\cos\theta_{BY}$ vs $\cos\theta_{T_B T_{\mathrm{ROE}}}$.

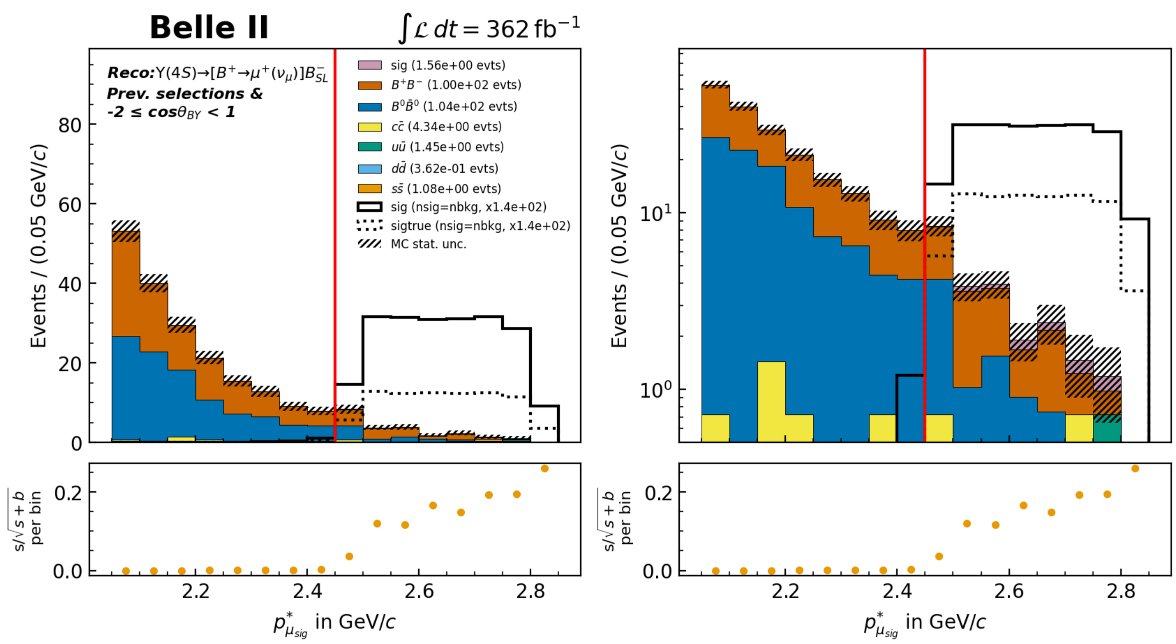
Figure 4.42. *(continued).*

Table 4.12. Yields and Efficiencies after $\cos\theta_{BY} \in$ [-2, 1).

Sample	Yield	Sel. Eff [%]	Cuml. Eff [%]
$B^+ \to \mu^+\nu_\mu$	1.5±1.2	93.45	0.89
B^+B^-	11.9±3.4	83.64	5.81×10^{-6}
$B^0\overline{B}^0$	7.4±2.7	50.89	3.81×10^{-6}
$c\overline{c}$	1.4±1.2	26.67	3.01×10^{-7}
$u\overline{u}$	0.7±0.9	18.18	1.25×10^{-7}
$d\overline{d}$	0	0	0
$s\overline{s}$	0.4±0.6	25.00	2.61×10^{-7}
tot. bkg.	21.8±4.7	55.10	1.25×10^{-6}
$\frac{S}{\sqrt{S+B}}$	0.32	-	-



(a) Proposed $\cos\theta_{BY}$ selections.



(b) Linear and log-scale distributions of $\Upsilon(4S)$ frame muon momentum.

Figure 4.43. Histograms of events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\rm SL}$, after $B_{\rm tag}$ $\cos\theta_{BY}$ selection. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates proposed selection.

B mesons largely remain, as both samples will contain real charged B_{tag} and thus physical $\cos\theta_{BY}$.

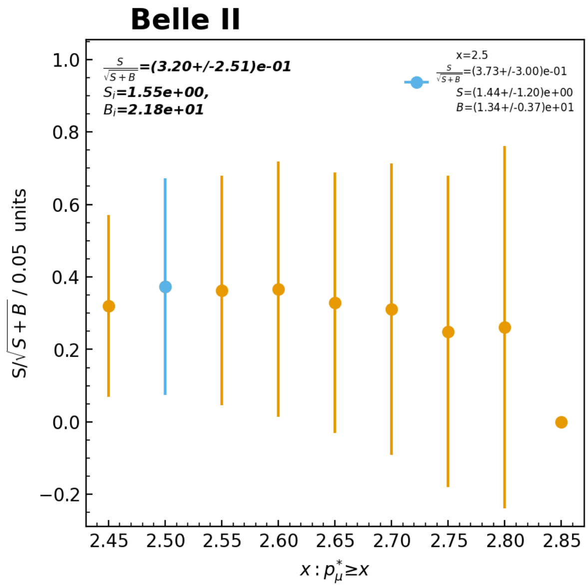
Redefining the signal region with a fourth selection$p_\mu^* \in$[2.5, 2.85) GeV/c

Table 4.13. Potential fourth selection signal region $p_\mu^* \in$[2.45,2.85) GeV/c, with previous selections $E_{\mathrm{ECL}}\in$[0,0.25) GeV, $\cos\theta_{T_{B_{\mathrm{tag}}}T_{\mathrm{ROE}}}$ $\in$[0,0.75), $\cos\theta_{B_{\mathrm{tag}}Y}$ $\in$[-2,1), ranked by $\frac{S}{\sqrt{S+B}}$,
Pre-selection values: S_i=1.5±1.2, B_i=22±5, $\mathcal{S}_i$=0.32±0.25.

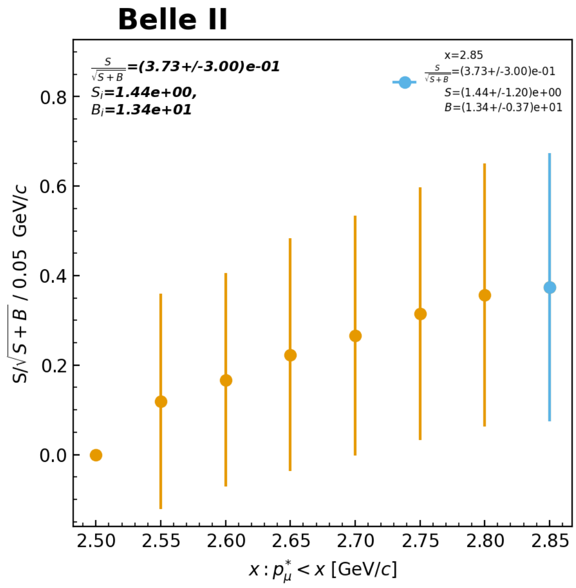
Selection	ssb	nsig	nbkg
$p_\mu^* \geq$2.506	0.37±0.30	1.4±1.2	13±4
$E_\mu^*\geq$2.5083	0.37±0.30	1.4±1.2	13±4
$E^*_{D^{(*)}\pi}$<3.7204	0.33±0.26	1.5±1.2	19±4
$p_{T,\mu}\geq$0.8791	0.33±0.28	1.3±1.1	14±4
$p^*_{T,\mu}\geq$0.5937	0.33±0.27	1.5±1.2	18±4
B_{tag} Decay Mode ID<3.08	0.33±0.26	1.5±1.2	18±4
$\theta_{\vec{p}_{\mathrm{miss}}}\geq$0.3562	0.33±0.26	1.5±1.2	19±4
$p_\mu^{\mathrm{lab}}\geq$2.0386	0.33±0.26	1.5±1.2	20±4
$\theta_\mu^{\mathrm{lab}}$<2.424	0.33±0.26	1.5±1.2	19±4
$\cos\theta_{\vec{p}_{\mathrm{miss}}}$<0.92	0.33±0.26	1.5±1.2	18±4

The results of exploring further selection criteria to optimise $\frac{S}{\sqrt{S+B}}$ in the p_μ^* signal region is presented in Table 4.13. The selection with the largest $\frac{S}{\sqrt{S+B}}$ motivates an increase to the lower bound on the p_μ^* signal region. Choosing alternate selections is not worthwhile, given that all other selection criteria listed (aside from the muon centre-of-momentum energy) have smaller $\frac{S}{\sqrt{S+B}}$ and are identical up to the second decimal place.

Increasing the lower bound on the p_μ^* signal region has been considered in intervals of 50 MeV/c, between 2.45 and 2.85 GeV/c. From the Figures-of-Merit plots in Figure 4.44, increasing the lower bound to $p_\mu^* \geq$2.5 GeV/c is implied. Visually, this is a minimal increase in $\frac{S}{\sqrt{S+B}}$ from $p_\mu^* \geq 2.45$ GeV/c, that overlaps the original bound within statistical error. However, the analysis will proceed with the change, based on the increase on the central value. As expected with the isolated signal events present in the large p_μ^* bins, no change to the p_μ^* upper bound is suggested. As the signal region has decreased from 40 MeV/c in width to 35 MeV/c, for symmetries sake the lower bound is increase in the sideband as well. The signal region is now $p_\mu^* \in$[2.5, 2.85) GeV/c and the sideband is now $p_\mu^* \in$[2.15, 2.5) GeV/c.



(a) Proposed: $p^*_\mu \geq 2.5$ GeV/c.



(b) Opposite bound: $p^*_\mu < 2.85$ GeV/c.

Figure 4.44. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for an additional selection on signal muon momentum in $\Upsilon(4S)$ frame. p^*_μ is varied in 0.05 GeV/c increments between [2.45, 2.85] GeV/c. Error bars are propagated from Poisson statistical error in the MC components.

The effect of redefining the signal region and sidebands of variable p^*_μ is presented in Figure 4.45. While the shape of the overlaid signal shape indicates the 2.45 GeV/c bin yields a non-negligible number of events relative to the other signal bins, the fractional signal contribution is much less discernible from $B\overline{B}$ background events relative to other bins.

Table 4.14 contains the yields after increasing the lower bound of the signal region. With 1.4±1.2 events in the new signal region, 0.1 signal events have been sacrificed to remove approximately 10 background events that were present in the old signal region.

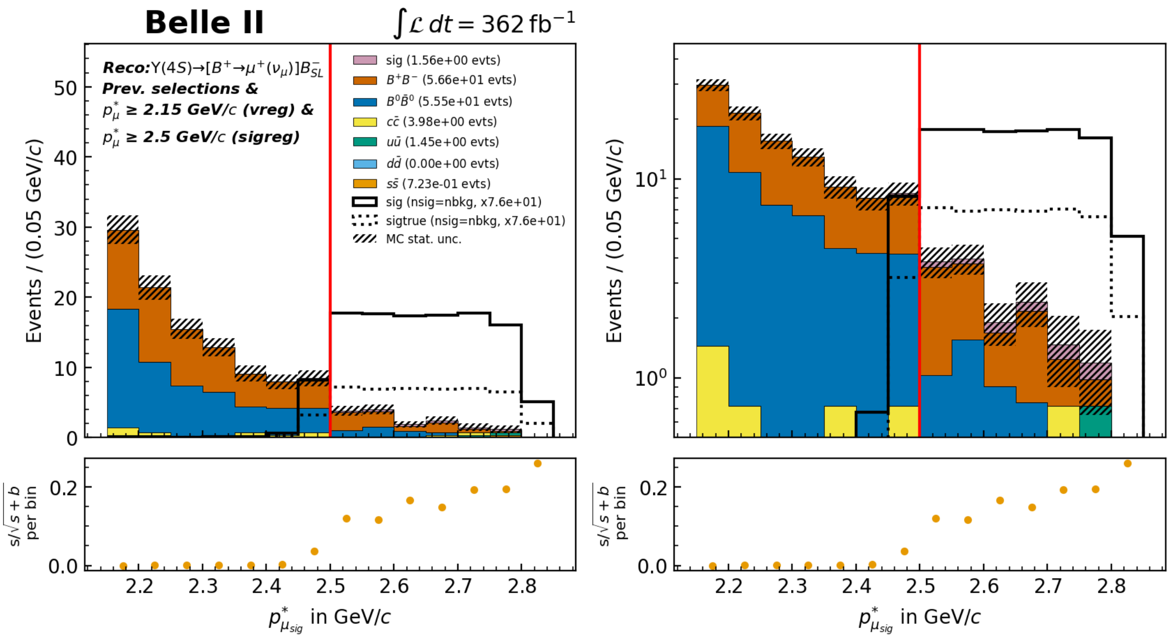


Figure 4.45. Histograms of events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$, after previous selections. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates proposed redefinition of $B^+ \to \mu^+\nu_\mu$ signal region.

Table 4.14. Yields and retentions after changing signal region to $p^*_\mu \in$[2.5, 2.85) GeV/c.

Sample	Yield	Sel. Eff. [%]	Cuml. Eff. [%]
$B^+ \to \mu^+\nu_\mu$	1.4±1.2	93.03	0.83
B^+B^-	7.7±2.8	65.22	3.79×10^{-7}
$B^0\overline{B}^0$	3.9±2.0	52.63	2.00×10^{-7}
$c\overline{c}$	0.7±0.9	50.00	1.50×10^{-8}
$u\overline{u}$	0.7±0.9	100.00	1.25×10^{-8}
$d\overline{d}$	0	0	0
$s\overline{s}$	0.4±0.6	100.00	2.61×10^{-8}
tot. bkg.	13.4±3.7	61.68	7.71×10^{-8}
$\frac{S}{\sqrt{S+B}}$	0.37	-	-

Fifth Selection: $\cos\theta_{\vec{p}_{\mathrm{miss}}} \in$[-0.9, 0.9)

The procedure is repeated to find selections that optimise $\frac{S}{\sqrt{S+B}}$, this time in the narrowed signal region. The results in Table 4.15 show that $\frac{S}{\sqrt{S+B}}$ is beginning to plateau. Moreover,

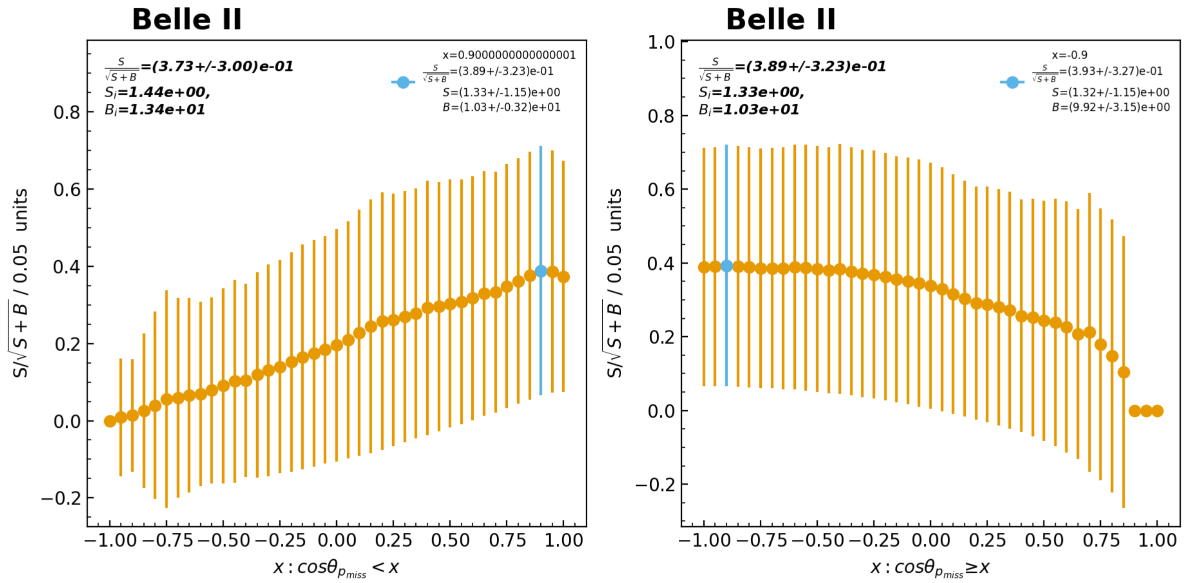
Table 4.15. Potential fifth selection, in improved signal region $p^*_\mu \in$[2.5,2.85) GeV/c, with previous selections $E_{\mathrm{ECL}}\in$[0,0.25) GeV, $\cos\theta_{T_{B_{\mathrm{tag}}}T_{\mathrm{ROE}}} \in$[0,0.75) and $\cos\theta_{B_{\mathrm{tag}}Y} \in$[-2,1), ranked by $\frac{S}{\sqrt{S+B}}$.
Pre-selection values:S_i=1.4±1.2, B_i=13±4, $\mathcal{S}_i$=0.37±0.30.

Selection	$\frac{S}{\sqrt{S+B}}$	S	B
$p_{T,\mu} \geq 0.8317$	0.40±0.34	1.2±1.1	8.5±2.9
$\theta_{\vec{p}_{\mathrm{miss}}} \geq 0.1841$	0.39±0.31	1.4±1.2	12.0±3.5
$p^*_{T,\mu} \geq 0.8717$	0.39±0.34	1.2±1.1	8.4±2.9
$\cos\theta_{\vec{p}_{\mathrm{miss}}} < 0.92$	0.39±0.32	1.4±1.2	10.8±3.3
$p_{z,Y} \geq$-0.1921	0.39±0.31	1.4±1.2	11.6±3.4
$\log_{10}\mathtt{kaonID}_{\mu_{\mathrm{sig}}}$ <-5.5416	0.39±0.32	1.4±1.2	11.1±3.3
$E^*_{D^{(*)}\pi} < 3.7275$	0.39±0.31	1.4±1.2	11.9±3.4
$p^*_{z,D^{(*)}\pi} \geq$-1.168	0.38±0.31	1.4±1.2	11.6±3.4
$\mathtt{muonID}_\mu \geq 0.988$	0.38±0.31	1.4±1.2	11.7±3.4
$\log_{10}\mathtt{muonID}_{\mu_{\mathrm{sig}}} \geq$-0.0019	0.38±0.31	1.4±1.2	11.4±3.4

the highest ranked selection is the transverse component of the muon momentum, which is inappropriate for selection. With the similarity in estimated significance in the next ranked variables that aren't related to p^*_μ, there is a choice of selection based on the angle of missing momentum and its cosine version, the longitudinal Y system momentum. An upper bound on $\cos\theta_{\vec{p}_{\mathrm{miss}}}$(i.e. the cosine polar angle of the missing momentum vector, measured in the lab frame) is chosen as it keeps a smaller number of background events.

Selections on $\cos\theta_{\vec{p}_{\mathrm{miss}}}$upper bounds are trialled in steps of 0.05, producing Figures-of-Merit in Figure 4.46. The maximum value of $\frac{S}{\sqrt{S+B}}$ is achieved for $\cos\theta_{\vec{p}_{\mathrm{miss}}} < 0.9$, a very slight increase from implementing no bound. This is consistent with what was seen in Figure 4.32, where signal events dropped suddenly in this region, while background events continued to increase. Searching for a potential lower bound results in a symmetric selection of $\cos\theta_{\vec{p}_{\mathrm{miss}}} \geq$ -0.9. Such a change is negligible but justified, as this was also previously seen to be a bin with an increased number of events.

The 2D histograms exploring potential correlation between $\cos\theta_{\vec{p}_{\mathrm{miss}}}$and the previous selection variables are produced in Figure 4.48b. Within the signal distribution of Figure 4.48b, there is a faint 'U' shaped structure within the p^*_μ signal region. With the signal muon and a missing neutrino originating from the same parent B meson, there is potential correlation between the muon's momentum and the angle at which neutrino is omitted. It is emphasised that $\cos\theta_{\vec{p}_{\mathrm{miss}}}$corresponds to the missing momentum not only

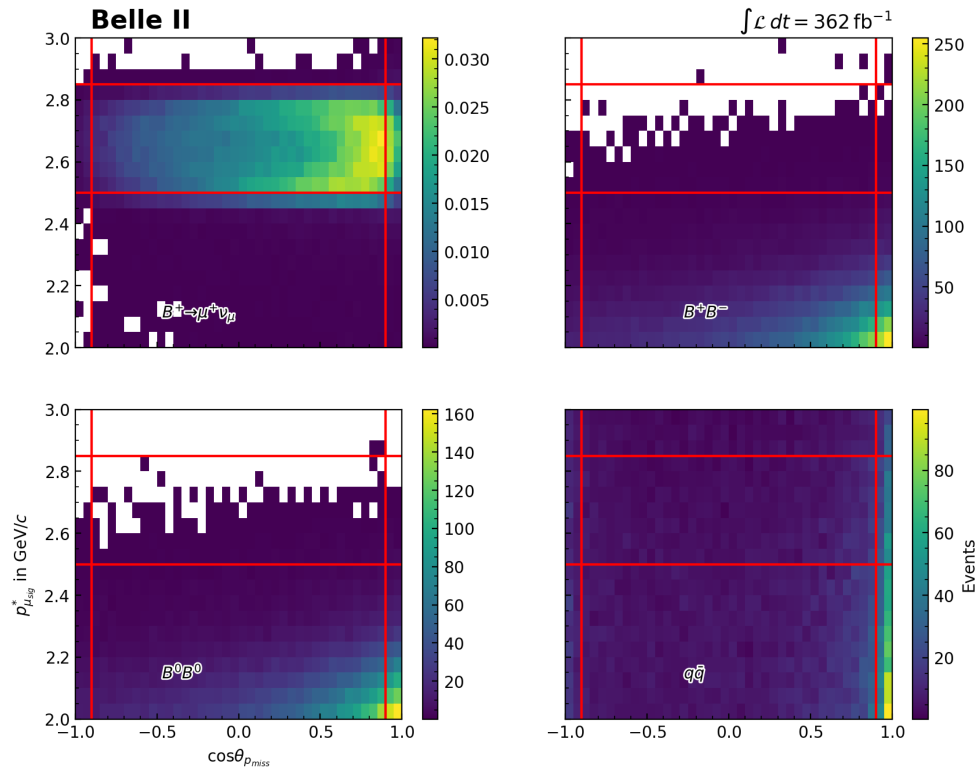


(a) Proposed: $\cos\theta_{\vec{p}_{\rm miss}} < 0.9$. **(b)** Opposite bound: $\cos\theta_{\vec{p}_{\rm miss}} \geq$ -0.9.

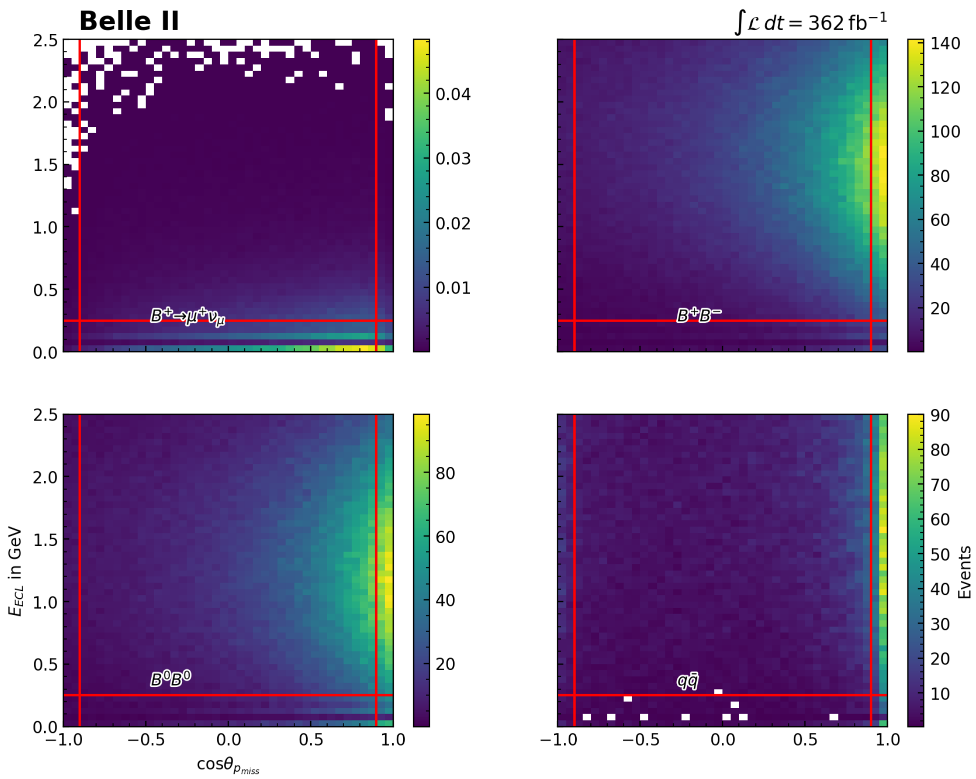
Figure 4.46. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for selections in the cosine polar angle of the missing momentum vector of the event. B direction. $\cos\theta_{\vec{p}_{\rm miss}}$ is varied 0.05 increments between 0 and 1. Error bars are propagated from Poisson statistical error in the MC components.

from the signal B, but combined with the neutrino from tag-side B as well, which will be much less correlated with the signal side muon. With rectangular selections, there is little less risk of 'shaping' the MC, especially as the correlated structure is not present in the background distributions.

The distribution in Figure 4.48a demonstrates the values of $\cos\theta_{\vec{p}_{\rm miss}}$ for the $\Upsilon(4S)$ reconstructed events in the signal region of p^*_μ with the previously detemined selections. Compared to the version without the $\frac{S}{\sqrt{S+B}}$ informed selections in Figure 4.32, there are significantly less events, though the overall trend remains. The outside bins of $\cos\theta_{\vec{p}_{\rm miss}}$ are to be omitted via the selection from the $\frac{S}{\sqrt{S+B}}$ procedure. Here, the bin at 0.9 (i.e. with total missing neutrino momentum near parallel to the beamline) contains a peak of background, including where the last $s\bar{s}$ event resides. The bin omitted at -1 removes a single non-signal charged event. The effect of removing all but a single event in continuum background is evident in Figure 4.48b, where all $d\bar{d}$ and $s\bar{s}$ are now removed completely. The signal event yields in the signal region remain within the range of the statistical uncertainties. Moreover, the p^*_μ =2.75 bin now also has the largest $\frac{S}{\sqrt{S+B}}$ with a sub-0.5 events total, evident from the log-scale.

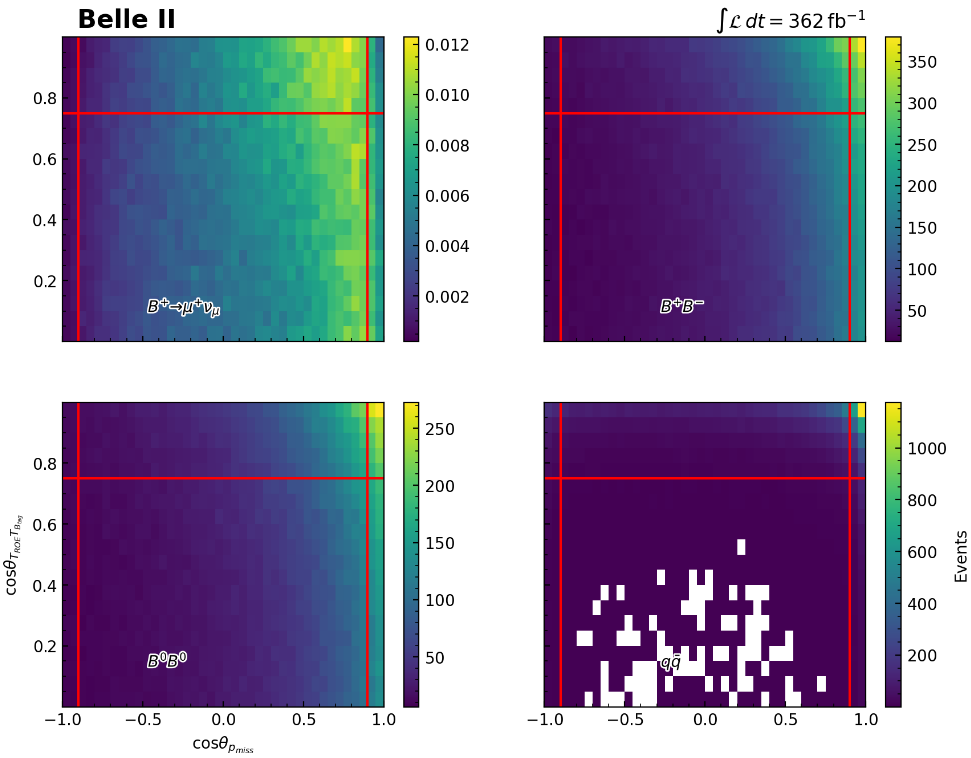


(a) $\cos\theta_{\vec{p}_\mathrm{miss}}$ vs p^*_μ.

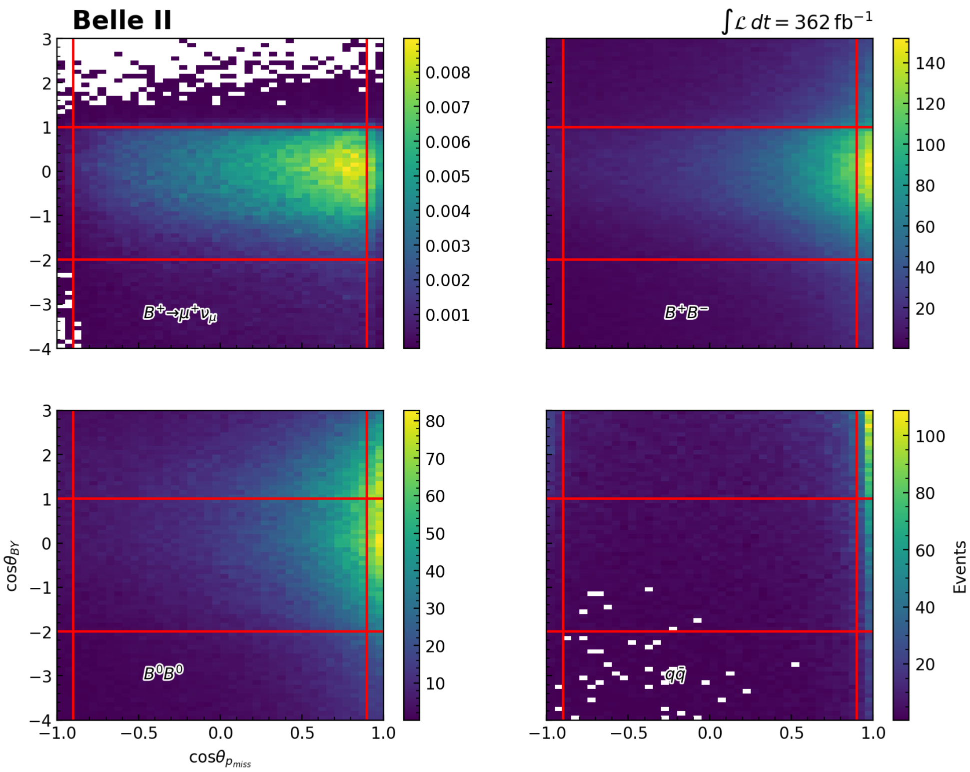


(b) $\cos\theta_{\vec{p}_\mathrm{miss}}$ vs E_ECL.

Figure 4.47. 2D histograms of $\cos\theta_{\vec{p}_\mathrm{miss}}$ against signal variable p^*_μ and prior determined selection variables, in different MC types. *Top left:* signal MC, *top right:* B^+B^- MC, *bottom left*: $B^0\overline{B}^0$, *bottom right*: $q\bar{q}$ for $q \in=\{u, d, s, c\}$. Proposed selection indicated by red lines. Only pre-selections applied, so that there are sufficient events to visualise correlations.

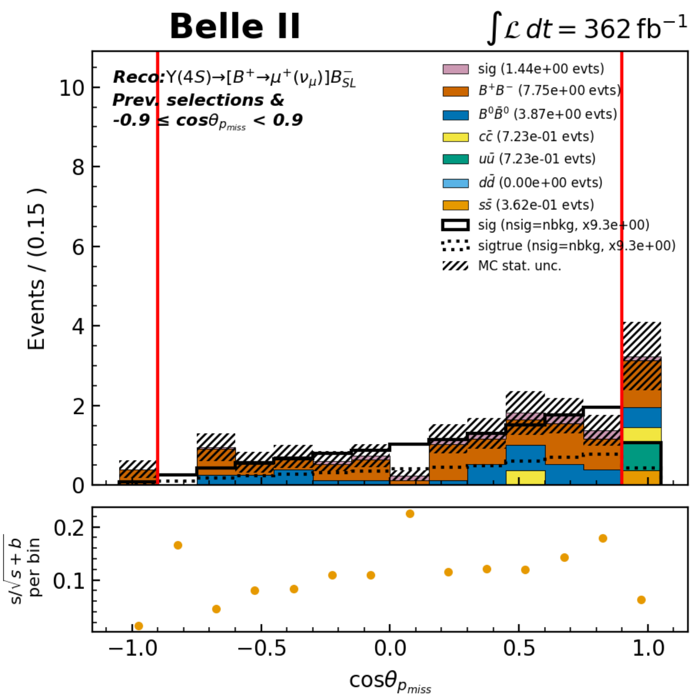


(c) $\cos\theta_{\vec{p}_{\mathrm{miss}}}$ vs $\cos\theta_{T_B T_{\mathrm{ROE}}}$.

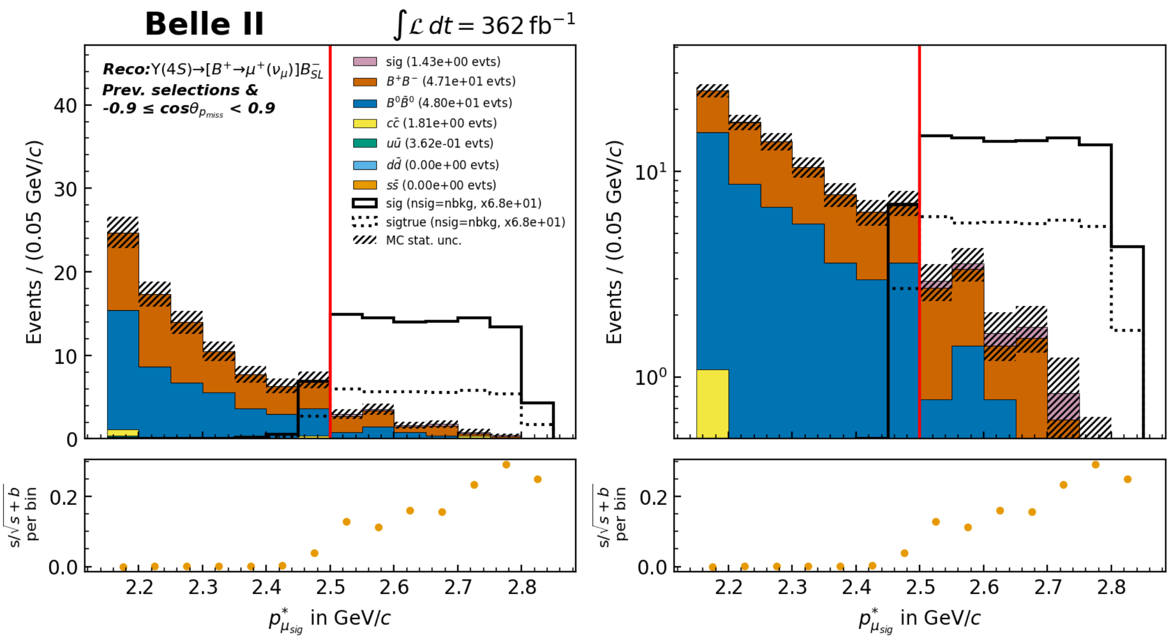


(d) $\cos\theta_{\vec{p}_{\mathrm{miss}}}$ vs $\cos\theta_{BY}$.

Figure 4.47. *(continued).*



(a) Proposed $\cos\theta_{\vec{p}_{\mathrm{miss}}}$ selection.



(b) Linear and log-scale distributions of $\Upsilon(4S)$ frame muon momentum.

Figure 4.48. Histograms of events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$, after $B_{\mathrm{tag}}\ \cos\theta_{BY}$ selection. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates proposed selection.

Table 4.16. Yields and Efficiencies after fifth selection $\cos\theta_{\vec{p}_{\text{miss}}} \in$ [-0.9, 0.9) GeV/c.

Sample	Yield	Sel. Eff. [%]	Cuml. Eff. [%]
$B^+ \to \mu^+\nu_\mu$	1.3±1.1	91.51	0.76
B^+B^-	6.2±2.5	80.00	3.03×10^{-6}
$B^0\overline{B}^0$	3.4±1.8	86.67	1.74×10^{-6}
$c\overline{c}$	0.4±0.6	50.00	7.52×10^{-7}
$u\overline{u}$	0	0	0
$d\overline{d}$	0	0	0
$s\overline{s}$	0	0	0
tot. bkg.	9.9±3.1	73.85	5.69×10^{-7}
$\frac{S}{\sqrt{S+B}}$	0.39	-	-

The number of remaining events remaining after $\cos\theta_{\vec{p}_{\text{miss}}}$ is restricted to [-0.9, 0.9) is presented in Table 4.16. There are now less than 10 background $B\overline{B}$ events in the signal region, as well as no light quark continuum.

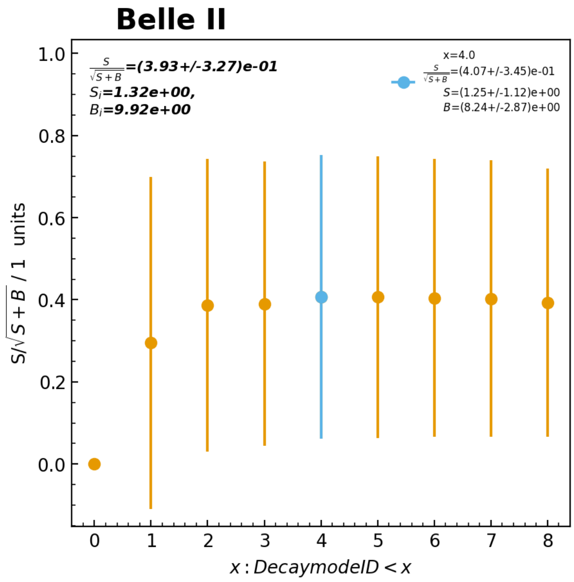
Sixth Selection: B_{tag} Decay Mode ID < 4

The results of the final application of $\frac{S}{\sqrt{S+B}}$ maximisation to reject background events in the signal region is provided in Table 4.17. Increasing the signal region bound again is ranked the highest in significance estimate, but is not pursued as it leaves less than one $B^+ \to \mu^+\nu_\mu$ event in the signal region. The next highest $\frac{S}{\sqrt{S+B}}$ is a selection on either the centre-of-momentum energy of the hadronic system within the B_{tag} or limiting the B_{tag} the decay modes. The numbers used to signify the semileptonic B_{tag} decay modes happen to be ordered according to their theoretical branching fractions, with the last four modes having identical values. These last four modes are also the four-body B_{tag} modes, each containing a pion. From the table, the removal of these modes seems to be equivalent to placing an upper bound on the hadronic energy by removing the pion.

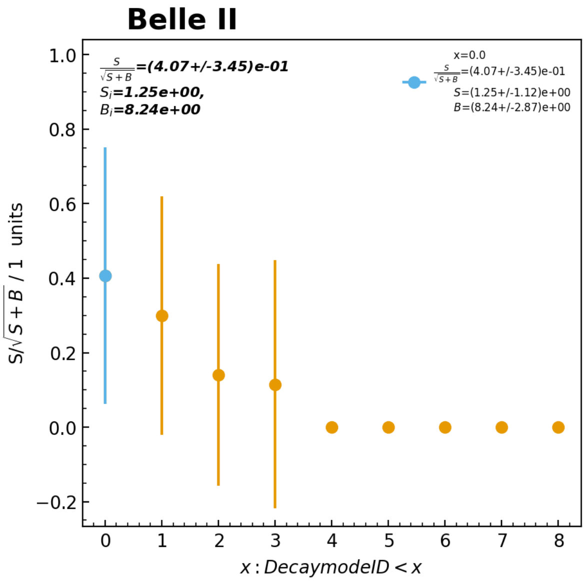
In the Figures-of-Merit in Figure 4.49, limiting the analysis to the three body semileptonic decay modes had not visibly increase the $\frac{S}{\sqrt{S+B}}$, though the selection has removed a background event in the signal region. There is no improvement in placing a lower bound on the decay mode ID, which is consistent with the large branching fraction of the $B^+ \to \overline{D}^0\ \ell^+\ \nu_\ell$ modes.

Table 4.17. Potential sixth selection with improved signal region $p^*_\mu \in$[2.5, 2.85) GeV/c, with previous selections $E_{\text{ECL}}\in$[0, 0.25) GeV, $\cos\theta_{T_{B_{\text{tag}}}T_{\text{ROE}}} \in$[0, 0.75), $\cos\theta_{B_{\text{tag}}Y} \in$[-2, 1) and $\cos\theta_{\vec{p}_{\text{miss}}}\in$[-0.9, 0.9), ranked by $\frac{S}{\sqrt{S+B}}$. Pre-selection values: S_i=1.3±1.1, B_i=10±3, $\mathcal{S}_i$=0.39±0.33.

Selection	$\frac{S}{\sqrt{S+B}}$	S	B
$p^*_\mu \geq$2.689	0.42 ±0.50	0.5±0.7	1.0±1.0
$E^*_\mu \geq$2.6911	0.42 ±0.50	0.5±0.7	1.0±1.0
$E^*_{D^{(*)}(\pi)}$<3.6115	0.41±0.34	1.3±1.1	8.6±2.9
B_{tag} Decay Mode ID<3.08	0.41±0.34	1.3±1.1	8.6±2.9
B_{tag} first daughter PDG code <16.88	0.40±0.34	1.3±1.1	9.1±3.0
$p^*_{T,\mu} \geq$0.5882	0.40±0.34	1.3±1.1	8.7±2.9
(no. of Neutral Hadrons in ROE)<4.02	0.40±0.33	1.3±1.1	9.3±3.0
$p_{T,\mu} \geq$0.451	0.40±0.33	1.3±1.1	9.3±3.0
$\cos^*\theta_{\mu_{\text{sig}}\ell_{\text{tag}}} \geq$-0.8387	0.39±0.33	1.3±1.1	9.6±3.1
Fox-Wolfram$R_2 \geq$0.0592	0.39±0.33	1.3±1.1	9.9±3.1



(a) Proposed: $B_{\rm tag}$ decay mode < 4.

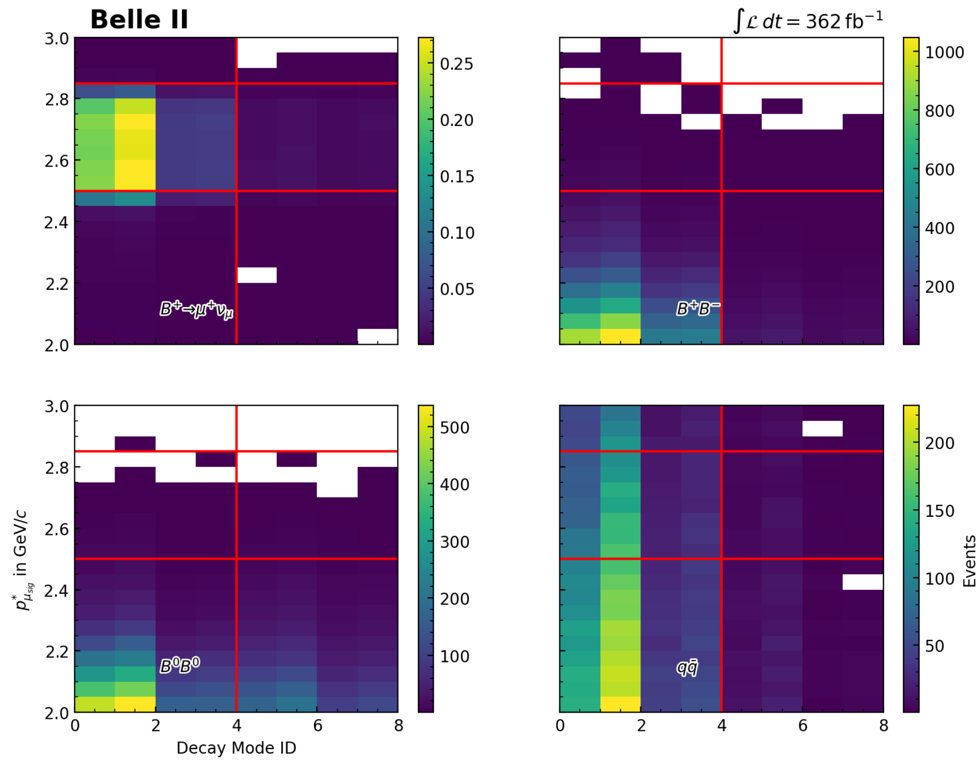


(b) Opposite bound: $B_{\rm tag}$ decay mode ≥ 0.

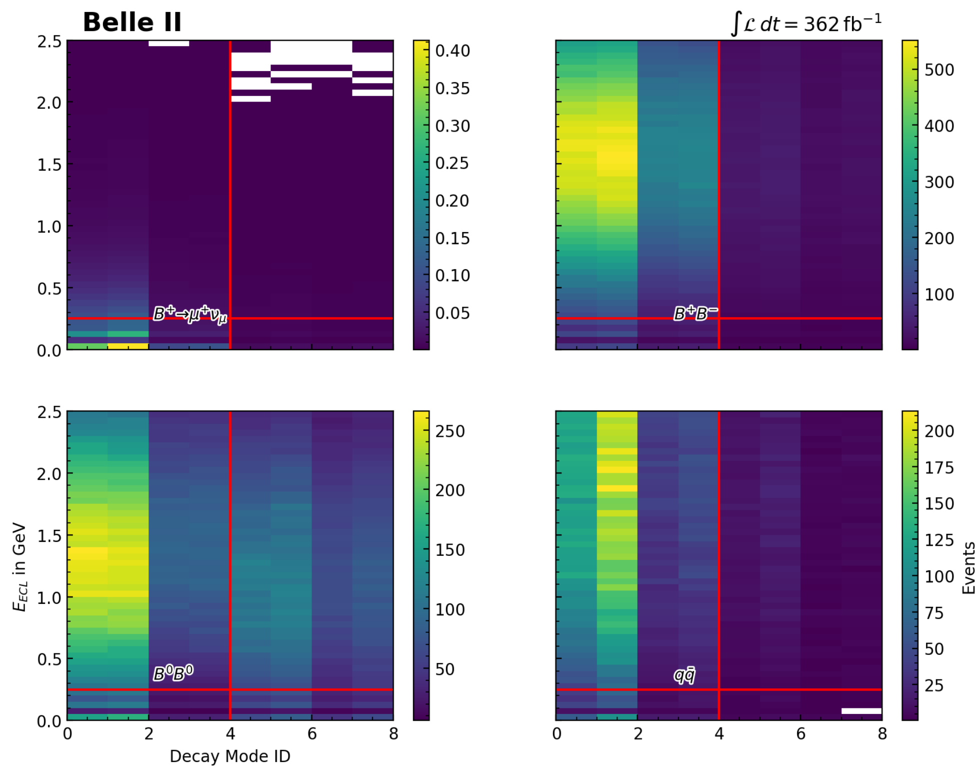
Figure 4.49. $\frac{S}{\sqrt{S+B}}$ Figures-of-Merit for selections in the $B_{\rm tag}$ decay mode. Error bars are propagated from Poisson statistical error in the MC components.

The 2D correlation histograms comparing the decay modes with each selection variable are illustrated in Figure 4.50. In this case, the correlation histograms are effectively tag maps, demonstrating how each observable varies per decay mode, similar to those used for investigating log-transformed signal probability in Figure 4.24. There are no obvious correlations, like the structure that was seen in the 2D histograms of $\cos\theta_{\vec{p}_{\rm miss}}$ and p^*_μ. However, among all MC samples, it seems that the latter four modes are most favoured by the $B^0\overline{B}^0$ modes, meaning their omission may lead to less $B^0\overline{B}^0$ events in the signal region.

Figure 4.51b contains the distribution of $B_{\rm tag}$ decay mode ID's used in $\Upsilon(4S)$ reconstruction, restricted to events in the p^*_μ signal region with currently decided selections. The four-body modes, have demonstrated a bias towards the $B^0\overline{B}^0$ events in the distribution prior to $\frac{S}{\sqrt{S+B}}$ selections in Figure 4.23. That the only four-body $B_{\rm tag}$ modes that remain continue to be mixed B modes validates their removal. Also, the signal events stacked on background components are now visible on linear scale with $B_{\rm tag}$ mode binning. With the removal of the $B^+ \to D^{(*)}\pi\ell^+\,\nu_\ell$ decay modes, the distribution of p^*_μ in the signal region can now be seen to be mostly B^+B^- background.

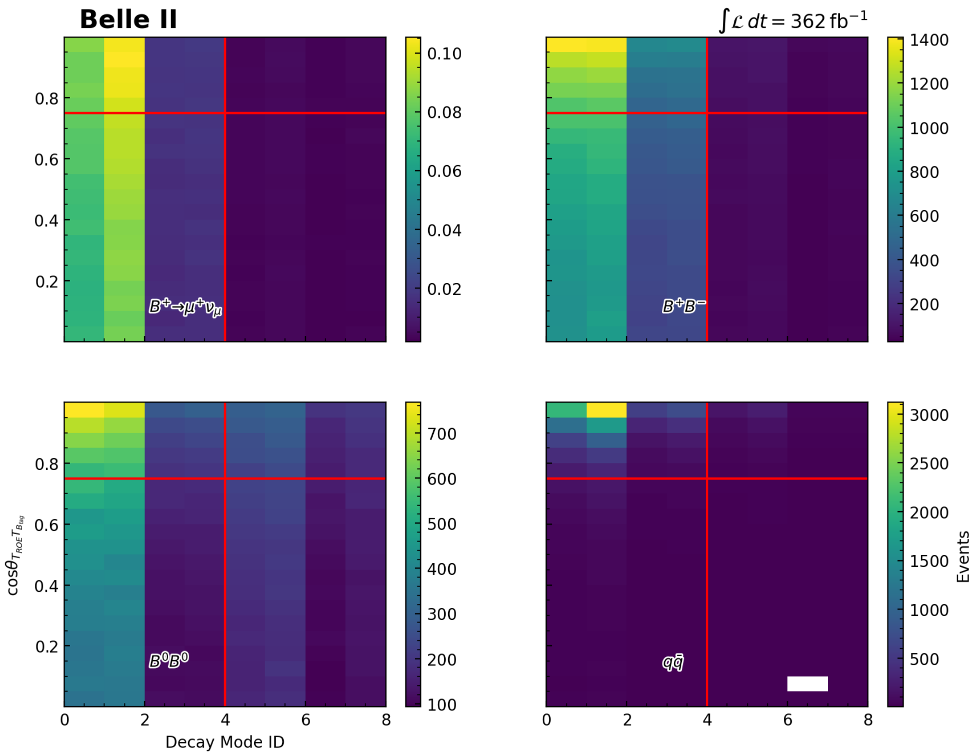


(a) B_{tag} decay mode vs p^*_μ.

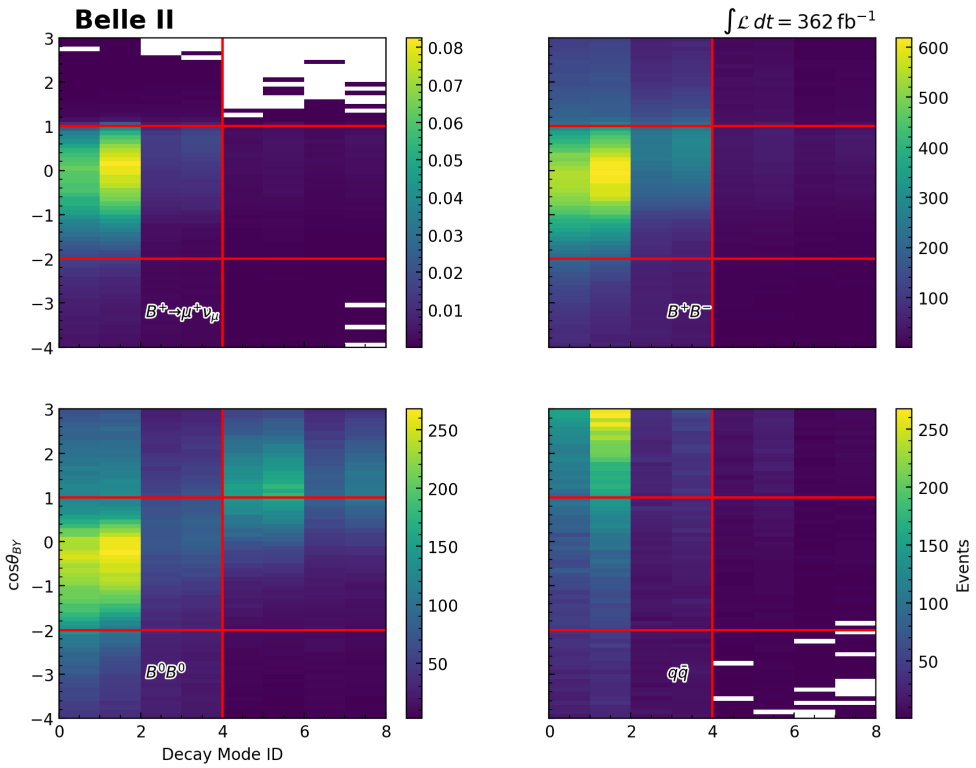


(b) B_{tag} decay mode vs E_{ECL}.

Figure 4.50. 2D histograms of B_{tag} decay mode against signal variable p^*_μ and prior determined selection variables, in different MC types. *Top left:* signal MC, *top right:* B^+B^- MC, *bottom left*: $B^0\overline{B}^0$, *bottom right*: $q\bar{q}$ for $q \in = \{u, d, s, c\}$. Proposed selections indicated by red lines. Only pre-selections applied, so that there are sufficient events to visualise correlations.

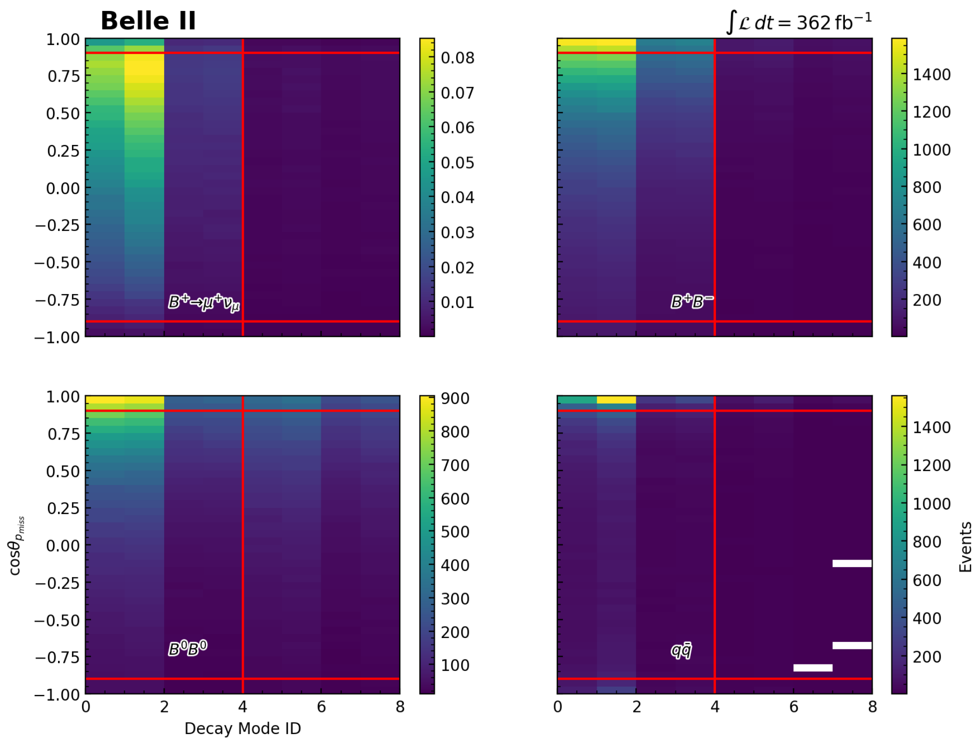


(c) B_{tag} decay mode vs $\cos\theta_{T_B T_{\mathrm{ROE}}}$.



(d) B_{tag} decay mode vs $\cos\theta_{BY}$.

Figure 4.50. *(continued).*



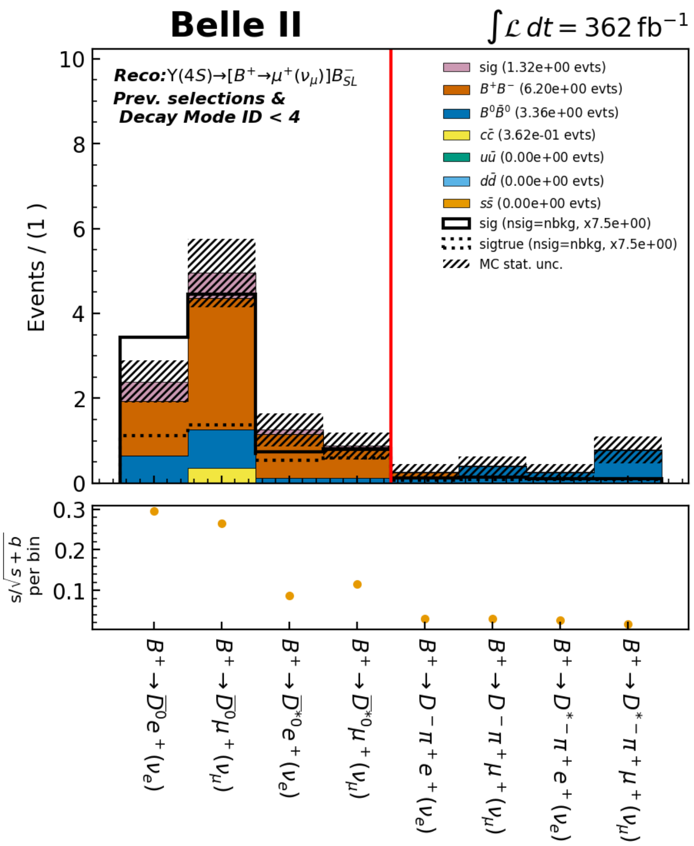
(e) B_{tag} decay mode vs $\cos\theta_{\vec{p}_{\text{miss}}}$.

Figure 4.50. *(continued).*

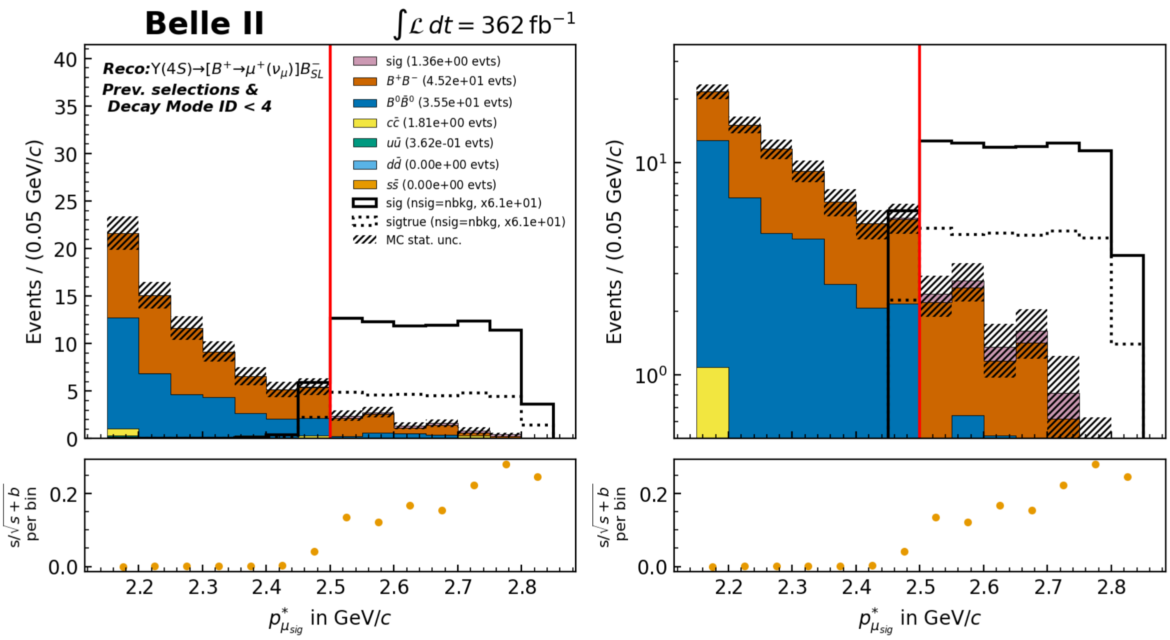
Table 4.18. Yields and Efficiencies after sixth selection criterion B_{tag} Decay Mode ID < 4.

Sample	Yield	Sel. Eff. [%]	Cuml. Eff. [%]
$B^+ \to \mu^+\nu_\mu$	1.3±1.1	95.12	0.72
B^+B^-	6.1±2.5	97.92	2.97×10^{-6}
$B^0\overline{B}^0$	1.8±1.3	53.85	9.35×10^{-7}
$c\overline{c}$	0.4±0.6	100.00	7.52×10^{-8}
$u\overline{u}$	0	0	0
$d\overline{d}$	0	0	0
$s\overline{s}$	0	0	0
tot. bkg.	8.2±2.9	83.07	4.73×10^{-7}
$\frac{S}{\sqrt{S+B}}$	0.41	-	-

Table 4.18 contain the yields after the final selection, restricting decay modes to the three-body decays. A further iteration of the $\frac{S}{\sqrt{S+B}}$ method was attempted, though they did not provide options that retained at least one scaled signal event. Thus, 1.3±1.1 signal events are left, against 8.2±2.9 background events. The signal component then



(a) Proposed B_{tag} decay mode selection.



(b) Linear and log-scale distributions of $\Upsilon(4S)$ frame muon momentum.

Figure 4.51. Histograms of events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$, after B_{tag} decay mode selection. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Red line indicates proposed selection.

represents approximately one-tenth of the events expected in this constrained parameter space. This is 0.7% of all $B^+ \to \mu^+\nu_\mu$ events at 362 fb^{-1}, and a reduction to 4.7×10^{-7}% of background events. The largest component is from non-signal B^+B^- decays.

4.2.4. Final analysis selection criteria and their effects on key variables

The final selections have been decided. Over 150 variables were trialled over multiple bounds, with the goal of obtaining selections that could optimise the number of $B^+ \to \mu^+\nu_\mu$ events over background events. Selections were decided by repeated calculation of a significance estimator after successive selection criteria, where the selections ultimately chosen were those that maximised $\frac{S}{\sqrt{S+B}}$ over the increasingly constrained parameter space. The $B^+ \to \mu^+\nu_\mu$ yield will be evaluated in the region defined by the ultimate selections, defined in Table 4.19

Table 4.19. Final analysis selection criteria decided by optimising the significance estimator, $\frac{S}{\sqrt{S+B}}$. Selections were decided after pre-selections of $E_{\text{ECL}} < 2.5$ GeV/c and zero tracks remaining in ROE were applied to the candidates of $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$ (and charge-conjugate) event reconstruction.

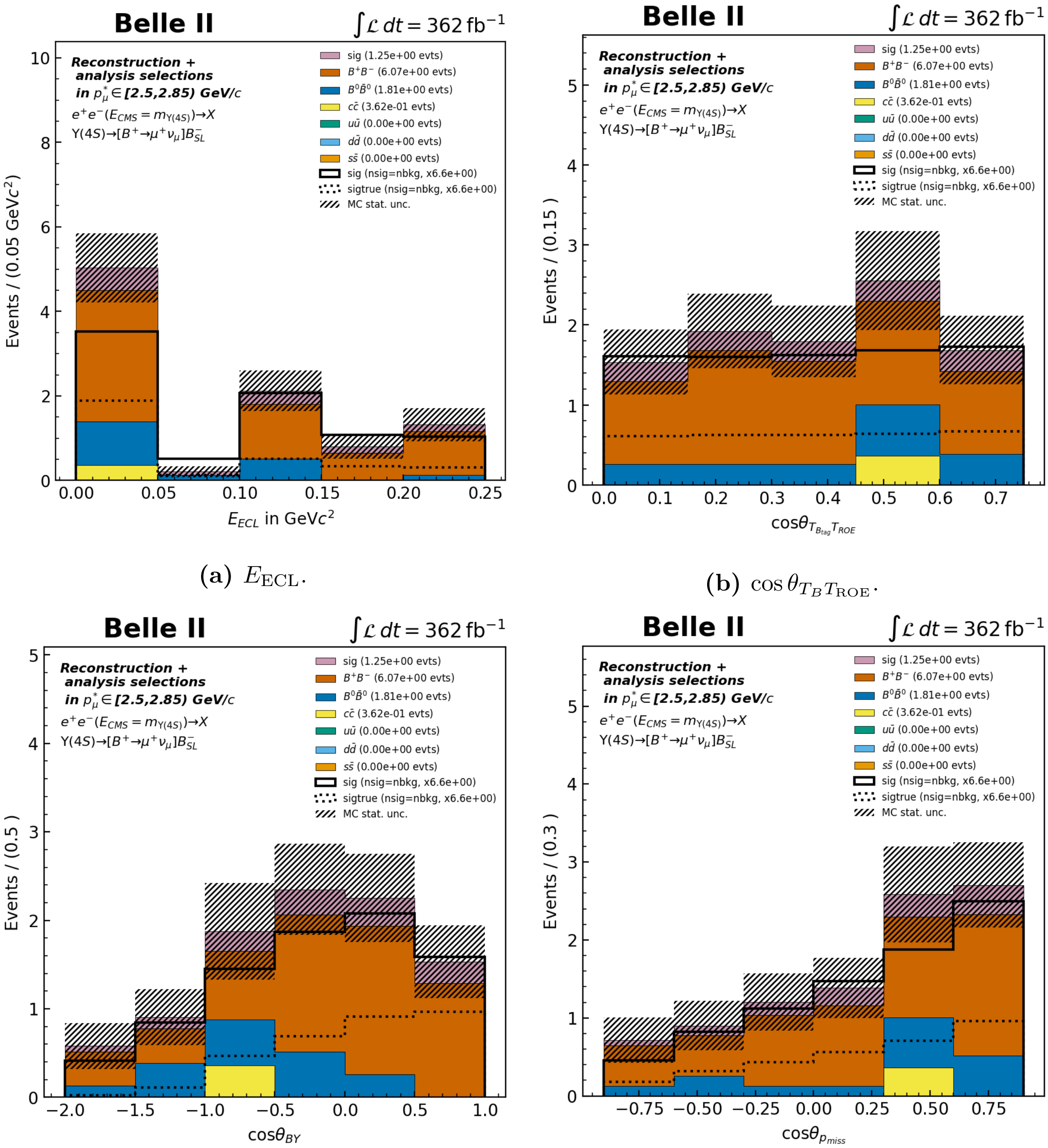
Selection	Variable Definition
$p^*_\mu \in [2.5,\ 2.85)$ GeV/c	Signal muon momentum magnitude in parent $\Upsilon(4S)$ reference frame
$E_{\text{ECL}} \in [0,\ 0.25)$ GeV	Energy of electromagnetic clusters remaining in the ECL after $\Upsilon(4S)$ reconstruction
$\cos\theta_{T_B T_{\text{ROE}}} \in [0,\ 0.75)$	Cosine angle between the B_{tag} thrust axis and the thrust axis of tracks and clusters remaining after $\Upsilon(4S)$ reconstruction.
B_{tag} $\cos\theta_{BY} \in$ [-1, 2)	Cosine angle (in $\Upsilon(4S)$ frame) between the total momentum of the non-neutrino B_{tag} products and the nominal B momentum direction, derived from the constrained $e^+e^- \to B\overline{B}$ kinematics.
B_{tag} $\cos\theta_{\vec{p}_{\text{miss}}} \in$ [-0.9, 0.9)	Cosine polar angle (in lab frame) between the forward beam axis and the missing momentum.
B_{tag} Decay Mode ID $\in$ {0,1,2,3}	B_{tag} semileptonic FEI decay modes: $B^+ \to \bar{D}^{(*)}\ell^+\nu_\ell$ (and charge conjugate), for $\ell = e, \mu$

The result of applying the selection criteria listed in Table 4.19 has already been seen for p^*_μ, having been presented in Figure 4.51b after the final B_{tag} decay mode selection. To further investigate how the selection criteria has affected other key variables, their distributions will now be presented with all analysis selections and $p^*_\mu \in [2.5, 2.85)$ GeV/c. It is noted that the distributions generally have low statistics (and thus wider bins) after the selections. Thus any observed histogram properties may be properties of such effects, rather than a reflection of underlying physics.

Additionally, the distributions will only contain MC events. Further systematic corrections must be applied to the events before the analysis can be considered finalised, and thus the blinding of collision data must still be respected. If data were incorporated at this stage, only partial event distributions could be presented, such as by combining selection criteria with the p^*_μ sideband region. This is done to prevent any inference about the full signal region from the alternative variables. Thus, to allow for a full inspection of the key variables, the following distributions are MC events only, with data to be incorporated into sideband regions in the next section.

The distributions in Figure 4.52 are those of the selection variables, for events in the p^*_μ signal region, with the analysis selection criteria applied. The number of events is substantially lesser than the number of background events prior to events election. The $B^+ \to \mu^+\nu_\mu$ component is visible in these plots, though they are overcome by the Poisson uncertainty in each bin. In this way, the statistical fluctuations now make up a bigger proportion of each bin relative to the pre-selection histograms of Figure 4.29, Figure 4.30, Figure 4.25 and Figure 4.32. The overlaid signal shape has been scaled close to that of the background yield and generally matches the histogram shape, indicating that background sculpting has taken place. This also demonstrates that tightening the selections derived from the $\frac{S}{\sqrt{S+B}}$ procedure any further will not further enhance the ratio of $B^+ \to \mu^+\nu_\mu$ events. It must be noted that such shape agreement could be a low statistic effect and may differ with further data collected at Belle II.

For the Figures individually, both the signal and background components peak at zero for the E_{ECL} distribution. There is a small number of events in the 0.05 E_{ECL} bin of Figure 4.52a, compared to those next to it. This is due to the ROE mask defined in Table 4.2, where energy thresholds on ECL clusters were either $E>$ 0.09, 0.16 GeV, according to where the cluster was recorded. Thus there is a small energy range which could be in the 0.05 E_{ECL} bin, from photons within 0.09 and 0.1 GeV. From this limit, the zero bin thus corresponds to events with no extra clusters and $E_{\text{ECL}} = 0$



(a) E_{ECL}.

(b) $\cos\theta_{T_B T_{\mathrm{ROE}}}$.

(c) B_{tag} $\cos\theta_{BY}$.

(d) $\cos\theta_{\vec{p}_{\mathrm{miss}}}$.

Figure 4.52. Selection variables: distributions of variables reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,] B^-_{\mathrm{SL}}$, for events surviving the decided selection criteria. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

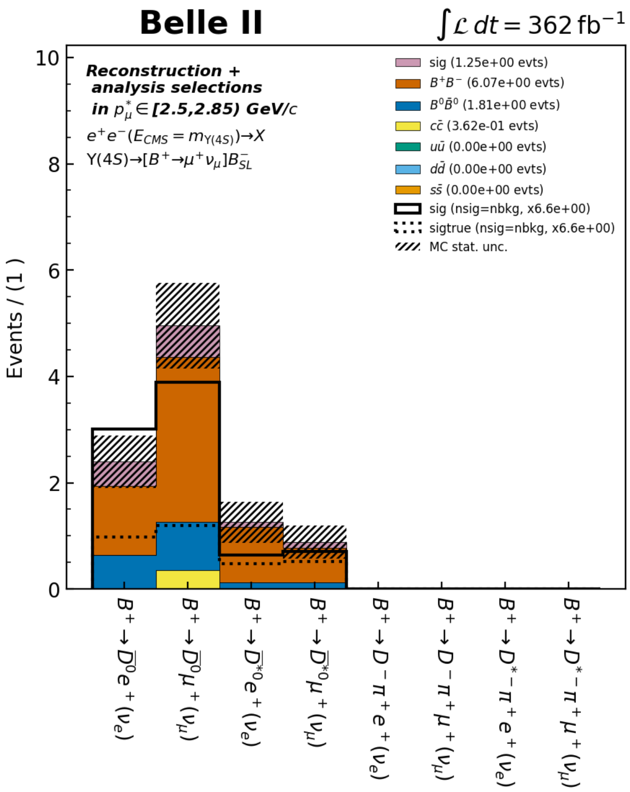
The constrained $\cos\theta_{T_B T_{\mathrm{ROE}}}$ distribution is generally flat, with the most of the peaking continuum events from the pre-selection distribution removed passed $\cos\theta_{T_B T_{\mathrm{ROE}}} \geq 0.75$. Less than a single $c\bar{c}$ event remains in the $\cos\theta_{T_B T_{\mathrm{ROE}}} = 0.45$ bin. This $c\bar{c}$ event is represented by a single unscaled event. The overlaid $B^+ \to \mu^+\nu_\mu$ shape appears to have lesser variation in $\cos\theta_{T_B T_{\mathrm{ROE}}}$ than that of the stacked histogram, though this too is an effect of the underlying number of unscaled MC events; the charged and mixed events are represented by $\mathcal{O}(100) >$ events, while $B^+ \to \mu^+\nu_\mu$ event are represented by $\mathcal{O}(10^4)$. For $\cos\theta_{BY}$, the analysis selections have resulted in more distinct peaks of the MC components than in the pre-selection distribution. The $c\bar{c}$ event occurs in the $\cos\theta_{BY}$ = -1 bin, while the maximum of the mixed component occurs between [-1, 0]. The charged component peaks on [-0.5, 0.5] and the $B^+ \to \mu^+\nu_\mu$ signal shape peaks in the $\cos\theta_{BY} = 0$ bin. The maximum of the correctly-truth matched $B^+ \to \mu^+\nu_\mu$ shape, indicated by the dotted histogram, is located in the $\cos\theta_{BY}$ =0.5 bin. Having $\cos\theta_{BY}$ close to 1 (i.e. θ_{BY} is closer to 0) is consistent with lesser missing momentum, something that was seen for the case of the hadronic FEI in Figure 3.16. For $\cos\theta_{\vec{p}_{\mathrm{miss}}}$, as one of the final selection criteria decided, the distribution does not admit many differences from the pre-selection version, aside from obvious removal of the background peak at $\cos\theta_{\vec{p}_{\mathrm{miss}}} = 0.9$.

The B_{tag} decay modes, as well as the Fox-Wolfram R_2, for events that survive the analysis selections are presented in Figure 4.53. The B_{tag} decay mode, as the final selection variable, also has matching signal shape with the stacked histogram. It too is largely unchanged from pre-selection, aside from the removal of the $B^0\overline{B}^0$-dominating four-body modes. The most popular tagging mode after the analysis selections is $B^+ \to \bar{D}^0\mu^+\nu_\mu$, the same as in Figure 4.23. No selection was ultimately made on the Fox-Wolfram R_2 variable, despite it often being used in continuum suppression in the previous $B^+ \to \mu^+\nu_\mu$ searches discussed in Section 1.4. This is mostly owed to $\cos\theta_{T_B T_{\mathrm{ROE}}}$ performing the same role that the Fox-Wolfram moments would play in exploiting the $q\bar{q}$-$B\overline{B}$ event shape differences. From Figure 4.53b, there are no further selections that could be made to reject $B^+ \to \mu^+\nu_\mu$ background events without also removing signal ones, with most events $R_2 < 0.5$. This R_2 range has also diminished, with events as high as R_2=0.8 in Figure 4.31.

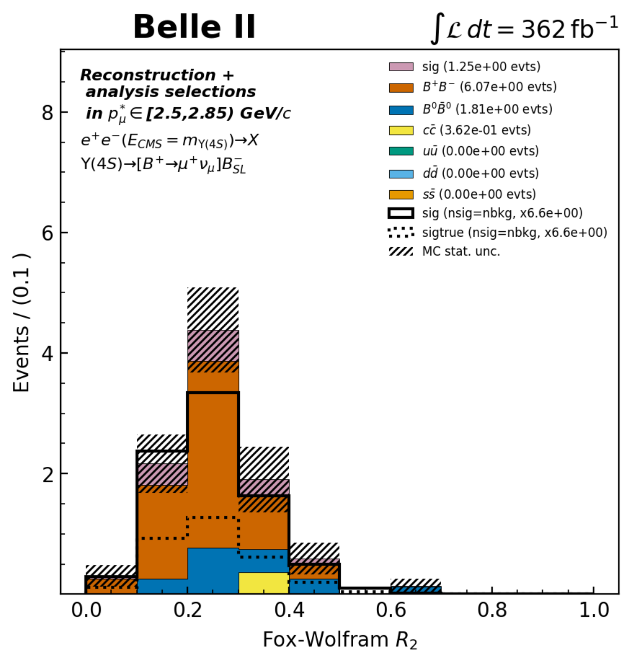
Figure 4.54 contains various B_{tag} variables that weren't involved in selection. The $\log_{10}(\texttt{sigProb})$ and p^*_ℓ distributions are the first to shown slight difference between signal shape and the general stacked MC shape, though this could be a low statistics effect. The selections have removed the peaking events in the lower values of $\log_{10}(\texttt{sigProb})$, that was seen in Figure 4.24. For Figure 4.54a, it seems that an additional lower bound could be made as $\log_{10}(\texttt{sigProb}) \geq$ -2, removing the final $c\bar{c}$ event and a single $B^0\overline{B}^0$ event. In trialling this selection however, the $B^+ \to \mu^+\nu_\mu$ yield falls below one, which was treated as the limit of the $\frac{S}{\sqrt{S+B}}$ selection procedure. Such a selection may be more useful once more Belle II data is acquired. For the zeroth B_{tag} daughter mass distribution, the stacked MC agrees with what was earlier seen in the B_{tag} decay mode histogram. The decay $B^+ \to \bar{D}^0\mu\nu_\mu$ was the most popular mode and the zeroth B_{tag} daughter mass shown to be most likely to occur alongside the $B^+ \to \mu^+\nu_\mu$ process is in the m_D=1.85 GeV/c bin. Unlike in the bins of other histograms shown in this section, there is a whole event expected in this bin. This is not seen in the D/D^* splitting diagram, with sub-1 event bin yields and a histogram shape that has not changed largely from Figure 4.26b.

The final two variables, the signal muon's muon identification probability and the opening angle between the signal muon and the tag-side lepton, are shown in Figure 4.55. Noting the logarithmic-scaled events of the `muonID` distribution in Figure 4.55a, the analysis selections resulted in most events with 0.95 or higher. While the `muonID` > 0.5 list was chosen to maximise the number of $B^+ \to \mu^+\nu_\mu$ events that could be reconstructed, it seems that the 95% probability list could have been used instead. In Figure 4.55b, the opening angle between the signal muon differs significantly from the pre-selection version in Figure 4.28b. The previous distribution has continuum events peaking at $\cos\theta^*_{\mu_{\text{sig}}\ell_{\text{tag}}} = \pm 1$, while $B\overline{B}$ events existed in the flat 'trough' between these peaks. With most of the continuum events removed by the selection criteria, the remaining events are now shaped towards the central values, with minima at the extreme values of $\cos\theta^*_{\mu_{\text{sig}}\ell_{\text{tag}}}$. In addition to the continuum events removed from the extremes (aside for the remaining $c\bar{c}$ event), there are minimal $B\overline{B}$ events residing there. These bins correspond to jet-like separations of the signal muon and tag-side lepton, and so it is possible that the small number of $B\overline{B}$ events with this opening angle would have been removed by $\cos\theta_{T_B T_{\text{ROE}}}$. Additionally, such events could be removed by $\cos\theta_{BY}$ selections, due to its dependence on with p^*_ℓ, which in turn affects the opening angle.

The review of the analysis criteria as applied to the key variables of the $B^+ \to \mu^+\nu_\mu$ search is now complete. In the following section, MC events will finally be compared with collision data in sidebands.

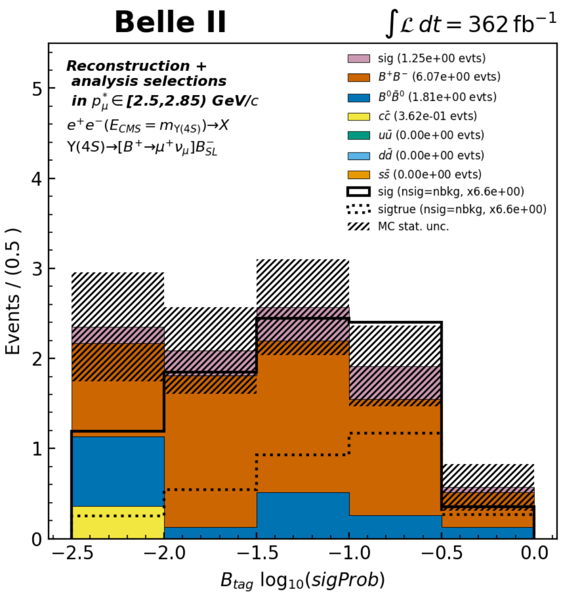


(a) B_{tag} decay mode.

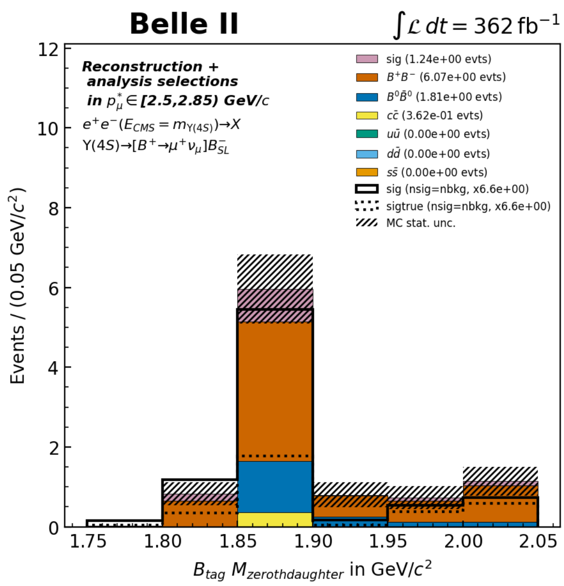


(b) Fox-Wolfram R_2.

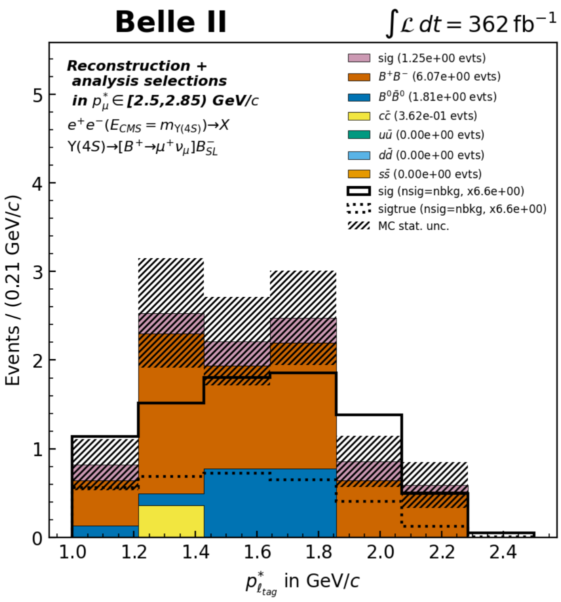
Figure 4.53. B_{tag} decay modes and Fox-Wolfram R_2: distributions of key variables reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$, for events surviving the decided selection criteria. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.



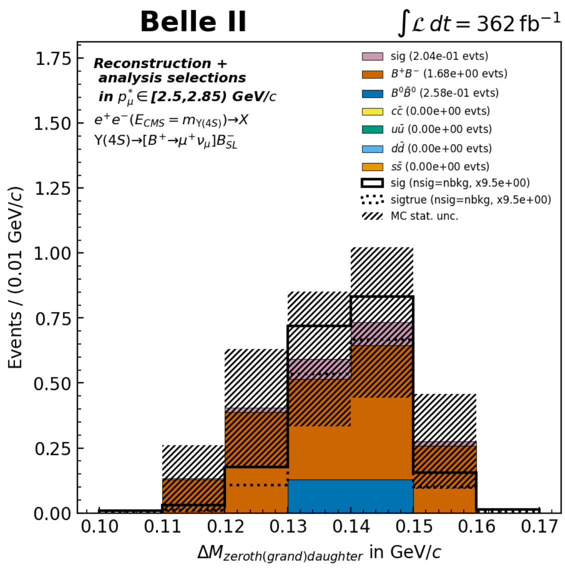
(a) B_{tag} $\log_{10}$(sigProb).



(b) B_{tag} zeroth daughter mass.

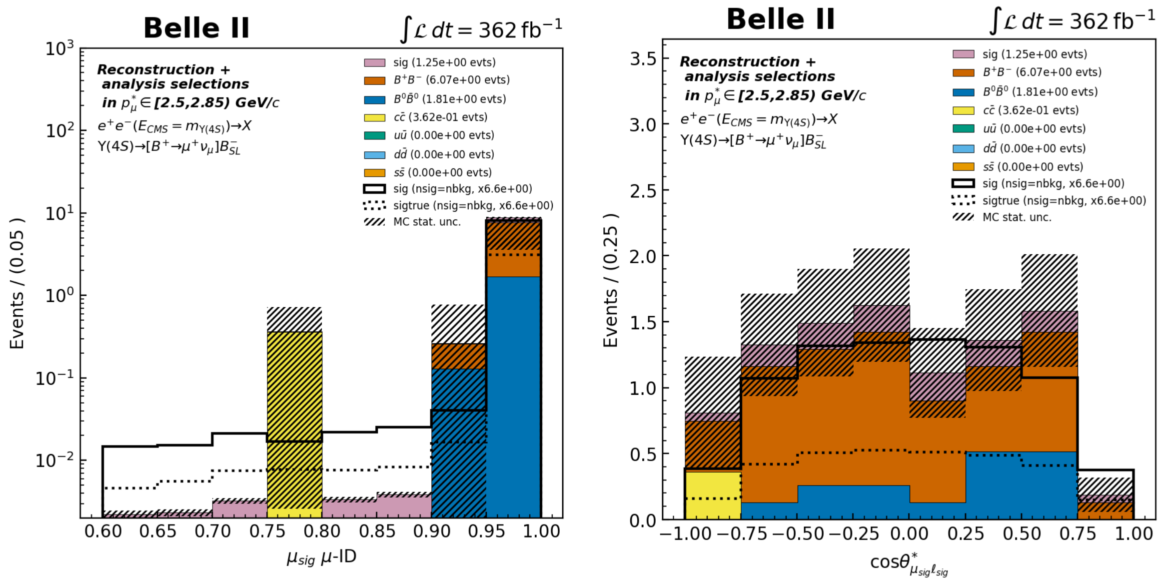


(c) Tag-side p^*_ℓ.



(d) D/D^* mass splitting.

Figure 4.54. Additional B_{tag} variables: distributions of key variables reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$, for events surviving the decided selection criteria. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.



(a) Signal muon `muonID`.

(b) μ_{sig}-ℓ_{tag} opening angle.

Figure 4.55. Muon variables: distributions of key variables reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$, for events surviving the decided selection criteria. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

4.3. Data/MC comparison

The final selection criteria to be applied in the search for the rare $B^+ \to \mu^+\nu_\mu$ decay mode has been decided, using MC-simulated e^+e^- collisions. The effect of the selection criteria in MC can be compared to the effect it has in real in Belle II e^+e^- collisions. The dataset used is the 362 fb^{-1} Run 1 Belle II dataset, recorded at the $\Upsilon(4S)$ resonance between 2019-2021. The Belle II collision dataset has already been described in Section 2.4.4. It is noted that there is no need for event scaling in data. The purpose of event scaling was to ensure that MC yields reflect what is possible in 362 fb^{-1}. The collision dataset is already reflective of 362 fb^{-1} of e^+e^- collisions.

The purpose of this section is not only to observe if the MC accurately reflects what is seen in data for events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\text{SL}}$ (and its charged conjugate). A data/MC ratio in the sidebands must be derived, both as a measure of how well data and MC yields agree, but as a potential correction for the number of events that will eventually be measured in the signal region.

A challenge in achieving the goal stated above is that MC simulations are ultimately approximations of the decays that occur in real e^+e^- collisions. Complete and precise knowledge of every outcome of e^+e^- decays at the $\Upsilon(4S)$ resonance and how to mathematically model their mechanisms is not possible. Notably, if it were, there would be no need for measuring such processes at the Belle II experiment! Thus, as the selection criteria has been tailored to MC and not data, the selections are potentially optimising on underlying features that may not be present in data. This can result in shape disagreement between the distributions seen previously and their counterparts from data. This also extends to semileptonic FEI B-tagging, which functions based on training charged B decay properties within MC. As was alluded to in Section 3.4.6, the semileptonic FEI may have learned features of charged B decays in MC which may not be present in a true dataset, thus resulting in a drop in the yield of tagged-B events in data.

It is imperative to have data-MC agreement, so that the MC components can be used to infer the underlying decays in data. To attempt to rectify any minor disagreement, correction factors can be derived to rescale MC yields towards that of data. Such corrections generally seek to match data and MC yields within statistical error, without artificially inflating one component over the other.

Some analyses substitute off-resonant data (i.e. below the energy threshold of $\Upsilon(4S)$ production, where only continuum events are possible) for continuum MC to mitigate disagreements. Currently, there is only 42 fb^{-1} of off-resonant data, which is too small a dataset for the rare $B^+ \to \mu^+\nu_\mu$ decay. Thus correction factors will be the procedure used to address data-MC issues.

Correction factors can be applied to MC to have them better reflect data, or to data directly, to account for measurement errors and better reflect the underlying physics. To best understand any data-MC disagreement that may be due to the analysis strategy, a range of already available correction factors are applied for all future distributions containing data. Specifically, Belle II recommends corrections due to the following; differences in lepton ID likelihood profiles in data and MC, differences in high-momentum tracking efficiency in data and MC, biases in data track momentum measurements, as well as track momentum differences between in data and MC. An additional correction is required to account for differences in signal selection in data and MC, especially as a result of FEI B-tagging. This requires calculation via a control mode, and so is delayed until Section 5.2.2. The correction will be applied to the signal efficiency directly. Moreover, the impact of the correction factors and their derivations will be documented in Section 5.2.

In this section, data-MC agreement will be examined for signal and other related analysis variables. The impact with and without `muonID` systematics corrections will be presented. With the data/MC ratio in the sideband playing a role in deriving the $B^+ \to \mu^+\nu_\mu$ yield, data-MC agreement is probed in two potential sidebands. These will be either in the p^*_μ sideband, or in the p^*_μ signal region with the E_{ECL} selection reversed (i.e, $E_{\mathrm{ECL}} \in [0.25,2.5)$ GeV). One of the sidebands will be chosen for use in the next stage of the analysis, where the obtained yields and efficiencies will act as inputs to the experimental derivation of the $B^+ \to \mu^+\nu_\mu$ branching fraction.

4.3.1. Data/MC agreement in signal variables

Throughout Section 4.2, it became obvious via the $\frac{S}{\sqrt{S+B}}$ significance estimation procedure that selections on p^*_μ and E_{ECL} would make the greatest impact on removing $B^+ \to \mu^+\nu_\mu$ backgrounds. As attributes closely related to the $B^+ \to \mu^+\nu_\mu$ process, data-MC agreement will primarily be judged in distributions of such attributes.

What follows is a visual inspection into the bin-by-bin data-MC agreement in distributions of p^*_μ and E_{ECL}. The distributions will include both the signal and sideband regions of the variable, for events with all other analysis selections applied. To not bias the future data corrections towards excess signal events in this region, there will be no data events plotted in the corresponding signal regions. Additionally, lepton ID corrections have been applied to MC, with their systematic uncertainties combined with the statistical uncertainty in quadrature for each bin. More discussion on the role of lepton ID corrections, their derivation and how systematic uncertainties are computed will be presented in Section 5.2.1. The MC stacked distribution is scaled as previously to expected yields in 362 fb^{-1}. Data points are assigned Poisson counting error. Each distribution is accompanied by a subplot of the absolute difference between data and MC, scaled by their errors summed in quadrature, to assist in judging agreement.

To begin, Figure 4.56a depicts the p^*_μ for muons from $B^+ \to \mu^+\nu_\mu$ reconstruction with all selection criteria applied, in signal MC, background MC and now data. Within the p^*_μ sidebands, there is reasonable agreement within error bars, between binned data points and the $B\overline{B}$ background dominated stack. The mean data-MC difference scaled is positive (i.e. MC yields may under estimate data yields), though it is less than one. Unfortunately, per-bin uncertainties continue to overwhelm signal MC muons. The data-MC agreement can be compared with that of E_{ECL} in Figure 4.56b. Here, the full set of selections has been made, including $p^*_\mu \in$[2.5,2.85) GeV/c. There is also reasonable agreement within the E_{ECL} sideband and thus reasonable mean scaled data-MC discrepancy, though this time negative (i.e. MC overestimating the data on average).

It was remarked at the beginning of Section 4.3 that lepton ID corrections (amongst others) were to be applied to future distributions involving MC, to enhance data-MC agreement. The distributions without the lepton ID corrections are provided in Figure 4.57 for comparison, showing slightly better mean bin agreement before corrections in p^*_μ and slightly worse mean bin agreement before corrections in E_{ECL}. Lepton corrections are made according to the kinematics of the lepton rather than a global scaling factor, so have not uniformly affected the different MC components. The biggest impact is on incorporating the systematic error into the bin-by-bin error. Lepton ID error has been summed in quadrature with the $\sqrt{N}$ error per bin, resulting in larger uncertainties when compared to the original histograms. It is noted that the systematic error from other correction factors have not yet been incorporated into the error bars.

Table 4.20 contains the yields per region, with and without the lepton ID corrections. Here, data and MC have had all other relevant corrections applied that were listed previously. These have been incorporated into the yields, but not yet into systematic uncertainty and thus the stated error is statistical only. Noting that the lepton ID scaling factors are applied to MC only, there is no difference in the data yield between the two sections of the table. As for the MC yield, the implementation of the corrections has resulted in a lesser number of events in all components and regions, aside from in the signal component. This has led to greater $\frac{\mathrm{data}}{\mathrm{MC}}$ ratios in both sidebands. It is a coincidence that the correction factors have increased the signal yield while decreasing background, as the factors are derived from an independent sample of muons by the Belle II Lepton ID group, to be discussed in Section 5.2.1. The lepton ID correction factors have improved data-MC agreement, as designed, where it seen that the $(\mathrm{data} - \mathrm{MC})/\sqrt{\sigma^2_{\mathrm{MC}} + \sigma^2_{\mathrm{data}}}$ metrics move closer to 1 in both the E_{ECL} and p^*_μ sidebands.

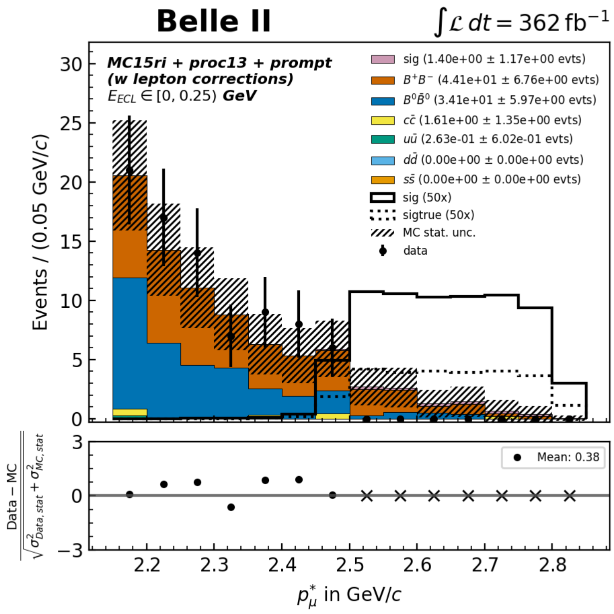
When making comparisons between the sidebands, it should not be expected that the yields agree as they are different datasets. What can be compared is data-MC agreement. The E_{ECL} sidebands have worse agreement than in p^*_μ sidebands, with data/MC ratios and scaled data-MC difference much larger than one. When combining the data/MC ratio from the sidebands with the MC yields in the signal region to form estimates of yields in data, both calculations are within statistical uncertainty of one another. The larger ratios of the E_{ECL} sideband results in larger estimates.

At this stage in the analysis, it would be natural to use the p^*_μ sideband for future calculations, due to the better data-MC agreement. The origin of the worse data-MC agreement in the E_{ECL} sideband could arise due to the presence of more $q\overline{q}$ events. It was previously mentioned at the beginning of the data-MC comparison section that producing continuum MC that accurately reflects data can be challenging. To investigate this hypothesis further, as well as to probe the p^*_μ sideband further, data-MC agreement will be investigated in other variables concerning the $B^+ \to \mu^+\nu_\mu$ process.

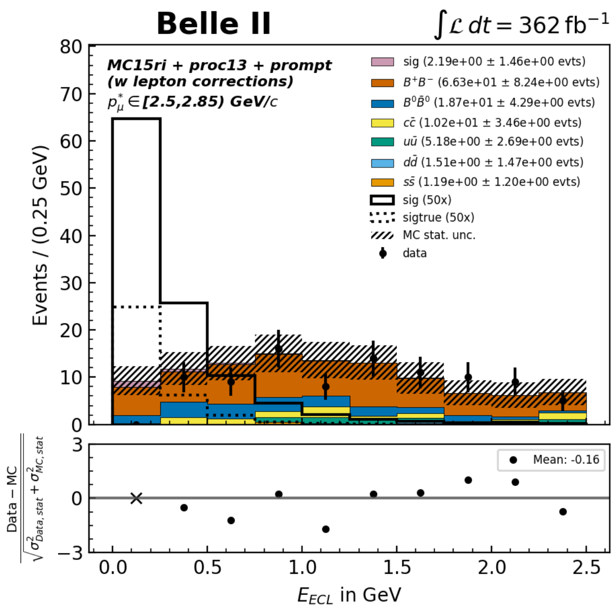
Ultimately, there will be no fitting procedure used to estimate the yield in data due to $B^+ \to \mu^+\nu_\mu$ processes. Due to the relatively small signal yields, 'cut-and-count' methods will be used. As such, bin-by-bin agreements in variables becomes less important than the total data-MC agreement in each region. Looking at data-MC agreement in other variables can still be useful, to potentially isolate variables for which disagreement persists in.

Table 4.20. Yields in signal region (SR) and sidebands (SB), before and after lepton ID corrections to MC. The p^*_μ sideband is $p^*_\mu \in$[2.15,2.5) GeV/c, with all selection criteria. The E_{ECL} sideband is $E_{\mathrm{ECL}}\in$[0.25,2.5) GeV, with all other selection criteria and $p^*_\mu \in$[2.5,2.85) GeV/c.

	no correction			corrected		
Sample	signal region	p^*_μ sideband	E_{ECL} sideband	signal region	p^*_μ sideband	E_{ECL} sideband
$B^+ \to \mu^+\nu_\mu$	1.25±1.12	0.11±0.33	0.87±0.94	1.29±1.14	0.11±0.33	0.90±0.94
B^+B^-	6.07±2.46	39.14±6.26	60.58±7.78	5.96±2.77	38.17±7.84	60.31±11.70
$B^0\overline{B}^0$	1.81±1.34	33.71±5.81	16.40±4.05	1.74±1.38	32.33±6.81	16.93±5.21
$c\overline{c}$	0.36±0.60	1.45±1.20	11.57±3.40	0.24±0.60	1.37±1.24	9.94±4.15
$u\overline{u}$	0	0.36±0.60	7.23±2.69	0	0.26±0.60	5.18±3.08
$d\overline{d}$	0	0	2.17±1.47	0	0	1.51±1.50
$s\overline{s}$	0	0	1.45±1.20	0	0	1.19±1.29
tot. bkg.	8.24±2.87	75±9	99.40±9.97	7.94±2.82	72.14±8.50	95.06±13.95
$\frac{S}{\sqrt{S+B}}$	0.41±0.34	0.01±0.04	0.09±0.09	0.43±0.35	0.01±0.04	0.09±0.10
MC	9.49±3.08	74.76±8.65	100.28±10.01	9.24±3.04	72.25±8.50	95.96±13.99
data	×	82.00±9.06	130.00±11.40	×	82.00±9.06	130.00±11.40
$\frac{\mathrm{data}}{\mathrm{MC}}$	×	1.10±0.18	1.30±0.17	×	1.13±0.21	1.35±0.23
$\frac{\mathrm{data-MC}}{\sqrt{\sigma^2_{\mathrm{MC}}+\sigma^2_{\mathrm{data}}}}$	×	0.58	1.96	×	0.70	1.89
$\mathrm{sig}_{\mathrm{SR}} \times \left(\frac{\mathrm{data}}{\mathrm{MC}}\right)_{\mathrm{SB}}$	-	1.37±1.25	1.62±1.47	-	1.47±1.31	1.75±1.55
$\mathrm{bkg}_{\mathrm{SR}} \times \left(\frac{\mathrm{data}}{\mathrm{MC}}\right)_{\mathrm{SB}}$	-	9.04±3.46	10.68±3.98	-	9.01±3.51	10.76±4.65

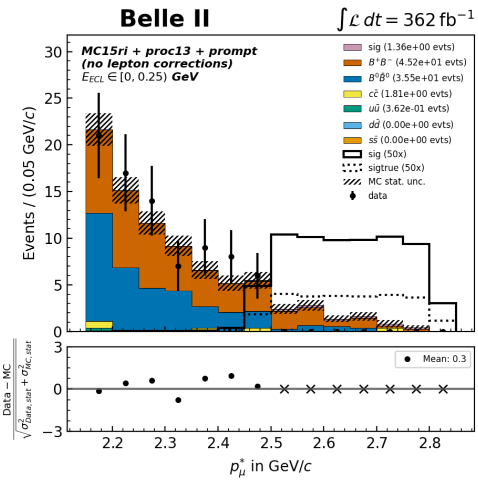


(a) p^*_μ, with $E_{\mathrm{ECL}} \in [0,0.25)$ GeV.

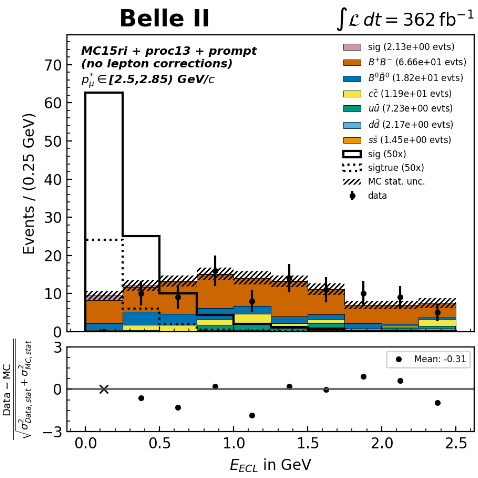


(b) E_{ECL}, with $p^*_\mu \in [2.5,2.85)$ GeV/c.

Figure 4.56. Distributions of $B^+ \to \mu^+\nu_\mu$ primary variables, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.



(a) p^*_μ, with $E_{\mathrm{ECL}} \in [0,0.25)$ GeV.



(b) E_{ECL}, with $p^*_\mu \in [2.5,2.85)$ GeV/c

Figure 4.57. Distributions of $B^+ \to \mu^+\nu_\mu$ primary variables, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$. There are no corrections applied due to differences in `muonID`, for comparison with Figure 4.56. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

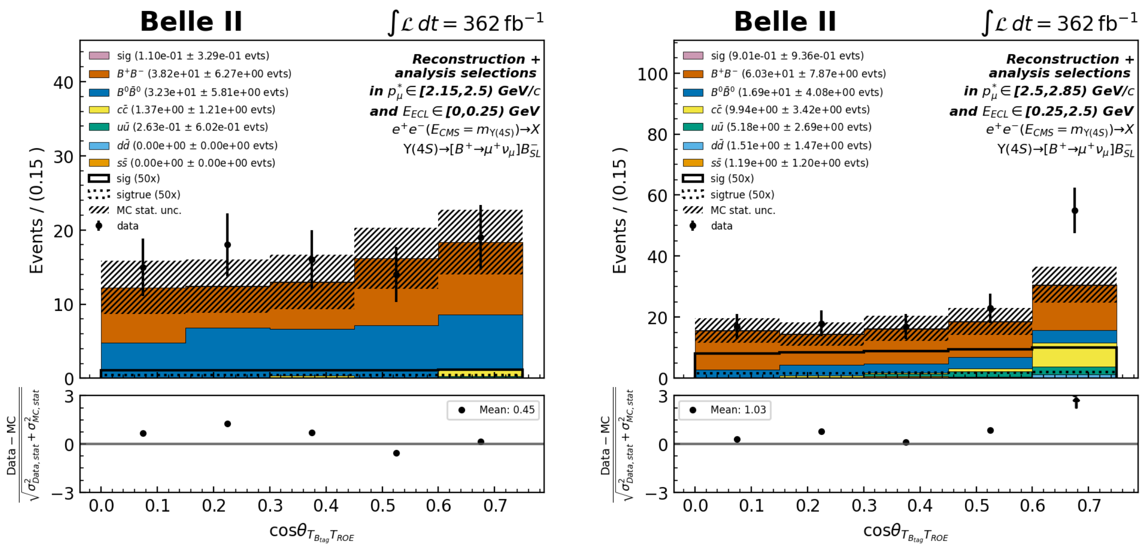
4.3.2. Data/MC agreement in other variables of interest

In the previous section, the data-MC agreement of distributions of p^*_μ and E_{ECL} were studied, so as to decide which may be appropriate for use in ultimately estimating the $B^+ \to \mu^+\nu_\mu$ branching fraction. While both distributions showed reasonable agreement after lepton ID corrections were incorporated, it was decided that the p^*_μ sideband would be used, as the variable admitted data/MC ratios and data-MC scaled differences closer to unity. It was also theorised that poorer agreement of the E_{ECL} sideband could arise from excess $q\overline{q}$ events in this region.

To continue to study sources of data-MC disagreement, distributions of the key variables used throughout this analysis are presented. They are compared in both signal region p^*_μ/sideband E_{ECL} and signal region E_{ECL}/sideband p^*_μ, thus avoiding unblinding the events where the signal yield in data is expected.

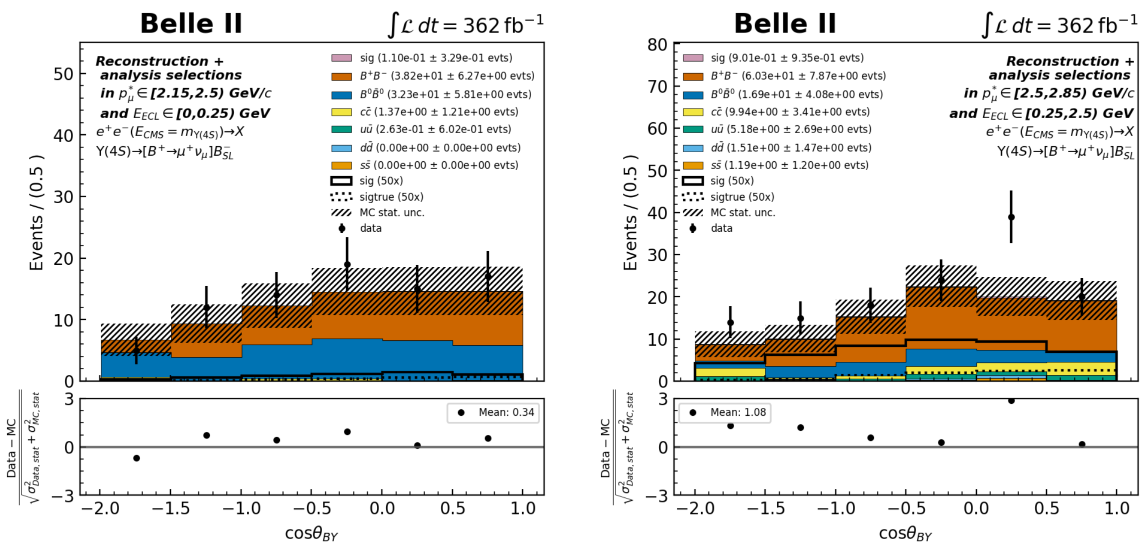
For Figures 4.58-4.68 contain the distributions of p^*_μ and E_{ECL} sidebands. The shapes of the variables histograms in p^*_μ and E_{ECL} sidebands generally agree, a result of sharing most selection criteria. When comparing data-MC agreement, the p^*_μ sidebands perform better than the E_{ECL} sidebands. The worst instances of disagreement in the E_{ECL} sideband distributions are in the following bins; $\cos\theta_{T_B T_{\mathrm{ROE}}} = 0.6$ (Figure 4.58b); $\cos\theta_{BY} = 0$ (Figure 4.59b); $\cos\theta_{\vec{p}_{\mathrm{miss}}} = 0.625$ (Figure 4.60b); the $B^+ \to \overline{D}^0\ell^+\nu_\ell$ decay modes (Figure 4.61b); $\log_{10}(\texttt{sigProb})$ = -2.5 (Figure 4.62b); B_{tag} daughter m= 1.8 GeV/c^2 (Figure 4.61b); and Fox-Wolfram R_2= 0.3 (Figure 4.68b). All of these bins contain $c\overline{c}$ and $u\overline{u}$ events, lending evidence to MC disagreements arising from $q\overline{q}$ mismodelling. Additionally, most of these bins are at the boundaries of the distributions, away from where $B^+ \to \mu^+\nu_\mu$ events exist, especially in the case of $\log_{10}(\texttt{sigProb})$. This suggests that such disagreements may have lesser impact on signal yields, as it is likely such $q\overline{q}$ events will be removed with the restoration of the full p^*_μ signal region.

Ultimately, data-MC agreement is not only desired in the sidebands, but in the signal region. Without unblinding the data, it will be unknown if the data-MC disagreement in the E_{ECL} sideband plots are a result of the continuum events with $E_{\mathrm{ECL}} \geq 0.25$ GeV, if it reflects lack of agreement in the p^*_μ signal region, or if it is a result of the larger number of events as a whole in the E_{ECL} sideband. A similar statement could be said about the p^*_μ sideband plots, if the good agreement is a reflection of the p^*_μ variable as a whole, or if this is due to the limited leftover ECL energy. The use of off-resonant data to better approximate continuum processes could provide a method of resolving this in future analyses.



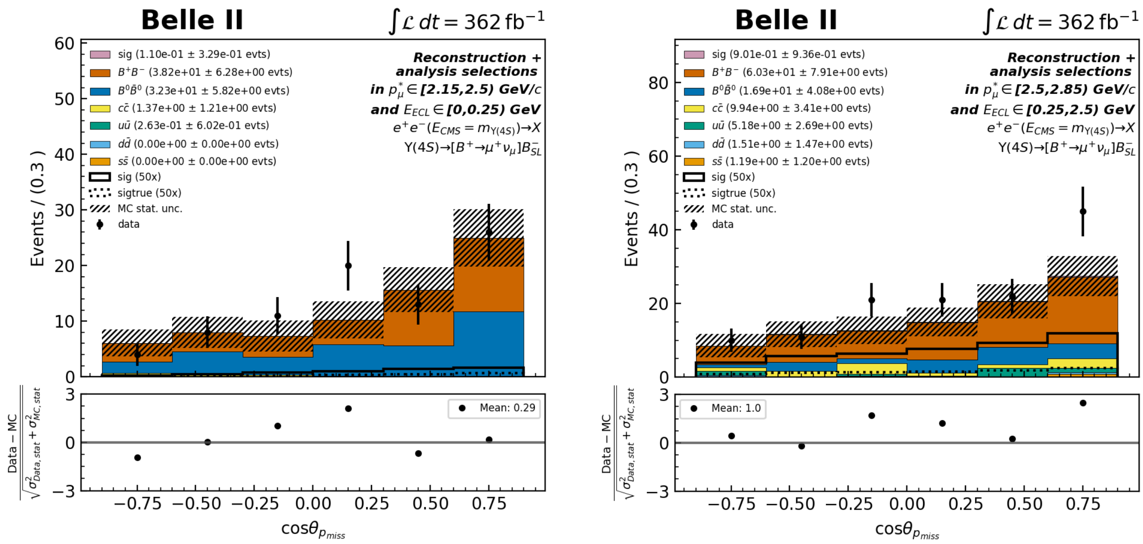
(a) p^*_μ sideband. **(b)** E_{ECL} sideband.

Figure 4.58. Distributions of $\cos\theta_{T_B T_{\mathrm{ROE}}}$, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.



(a) p^*_μ sideband. **(b)** E_{ECL} sideband.

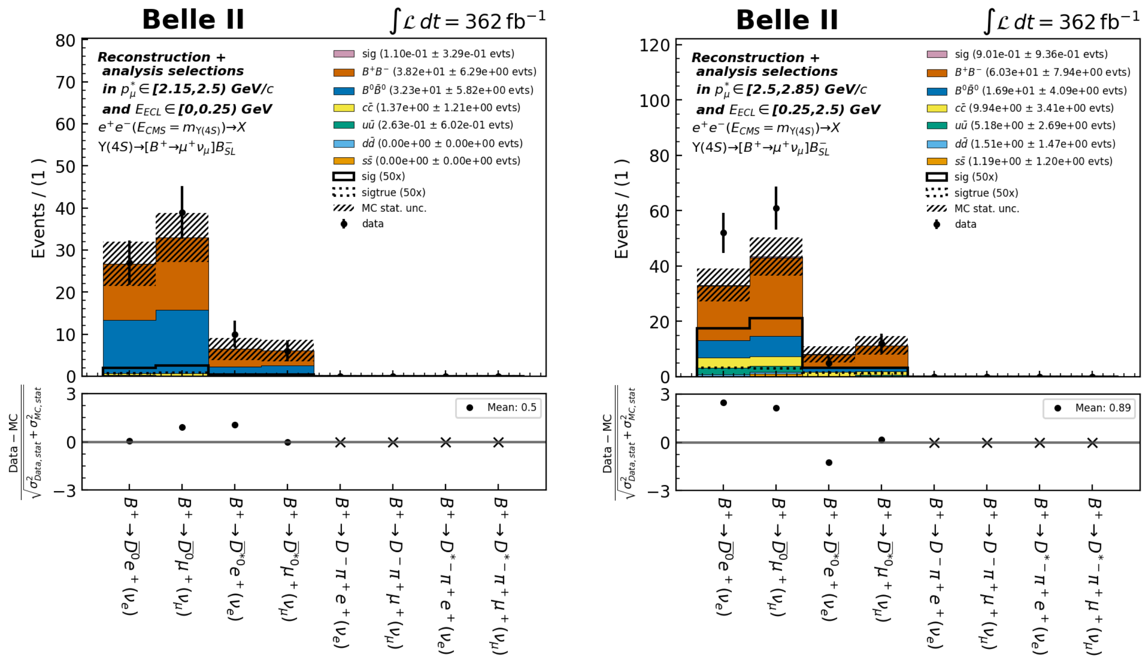
Figure 4.59. Distributions of $\cos\theta_{BY}$, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.



(a) p^*_μ sideband.

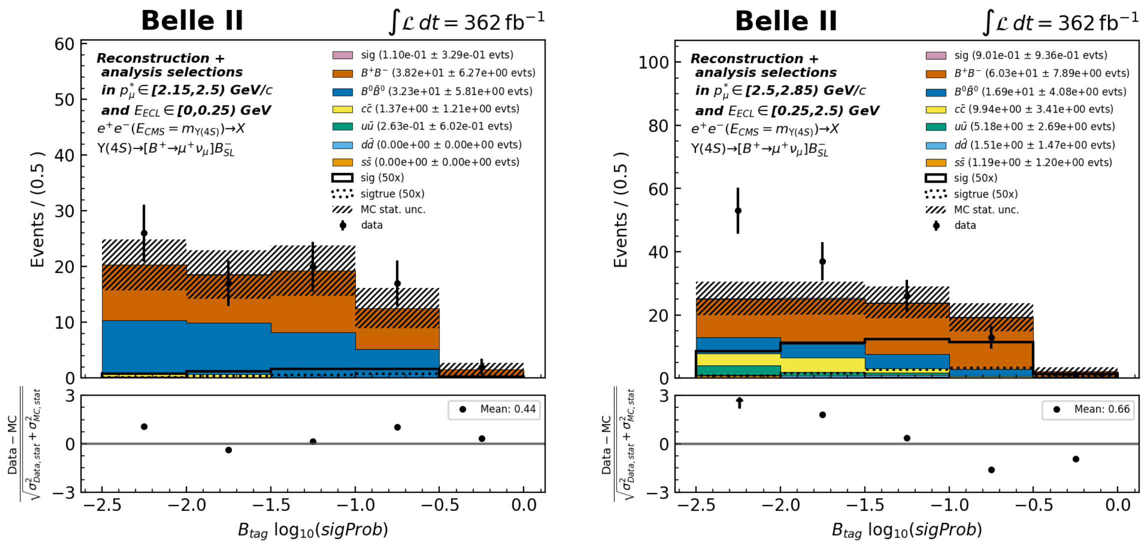
(b) E_{ECL} sideband.

Figure 4.60. Distributions of $\cos\theta_{\vec{p}_{\mathrm{miss}}}$, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.



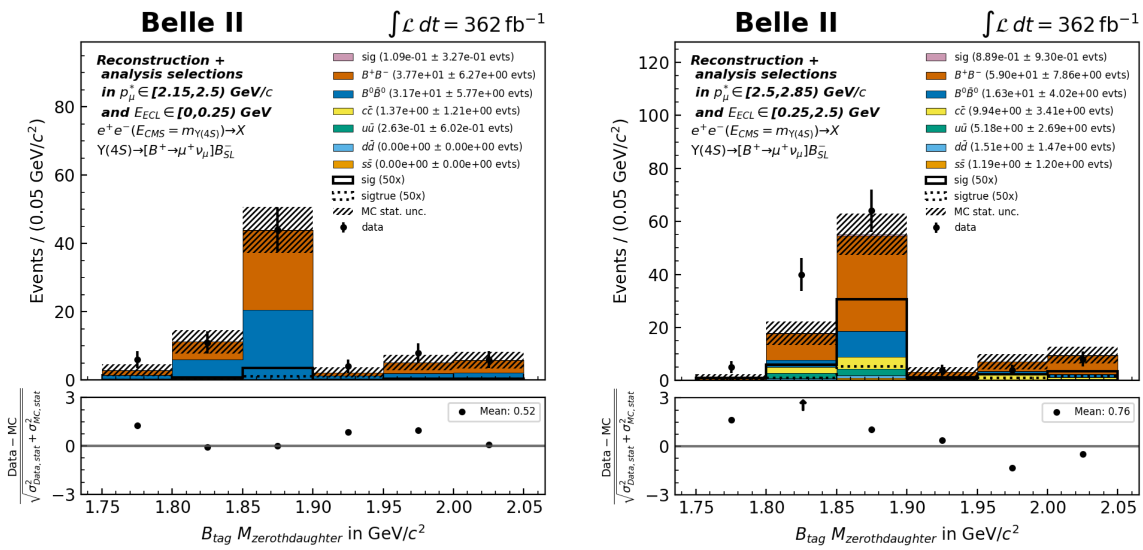
(a) p^*_μ sideband. (b) E_{ECL} sideband.

Figure 4.61. Distributions of B_{tag} decay mode ID, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\ ]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.



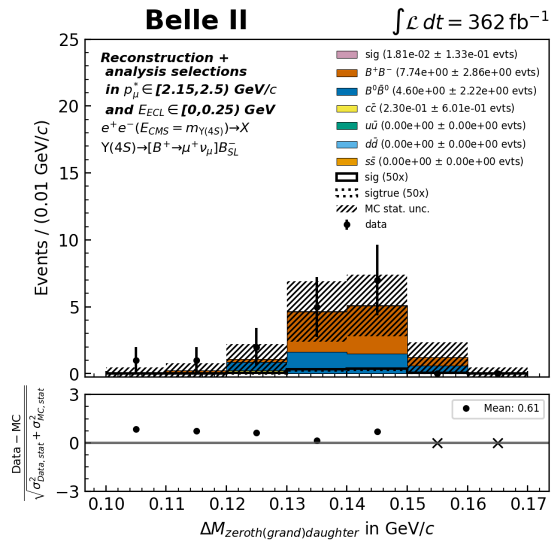
(a) p^*_μ sideband. **(b)** E_{ECL} sideband.

Figure 4.62. Distributions of $\log_{10}$(`sigProb`), from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

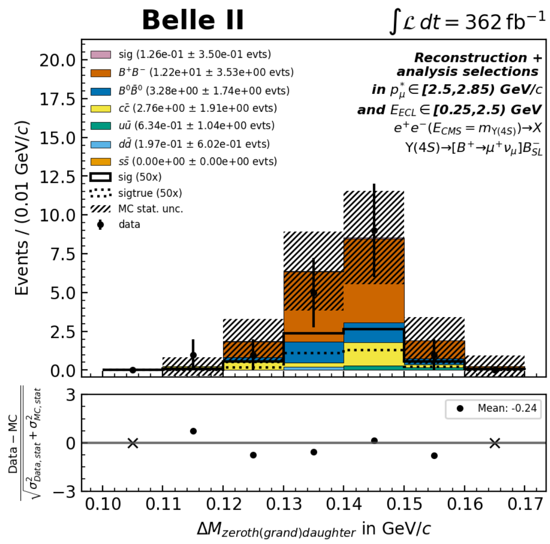


(a) p^*_μ sideband. **(b)** E_{ECL} sideband.

Figure 4.63. Distributions of B_{tag} zeroth daughter mass, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

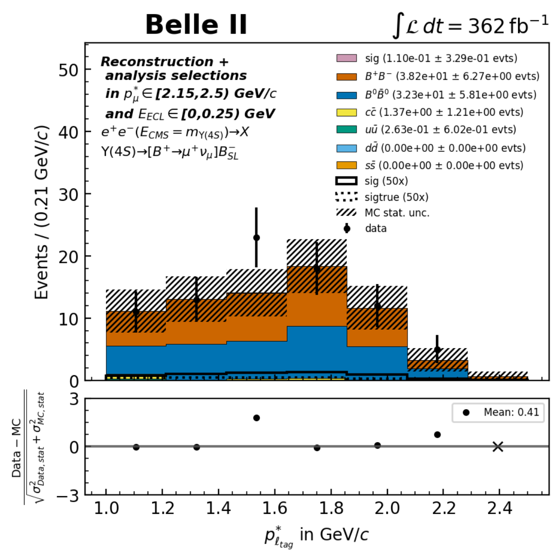


(a) p^*_μ sideband.

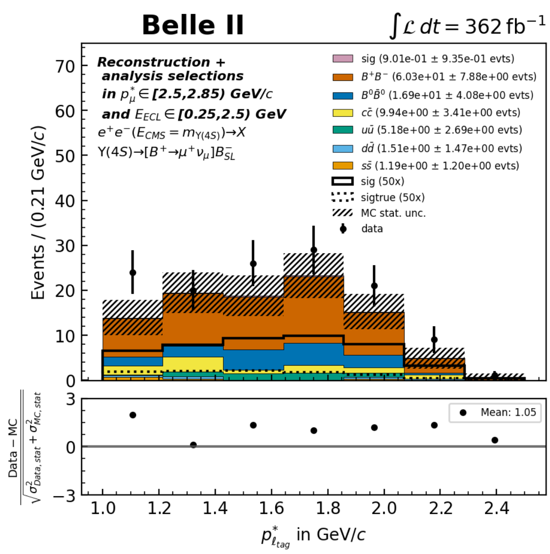


(b) E_{ECL} sideband.

Figure 4.64. Distributions of B_{tag} D^*/D mass-splitting, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

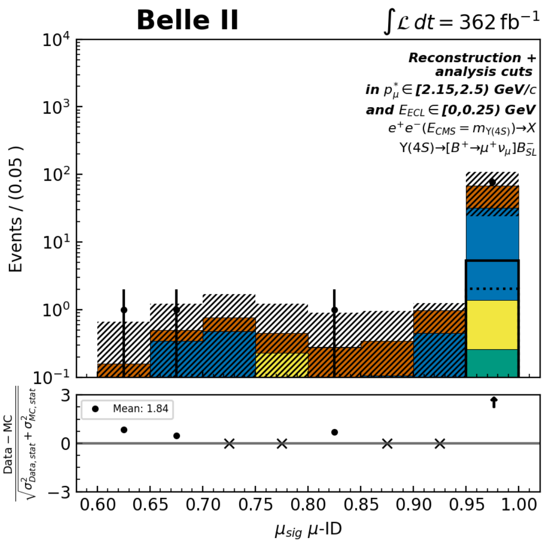


(a) p^*_μ sideband.

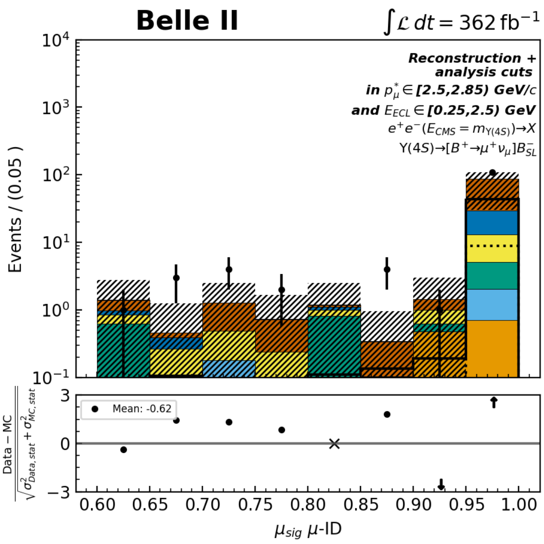


(b) E_{ECL} sideband.

Figure 4.65. Distributions of B_{tag} tag-side lepton momentum p^*_ℓ, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

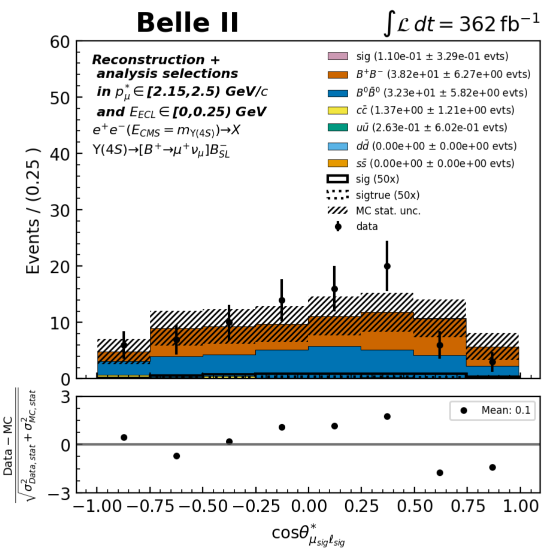


(a) p^*_μ sideband.

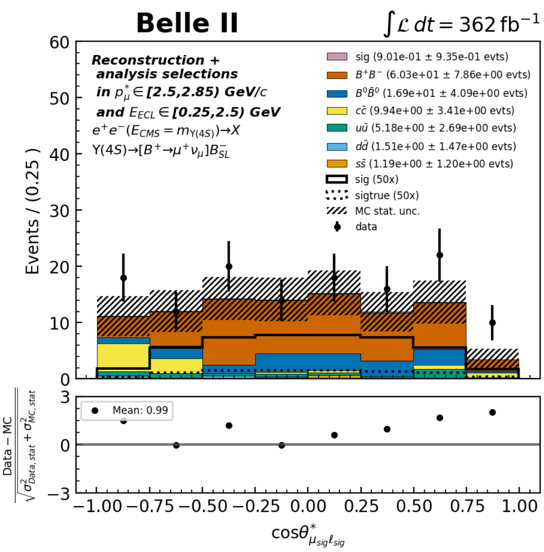


(b) E_{ECL} sideband.

Figure 4.66. Distributions of signal muon `muonID`, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

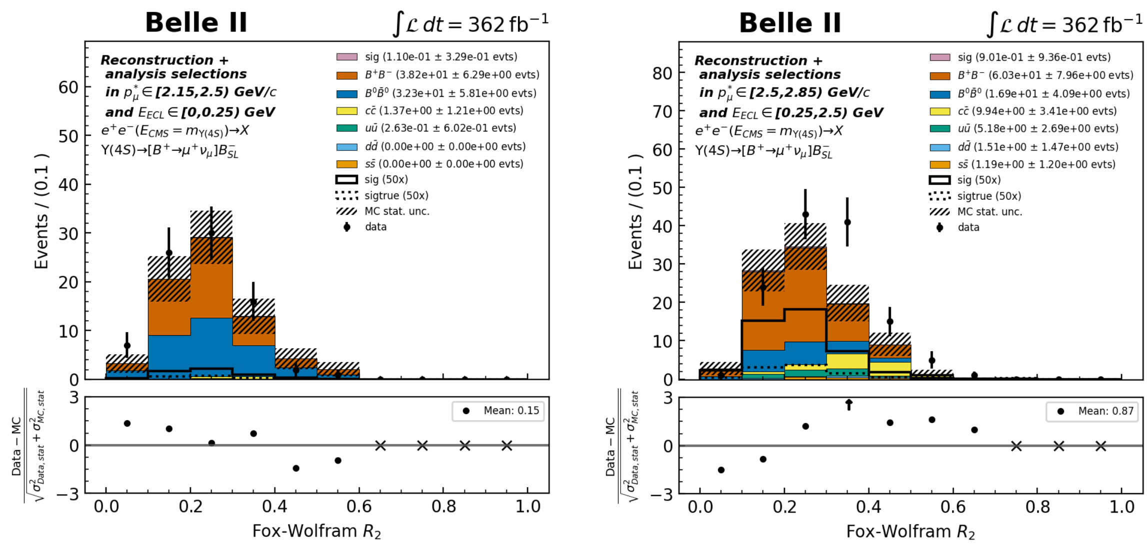


(a) p^*_μ sideband.



(b) E_{ECL} sideband.

Figure 4.67. Distributions of the cosine opening angle between the signal muon and the tag-side lepton, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.



(a) p^*_μ sideband.　　**(b)** E_{ECL} sideband.

Figure 4.68. Distributions of the Fox-Wolfram R_2 ratio, from MC and data events reconstructed as $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ in differing sidebands. Corrections due to differences in `muonID` have been incorporated, both in scaling factors and MC uncertainties. Stacking is according to event type, each scaled to 362 fb^{-1}. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly. Events from collision data are overlaid.

4.4. Summary

This Chapter began with a set of best-candidate B-tagged events in signal, background $B\overline{B}$ and $q\overline{q}$ MC. The goal was to perform $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$ reconstruction with these tagged B events. The properties of muons in the $B^+ \to \mu^+\nu_\mu$ decay were studied, where it was found that the muon with the highest $\Upsilon(4S)$ frame momentum was most often the signal muon in $B^+ \to \mu^+\nu_\mu$ MC. After this, the properties of events that admitted suitable signal muon and tagged B mesons were investigated by looking at distributions of key variables, both before and after reconstruction. Knowledge of these distributions provided intuition of the differences between the observables of signal events and those of background ones. This knowledge also helped when seeking to define signal variables, signal regions and selection criteria that optimised the estimated significance of the small number of $B^+ \to \mu^+\nu_\mu$ events in the very large number of background events. Finally, after the set of selection criteria was decided one, the distributions of data compared to their counterparts in MC could be analysed, outside of signal regions. After implementing corrections for `muonID` selection, reasonable agreement was found in the region $p^*_\mu \in$[2.15,2.5) GeV/c with a data/MC ratio of 1.13±0.21. There was lesser agreement in the signal region of $p^*_\mu \in$[2.5,2.85) GeV/c (with reversed E_{ECL} selection of $E_{\mathrm{ECL}}\in$[0.25,2.5) GeV, with data/MC ratio of 1.35±0.23.

In the next chapter, the preliminary calculations of the $B^+ \to \mu^+\nu_\mu$ branching fraction will be performed. The source of data/MC corrections already applied in Section 4.3 will be presented, including propagation of their systematic uncertainties. This will finally allow for calculation of confidence intervals on the calculation of the $B^+ \to \mu^+\nu_\mu$ branching fraction, which will be interpreted within the context of previous $B^+ \to \mu^+\nu_\mu$ searches, as well as potential for New Physics contributions.

Before moving on, the key data from this Chapter that will go on to be used in evaluating the branching fraction is summarised in Table 4.21.

Table 4.21. Yields and efficiencies in p^*_μ signal region and sidebands to be used in evaluation of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$. Uncertainties are derived by treating yields with Poisson statistics and propagating them to obtain composite values. They are also expressed as percentages of their corresponding values in the $\%\Delta N_{\text{stat}}$ column.

Observable	Value	$\%\Delta N_{\text{stat}}$	Definition
$N^{\text{data}}_{\text{SR,all}}$	$\times$	$\times$	Total no. of events in data in signal region
$N^{\text{MC}}_{\text{SR,bkg}}$	7.94±2.82	31.01	No. of events in MC due to background in signal region
$N^{\text{data}}_{\text{SB,all}}$	82.00±9.06	11.04	Total no. of events in data in sideband region
$N^{\text{MC}}_{\text{SB,all}}$	72.25±8.50	11.37	No. of $B^+ \to \mu^+\nu_\mu$ MC events used in full 362 fb^{-1}
$N^{\text{data}}_{\text{SR,bkg}}$	9.01±3.51	38.57	No. of events due to background in signal region (Estimated as $N^{MC}_{\text{bkg}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}}$)
$N^{\text{MC}}_{\text{SR(gen)},B^+\to\mu^+\nu_\mu}$	$(5.96\pm0.02)\times10^4$	0.33	No. of generated events in signal region due to $B^+ \to \mu^+\nu_\mu$ (i.e. unscaled)
$N^{\text{MC}}_{\text{gen},B^+\to\mu^+\nu_\mu}$	8.00×10^6	-	No. of generated events in signal MC region due to $B^+ \to \mu^+\nu_\mu$
$\epsilon_{\text{sig+tag}}$	$0.75\pm(3.05\times10^{-3})$ %	6.99×10^{-5}	(Estimated as $\frac{N^{\text{MC}}_{\text{SR(gen},B^+\to\mu^+\nu_\mu}}{N^{\text{MC}}_{\text{gen}},B^+\to\mu^+\nu_\mu}$). Cumulative signal efficiency in MC (%)

⋈ ⋈ ⋈

Chapter 5.

Evaluating $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ with systematic uncertainty.

Having previously evaluated the yields of simulated and recorded Belle II collision data after $B^+ \to \mu^+\nu_\mu$ event selection, Chapter 5 will use these values to evaluate the branching fraction of $B^+ \to \mu^+\nu_\mu$ according to potential signal yields.
To estimate a systematic uncertainty on the yields, contributing factors such as lepton identification and track momentum, are considered.
The more common B decay, $B^+ \to \overline{D}^0\ \pi^+$, is reconstructed using the same selection criteria as $B^+ \to \mu^+\nu_\mu$ and utilised as a control mode. This control mode is used to ensure the selection efficiency of $B^+ \to \mu^+\nu_\mu$ events in MC closer resembles what would be expected in data.
The method of Feldman-Cousins is used alongside propagated errors and the corrected yields to determine a 90% confidence interval on potential signal yields.
With these corrected yields and confidence intervals, a range of upper limits on the $B^+ \to \mu^+\nu_\mu$ branching fraction can be produced. A nominal branching fraction will also be evaluated as the ratio of generic B decays to $B^+ \to \mu^+\nu_\mu$ ones in the full 362 fb^{-1} Belle II dataset.
The branching fraction and its confidence interval are then compared to previous results. The implications this has on flavour physics parameters will be discussed.

⋈ ⋈ ⋈

5.1. The strategy of evaluating $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ from yields in MC and data

In the previous chapter, $B^+ \to \mu^+\nu_\mu$ events in simulated and recorded Belle II collisions were investigated. This was done by reconstructing $\Upsilon(4S) \to [B^+ \to \mu^+\nu_\mu\,]B^-_{\mathrm{SL}}$, with the application of strict selection criteria to reject the large amount of non-$B^+ \to \mu^+\nu_\mu$ events retained throughout the analysis. This was done so that analysis yields could be measured in a signal-enhanced region. Ultimately, the number of $B^+ \to \mu^+\nu_\mu$ decays that specifically occurred alongside semileptonic FEI charged B decay modes in the narrowed parameter space is not a useful observable to compare with previous searches, given the differing tagging and analysis strategies. A more useful observable is the branching fraction (i.e. the rate of $B^+ \to \mu^+\nu_\mu$ decays compared to generic charged B decays), a quantity that is independent of partner B decay, selection criteria or the integrated luminosity measured. The $B^+ \to \mu^+\nu_\mu$ branching fraction, $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$, can be calculated using the yields previously measured in signal MC, background MC and data. To these quantities, selection efficiencies can be applied to estimate the number of $B^+ \to \mu^+\nu_\mu$ events that were present before tagging or selection criteria, thus evaluating the fraction of unconstrained $B^+ \to \mu^+\nu_\mu$ decays. The yields pertinent to the derivation of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ were presented at the end of the previous Chapter, in Table 4.21.

The relationship between $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ and the yields of the analysis will now be derived. This begins by considering the branching fraction as equivalent to the number of $B^+ \to \mu^+\nu_\mu$ decays that have taken place in the 362 fb^{-1} dataset, in ratio to the number of charged B decays generally.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{\Gamma(B^+ \to \mu^+\nu_\mu)}{\Gamma(B^+)} = \frac{N^{\mathrm{data}}_{362\,\mathrm{fb}^{-1},B^+\to\mu^+\nu_\mu}}{N^{\mathrm{data}}_{362\,\mathrm{fb}^{-1},B^+}} \tag{5.1}$$

In an electron-positron collision to charged B pairs, both B mesons are capable of decaying to the signal decay of interest (or its charge conjugate, as determined by the charge of the B). Thus, the number of charged B mesons in the denominator of Eq. (5.1) is double the number of $e^+e^- \to B^+B^-$ processes expected in the 362 fb^{-1} data set i.e. $N^{\mathrm{data}}_{362\mathrm{fb}^{-1},B^+} = 2N^{\mathrm{data}}_{362\mathrm{fb}^{-1},B^+B^-}$. The number of $e^+e^- \to B^+B^-$ processes can be calculated from the number of $e^+e^- \to B\overline{B}$ events, which is inclusive of both B^+B^- and $B^0\overline{B}^0$ events. The latter, denoted by $N^{\mathrm{data}}_{362\,\mathrm{fb}^{-1},B\overline{B}}$, is a carefully measured quantity by Belle II. To find the number of B^+B^- events from $B\overline{B}$, the ratio of decay widths ($\frac{\Gamma(\Upsilon(4S)\to B^+B^-)}{\Gamma(\Upsilon(4S)\to B\overline{B})}$)

can be applied as a multiplicative factor. Additionally, under the assumption that $\mathcal{B}(\Upsilon(4S) \to B\overline{B})$=1, the decay width ratio becomes equal to $f^{+-} = \frac{\Gamma(\Upsilon(4S) \to B^+B^-)}{\Gamma(\Upsilon(4S))}$, and thus $N^{\rm data}_{362{\rm fb}^{-1},B^+B^-} = f^{+-}N^{\rm data}_{362{\rm fb}^{-1},B\overline{B}}$ The assumption of $\mathcal{B}(\Upsilon(4S) \to B\overline{B}) = 1$ was introduced in Section 2.2.2 and has been used already in this analysis. The f^{+-} parameter itself is taken from HFLAV averages of a range of CLEO, Belle and BaBar results [114]. The quantity $N^{\rm data}_{362{\rm fb}^{-1},B^+}$ can therefore be redefined to become the denominator in the following equation.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{N^{\rm data}_{362\,{\rm fb}^{-1},B^+\to\mu^+\nu_\mu}}{2f^{+-}N^{\rm data}_{362\,{\rm fb}^{-1},B\overline{B}}} \tag{5.2}$$

The quantity $N^{\rm data}_{362\,{\rm fb}^{-1}B^+\to\mu^+\nu_\mu}$(i.e. the total number of $B^+ \to \mu^+\nu_\mu$ in the 362 fb^{-1} dataset) can be extrapolated from the data yield of $B^+ \to \mu^+\nu_\mu$ events in the constrained parameter space of the signal region, ($N^{\rm data}_{{\rm SR},B^+\,\to\,\mu^+\nu_\mu}$). The calculation of this quantity from the analysis will be explained below, though once evaluated it can be scaled up from a yield in the signal region to the expected number in the full dataset. This is done by multiplying by the reciprocal of the signal efficiency, where $1/\epsilon_{\rm sig+tag} = N^{\rm MC}_{362\,{\rm fb}^{-1},B^+\to\mu^+\nu_\mu}/N^{\rm MC}_{{\rm SR},B^+\,\to\,\mu^+\nu_\mu}$. The cumulative signal efficiency has already been calculated for this analysis, as the ratio of MC $B^+ \to \mu^+\nu_\mu$ events in the signal region to the $B^+ \to \mu^+\nu_\mu$ events expected in 362 fb^{-1}. Specifically, $N^{\rm MC}_{362\,{\rm fb}^{-1},B^+\to\mu^+\nu_\mu}$ was evaluated in Table 2.6, from the Standard Model branching fraction in Eq. (1.7). Using the signal efficiency to scale the number of events from signal region to the unconstrained dataset assumes that the efficiency with which $B^+ \to \mu^+\nu_\mu$ has been identified in MC matches data. It was purported in Section 4.3 that this may not be true. For this reason, a data-driven analysis of the control mode $B^+ \to \overline{D}^0\pi^+$ will be performed to create a correction factor for $\epsilon_{\rm sig+tag}$ later in Section 5.2.2. Thus, substituting the numerator of Eq. (5.2) with $N^{\rm data}_{{\rm SR},B^+\,\to\,\mu^+\nu_\mu}/\epsilon_{\rm sig+tag}$ leads to the following equation.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{N^{\rm data}_{{\rm SR},B^+\,\to\,\mu^+\nu_\mu}}{2f^{+-}N^{\rm data}_{362\,{\rm fb}^{-1},B\overline{B}} \times \epsilon_{\rm sig+tag}} \tag{5.3}$$

The parameter $N^{\rm data}_{{\rm SR},B^+\,\to\,\mu^+\nu_\mu}$ can now be broken down. As has been mentioned through out this thesis, the signal yield will be estimated via a ‘cut-and-count’ analysis, and not a fitting procedure. Therefore, $N^{\rm data}_{{\rm SR},B^+\,\to\,\mu^+\nu_\mu}$ will be calculated via the full yield of data in the signal region ($N^{\rm data}_{\rm SR,all}$), subtracting the expected background contribution to data in the signal region, $N^{\rm data}_{\rm SR,bkg}$. $N^{\rm data}_{\rm SR,all}$ is the quantity that has been blinded throughout the previous Chapter. Thus, the numerator of Eq. (5.3) is reexpressed as the equation

Table 5.1. External values to be used in evaluation of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ via Eq. (5.5).

Observable	Value	Source	Definition
$N^{\text{data}}_{362\,\text{fb}^{-1},B\bar{B}}$	$(387.1 \pm 5.6)\times10^6$	Belle II Collaboration	Number of $e^+e^- \to B\bar{B}$ events at Belle II [115]
f^{+-}	0.514 ± 0.006	Produced by HFLAV group for PDG	Ratio of $\Upsilon(4S) \to B^+B^-$ to $\Upsilon(4S)$ decay widths [114].

that follows.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{N^{\text{data}}_{\text{SR,all}} - N^{\text{data}}_{\text{SR,bkg}}}{2f^{+-}N^{\text{data}}_{362\,\text{fb}^{-1},B\bar{B}} \times \epsilon_{\text{sig+tag}}} \tag{5.4}$$

The method of producing $N^{\text{data}}_{\text{SR,bkg}}$ was presented in Table 4.20. With the assumption that the data/MC ratio in the p^*_μ sidebands ($N^{\text{data}}_{\text{SB,all}}/N^{\text{MC}}_{\text{SB,all}}$) would be equal to the data/MC ratio in the p^*_μ signal region, the known yield of background MC in the signal region ($N^{\text{MC}}_{\text{SR,bkg}}$) can be scaled up to the equivalent value in data.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{N^{\text{data}}_{\text{SR,all}} - (N^{\text{MC}}_{\text{SR,bkg}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}})}{2f^{+-}N^{\text{data}}_{362\,\text{fb}^{-1},B\bar{B}} \times \epsilon_{\text{sig+tag}}} \tag{5.5}$$

Thus with Eq. (5.5), an experimental expression for transforming the data and MC yields from this analysis into and analysis-agnostic quantity, $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$. The nominal values of most of these yields are already listed in Table 4.20. The external values discussed can be substituted from Table 5.1.

To evaluate $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ completely, the concealed quantity $N^{\text{data}}_{\text{SR,all}}$ is required. However, as the systematic uncertainty still requires estimation, this yield will continue to be kept blind. This avoids temptations to bias uncertainties towards a non-zero $B^+ \to \mu^+\nu_\mu$ branching fraction.

It is therefore the goal of the next section to understand the data/MC corrections that have been applied to the samples used, as well as to quantify the uncertainties in the measurement of data and MC yields due to the analysis strategy. A summary of the sources of systematic uncertainty that will be explored in the data/MC yields is presented in Table 5.1. This will then assist in setting confidence intervals on the eventual calculation of Eq. (5.5).

To be able to 'unblind' data within the Belle II experiment requires a series of cross-checks by the Belle II Collaboration. Unfortunately, such checks did not converge in time for delivery of this thesis. As such, the quantity $N^{\text{data}}_{\text{SR,all}}$ will not be unblinded by the end of this work. Instead, the $B^+ \to \mu^+\nu_\mu$ branching fraction will be produced for a range of possible signal region yields in collision data with accompanying confidence intervals in Section 5.4.

5.2. Deriving systematic errors from data/MC corrections and the analysis strategy

In the derivation of the expression Eq. (5.5), a method of evaluating $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ from the yields in data and MC was obtained. The external values to be used in the calculation were also identified, which were provided with their own uncertainties. So far, quoted uncertainties in measurements have been almost entirely statistical. In the interest of establishing errors on the eventual calculation of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ that truly reflect the analysis strategy, systematic uncertainties must be quantified as well. Systematic uncertainties attempt to encode measurement error. They are often evaluated by varying the underlying quantities by an accepted amount, to see how robust a measurement is against incidental changes. A key source of systematic uncertainty is in correction factors, which seek to either correct MC to better reflect data or to correct data measured with experimental error, to better reflect the underlying physics. These corrections are derived from studies of well-known decay processes in Belle II data and were applied to have the best match between reconstruction in MC and in collision data.

In this section, each source of systematic uncertainty will be reviewed. The strategies to estimate the systematic uncertainty for each source or within each correction factor will be presented. The uncertainty sources and the affected observable are summarised in Table 5.2. Notably, some systematic uncertainty can be avoided in the signal efficiency $\epsilon_{\text{sig+tag}}$, by evaluating the quantity in the unscaled signal events. This is justified as the net scaling factors to account for the integrated luminosity are ultimately cancelled out within the ratio. Using unscaled signal events avoids error contributions from the theoretical Standard Model branching fraction.

Previously in this work, there was an attempt to quantify a systematic uncertainty due to the $\frac{S}{\sqrt{S+B}}$ signal optimisation procedure in determining selection criteria. The magnitude of such uncertainties were much larger than the contributions described soon. This added doubt to the method of doing of their uncertainty estimation. As such, their discussion is omitted from this Chapter and presented in Appendix A.

Table 5.2. Sources of systematic error in data and MC yields.

Quantity	Error Sources
$N^{\mathrm{MC}}_{\mathrm{SR,bkg}}$	Lepton ID correction, mid-to-high momentum tracking efficiency
$N^{\mathrm{MC}}_{\mathrm{SB,all}}$	Lepton ID correction, mid-to-high momentum tracking efficiency
$N^{\mathrm{data}}_{\mathrm{SB,all}}$	Track momentum correction
$\epsilon_{\mathrm{sig+tag}}$	Data/MC correction from control mode
	(From $N^{\mathrm{MC}}_{\mathrm{SR,(unscaled)}B^+\to\mu^+\nu_\mu}$): Lepton ID correction, mid-to-high momentum tracking efficiency

5.2.1. Data/MC corrections from Belle II detector and their associated systematic errors

As was discussed in Section 4.3, MC simulations are modelled based on the current understanding of e^+e^- collision phenomena and Belle II detector response. The following Sections discuss the sources of corrections that have been incorporated into visualisations and yields of data and MC since Section 4.3, as well as the derivation of new systematic uncertainties. Specifically, these follow the recommendations of the Belle II Performance group. Internally this is referred to as the 'Moriond 2023' recommendations, applicable to the MC15ri and proc13+prompt MC and data described in Section 2.4.4.

MC correction and systematic error: Lepton ID

The set of signal muons used in this $B^+ \to \mu^+\nu_\mu$ search were subject to a muon identification probability selection criterion of `muonID` > 0.5. The explanation of how particle identification probabilities are derived was presented in Section 2.3.2. Specifically, it was stated that the underlying likelihood profiles used to create the charged particle probabilities can be created analytically, from control data samples or from detector simulations. Depending on the technique used, this can lead to different charged particle selection efficiencies according to if they are applied to MC or collision events.

With the `muonID`>0.5 selection in this analysis, it is important to account for the differences in the performance of this selection in data and MC. Such studies are performed by the Belle II Lepton ID Group, whose methods are outlined in Reference [77]. Discussions of electron correction efforts are omitted, as they are not relevant to the $B^+ \to \mu^+\nu_\mu$ search.

Two effects in muon selection via `muonID` require correction: data-MC differences in the yield of muons selected as well as data-MC differences in the number of hadrons (i.e. π and K) mistaken as muons. Being able to correct for differences in hadron-muon fake rates in data and MC is especially important given the small difference in pion/muon masses.

The differences in muon selection efficiency is measured for tracks surviving a selection of `muonID` > 0.5, data and MC. Such tracks are then used to reconstruct $e^+e^- \to \mu^+ \mu^- (\gamma)$, $e^+e^- \to e^+e^-\mu^+\mu^-$ and $J/\psi \to \ell^+ \ell^-$ processes at $\sqrt{s}$ =10.85 GeV. The modes chosen provide a pure and easily constrained sample of muons in data. Each process covers a unique combination of lab-frame momentum and polar angle muons, so that a correction can be applied for the wide range of muon momenta in analyses.

For the data/MC differences in the rate of pions selected as muons, the selection of `muonID` > 0.5 is applied again, with surviving tracks used to reconstruct $K^0_S \to \pi^+ \pi^+$ and $e^+e^- \to \tau^+ \tau^-$ processes. For the case of the tau pair mode, further selections are applied to limit the analysis to taus that decay to a single charged particle (with missing neutrinos) as a 'τ-tag'. The second tau decay is limited to cases where three charged pions are produced, alongside missing neutrinos. Meanwhile, for kaon-muon fake rate corrections, the process $D^{*+} \to [\overline{D}^0 \to K^- \pi^+]\pi^+$ is reconstructed. The track chosen to constitute the kaon can be distinguished from the other tracks due to its charge being the opposite of the pion that originates from D^{*+}. This pion is often referred to as a *slow pion*, as its momentum is notably smaller than that of the pion originating from the $\overline{D}^0$.

The ultimate correction factors are constructed as ratios of efficiencies and fake rates between data and MC, for bins of lab-frame momentum and polar angle. For the analyst, the factors are applied as event weights, according to the truth identity of the muon track in MC. When the weights of all events of the selection are summed, they reflect the appropriate number of fake and true muons to the same `muonID` selection in data. Thus by definition, corrections are only MC.

Lepton ID corrections have already been applied to each MC event in this analysis, according to the lab-frame momentum and polar angle of the signal muon. The impact of the reweighting is seen in the distributions of p^*_μ and E_{ECL} given in Figure 4.57. The weightings per event were provided with statistical and systematic errors, which now need to be propagated to systematic uncertainty on the sum of the weights.

To estimate the systematic uncertainty due to this reweighting, the current recommendation is to use the bootstrapping statistical procedure, also referred to generating 'toy MC'. For each event, two sets of 200 random numbers are generated, from Gaussian distributions centred on zero. One set has width equal to the systematic error, and the other has width equal to the statistical error. These two sets are summed together to obtain a single set of 200 values. The original event weight is added to each of the 200 values. This yields 200 variations of event weights. Statistically, this has emulated re-running the lepton ID correction procedure 200 times to obtain a spread of possible weights for an event. The current standard within Belle II is to perform this procedure using the `PIDvar` [116] package.

For the events remaining after selection in the MC dataset, the n^{th} weight variation of each event are summed together, for all n in the 200 variations. This is to say, the first weight variation of all events are summed together, then the second weight variations of all events are summed separately, and so on. This results in 200 variations on the total event yield, where the variations are due to uncertainty in the correction factors method. The standard deviation of the 200 variations can be treated as the systematic uncertainty of the `muonID` correction. The results of this are presented in Table 5.3, for all MC components.

From Table 5.3, it is seen that the percentage systematic uncertainties due to the `muonID` corrections are smaller than that of the Poisson statistical uncertainty. Table 5.4 summarises the yields and their systematic uncertainty propagated to the inputs of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$. The largest effect is on the reconstruction efficiency, with 5.02% systematic uncertainty. This is understandable, as this input is derived from the number of (unscaled) $B^+ \to \mu^+\nu_\mu$ events in the signal region, which is directly impacted by the `muonID` > 0.5 selection on the signal muon.

Table 5.3. Systematic uncertainty due to lepton ID data/MC corrections, for yields in p^*_μ signal (SR) and sideband (SB) regions. Statistical uncertainty is included for comparison. Yields are formatted as $N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$, where statistical error is $\sqrt{N}$.

Sample	Reg.	Yield	$\%\Delta N_{\text{stat}}$	$\%\Delta N_{\text{sys}}$	Reg.	Yield	$\%\Delta N_{\text{stat}}$	$\%\Delta N_{\text{sys}}$
$B^+ \to \mu^+\nu_\mu$	SR	1.29±1.14±0.06	87.92	5.02	SB	$0.11\pm0.33\pm3.89\times10^{-3}$	301.70	3.54
B^+B^-	SR	5.96±2.44±0.26	40.97	4.33	SB	38.17±6.18±0.84	16.18	2.19
$B^0\overline{B}^0$	SR	1.74±1.32±0.71	75.73	4.10	SB	32.33±5.69±0.47	17.59	1.46
$c\overline{c}$	SR	0.24±0.49±0.02	203.91	9.60	SB	1.37±1.17±0.10	85.33	7.24
$u\overline{u}$	SR	0	0	0	SB	0.26±0.51±0.02	194.99	7.22
$d\overline{d}$	SR	0	0	0	SB	0	0	0
$s\overline{s}$	SR	0	0	0	SB	0	0	0
tot. bkg.	SR	7.94±2.82±0.37	29.68	4.69	SB	72.14±8.49±1.28	11.36	1.77
MC	SR	9.24±3.04±0.28	32.01	2.99	SB	72.25±8.50±0.97	11.37	1.34
data	SR	×	×	×	SB	82.00±9.06	11.04	-

Table 5.4. Inputs to $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ affected by systematic uncertainty in lepton ID. Lepton corrections are not applied to data samples.

Observable	$N \pm \Delta N_{\text{sys}}$	$\%\Delta N_{\text{sys}}$
$N^{\text{MC}}_{\text{SR,bkg}}$	7.94±0.37	4.69
$N^{\text{MC}}_{\text{SB,all}}$	72.25±0.97	1.34
$N^{\text{MC}}_{\text{SR(gen)},B^+\to\mu^+\nu_\mu}$	$(5.96\pm0.30)\times10^4$	5.02
$\epsilon_{\text{sig+tag}} \left(= \frac{N^{\text{MC}}_{\text{SR(gen)},B^+\to\mu^+\nu_\mu}}{N^{\text{MC}}_{\text{gen},B^+\to\mu^+\nu_\mu}}\right)$	$(7.45\pm0.37)\times10^{-3}$	5.02
$N^{\text{data}}_{\text{SR,bkg}} \left(= N^{\text{MC}}_{\text{SR,bkg}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}}\right)$	9.01±0.44	4.88

Data correction and systematic error: track momentum scale due to magnetic field modelling

The invariant masses of hadrons such as $D^{+/0}$, J/ψ, K^0_S, Λ^+_c were previously reported to be consistently lower than their PDG value, when measured in early Belle II data. Reconstruction of such hadrons use only charged final state particles i.e. e, μ, π, K and p and charge conjugates. After investigating a number of potential causes, a study on the reconstructed hadron masses was performed using a sample of data where the charged track trajectories were derived using with an alternate magnetic field map. The magnetic field map (i.e. the value of the $\vec{B}$-field at each point in the Belle II detector volume) is usually derived from simulation. It was posited that an inaccurate magnetic field map could lead to poor track-fitting and thus biased charged particle momentum used in the invariant mass reconstruction. Using a magnetic field map that was created within the physical detector before the central drift chamber (CDC) was inserted improved the agreement of hadron invariant masses reconstructed in data, though not perfectly.

Preparing the track momentum scaling factors for each data processing is the responsibility of the Belle II Tracking Group. They are produced by measuring the D^0 mass in $D^{*0} \to [\overline{D}^0 \to K^- \pi^+]\pi^+$ decays in data. A variety of trial scaling factors were applied to the kaon and pion momenta used in the reconstruction. The ultimate scaling factor is derived by interpolating to the factor that best fit the PDG value of $\overline{D}^0$ mass. The original study and derivation of corrections is documented in Reference [117].

The corrections appropriate to the data implemented in this $B^+ \to \mu^+\nu_\mu$ search have already been applied and are listed in Table 5.5, which were recommended in the presentation of Ref. [118]. As the data in the $B^+ \to \mu^+\nu_\mu$ signal region is blinded, such corrections can only be applied to the sideband. Unlike the lepton ID corrections, the scaling factor is the same for all tracks (i.e. 'global'). This is independent of particle hypothesis or track parameters, but dependent on which calibration phase the data is from (i.e. the prompt stage or reprocessing stage discussed in Section 2.4.4). To evaluate the systematic uncertainty on this correction, the yields when the scale factor is varied to its maximum and minimum within the provided uncertainties are derived. The half of the absolute difference in the yields is accepted as the systematic uncertainty. The result of this is listed in Table 5.6. The narrow uncertainty range and the correction factor very close to unity results in a very small overall uncertainty in the number of events in data, with most datasets demonstrating no change when varying the scaling applied. It is only in the two largest datasets, proc 13 and bucket 30 that contains a change in

Table 5.5. Table of track momentum scalings, as applied to data.

Dataset	Momentum Scale Factor
prompt (buckets 26-36)	$0.99990^{+0.00033}_{-0.00047}$
proc 13	$0.99984^{+0.00043}_{-0.00066}$

Table 5.6. Systematic error due to track momentum correction in data (p^*_μ sideband only). Yields are formatted as $N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$, where statistical error is $\sqrt{N}$.

Dataset	Region	Yield	$\%\Delta N_{\text{stat}}$	$\%\Delta N_{\text{sys}}$
bucket 26	SB	1.00±1.00±0.00	100.00	0
bucket 28	SB	5.00±2.24±0.00	44.72	0
bucket 29	SB	6.00±2.45±0.00	40.82	0
bucket 30	SB	7.00±2.65±0.50	37.80	7.14
bucket 31	SB	7.00±2.65±0.00	37.80	0
bucket 32	SB	4.00±2.00±0.00	50.00	0
bucket 33	SB	6.00±2.45±0.00	40.82	0
bucket 35	SB	6.00±2.45±0.00	40.82	0
bucket 36	SB	3.00±1.73±0.00	57.74	0
proc 13	SB	37.00±6.08±1.00	16.44	2.70
total	SB	82.00±9.06±1.5	11.04	1.83

events. The percentage uncertainty due to the track momentum systematic is a third that of the Poisson error. The impact of the track momentum scaling on the data yield in the sideband, and further propagated to the estimate of the signal region data yield due to background events is presented in Table 5.7. As the correction is only applied to data and not MC yields, the percentage uncertainty is unchanged in the evaluated background data yield. Moreover, it would be expected that the data yield in the signal region, $N^{\text{data}}_{\text{SR,all}}$, would also acquire the track momentum scaling systematic uncertainty, though this cannot be evaluated until after unblinding.

Table 5.7. Inputs to $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ affected by systematic uncertainty in track momentum scaling. Momentum corrections are not applied to MC samples.

Observable	$N \pm \Delta N_{\text{sys}}$	$\%\Delta N_{\text{sys}}$
$N^{\text{data}}_{\text{SB,all}}$	82.00±1.50	1.83
$N^{\text{data}}_{\text{SR,bkg}} \left(= N^{\text{MC}}_{\text{SR,bkg}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}}\right)$	9.01±0.16	1.83
$N^{\text{data}}_{\text{SR,all}}$	×	×

MC systematic error: Mid-to-high momentum tracking efficiency

The role of tracking in determining the momentum of charged particles from signals in subdetectors was presented in Section 2.3.2. A constructed track can be considered more reliable if it leaves more than 20 hits in the central drift chamber. Thus for a multitrack event, there could be hundreds of hits in which to establish tracks, leading to substantial probability of combinatoric backgrounds i.e. fake tracks.

Similar to the previous discussion of `muonID` likelihood, the performance of tracking algorithms can differ in data and simulation, if track hits are not reliably modelled in MC. This also can impact the rate that fake tracks are established. With the signal muon being derived from a charged track, data/MC discrepancies in track finding must be considered.

The procedure of quantifying data/MC agreement in track-finding is detailed in Ref. [119] and is based on the BaBar method [120]. The application of the procedure to the current data and MC production is Ref. [121]. Similar to the case of measuring pion-muon fake rate, samples of $e^+e^- \to \tau^+\,\tau^-$ processes are used. Selection criteria limits the decays studied to cases of where one tau decays as $\tau^\pm \to \ell^\pm\nu_\ell\nu_\tau$ while the second decays hadronically to three charged pions and any number of neutral pions. The pionic tau is treated as the 'tagging'-tau, while the other is the 'signal'-tau. Once the pionic tau is found, an additional track is guaranteed to have occurred, originating from the leptonic tau. If this track is not found, this can be treated as failure of the tracking algorithm. Thus, the tracking efficiency can be measured as the ratio between the number of events where the leptonic tau is found, against all events studied.

To then estimate the rates at which false tracks are identified by the tracking algorithm in data and MC, the same pionic and leptonic taus modes are used. An additional pion is added to the reconstruction, resulting in five tracks required. This is performed in a sample where it is known that one of the taus create additional pions and one where

Table 5.8. Systematic error due to tracking efficiency data/MC discrepancy, for yields in p^*_μ signal (SR) and sideband (SB) regions. Yields are formatted as $N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$, where statistical error is $\sqrt{N}$. (*) indicates that $N_{\text{sys}} < 0.01$.

Sample	Reg.	Yield	$\%\Delta N_{\text{stat}}$	$\%\Delta N_{\text{sys}}$	Reg.	Yield	$\%\Delta N_{\text{stat}}$	$\%\Delta N_{\text{sys}}$
$B^+ \to \mu^+\nu_\mu$	SR	1.29±1.14±0.01*	87.92	0.24	SB	0.11±0.33±0.01*	301.70	0.24
B^+B^-	SR	5.96±2.44±0.01	40.97	0.24	SB	38.17±6.18±0.09	16.18	0.24
$B^0\overline{B}^0$	SR	1.74±1.32±0.01*	75.73	0.24	SB	32.33±5.69±0.08	17.59	0.24
$c\overline{c}$	SR	0.24±0.49±0.01*	203.91	0.24	SB	1.37±1.17±0.01*	85.33	0.24
$u\overline{u}$	SR	0	0	0	SB	0.26±0.51±0.01*	194.99	0.24
$d\overline{d}$	SR	0	0	0	SB	0	0	0
$s\overline{s}$	SR	0	0	0	SB	0	0	0
tot. bkg.	SR	7.94±2.82±0.02	29.68	0.24	SB	72.14±8.49±0.17	11.36	0.24
MC	SR	9.24±3.04±0.02	32.01	0.24	SB	72.25±8.50±0.17	11.37	0.24
data	SR	×	×	×	SB	82.00±9.06	11.04	-

there is only the previous four tracks. If the additional particle is found in the sample where there is no true additional pion, this indicates that a fake track has been identified by the tracking algorithm.

There are currently no recommendations to apply any corrections due to differences between fake track identification in data and MC. As for tracking efficiency, it was found that there was a 0.061±0.027(stat)±0.231(sys) % discrepancy between data and MC. As the systematic uncertainty dominates the central value, it is advised to not apply a correction, but to apply a 0.24% systematic uncertainty per-track on the event yields in MC. As it is only the signal muon within the mid-to-high momentum range, 0.24% in uncertainty is assigned due to tracking efficiency. The resulting systematic uncertainty in each data component used in this analysis is presented in Table 5.8. A summary of the relevant $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ parameters that acquire this systematic uncertainty are listed in Table 5.9.

Table 5.9. Inputs to $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ affected by systematic uncertainty in track efficiency data/MC discrepancy. Systematic is not assigned to data samples.

Observable	$N \pm \Delta N_{\text{sys}}$	$\%\Delta N_{\text{sys}}$
$N^{\text{MC}}_{\text{SR,bkg}}$	7.94±0.02	0.24
$N^{\text{MC}}_{\text{SB,all}}$	72.25±0.17	0.24
$N^{\text{MC}}_{\text{SR(gen)},B^+ \to \mu^+\nu_\mu}$	(5.96±0.01)×10^4	0.24
$\epsilon_{\text{sig+tag}} \left(= \frac{N^{\text{MC}}_{\text{SR(gen)},B^+ \to \mu^+\nu_\mu}}{N^{\text{MC}}_{\text{gen},B^+ \to \mu^+\nu_\mu}} \right)$	(7.45±0.02)×10^{-3}	0.24
$N^{\text{data}}_{\text{SR,bkg}} \left(= N^{\text{MC}}_{\text{SR,bkg}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}} \right)$	9.01±0.03	0.34

5.2.2. Data/MC corrections in $\epsilon_{\text{tag+sig}}$ via control mode, $B^+ \to \bar{D}^0\ \pi^+$, and its associated systematic error

In the procedure of deriving an experimental expression for the $B^+ \to \mu^+\nu_\mu$ branching fraction in Section 5.1, an assumption about the $B^+ \to \mu^+\nu_\mu$ signal efficiency was necessary. Specifically, this was that the signal efficiency derived from MC was equivalent to the ratio of $N^{\text{data}}_{\text{SR},B^+ \to \mu^+\nu_\mu}/N^{\text{data}}_{362\text{fb}^{-1},B^+ \to \mu^+\nu_\mu}$, i.e. the ratio of $B^+ \to \mu^+\nu_\mu$ events in data between the signal region and the unconstrained dataset. This is equivalent to assuming that the signal efficiency of identifying $B^+ \to \mu^+\nu_\mu$ events in MC matches that within collision data. Sufficient cases of data-MC disagreement have been discussed in this Chapter to understand why such an assumption could be problematic in the derivation of a $B^+ \to \mu^+\nu_\mu$ branching fraction. An additional issue is that the FEI B-tagging is a key technique within this analysis, but that the analysis so far has not accounted for any differences in data and MC in the FEI application.

Table 5.10. Comparison of Belle II charged and $B^+ \to \bar{D}^0\pi^+$ MC used in this analysis. The number of events in $B^+ \to \bar{D}^0\pi^+$ sample is derived from the branching fraction specified in the '.dec' file and the number of charged events. The $B^+ \to \bar{D}^0\pi^+$ scaling factor is the B^+B^- factor multiplied by the ratio of PDG and '.dec' file branching fraction. See Table 2.6 for other background types used.

Sample ($e^+e^- \to X$)	Cross-section [nb]	Sample lumi. [fb^{-1}]	No. of events (in sample)	Scale factor	No. of events (in 362 fb^{-1})
B^- [$B^+ \to \bar{D}^0\pi^+$]	-	2800	7.40×10^6	1.27×10^{-1}	9.43×10^5
B^+B^-	0.5654	2800	1.58×10^9	1.29×10^{-1}	2.05×10^8

In order to address the issues above, the MC-derived $B^+ \to \mu^+\nu_\mu$ signal efficiency will be calibrated to match the efficiency of identifying $B^+ \to \mu^+\nu_\mu$ events in collision data. To avoid biasing the analysis, the calibration of the tagging and signal selection efficiency will be derived based on a data-MC ratio of efficiencies within an alternate decay with similar kinematics. This alternate decay is referred to as the control mode. The MC-based signal efficiency of the control mode will be produced in the same manner as in the $B^+ \to \mu^+\nu_\mu$ reconstruction. Specifically, this is the ratio of the number of expected control mode decays in the 362 fb^{-1} dataset to the number of control mode decays obtained in the signal region, A data-driven signal efficiency for the control mode will be produce by evaluating $\epsilon_{\mathrm{sig+tag}}$ from Eq. (5.5). This done by substituting in the known control mode branching fraction, as well as other necessary yields from control mode data and MC. The ratio of the two signal efficiencies can then be treated as a correction to the $B^+ \to \mu^+\nu_\mu$ MC-based signal efficiency.

As for the choice of control mode decay, various examples were presented when discussing prior $B^+ \to \mu^+\nu_\mu$ searches in Section 1.4. In the 2020 Belle search, the $B^+ \to \bar{D}^0\pi^+$ process was used to validate the inclusive-tagging strategy, where ultimately no correction was needed as the data-based and MC-based efficiencies agreed. The 2015 Belle search used the predecessor to the FEI to obtain hadronically-tagged B mesons. Here, correction factors were produced using $B^+ \to \bar{D}^0\ell^+\nu_\ell$ as the control mode, alongside the tagged B mesons.

For the analysis of this thesis, $B^+ \to \bar{D}^0\pi^+$ is chosen, so as to be able to exploit the similarities between this mode and $B^+ \to \mu^+\nu_\mu$ via the two-body decay features discussed in Section 1.2.1. Feynman diagrams of this process are depicted in Figure 5.1 Additionally, the $B^+ \to \bar{D}^0\pi^+$ branching fraction is larger than that of $B^+ \to \mu^+\nu_\mu$ and precisely known, $\mathcal{B}(B^+ \to \bar{D}^0\pi^+)$=$(4.61{\pm}0.1){\times}10^{-3}$ [4], compared with the SM branching fraction $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ =$(4.24{\pm}0.24){\times}10^{-7}$ that was derived in Eq. (5.5). The larger branching fraction of $B^+ \to \bar{D}^0\pi^+$ ensures there will be plenty of events in the 362 fb^{-1} Belle II dataset which can be used to produce a statistically significant data-MC calibration factor for the $B^+ \to \mu^+\nu_\mu$ signal selection strategy.

$B^+ \to \bar{D}^0\pi^+$ reconstruction method

To perform the control mode analysis, $B^+ \to \bar{D}^0\pi^+$ is reconstructed in the same dataset described in Section 2.4.4, (excluding the $B^+ \to \mu^+\nu_\mu$ MC), with semileptonic FEI B-tagging. The charged MC events generated with at least one B meson proceeding as

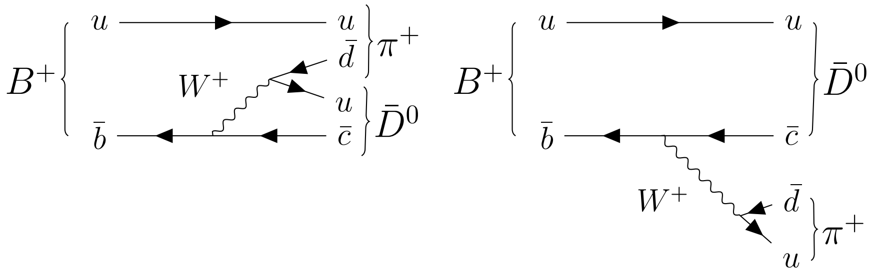


Figure 5.1. Feynman diagrams of the $B^+ \to \bar{D}^0\pi^+$ process.

$B^+ \to \bar{D}^0\pi^+$ are removed from the charged sample prior to reconstruction. The events removed from the charged sample are separately used as the $B^+ \to \bar{D}^0\pi^+$ signal MC for the reconstruction. The weighting of this MC component to reflect the number events expected in 362 fb^{-1} can be derived from the charged B component scale factor listed in Table 2.6. An additional ratio is multiplied into the scaling, to account for the difference between the PDG branching fraction and the branching fraction used in `EvtGen` event generation.

The pion in $B^+ \to \bar{D}^0\pi^+$ will be treated in analogue to the muon of $B^+ \to \mu^+\nu_\mu$. Consequently, the signal variable then becomes the $\Upsilon(4S)$ frame momentum of the pion in the $B^+ \to \bar{D}^0\pi^+$ decay, denoted by p^*_π. The $\bar{D}^0$ meson plays the role of the neutrino, which will not be used in event selection after its reconstruction. The process $B^+ \to \bar{D}^0\pi^+$ is reconstructed as $\Upsilon(4S) \to [B^+ \to [\bar{D}^0 \to K^+\pi^-]\pi^+]B^-_{\rm SL}$ (and charge conjugate), via `basf2`. The particle lists of kaons and pions used in reconstruction are created using the selections in Table 5.11. The signal pion (i.e. the pion that occurs directly from B^+, rather than the pion that occurs from $\bar{D}^0$) is chosen according to highest p^*_π of the event, similar to the muon being chosen according to p^*_μ. If there are remaining pions in the event, the $\bar{D}^0$ can then be reconstructed using the secondary second pion via $\bar{D}^0 \to K^+\pi^-$, with a pre-selection of $1.7\,\mathrm{GeV}/c^2 < m_{\bar{D}^0} < 1.95\,\mathrm{GeV}/c^2$. This pre-selection was chosen due its performance as a pre-selection for D mesons within semileptonic FEI. After combining the best signal pion with $\bar{D}^0$, the resulting best signal B candidate is chosen according which candidate has the $\bar{D}^0$ reconstructed mass closes to the nominal mass. After combining this signal B meson with the $B_{\rm tag}$ from the semileptonic FEI as $\Upsilon(4S) \to B^+_{sig}B^-_{\rm SL}$ (and charged conjugate), best $\Upsilon(4S)$ candidate of each event is chosen according to which $\Upsilon(4S)$ contains the $B_{\rm tag}$ with the highest FEI signal probability. All other steps of the `basf2` steering described in Section 2.4.3 for the $B^+ \to \mu^+\nu_\mu$ process are also performed in the control mode, such as the construction of the ROE object.

Table 5.11. Selections implemented in pions and kaons used in $B^+ \to \bar{D}^0\pi^+$. These are recommendations by Belle II's HadronID performance group.

Hadron selection (from track recommendations)	**Explanation**
`dr` < 0.5 *and* `abs(dz)` < 3 (cm)	Maximum transverse and z-distance of track origin from interaction point, ensures muon originates from $\Upsilon(4S)$ event.
`pt` > 0.2 (GeV/c)	Minimum transverse momentum, ensures $\mu_{\rm sig}$ has enough outwards momentum to reach later subdetectors like the KLM.
`thetaInCDCAcceptance == True`	Only accept signal muons produced at polar angle θ that can reach CDC subdetector, $\theta \in (17°, 150°)$.
Hadron selection (required to obtain hadron ID corrections)	**Explanation**
`kaonID` > 0.6 *or* `pionID` > 0.6	Minimum threshold on tracks implemented in particle list, must have 0.60 likelihood of being desired particle (i.e. K or π) over other charged particles.
$0.3 \geq$ `p` ≥ 4.5 *and* $0.4 \geq$ `theta` ≥ 2.6 (GeV/c and rad)	Magnitude and angular ranges for which data/MC corrections exist.
`nCDCHists` > 20	Range of hits in the CDC for which data/MC corrections exist.

To properly calibrate the semileptonic-tagging and signal selection strategy of the $B^+ \to \mu^+\nu_\mu$ analysis, the events remaining after $B^+ \to \bar{D}^0\pi^+$ reconstruction must under go the same selection criteria as in the $B^+ \to \mu^+\nu_\mu$ signal reconstruction. Therefore, the selection criteria derived via the $\frac{S}{\sqrt{S+B}}$ optimisation for the $B^+ \to \mu^+\nu_\mu$ case in Section 4.2.3 is applied to the $B^+ \to \bar{D}^0\pi^+$ data and MC. One key change that must be made is in the definition of the $B^+ \to \bar{D}^0\pi^+$ signal region. It was noted in the derivation of the signal muon momentum in the B-frame in Section 1.2.1 that if the neutrino were massive, this would alter the expression in Eq. (1.17) and thus result in a lower momentum for the signal muon. As $B^+ \to \bar{D}^0\pi^+$ is a two-body decay with two massive particles, it is therefore expected that the signal pion momentum in the $B^+ \to \bar{D}^0\pi^+$ decay has a lower magnitude than what is seen for the signal muon. The magnitude of the pion momentum in the signal-B frame can be derived by substituting D for ν, and π for μ in Eq. (1.15), resulting in $p_\pi^B = 2.31\,\text{GeV}/c < p_\mu^B = 2.64\;\text{GeV}/c$. Taking into account the smearing of the singular B-frame momentum when measured in the $\Upsilon(4S)$ frame, the bounds of the signal and sideband regions defined for the control mode must be altered.

After consulting the distribution of p^*_π in Figure 5.2 (which will be discussed further in the next section), $p^*_\pi \in$[2.15,2.5) GeV/c is defined as the signal region appropriate for the control mode and $p^*_\pi \in$[1.8,2.15) GeV/c is treated as the related sideband. To attempt to match the $B^+ \to \mu^+\nu_\mu$ case, the intervals were chosen such that their widths matched the $B^+ \to \mu^+\nu_\mu$ regions, despite being shift lower by 350 MeV/c. Notably, the interval of momenta used as the signal region for $B^+ \to \bar{D}^0\pi^+$ is identical to the region used as the sideband of $B^+ \to \mu^+\nu_\mu$. This may lead to lower efficiency in the control mode, as the selection criteria for $B^+ \to \mu^+\nu_\mu$ was optimised for the [2.5,2.85) GeV/c interval. To investigate this, the following section will look at the distributions of $B^+ \to \bar{D}^0\pi^+$ reconstructed events.

$B^+ \to \bar{D}^0\pi^+$ distributions

To assess the $B^+ \to \bar{D}^0\pi^+$ reconstruction in data and MC, distributions of such events are presented after selection criteria is applied. Additionally, all data-MC corrections that were applied to the $B^+ \to \mu^+\nu_\mu$ sample have also been applied to the $B^+ \to \bar{D}^0\pi^+$ reconstruction. With hadrons instead of leptons in the control mode's signal B decay, hadron-ID corrections take the place of lepton-ID corrections. The origin of hadron-ID corrections will be discussed in Section 5.2.2. Error bars are thus inclusive of Poisson statistical error, as well as error from hadron-ID.

A key change from the presentation of $B^+ \to \mu^+\nu_\mu$ data-MC distributions is that there will be no blinding of the $B^+ \to \bar{D}^0\pi^+$ signal region. With the event selection and data-MC correction study fixed from the main $B^+ \to \mu^+\nu_\mu$ analysis, there is no need to conceal the $B^+ \to \bar{D}^0\pi^+$ signal region. More importantly though is that the data yield in $B^+ \to \bar{D}^0\pi^+$ signal region is required to define the data/MC efficiency scaling factor for the $B^+ \to \mu^+\nu_\mu$ signal efficiency $\epsilon_{\text{sig+tag}}$, in its contribution to the $B^+ \to \mu^+\nu_\mu$ branching fraction. Thus, Figure 5.2 demonstrates the $\Upsilon(4S)$-frame signal pion momentum for events in data and MC, in both the sideband and signal regions. Contrasting with the $B^+ \to \mu^+\nu_\mu$ counterpart in Figure 4.56a, there is a much larger amount of $B^+ \to \bar{D}^0\pi^+$ events visible in the signal region, owing to the larger $B^+ \to \bar{D}^0\pi^+$ branching fraction. The overlaid signal shape, scaled identically to the $B^+ \to \bar{D}^0\pi^+$ signal yield, matches that of the overlaid signal muon shape seen in Figure 4.56a. This indicates that the

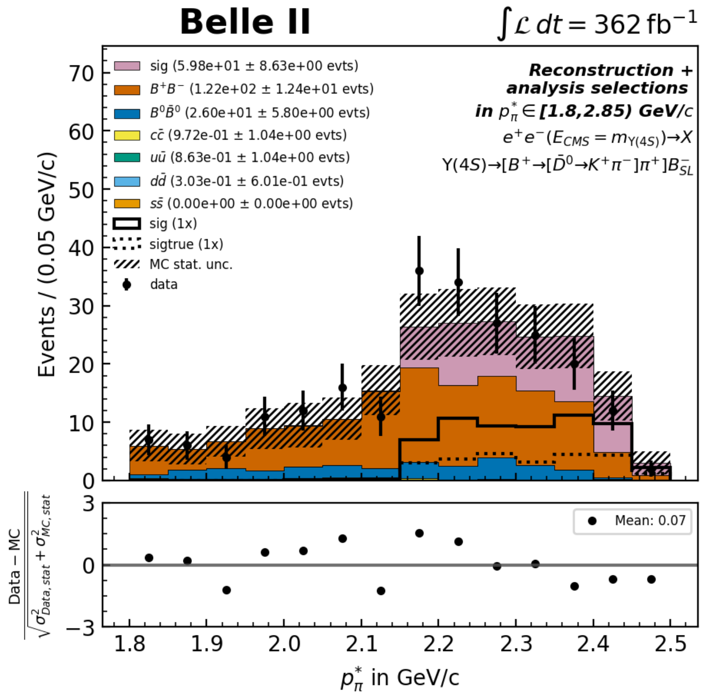
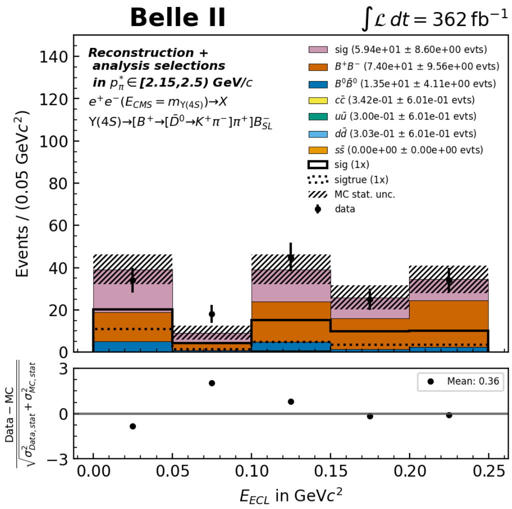


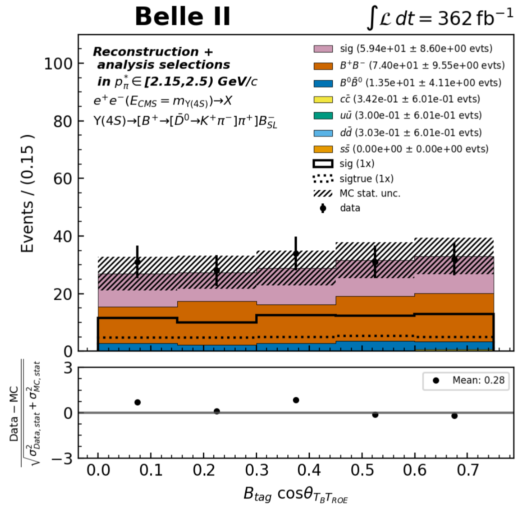
Figure 5.2. Signal pion $\Upsilon(4S)$-frame momentum for events reconstructed as $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+]B^-_{\text{SL}}$, with the same selection criteria as was used for the $B^+ \to \mu^+\nu_\mu$ reconstruction in Chapter 4. Stacking is according to event type, each scaled to $362\,\text{fb}^{-1}$. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

substitution of the signal pion in $B^+ \to \bar{D}^0\pi^+$ appropriately models the muon signal. Moreover, the shape of the stacked MC matches that of the data well within error bars, indicating the remaining charged and mixed MC reliably model the events in data. Differences between the original p^*_μ distribution and this p^*_π distribution includes the steady number of events in the $B^+ \to \bar{D}^0\pi^+$ sideband, while the sideband of $\Upsilon(4S)$-frame muon momentum gradually decreases. Notably, any $q\bar{q}$ contributions are overwhelmed by $B\overline{B}$ events, whereas the low statistics within the $B^+ \to \mu^+\nu_\mu$ version make $c\bar{c}$ events visible in Figure 4.56a. Regardless, it is demonstrated the $B^+ \to \mu^+\nu_\mu$ selection criteria are successful in isolating $B^+ \to \bar{D}^0\pi^+$ events from $B\overline{B}$ backgrounds when looking at the signal pion momentum.

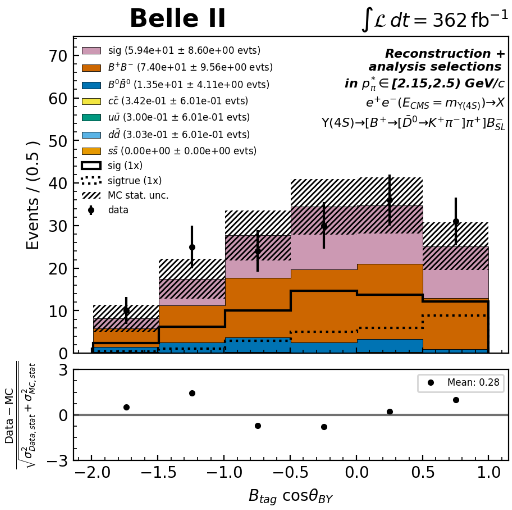
To complete the examination of $B^+ \to \bar{D}^0\pi^+$ reconstruction in data and MC, distributions of the same variables used to assess the quality of the $B^+ \to \mu^+\nu_\mu$ reconstruction will be consulted. In the plots of remaining variables investigated for the control mode, the full set of selection criteria derived for the $B^+ \to \mu^+\nu_\mu$ analysis will be applied to the $B^+ \to \bar{D}^0\pi^+$ events, with the p^*_π limited to the control mode signal region. Binning



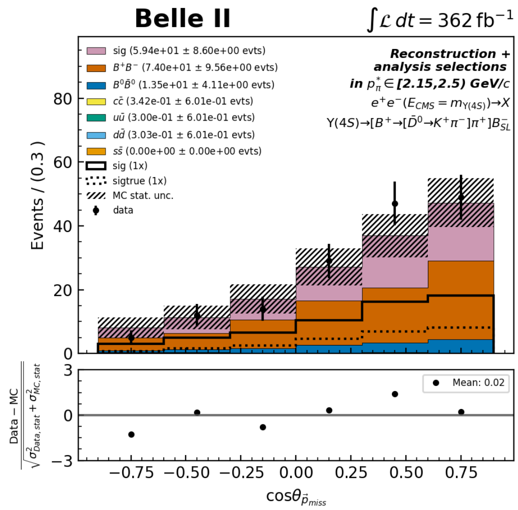
(a) Energy remaining in ECL after $\Upsilon(4S)$ reconstruction.



(b) Cosine angle between B_{tag} thrust axis and thrust axis derived from all other particles in the event.



(c) Cosine angle between visible particles of semileptonic B_{tag} and the nominal B momentum.



(d) Cosine polar angle between the beamline and the total visible momentum of the event.

Figure 5.3. Selection variables: distributions of key variables reconstructed via the control mode $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+]B^-_{\text{SL}}$. Stacking is according to event type, each scaled to $362\,\text{fb}^{-1}$. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.

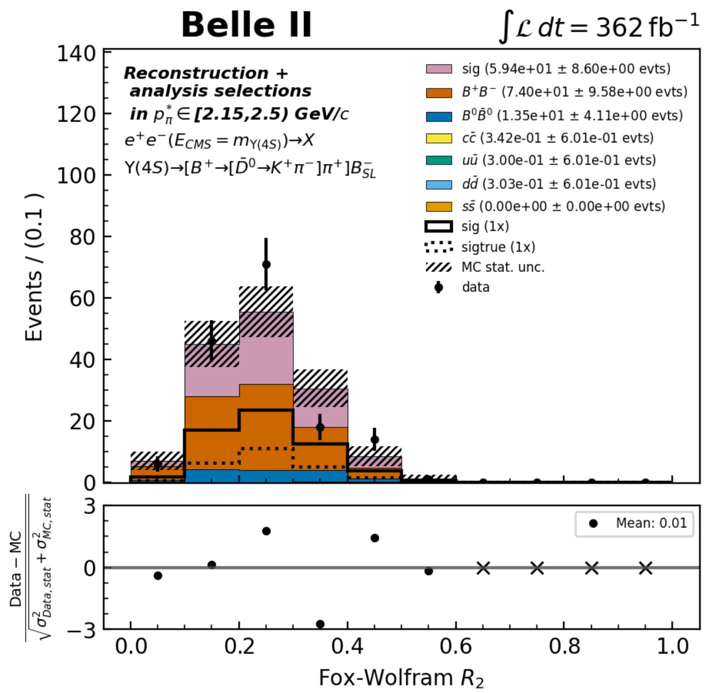
has been chosen to allow comparison with the MC-only signal region $B^+ \to \mu^+\nu_\mu$ distributions in Section 4.2.4. Comparison generally cannot be made with the data-MC distributions in Section 4.3.2, as these were made for events in sidebands of p^*_μ or E_{ECL}, so as to keep the $B^+ \to \mu^+\nu_\mu$ signal region blind in data.

The set of plots in Figure 5.3 contain the variables used in the selection criteria derived for the $B^+ \to \mu^+\nu_\mu$ analysis and now used in the $B^+ \to \bar{D}^0\pi^+$ reconstruction. Ignoring the differences owed to the yields of the two decay modes, the MC shapes in the control mode generally match that of the $B^+ \to \mu^+\nu_\mu$ case in Figure 4.52. The only difference is in the distribution of the unused energy in the ECL distribution, when compared between the $B^+ \to \bar{D}^0\pi^+$ mode in Figure 5.3a, and the $B^+ \to \mu^+\nu_\mu$ version in Figure 4.52a. For $B^+ \to \bar{D}^0\pi^+$, the zero energy bin is level with that of the following bins, (aside from the lack of events in the adjacent bin originating from the ROE mask threshold), while for $B^+ \to \mu^+\nu_\mu$ the zeroth bin has the largest number of events. This indicates that it is more likely for the $B^+ \to \bar{D}^0\pi^+$ reconstruction method to leave extra energy in the ECL than in $B^+ \to \mu^+\nu_\mu$ reconstruction. This could originate from $B^+ \to \bar{D}^0\pi^+$ containing more particles on the signal side, thus a higher probability of radiative photons producing extra clusters that are missed from the $\Upsilon(4S)$ reconstruction.

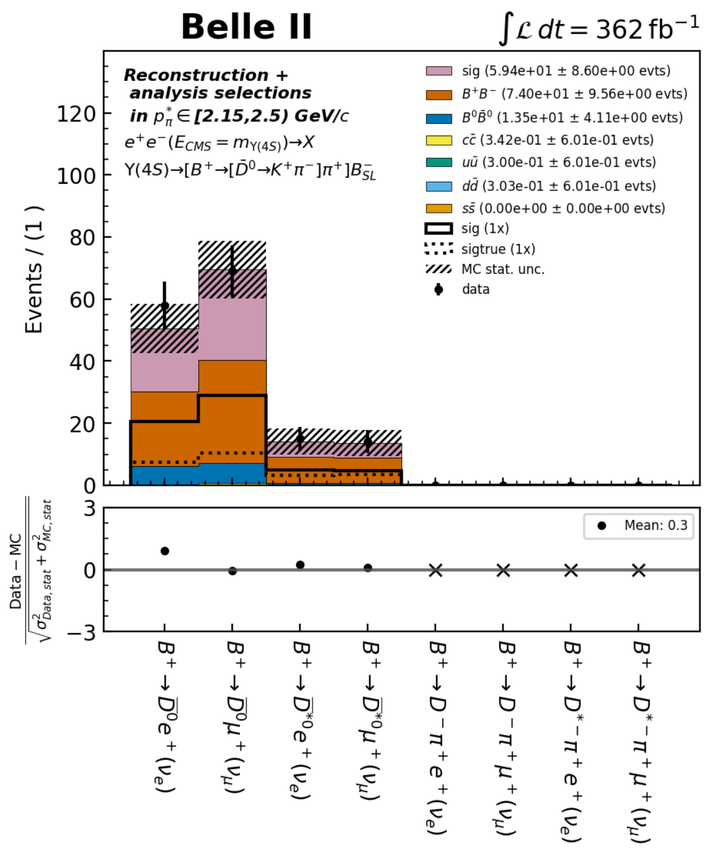
The next set of variables to present are in Figure 5.4, contains distributions of the final selection variable, the B_{tag} decay mode, in addition to the Fox-Wolfram R_2. Both match the trends set by $B^+ \to \mu^+\nu_\mu$ events in Figure 4.53.

The distributions of Figure 5.5 are of variables specific to the $B^+ \to \bar{D}^0\pi^+$ mode. Figure 5.5a contains the values of M_{bc} for signal B mesons obtained in reconstructing $B^+ \to \bar{D}^0\pi^+$ events in data and MC. The beam-constrained mass M_{bc} was introduced in Section 2.2.2, as a substitute for the invariant mass of a B candidate. As $B^+ \to \bar{D}^0\pi^+$ is a hadronic decay, the full B four-vector can be used to derive M_{bc} close to the nominal B mass. Such a constraint is not available in the $B^+ \to \mu^+\nu_\mu$ decay. The M_{bc} distribution for the control mode peaks at the nominal B mass of $5.28\,\mathrm{GeV}/c^2$, indicating that the procedure described in Section 5.2.2 yields well reconstructed signal B mesons. The narrow width, and thus accuracy, of the signal B meson M_{bc} peak can be compared to the M_{bc} distributions of hadronic FEI-tagged B mesons obtained from charged MC in Figure 3.7a.

Figure 5.5b contains the distribution of the difference between beam-constrained energy and the reconstructed energy, for $B^+ \to \bar{D}^0\pi^+$ reconstructed-events. The variable ΔE was introduced alongside M_{bc} in Section 2.2.2. Whilst also more narrow than the

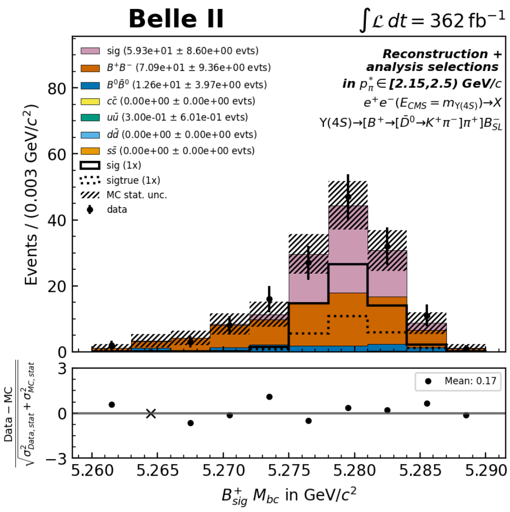


(a) Fox-Wolfram R_2 of events.

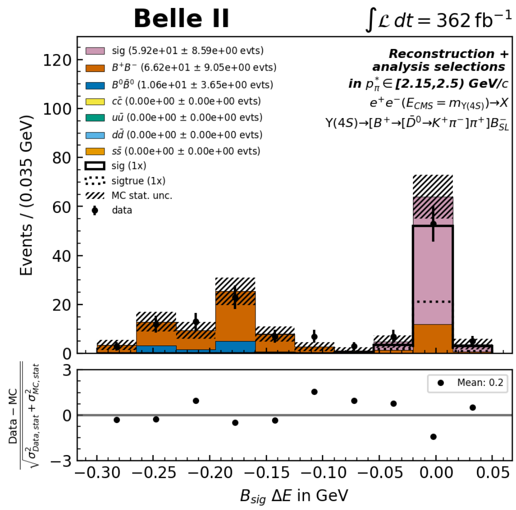


(b) Tag-side B FEI decay mode.

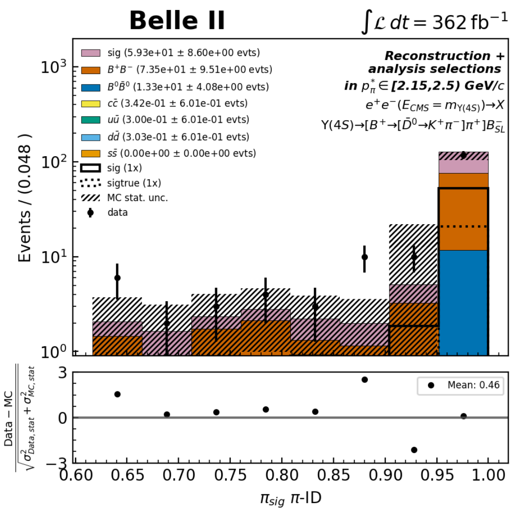
Figure 5.4. B_{tag} decay modes and Fox-Wolfram R_2: distributions of key variables reconstructed via the control mode $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+]B^-_{\mathrm{SL}}$. Stacking is according to event type, each scaled to $362\,\mathrm{fb}^{-1}$. 'Sigtrue' indicates the proportion of the scaled signal component that has been reconstructed correctly.



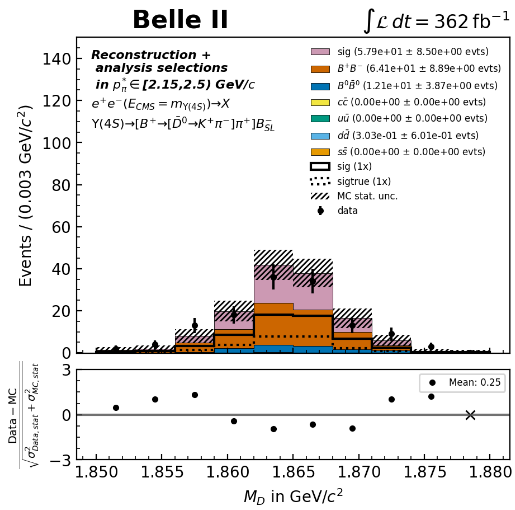
(a) Signal-side B beam-constrained mass.



(b) Signal-side B energy difference between measured and beam-constrained energies.



(c) Pion identification probability of signal pion.



(d) D from $B^+ \to \bar{D}^0\pi^+$ (and charge conjugated) mass.

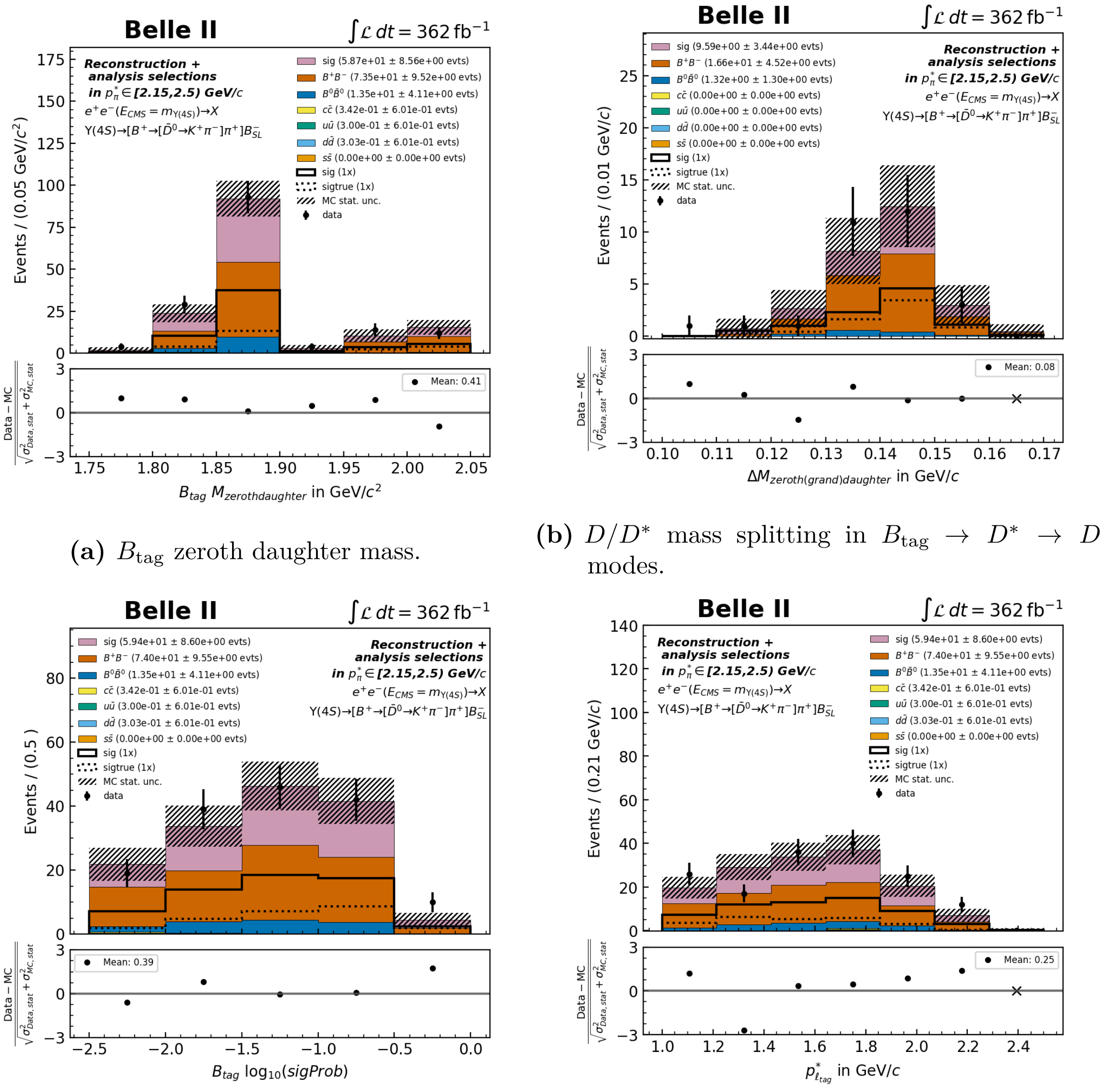
Figure 5.5. $B^+ \to \bar{D}^0\pi^+$ specific variables: distributions of key variables reconstructed via the control mode $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+\,]B^-_{\mathrm{SL}}$. Stacking is according to event type, each scaled to $362\,\mathrm{fb}^{-1}$. ‘Sigtrue’ indicates the proportion of the scaled signal component that has been reconstructed correctly.

hadronic FEI counterpart in Figure 3.7b, it also more asymmetrical. The Figure contains two structures, one centred on the nominal $\Delta E = 0\,\text{GeV}$ for correctly reconstructed B mesons, as well as a wider structure with $\Delta E < 0\,\text{GeV}$. All signal events are within 5 MeV of ΔE =0. Remaining charged events are split between the central peak and the lower peak, while all mixed events are in the lower peak. This further indicates that the energy of reconstructed signal B mesons is consistent with constraints provided by $\Upsilon(4S)$ $\to B\overline{B}$ kinematics at Belle II.

With the pion of $B^+ \to \bar{D}^0\pi^+$ taking on the role of the muon in $B^+ \to \mu^+\nu_\mu$, one can also examine the particle identity probability of the signal pion. Instead of consulting `muonID`, the variable relevant to the signal pion is `pionID`, which was used to obtain tracks with a 60% or greater probability of being a true pion. This same selection was used to obtain the pion daughter in the D reconstruction. Comparison with the MC-only `muonID` distribution for the signal muon in Figure 4.55a is challenging due to the logarithmic event-scale and the larger $B^+ \to \bar{D}^0\pi^+$ yield. However, it can be seen that both Figures contain the bulk of their events in `pionID` > 0.9 or `muonID` > 0.9, again demonstrating the suitability of the signal pion in modelling the muon of $B^+ \to \mu^+\nu_\mu$.

The last control-mode-only variable to inspect is the mass of the signal-side D meson, presented in Figure 5.5d. With the signal D chosen to be the candidate with the smallest reconstructed mass difference from the nominal mass, this has resulted in all signal D meson candidates close to the nominal D mass of $1.86\,\text{GeV}/c^2$. The bins closer to the nominal D mass is also where the signal $B^+ \to \bar{D}^0\pi^+$ MC component peaks, verifying that the D reconstruction technique is sound for the control mode.

Figure 5.6 contains the final distributions to be examined for the control mode, those concerning the semileptonic FEI B_{tag} quality. The same tagging strategy used in the $B^+ \to \mu^+\nu_\mu$ signal analysis has been used in this $B^+ \to \bar{D}^0\pi^+$ control mode analysis, so it is not expected that the control mode B_{tag} variables show considerable differences with those in the $B^+ \to \mu^+\nu_\mu$ analysis. The width of the signal-side D mass peak in Figure 5.5d can be compared to that of the tag-side D mass peak from the B_{tag} zeroth daughter mass plot in Figure 5.6a, where the best D meson candidate within the semileptonic FEI is also chosen based on smallest absolute mass difference. The D/D^* mass-splitting distribution Figure 5.6b for the tagged-B mesons in the control mode reconstruction generally matches the shape of the tagged-B meson shown in Figure 4.54d corresponding to the $B^+ \to \mu^+\nu_\mu$ analysis. There is minor shape difference in the B_{tag} signal probability plots between the $B^+ \to \bar{D}^0\pi^+$ version(Figure 5.6c) and the $B^+ \to \mu^+\nu_\mu$ version (Figure 4.54a). Specifically, the yield of the -2.5 bin contains the



(a) B_{tag} zeroth daughter mass.

(b) D/D^* mass splitting in $B_{\text{tag}} \to D^* \to D$ modes.

(c) B_{tag} $\log_{10}$ signal probability.

(d) Tag-side lepton momentum in $\Upsilon(4S)$ frame.

Figure 5.6. Additional B_{tag} variables: distributions of key variables reconstructed via the control mode $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+]B^-_{\text{SL}}$. Stacking is according to event type, each scaled to $362\,\text{fb}^{-1}$. ‘Sigtrue’ indicates the proportion of the scaled signal component that has been reconstructed correctly.

maximum of the mixed and $c\bar{c}$ events for B_{tag} used alongside $B^+ \to \mu^+\nu_\mu$ reconstruction, while this same bin contains a minimum of mixed background for $B^+ \to \bar{D}^0\pi^+$. Whether this is a low-statistics effect within the $B^+ \to \mu^+\nu_\mu$ sample or indicative of a difference in B-tagging within $B^+ \to \bar{D}^0\pi^+$ reconstructed events will need to be monitored when further Belle II data is available. There is a similar difference in the tag-side lepton momentum, between the $B^+ \to \bar{D}^0\pi^+$ and $B^+ \to \mu^+\nu_\mu$ reconstructions, in Figure 5.6d and Figure 4.54c respectively. This difference can also be attributed to the bin where a $c\bar{c}$ event resides, where the leptons tagging the $B^+ \to \bar{D}^0\pi^+$ decay gradually increases from 1 to 1.8 GeV/c momentum, while those that tag $B^+ \to \mu^+\nu_\mu$ sharply increase from the zeroth bin to the first bin. Again, this may be the result of lower statistics within the $B^+ \to \mu^+\nu_\mu$ signal region, as opposed to the $B^+ \to \bar{D}^0\pi^+$ one. A shape difference in the tag-side lepton momentum may also be expected given the requirement that the tag-side lepton must have lower momentum than that of the signal side particle. The selection $p^*_\mu > p^*_\ell$ prior to signal side reconstruction was found to be effective for $B^+ \to \mu^+\nu_\mu$ in Section 4.1.2. This is potentially less effective for $B^+ \to \bar{D}^0\pi^+$, with the pionic signal region beginning at 2.15 GeV/c having more overlap with the tag-side lepton momentum range.

The distributions of key variables concerning the $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+\,]B^-_{\text{SL}}$ (and charge conjugate) process have been presented, demonstrating good agreement between histograms created via collision data and those in MC-simulated events. This indicates that the MC used in $B^+ \to \bar{D}^0\pi^+$ reconstruction accurately models data in both the sidebands and signal region of the signal pion $\Upsilon(4S)$ frame momentum. In addition to good agreement in the signal variable, data-MC agreement was also seen when probing the data on the basis of signal-side and tag-side variables, indicating that $B^+ \to \mu^+\nu_\mu$ reconstruction strategy in the signal and tag B mesons can be extended to that of the $B^+ \to \bar{D}^0\pi^+$ mode. With this determined, the control mode analysis can move on to determining the yields necessary to evaluate the data-MC efficiency correction to the $B^+ \to \mu^+\nu_\mu$ signal efficiency, knowing that the results from $B^+ \to \bar{D}^0\pi^+$ can be accurately to make inferences about $B^+ \to \mu^+\nu_\mu$.

$B^+ \to \bar{D}^0\pi^+$ results

Table 5.12 contains the event yields of the $B^+ \to \bar{D}^0\pi^+$ reconstruction with the $B^+ \to \mu^+\nu_\mu$ selection criteria applied, in the signal and sideband regions shifted to suit the signal pion momentum. The distributions presented in the previous section demonstrated

Table 5.12. Yields in signal region and sideband region for the control mode $B^+ \to \bar{D}^0\pi^+$. All corrections, including those from hadron ID have been applied. Error quoted is derived from Poisson statistical uncertainty.

Sample	SR	SB
$B^+ \to \bar{D}^0\pi^+$	59.35±7.7	0.41±0.64
B^+B^-	73.96±8.6	48.46±6.96
$B^0\overline{B}^0$	13.53±3.68	12.5±3.53
$c\bar{c}$	0.34±0.58	0.63±0.79
$u\bar{u}$	0.3±0.55	0.56±0.75
$d\bar{d}$	0.3±0.55	0
$s\bar{s}$	0	0
tot. bkg.	88.44±9.4	62.15±7.88
$\frac{S}{\sqrt{S+B}}$	4.88±0.53	0.05±0.08
MC	147.79±12.16	62.55±7.91
data	156.0±12.49	67.0±8.19
$\frac{\text{data}}{\text{MC}}$	1.06±0.12	1.07±0.19
$\frac{\text{data-MC}}{\sqrt{\sigma^2_{\text{MC}}+\sigma^2_{\text{data}}}}$	0.47±1.0	0.39±1.0
$\text{sig}_{\text{SR}} \times \left(\frac{\text{data}}{\text{MC}}\right)_{\text{SB}}$	63.57±13.89	-
$\text{bkg}_{\text{SR}} \times \left(\frac{\text{data}}{\text{MC}}\right)_{\text{SB}}$	94.73±19.46	-

that there would be a much larger $B^+ \to \bar{D}^0\pi^+$ signal yield than what is expected in $B^+ \to \mu^+\nu_\mu$, as desired in its role as a control mode. Within the signal region, 59.35 $B^+ \to \bar{D}^0\pi^+$ signal events have been obtained, with 88.44 background events, consisting mostly of charged $B\overline{B}$ events. From this, a compelling significance estimate of $\frac{S}{\sqrt{S+B}}$ =4.88 is obtained. Stemming from the continued data-MC agreement in the distributions of the previous section, the data/MC ratios are well within agreement of statistical uncertainties between the two regions. With both ratios being equivalent with unity, applying the data/MC ratio of the sideband to the MC signal and background yields returns yields comparable with the original MC values.

Table 5.13 contains the yields from Table 5.12 that will be used in the calculation of the data-driven $B^+ \to \bar{D}^0\pi^+$ signal efficiency. This table also contains the first calculation of the MC-driven $B^+ \to \bar{D}^0\pi^+$ signal efficiency, evaluated using the number of generated $B^+ \to \bar{D}^0\pi^+$ events. This value is lower than the MC-driven signal efficiency of $B^+ \to \mu^+\nu_\mu$, which can be expected given the selection criteria was optimised for the $B^+ \to \mu^+\nu_\mu$ analysis and not $B^+ \to \bar{D}^0\pi^+$. The table also contains the statistical uncertainty of these

Table 5.13. Final yields with statistical uncertainty and proportional systematic uncertainty, in signal region and sideband region for control mode $B^+ \to \bar{D}^0\pi^+$. Composite values indicated by (*).

	$N \pm \Delta N_{\text{stat}}$	$\%\Delta N_{\text{stat}}$
$N^{\text{data}}_{\text{SR,all}}$	156.00±12.49	8.01
$N^{\text{MC}}_{\text{SR,bkg}}$	88.44±9.4	10.63
$N^{\text{data}}_{\text{SB,all}}$	67.0±8.19	12.22
$N^{\text{MC}}_{\text{SB,all}}$	62.55±7.91	12.64
$N^{\text{data}}_{\text{SR,bkg}}$ *	94.73±19.46	20.55
$N^{\text{MC}}_{\text{SR(gen),sig}}$	466.02±21.59	4.63
$\epsilon^{\text{MC}}_{\text{sig+tag}}$ *	$(3.18\pm0.15)\times10^{-5}$	4.63

yields, obtained as $\sqrt{N}$ Poisson uncertainty and propagated accordingly for the composite values of $N^{\text{data}}_{\text{SR,bkg}}$ and $\epsilon^{\text{MC}}_{\text{sig+tag}}$. Considering the uncertainties as percentages relative to the yields, the statistical uncertainties are generally lesser than those of $B^+ \to \mu^+\nu_\mu$, which were presented in Table 4.21. This can be expected due to the smaller yields in $B^+ \to \mu^+\nu_\mu$ at 362 fb^{-1}, and the small-N behaviour of the percentage statistical uncertainty derived as $\frac{\sqrt{N}}{N}$.

$B^+ \to \bar{D}^0\pi^+$ efficiency calculation

With the yields of the $\Upsilon(4S) \to [B^+ \to \bar{D}^0\pi^+]B_{\text{SL}}$ control mode reconstruction obtained, the data/MC correction to be applied to the $B^+ \to \mu^+\nu_\mu$ signal efficiency can be constructed.

The first term required is the MC-based selection efficiency of the $B^+ \to \bar{D}^0\pi^+$ mode. This was evaluated in Table 5.13, in a similar manner to how the selection efficiency was calculated for $B^+ \to \mu^+\nu_\mu$ in Table 4.21. In this case, $N^{\text{MC}}_{\text{gen},B^+\to\bar{D}^0\pi^+}$, the number of generated $B^+ \to \bar{D}^0\pi^+$ decays, is treated as double the number of events which contain at least one $B^+ \to \bar{D}^0\pi^+$ candidate in Table 5.10. This is done as both B mesons of the event were allowed to proceed as $B^+ \to \bar{D}^0\pi^+$ (and charge conjugate) in the charged MC sample from which the signal MC was created. This factor of 2 was not needed in the $B^+ \to \mu^+\nu_\mu$ sample, as only one B was fixed to proceed as $B^+ \to \mu^+\nu_\mu$ (or charged conjugate) within the same event. This MC-based selection efficiency is denoted as $\epsilon^{\text{MC}}_{\text{sig+tag},B^+\to\bar{D}^0\pi^+}$ and is explicitly evaluated in the following equation, where errors are propagated statistical uncertainty.

$$\epsilon^{\mathrm{MC}}_{\mathrm{sig+tag},B^+\to\bar{D}^0\pi^+} = \frac{N^{\mathrm{MC}}_{\mathrm{SR(gen)},B^+\to\bar{D}^0\pi^+}}{N^{\mathrm{MC}}_{\mathrm{gen},B^+\to\bar{D}^0\pi^+}} \tag{5.6}$$

$$\epsilon^{\mathrm{MC}}_{\mathrm{sig+tag},B^+\to\bar{D}^0\pi^+} = (3.18 \pm 0.15) \times 10^{-5} \tag{5.7}$$

A data-based selection efficiency of $B^+ \to \bar{D}^0\pi^+$ can also be derived using an alternate, independent method. The expression for experimental branching fractions in Eq. (5.5) can be rearranged for $\epsilon_{\mathrm{sig+tag}}$. Relabelling the indicies to suit the $B^+ \to \bar{D}^0\pi^+$ control mode, the equation becomes the following.

$$\epsilon^{\mathrm{data}}_{\mathrm{sig+tag}} = \frac{N^{\mathrm{data}}_{\mathrm{SR,all}} - N^{\mathrm{data}}_{\mathrm{SR,bkg}}}{\mathcal{B}(B^+ \to \bar{D}^0\pi^+) \times 2f^{+-}N^{\mathrm{data}}_{362\mathrm{fb}^{-1},B\bar{B}}} \tag{5.8}$$

Both $N^{\mathrm{data}}_{\mathrm{SR,all}}$ and $N^{\mathrm{data}}_{\mathrm{SR,bkg}}$ for the control mode were evaluated in Table 5.13. The branching fraction of $B^+ \to \bar{D}^0\pi^+$ is known to be $\mathcal{B}(B^+ \to \bar{D}^0\pi^+)$=$4.61\times10^{-3}$ [4]. The remaining parameters in Eq. (5.8) were summarised in Table 5.1. Thus, the selection efficiency derived directly from the expected $B^+ \to \bar{D}^0\pi^+$ branching fraction and yields from the control mode analysis evaluates to the following expression.

$$\epsilon^{\mathrm{data}}_{\mathrm{sig+tag},B^+\to\bar{D}^0\pi^+} = (3.34 \pm 1.26) \times 10^{-5} \tag{5.9}$$

By fixing the physical $B^+ \to \bar{D}^0\pi^+$ branching fraction and using yields from the control mode, it is found that this data-based efficiency is in agreement with the MC-based efficiency from Eq. (5.7). It is noted that the statistical uncertainty is much larger in the data-based result in Eq. (5.9) than the MC-based result. This can be attributed to the data-based result being created from multiple individual quantities, whose uncertainties had to be combined according to the equation, whereas the MC-based result only relied on one yield.

With near-equal efficiencies, this suggests that the signal selection strategy has mostly equivalent sensitivity to events in both data and MC reconstructed as $B^+ \to \bar{D}^0\pi^+$, with a slight underestimation of the signal events in data by the MC modelling. As the control mode dynamics have been seen to match that of the $B^+ \to \mu^+\nu_\mu$, it is assumed that a similar relationship in data and MC efficiencies will be present in the signal region of the $B^+ \to \mu^+\nu_\mu$ analysis. To thus ensure that the MC-based signal efficiency of the

$B^+ \to \mu^+\nu_\mu$ signal does not underestimate the efficiency expected in data, the ratio of the data-based and MC-based $B^+ \to \bar{D}^0\pi^+$ efficiencies will be used to form a correction factor. The ratio of the $B^+ \to \bar{D}^0\pi^+$ efficiencies evaluate to the following value, where uncertainties quoted are also Poisson statistical uncertainties.

$$\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}} = 1.05 \pm 0.40 \tag{5.10}$$

The correction factor to the selection efficiency derived from the $B^+ \to \bar{D}^0\pi^+$ control mode and previously measured $B^+ \to \bar{D}^0\pi^+$ branching fraction has been evaluated. Before using this correction to calibrate the previously measured $B^+ \to \mu^+\nu_\mu$ MC-based signal efficiency, the systematic contributions to this correction will be discussed.

$B^+ \to \bar{D}^0\pi^+$ systematic uncertainty considerations

In order to make the best comparison between data and MC in the $B^+ \to \bar{D}^0\pi^+$ control mode reconstruction, it was necessary to apply the same set of Belle II-recommended corrections that were applied to the $B^+ \to \mu^+\nu_\mu$ reconstruction. These corrections bring their own systematic uncertainty into the evaluation of $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$. Systematic uncertainty due to the $\frac{S}{\sqrt{S+B}}$ optimisation, as well as the statistical uncertainty can also be considered for this ratio. To quantify these uncertainties in $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$, the same methods used to evaluate the uncertainty in $B^+ \to \mu^+\nu_\mu$ yields in Section 5.2.1 have been applied to the yields of the $B^+ \to \bar{D}^0\pi^+$ reconstruction. The resulting systematic uncertainties, separated by source and affected measurement is presented in Table 5.14.

The only altered data-MC correction technique was to implement hadron-ID specific corrections to the signal pion, rather than the lepton-ID corrections relevant to the signal muon of $B^+ \to \mu^+\nu_\mu$. The hadron-ID corrections compensate for differences in the performance of the hadron-ID likelihoods in data and MC events, by supplying an event scaling to MC events, similar to what was described for the case of lepton-ID in Section 5.2.1. The event scalings used in the control mode analysis were the product of three weightings, given the `kaonID` > 0.6 selection applied to the kaon, and `pionID` > 0.6 applied to the two separate pions in the $B^+ \to [\bar{D}^0 \to K^+\pi^-]\pi^+$ final state. The individual weightings are derived using studies of $D^{*+} \to [D^0 \to K^+\pi^-]\pi^+$ [122] and $K^0_S \to \pi^+\pi^-$ [123] in tightly constrained data and MC events, performed by the Belle II Hadron ID group. The weightings are applied according to each candidate hadron's

Table 5.14. Summary of uncertainties in yields from $B^+ \to \bar{D}^0\pi^+$ control mode reconstruction. Statistical uncertainty is denoted as ΔN_{stat}. Systematic uncertainty due to hadron-ID is denoted as ΔN_{HID}, due to uncertainty to track momentum scaling and magnetic field modelling denoted as $\Delta N_{\text{mom.scale}}$, and due to mid-to-high momentum tracking efficiency as $\Delta N_{\text{mom.eff.}}$.

	N	ΔN_{stat}	$\%\Delta N_{\text{stat}}$	ΔN_{HID}	$\%\Delta N_{\text{HID}}$
$N^{\text{data}}_{\text{SR,all}}$	156.00	12.49	8.01	-	-
$N^{\text{MC}}_{\text{SR,bkg}}$	88.44	9.40	10.63	2.06	2.33
$N^{\text{data}}_{\text{SB,all}}$	67.00	8.19	12.22	-	-
$N^{\text{MC}}_{\text{SB,all}}$	62.55	7.91	12.64	6.77	10.82
$N^{\text{data}}_{\text{SR,bkg}}{}^*$	94.73	19.46	20.55	10.48	11.07
$N^{\text{MC}}_{\text{SR(gen),sig}}$	466.02	21.59	4.63	2.06	0.44
$\epsilon^{\text{MC}*}$	3.18×10^{-5}	0.15×10^{-5}	4.63	0.01×10^{-5}	0.44

	$\Delta N_{\text{mom.scale}}$	$\%\Delta N_{\text{mom.scale}}$	$\Delta N_{\text{mom.eff.}}$	$\%\Delta N_{\text{mom.eff.}}$
$N^{\text{data}}_{\text{SR,all}}$	1	0.64	-	-
$N^{\text{MC}}_{\text{SR,bkg}}$	-	-	0.21	0.24
$N^{\text{data}}_{\text{SB,all}}$	0.5	0.75	-	-
$N^{\text{MC}}_{\text{SB,all}}$	-	-	0.15	0.24
$N^{\text{data}}_{\text{SR,bkg}}{}^*$	0.71	0.75	0.32	0.34
$N^{\text{MC}}_{\text{SR(gen),sig}}$	-	-	1.12	0.24
$\epsilon^{\text{MC}*}$	-	-	0.01×10^{-5}	0.24

lab-frame momentum and polar angle, using a combination of the Belle II `PIDvar` [116] and `syscorrfw` [124] packages.

From Table 5.14, it is seen that the largest source of percentage uncertainty is due to the statistical uncertainty. The track momentum scaling applied to events in data and the mid-to-high momentum tracking efficiency corrections contribute sub-percent systematic uncertainties to the $B^+ \to \bar{D}^0\pi^+$ yields. The statistical and hadron-ID provides the largest sources of uncertainties. Notably, it is the quantity $N^{\text{data}}_{\text{SR,bkg}}$ that generally has the largest percentage uncertainty when compared to the other quantities for the same source. This can be attributed to its nature as a composite quantity, where its uncertainty is propagated from the yields used in its construction.

The final quantities and $B^+ \to \bar{D}^0\pi^+$ yields used to evaluate the data/MC efficiency ratio $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$ are presented in Table 5.15. Uncertainties are separated into two groups: statistical uncertainty and detector-related systematic uncertainties.Thus the final data/MC

Table 5.15. Final measurements for $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$.

Measurement	$N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$
$N^{\text{data}}_{\text{SR,all}}$	156.00$\pm$ 12.49 $\pm$ 1.00
$N^{\text{data}}_{\text{SR,bkg}}$ *	94.73$\pm$ 19.46 $\pm$ 10.51
ϵ^{MC}*	$(3.18 \pm 0.15 \pm 0.02)\times10^{-5}$
ϵ^{data}*	$(3.34 \pm 1.26 \pm 0.58)\ \times10^{-5}$
$\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$*	1.05 $\pm$ 0.40 $\pm$ 0.18

Table 5.16. Summary of systematic uncertainties and sources in evaluation of the data-MC efficiency ratio derived in $B^+ \to \bar{D}^0\pi^+$ processes.

Source	$\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$ [%]
`hadronID` correction	17.11
Track momentum scale	2.00
Tracking efficiency	0.57
$\mathcal{B}(B^+ \to \bar{D}^0\pi^+)$	2.17
f^{+-}	1.17
$N_{B\bar{B}}$	1.45

efficiency correction is quoted as 1.05$\pm$0.40 (stat) $\pm$ 0.18 (sys). With the decay width ratio (f^{+-}) and number of $B\overline{B}$ events measured in the Belle II dataset ($N^{\text{data}}_{362\text{fb}^{-1}}$) being used in the derivation of ϵ^{data}, the uncertainties of the parameters as quoted in Table 5.1 are treated as additional systematic uncertainties, thus propagated into ΔN_{sys} of ϵ^{data} and $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$. The uncertainty quoted on the $B^+ \to \bar{D}^0\pi^+$ branching fraction is treated similarly. Moreover, the impact of each systematic uncertainty contribution propagated and expressed as a percentage of $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$ is presented in Table 5.16. The largest contribution is in the hadron-ID correction, potentially due to the three hadrons that require correction in the final state.

The ultimate goal of performing the control mode reconstruction is to perform a data-based correction on the $B^+ \to \mu^+\nu_\mu$ signal efficiency. With the correction factor calculated with its systematic and statistical contributions, the factor can be incorporated into the $B^+ \to \mu^+\nu_\mu$ signal efficiency calculated in Table 4.21. The correction factor is applied according to the equation below, where quoted uncertainties are currently from the $B^+ \to \bar{D}^0\pi^+$ error sources only.

$$\epsilon^{\text{data}}_{\text{sig+tag},B^+\to\mu^+\nu_\mu} = \epsilon^{\text{MC}}_{\text{sig+tag},B^+\to\mu^+\nu_\mu}\left(\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}\right)_{B^+\to\bar{D}^0\pi^+} \tag{5.11}$$

$$= (7.83 \pm 2.98\ (\text{stat})\ \pm 1.37\ (\text{sys})\ ) \times 10^{-3} \tag{5.12}$$

A data-sensitive $B^+ \to \mu^+\nu_\mu$ signal efficiency has been obtained, by calibrating the signal selection strategy against a well known particle decay, $B^+ \to \bar{D}^0\pi^+$. This new $B^+ \to \mu^+\nu_\mu$ signal efficiency will now more closely represent the ratio of $B^+ \to \mu^+\nu_\mu$ events in data that are selected to those in the whole $362\,\text{fb}^{-1}$ dataset, rather than only representing this ratio in MC events as it did before. In the next section, the uncertainties from the control mode correction factor can be combined with the uncertainties in the $B^+ \to \mu^+\nu_\mu$ signal efficiency from the $B^+ \to \mu^+\nu_\mu$ yields directly, to obtain the final quantities needed to evaluate the $B^+ \to \mu^+\nu_\mu$ branching fraction.

5.2.3. Summary of systematic uncertainties and sources

With the final correction applied, the full set of systematic uncertainties discussed throughout Section 5.2 as applied to the parameters of the $B^+ \to \mu^+\nu_\mu$ branching fraction can be summarised. A table of the uncertainties derived and their sources is listed in Table 5.17. The table focuses on the two main quantities derived in the $B^+ \to \mu^+\nu_\mu$ search, prior to unblinding: the number of events in data estimated to originate from non-$B^+ \to \mu^+\nu_\mu$ processes in the signal region ($N^{\text{data}}_{\text{SR,bkg}}$) and the data-calibrated cumulative signal efficiency ($\epsilon^{\text{data}}_{\text{sig+tag},B^+\to\mu^+\nu_\mu}$)

In Table 5.17, the systematic uncertainty is expressed as a percentage of the observable the factor effects, presented according source and observable. For $\epsilon_{\text{sig+tag}}$, the first few sources are those affecting the MC-based $B^+ \to \mu^+\nu_\mu$ selection efficiency. While these same sources can also affect the data/MC efficiency correction from $B^+ \to \bar{D}^0\pi^+$, these contributions are expressed instead within the final listing of the table. Thus, with no track momentum scaling needed in the calculation of the MC-based $B^+ \to \mu^+\nu_\mu$ selection efficiency, there is no contribution listed in this entry of the table. The systematic component of the control mode efficiency correction is the quadrature sum of uncertainties from `hadronID` correction, tracking momentum scale correction and tracking efficiency correction, as derived in $B^+ \to \bar{D}^0\pi^+$. The statistical uncertainties in the control mode

Table 5.17. Summary of statistical and systematic uncertainties and sources in evaluation of the cumulative signal efficiency ($\epsilon_{\text{sig+tag}}$) and the number of signal region events in data estimated to be from background events ($N^{\text{data}}_{\text{SR,bkg}}$).

Source	$\epsilon_{\text{sig+tag}}$ [%]	$N^{\text{data}}_{\text{SR,bkg}}$ [%]
Statistical	4.10	38.98
Particle-ID correction (`muonID`)	5.02	4.88
Track momentum scale	-	1.83
Tracking efficiency	0.24	0.34
Control mode efficiency correction (sys.)	17.47	-
Control mode efficiency correction (stat.)	38.02	-

efficiency correction are separated from the main systematic to emphasise their large magnitudes. Additionally, as the control mode was implemented to purely apply a data-MC correction to the signal efficiency, it has no effect on $N^{\text{data}}_{\text{SR,bkg}}$, thus also having no contributions in these entries of the table.

It is seen in Table 5.17 that the largest proportional uncertainty is from the control modes. For $N^{\text{data}}_{\text{SR,bkg}}$, the next largest proportional uncertainty contribution is statistical. The dominance of the proportional statistical uncertainty stems from the small-N behaviour of the $\sqrt{N}/N$ fraction. It is thus expected to be large for $N^{\text{data}}_{\text{SR,bkg}}$ both as a quantity composed of multiple independent yields and the small-scale of the dataset after the selections required to identify $B^+ \to \mu^+\nu_\mu$ events. This proportional uncertainty will be improved in future $B^+ \to \mu^+\nu_\mu$ searches, as more data is recorded at Belle II. For similar reasons, the statistical uncertainty from the control mode efficiency correction is also the next largest source within $\epsilon_{\text{sig+tag}}$. Notably, the statistical uncertainty stemming from the uncorrected $B^+ \to \mu^+\nu_\mu$ signal efficiency is more comparable to the sources of systematic uncertainty, such as from the `muonID` correction. This can be attributed to its evaluation from the large set of dedicated signal MC, even after signal selection, rather than the small values of the yield after being scaled to $362\,\text{fb}^{-1}$. The remaining systematic uncertainties originating from Belle II recommendations are more acceptable, all contributing around 5% or lower.

It is noted from Eq. (5.5) that the $B^+ \to \mu^+\nu_\mu$ MC yield in the p^*_μ signal region will not be used directly to derive the corresponding data yield i.e. $N^{\text{data}}_{\text{SR,sig}}$ is derived as $N^{\text{data}}_{\text{SR,all}}$ - $N^{\text{data}}_{\text{SR,bkg}}$). It is useful, however, to consider what signal yield is predicted from the

Table 5.18. Final measurements for $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$.

Measurement	$N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$
$\epsilon_{\text{sig+tag}}$	$(7.83{\pm}0.03{\pm}3.30){\times}10^{-3}$
$N^{\text{data}}_{\text{SR,bkg}}$	9.01±3.51±0.47

Table 5.19. Estimate of data yield in signal region attributed to $B^+ \to \mu^+\nu_\mu$ events.

Measurement	$N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$
$N^{\text{data}}_{\text{SR,sig(est)}} = N^{\text{MC}}_{\text{SR,sig}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}}$	1.47 ± 1.31 ± 0.08
$N^{\text{data}}_{\text{SR,sig(est)}} = 2f^{+-}N^{\text{data}}_{362\text{fb}^{-1},B\bar{B}} \times \epsilon^{\text{data}}_{\text{sig+tag}} \times \mathcal{B}(B^+ \to \mu^+\nu_\mu)_{\text{SM}}$	1.32± 0.01 ±0.56

MC and thus likely to be measured once the unblinding of the $N^{\text{data}}_{\text{SR,all}}$ takes place. Such an estimation will also be used in the following section to predict confidence intervals. Table 5.19 presents estimates of the $B^+ \to \mu^+\nu_\mu$ yield in data within the signal region, evaluated in two methods. The first row contains the signal yield estimated in a similar method to how $N^{\text{data}}_{\text{SR,bkg}}$ was obtained, from the signal MC prediction scale by the data/MC ratio from the p^*_μ sideband. This yields 1.47 predicted signal events. From this quantity, it is made obvious why this yield is not implemented as $N^{\text{data}}_{\text{SR,sig}}$ within the calculation of the branching fraction, given the large statistical uncertainty relative to the central value. Additionally, using such a yield in the evaluation of the $B^+ \to \mu^+\nu_\mu$ branching fraction cannot account for non-SM excesses that may be revealed in unblinding. As a cross-check of this prediction, an additional method is used to estimate the $B^+ \to \mu^+\nu_\mu$ yield in the 362 fb^{-1} dataset, for the case where the experimental branching fraction would match the SM prediction from Eq. (1.7). Both predictions are within uncertainties of one another, indicating the soundness of the analysis procedure .

With the source of uncertainties understood in the final measurements required to evaluate $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$, Table 5.18 presents the final values, with sources combined in quadrature to obtain total uncertainties. The systematic contribution consists of all systematic sources, including those from the $B^+ \to \mu^+\nu_\mu$ sample, as well as the statistical and systematic contributions from the $B^+ \to \bar{D}^0\pi^+$ control mode in $\epsilon_{\text{sig+tag}}$. The statistical uncertainty is from the $B^+ \to \mu^+\nu_\mu$ sample alone. In the next section, these values and their uncertainties will be used to establish confidence intervals on a range of yields that could be seen during unblinding. These will be then propagated to limits on final $B^+ \to \mu^+\nu_\mu$ branching fractions.

5.3. Evaluating confidence intervals via the method of Feldman and Cousins

The number of signal region events in data estimated to be from non-$B^+ \to \mu^+\nu_\mu$ processes ($N^{\text{data}}_{\text{SR,bkg}}$) has been evaluated, as well as the data-calibrated signal efficiency ($\epsilon_{\text{sig+tag}}$). The effect of the systematic and statistical uncertainty in both have also been quantified. From the numerator of Eq. (5.4), the number of signal region events in data estimated to be from $B^+ \to \mu^+\nu_\mu$ events will be taken to be the difference between unblinded number of data events in the signal region and the background estimate, i.e. $N^{\text{data}}_{\text{SR},B^+\to\mu^+\nu_\mu}$=$N^{\text{data}}_{\text{SR,all}}$ - $N^{\text{data}}_{\text{SR,bkg}}$. While it is an option to propagate the uncertainties to understand potential variation in the signal yield in data, a confidence limit will instead be defined. This follows the convention set by previous $B^+ \to \mu^+\nu_\mu$ searches, presented in Figure 1.10. To establish the confidence limits, the method of Feldman-Cousins is implemented [109]. The exact implementation of the method follows that described in Reference [125], pertaining to the 2010 BaBar $B^+ \to \mu^+\nu_\mu$ search. The Feldman-Cousins method is well suited to cases with small signal, as is expected in the relative number of $B^+ \to \mu^+\nu_\mu$ to non-$B^+ \to \mu^+\nu_\mu$ events seen in MC. Additionally, the method is identical for deriving a one-sided limit in the event of a null-result (i.e. no $B^+ \to \mu^+\nu_\mu$ events) or a two-sided interval for error setting if the yield is statistically significant.

5.3.1. Method outline

In preparation for unblinding the signal region in data, a prediction of the significance of potential signal yields in data is produced, given the estimate of the background yield in Table 5.18. The following section will describe how the method of Feldman and Cousins is used to estimate the confidence intervals on possible signal yields. In short, the method will use ensembles of 'toy-MC' to simulate the measurement of a signal yield from background yield and the number of unblinded observations. The number of unblinded observations will be constructed from an underlying 'true' number of $B^+ \to \mu^+\nu_\mu$ events. This allows a derivation of likelihood that a given measured signal yield can occur, given the underlying 'true' signal yield. From this, the confidence interval is established as the range of 'true' signal yields that are most likely to result in the given measured signal yield. The specific process will now be applied and described in more detail.

To begin, an underlying 'true' number of $B^+ \to \mu^+\nu_\mu$ events in the signal region is modelled. This will be denoted as $N_{\text{sig,true}}$. The value $N_{\text{sig,true}}$ is set to be the mean of a Poisson distribution. 10000 samples are generated from the distribution, simulating 10000 statistical variations on a measurement of $B^+ \to \mu^+\nu_\mu$ yield.

The number of background events that occur within the same event as the toy signal yields is modelled using the background yield of the $B^+ \to \mu^+\nu_\mu$ search. Taking the measurement of $N^{\text{data}}_{\text{SR,bkg}}$= 9.01 ± 3.51 ±0.47 from Table 5.18 to be representative of a 'true' number of background events, a random number is created based on a Gaussian distribution centred on the nominal value of $N^{\text{data}}_{\text{SR,bkg}}$ and width equal to the statistical error. This simulates the measurement subject to statistical fluctuations. This single number is then used as the centre of a second Gaussian distribution with width set to the systematic error. 10000 samples are taken from this distribution, where this time variations represent the combined uncertainties in the measurement method, the sources of which were discussed throughout Section 5.2. Summing the two sets of numbers creates 10000 variations of the blinded quantity, the total number of events yielded in the signal region.

To simulate the derivation of the signal yield from the blinded yield and the estimated background, the mean of the second background Gaussian is treated as the most likely measurement of the background derived in the analysis. Subtracting this value from the set of 10000 signal region yields thusly models 10000 measurements of $N^{\text{data}}_{\text{SR,sig}}$, each measurement denoted as $N_{\text{sig,meas.}}$. A single-Gaussian fit can then be performed to the toy measurements of $N^{\text{data}}_{\text{SR,sig}}$. The fitted mean is treated as the most probable measured value of $N^{\text{data}}_{\text{SR,sig}}$ of the original implemented 'true' number of $B^+ \to \mu^+\nu_\mu$ events. Example distributions are presented in Figure 5.7.

By repeating the procedure described above for multiple possible $N_{\text{sig,true}}$, a 2D probability distribution can be constructed from the Gaussians of fitted signal yields. This is demonstrated in Figure 5.8a, where each row is the Gaussian distribution of measured toy MC signal yields for the given 'true' signal yield. By normalising the measured signal Gaussian, the probability of measuring a specific signal yield given an underlying 'true' signal yield (i.e. $\Pr(N_{\text{sig,meas.}}|N_{\text{sig,true}})$) can be interpreted. Figure 5.8b reproduces the same 2D distribution, where shading is now according to how many standard deviations away each $N_{\text{sig,meas}}$ is from the signal yield most probable to occur

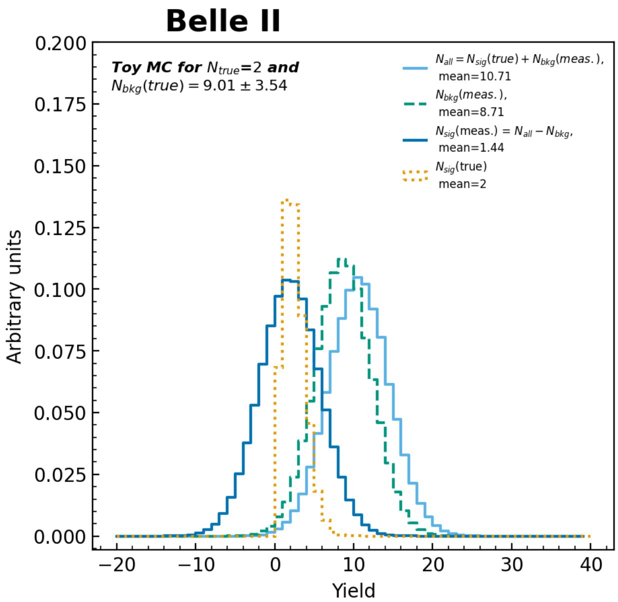


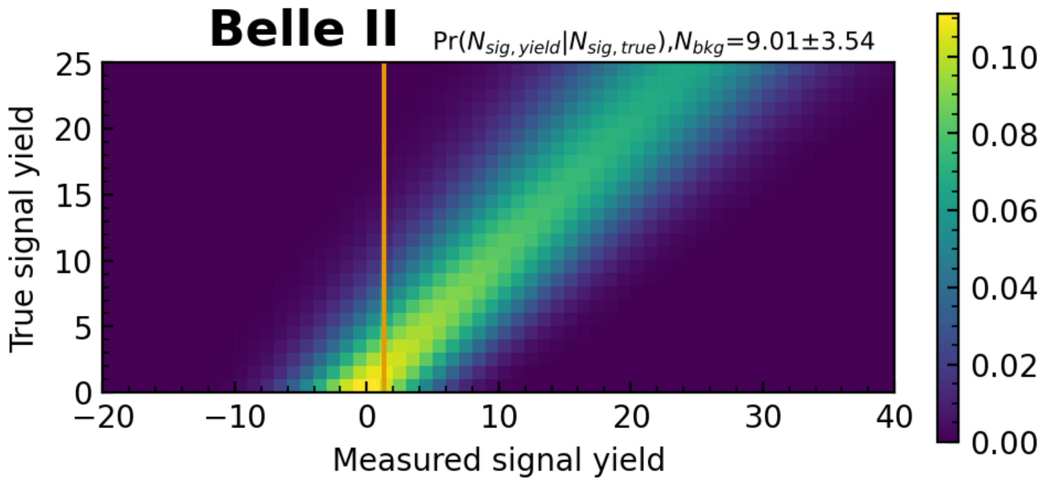
Figure 5.7. Example distributions of toy MC generated as part of the Feldman-Cousins method. Distributions include that of the statistical variations in signal yield (dotted), the systematic variations in background yield (dashed), the combined signal region yield (solid, light-blue), and the resultant measured signal yield (solid, dark-blue).

given $N_{\text{sig,true}}$. As expected, this plot indicates the most probable measurements are those that close the true yield. The vertical line in both of these distributions indicates the signal yield expected in this analysis once unblinding takes place. This value was calculated in Table 5.19, under the assumption that only SM processes are contributing.

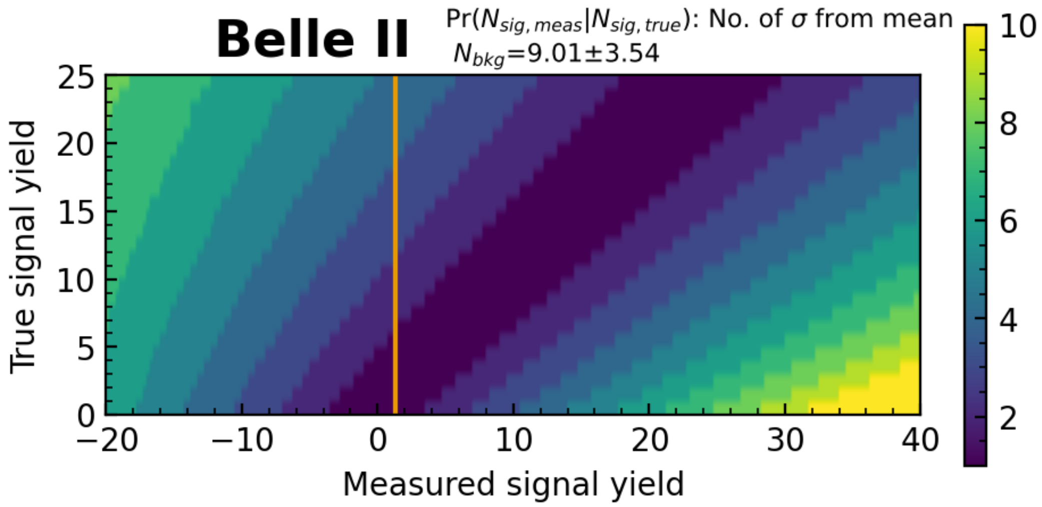
The probabilities in Figure 5.8a will now be used to construct the confidence intervals, signifying the range of $N_{\text{sig,true}}$ that will be accepted as compatible with measurement $N_{\text{sig,meas.}}$. To determine which $N_{\text{sig,true}}$ are to be accepted, an ordering principle is required so that $N_{\text{sig,true}}$ are admitted into the interval around $N_{\text{sig,meas.}}$ in order of descending likelihood. As was seen in Section 1.4 when investigating previous $B^+ \to \mu^+\nu_\mu$ searches, it is convention to establish the confidence interval for a given $N_{\text{sig,meas.}}$ such that the sum of probabilities $\Pr(N_{\text{sig,meas.}}|N_{\text{sig,true}})$ for agreeing $N_{\text{sig,true}}$ in the interval is at least 90%.

The Feldman-Cousins method suggests adding $N_{\text{sig,true}}$ to the interval according to an ordering principle on the probability ratio R, defined with the following equation.

$$R = \frac{\Pr(N_{\text{sig,meas.}}|N_{\text{sig,true}})}{\Pr(N_{\text{sig,meas.}}|N_{\text{sig,best}})} \tag{5.13}$$

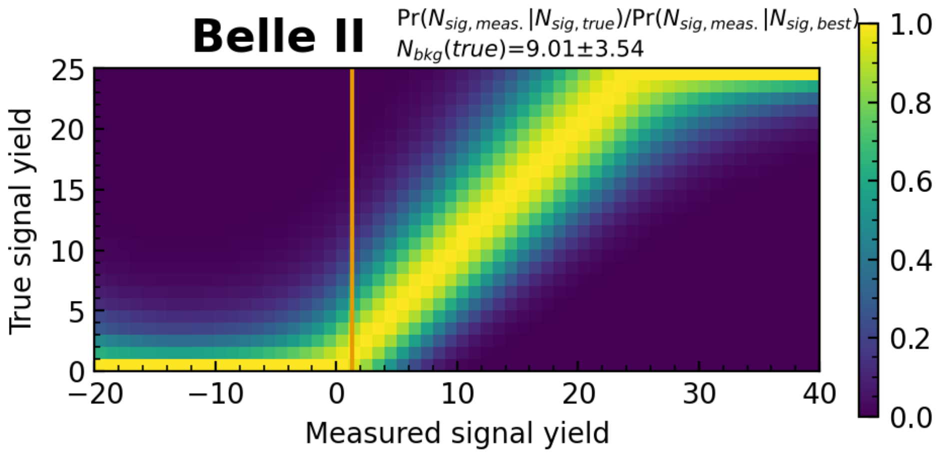


(a) z-axis is probability.

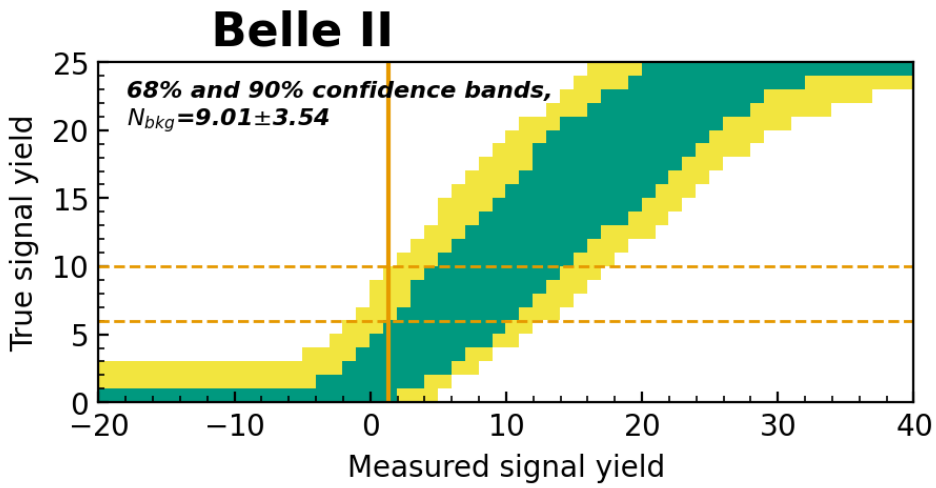


(b) z-axis is no. of standard deviations measured signal yield is away from most probable measured signal yield for a given 'true' yield.

Figure 5.8. 2D distributions of probabilities of measuring a signal yield, given an underlying number of 'true' signal events. The vertical line indicates the total number of signal events expected in data from this analysis, $N_{\text{sig,meas.}}$=1.47.



(a) z-axis is R.



(b) 68% and 90% confidence intervals of $N_{\mathrm{sig,true}}$ for a variety of $N_{\mathrm{sig,meas.}}$.

Figure 5.9. Distributions of $N_{sig,meas.}$ and $N_{\mathrm{sig,true}}$.

From this equation, $N_{\mathrm{sig,best}}$ denotes the $N_{\mathrm{sig,true}}$ that best supports the measurement i.e. $N_{\mathrm{sig,true}}$ that maximises $\Pr(N_{\mathrm{sig,meas.}}|N_{\mathrm{sig,true}})$. Graphically, this is the bin with the largest probability for a given x=$N_{\mathrm{sig,meas.}}$ in Figure 5.8a. Thus, the quantity R is ratio of how well a theorised $N_{\mathrm{sig,true}}$ matches the measured result, against the $N_{\mathrm{sig,true}}$ that is most compatible with the measurement. The corresponding map of R for the variety of measured and true signal yields constructed for the toy MC in this Section is presented in Figure 5.9a. Generally, the highest values of R occur coincide with the Gaussian mean of $N_{\mathrm{sig,meas.}}$.

The final step to constructing the confidence intervals is to apply the ordering principle. Figure 5.9b has been obtained by adding candidate $N_{\mathrm{sig,true}}$ to the interval of each $N_{\text{sig,meas.}}$, in order of their descending R. Each time, the corresponding bin of $\Pr(N_{\text{sig,meas.}}|N_{\text{sig,true}})$ is added to the total, until the desired minimum significance is achieved. The final bin added marks the beginning of the confidence band. This creates the edges of the confidence bands seen in Figure 5.9b, where both 68% and 90% intervals have been specified. Once the total data yield in the signal region is unblinded and the signal is evaluated, a vertical line is drawn for the extracted value through the distribution to determine its confidence. As a demonstration, a line has been drawn through $N_{\text{sig,meas.}}$=1.47, the predicted signal yield of data from Table 5.19.

5.3.2. Results

As stated throughout this section, unblinding of the data in the signal region is left for future work. In absence of the exact data yield in the signal region (i.e. $N^{\text{data}}_{\text{SR,all}}$), a range of confidence intervals for multiple possible signal yields ($N^{\text{data}}_{\text{SR,sig}}$) will be presented. To do so, a range of $N^{\text{data}}_{\text{SR,all}}$ will be considered. Such values will be integer valued, as would be expected in the measured data yield. The background yield estimate for this dataset, $N^{\text{data}}_{\text{SR,bkg}}$=9.01, is subtracted from each possible value of $N^{\text{data}}_{\text{SR,all}}$ to obtain a corresponding $N^{\text{data}}_{\text{SR,sig}}$. The confidence limit on each derived $N^{\text{data}}_{\text{SR,sig}}$ is then extracted from Figure 5.9b, by observing the boundaries of the required confidence band. The range of varied data yields $N^{\text{data}}_{\text{SR,all}}$ were centred on the predicted signal yield from this analysis being $N^{\text{data}}_{\text{SR,sig}}$=1.47. These variations then examines the case of zero $B^+ \to \mu^+\nu_\mu$ events, as well as an excess of $B^+ \to \mu^+\nu_\mu$ events. Once the limit on the derived $N^{\text{data}}_{\text{SR,sig}}$ for each $N^{\text{data}}_{\text{SR,all}}$ is extracted, it can be substituted into Eq. (5.3) to finally obtain the related limits on the $B^+ \to \mu^+\nu_\mu$ branching fraction.

Table 5.20 contains the table of 90% confidence limits on potential signal yields in data and on the corresponding branching fractions, as derived from the Feldman-Cousins method. With the chosen range of of potential $N^{\text{data}}_{\text{SR,all}} \in$ [2,16], this results in confidence limits being presented for the equivalent range $N^{data}_{SR,sig} \in$ [-7.01,6.99]. It is seen that the method of Feldman and Cousins provides upper bounds for a majority of the trial yields, rather than two-sided confidence intervals. Thus this analysis can only reject a null-result (i.e. reject a non-zero $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ or signal yield) with 90% confidence if 15 or more events are produced in the signal region. This corresponds to ≈5.99 signal events. Such a value is not within the uncertainties of the predicted signal yield in data, which was evaluated in Table 5.19. It is notable that near all provided limits, as

Table 5.20. Table of 90% confidence intervals extracted from Figure 5.9b, for varying $N^{\text{data}}_{\text{SR,all}}$. Evaluated with $N^{\text{data}}_{\text{SR,bkg}}$=9.01 and $N^{\text{data}}_{\text{SR,sig}}$=$N^{\text{data}}_{\text{SR,all}}$ - $N^{\text{data}}_{\text{SR,bkg}}$. Selection based uncertainties are quoted for completeness, though they require evaluation through a more sophisticated procedure.

$N^{\text{data}}_{\text{SR,all}}$	$N^{\text{data}}_{\text{SR,sig}}$	$N^{\text{data}}_{\text{SR,sig}}$ CL	$\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ CL
2	-7.01	$N^{\text{data}}_{\text{SR,sig}} < 3$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 9.63\times10^{-7}$
3	-6.01	$N^{\text{data}}_{\text{SR,sig}} < 3$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 9.63\times10^{-7}$
4	-5.01	$N^{\text{data}}_{\text{SR,sig}} < 3$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 9.63\times10^{-7}$
5	-4.01	$N^{\text{data}}_{\text{SR,sig}} < 4$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.28\times10^{-6}$
6	-3.01	$N^{\text{data}}_{\text{SR,sig}} < 4$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.28\times10^{-6}$
7	-2.01	$N^{\text{data}}_{\text{SR,sig}} < 5$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.60\times10^{-6}$
8	-1.01	$N^{\text{data}}_{\text{SR,sig}} < 6$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.93\times10^{-6}$
9	-0.01	$N^{\text{data}}_{\text{SR,sig}} < 7$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 2.25\times10^{-6}$
10	0.99	$N^{\text{data}}_{\text{SR,sig}} < 9$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 2.89\times10^{-6}$
11	1.99	$N^{\text{data}}_{\text{SR,sig}} < 10$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29\times10^{-6}$
12	2.99	$N^{\text{data}}_{\text{SR,sig}} < 11$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.53\times10^{-6}$
13	3.99	$N^{\text{data}}_{\text{SR,sig}} < 12$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.85\times10^{-6}$
14	4.99	$N^{\text{data}}_{\text{SR,sig}} < 13$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 4.17\times10^{-6}$
15	5.99	$1 < N^{\text{data}}_{\text{SR,sig}} < 15$	$3.21\times10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 4.81\times10^{-6}$
16	6.99	$2 < N^{\text{data}}_{\text{SR,sig}} < 16$	$6.42\times10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 5.13\times10^{-6}$

either upper limits or two-sided limits, are inclusive of the Standard Model evaluation of $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ =4.24×10^{-7}. The final row presents the intervals if 16 events were recorded in the signal region, which excludes the SM calculation of the branching fraction. Equivalent to ≈6.99 events, this is over four times larger than what is predicted by this analysis. Such an excess in measured $B^+ \to \mu^+\nu_\mu$ events could indicate the presence of beyond Standard Model mechanisms in the $B^+ \to \mu^+\nu_\mu$ process. On the otherhand, it is noted that the branching fraction upper limit plateaus at $N^{\text{data}}_{\text{SR,all}} = 4$ or less, well below even the expected number of background events expectedo. This minimum upper limit, at $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 9.63\ \times10^{-7}$, still exceeds the theoretical SM branching fraction. The analysis therefore is not sensitive enough to constrain the $B^+ \to \mu^+\nu_\mu$ branching fraction to be below that of the SM. Such an effect could be seen if beyond SM processes destructively intefered with the SM process.

While 68% confidence is not the accepted benchmark by previous $B^+ \to \mu^+\nu_\mu$ searches, the table of corresponding 68% confidence limits indicated by Figure 5.9b are presented

Table 5.21. Table of 68% confidence intervals extracted from Figure 5.9b, for varying $N^{\text{data}}_{\text{SR,all}}$. Evaluated with $N^{\text{data}}_{\text{SR,bkg}}$=9.01 and $N^{\text{data}}_{\text{SR,sig}}$=$N^{\text{data}}_{\text{SR,all}}$ - $N^{\text{data}}_{\text{SR,bkg}}$.

$N^{\text{data}}_{\text{SR,all}}$	$N^{\text{data}}_{\text{SR,sig}}$	$N^{\text{data}}_{\text{SR,sig}}$ CL	$\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ CL
2	-7.01	$N^{\text{data}}_{\text{SR,sig}} < 1$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.21{\times}10^{-7}$
3	-6.01	$N^{\text{data}}_{\text{SR,sig}} < 1$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.21{\times}10^{-7}$
4	-5.01	$N^{\text{data}}_{\text{SR,sig}} < 1$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.21{\times}10^{-7}$
5	-4.01	$N^{\text{data}}_{\text{SR,sig}} < 1$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.21{\times}10^{-7}$
6	-3.01	$N^{\text{data}}_{\text{SR,sig}} < 2$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 6.42{\times}10^{-7}$
7	-2.01	$N^{\text{data}}_{\text{SR,sig}} < 2$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 6.42{\times}10^{-7}$
8	-1.01	$N^{\text{data}}_{\text{SR,sig}} < 3$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 9.63{\times}10^{-7}$
9	-0.01	$N^{\text{data}}_{\text{SR,sig}} < 4$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.28{\times}10^{-6}$
10	0.99	$N^{\text{data}}_{\text{SR,sig}} < 5$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.60{\times}10^{-6}$
11	1.99	$N^{\text{data}}_{\text{SR,sig}} < 6$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 1.93{\times}10^{-6}$
12	2.99	$1 < N^{\text{data}}_{\text{SR,sig}} < 7$	$3.21{\times}10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 2.25{\times}10^{-6}$
13	3.99	$1 < N^{\text{data}}_{\text{SR,sig}} < 9$	$3.21{\times}10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 2.89{\times}10^{-6}$
14	4.99	$2 < N^{\text{data}}_{\text{SR,sig}} < 10$	$6.42{\times}10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.21{\times}10^{-6}$
15	5.99	$2 < N^{\text{data}}_{\text{SR,sig}} < 11$	$6.42{\times}10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.53{\times}10^{-6}$
16	6.99	$3 < N^{\text{data}}_{\text{SR,sig}} < 12$	$9.63{\times}10^{-7} < \mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.85{\times}10^{-6}$

in Table 5.21. In this case, the null-result can be rejected with 68% confidence if 12 events are observed in the signal region, equivalent to 2.99 signal events. The SM branching fraction is excluded if 14 events are yielded upon unblinding, equivalent to ≈4.99 events. For $N^{\text{data}}_{\text{SR,all}} \leq$ -4.01, the 68% confidence limit similarly excludes the SM branching fraction. Figure 5.10 presents the 68% and 90% confidence limits of Table 5.20 and Table 5.20 graphically. Here, the total data yield and equivalent signal yields required to exclude the $B^+ \to \mu^+\nu_\mu$ SM from being the sole contribution to this analysis at 68% and 90% confidence is made more obvious.

Thus, using the method of Feldman and Cousins, a series of confidence limits on both signal region yields and $B^+ \to \mu^+\nu_\mu$ branching fractions have been produced. To be able to compare the $B^+ \to \mu^+\nu_\mu$ signal strategy with previous searches, $B^+ \to \mu^+\nu_\mu$ branching fraction upper limit compatible with the predicted signal yield $N^{\text{data}}_{\text{SR,sig}}$=1.47 will be used. It was indicated in the Feldman-Cousins curve of Figure 5.9b that using $N^{\text{data}}_{\text{SR,sig}} < 10$ is the appropriate limit for this singal yield. Therefore, the following 90%

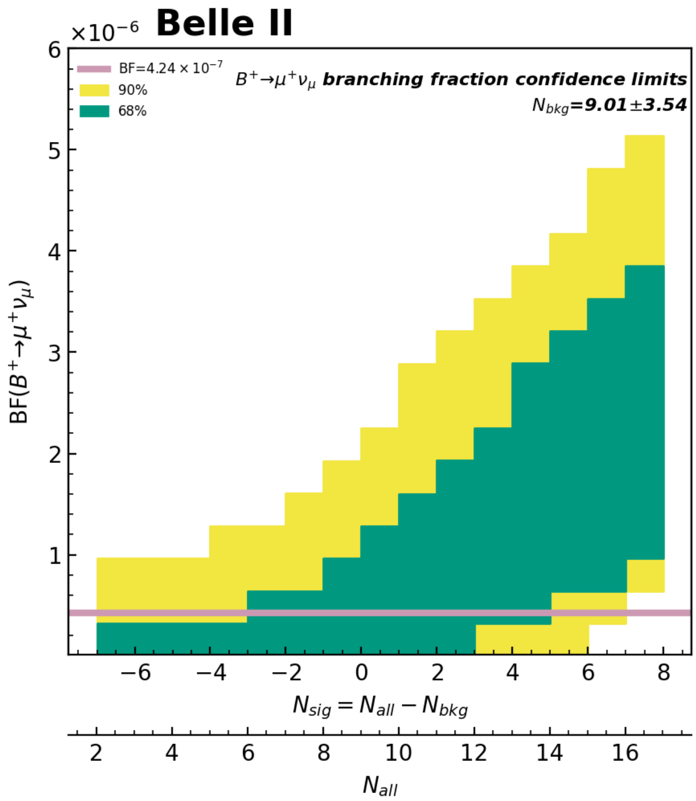


Figure 5.10. 68% and 90% confidence bands on the $B^+ \to \mu^+\nu_\mu$ branching fraction. The horizontal line indicates the SM $B^+ \to \mu^+\nu_\mu$ branching fraction, where $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = 4.24{\times}10^{-7}$.

confidence limit on the branching fraction will be adopted.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29 \times 10^{-6} \tag{5.14}$$

Figure 5.11 presents the 90% confidence limits on $B^+ \to \mu^+\nu_\mu$ branching fractions derived from the 20 years of previous searches, each of which were discussed in Section 1.4. The limit predicted from this analysis is placed at $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29 \times 10^{-6}$. This prediction does not improve upon the previous upper bound set by Belle in the 2020 search, as it exceeds the limits placed on all previous Belle searches. For all except the Belle 2007 search, this can be expected given they use a larger 711 fb^{-1} dataset, just under double what was used in this analysis. The challenge that of the size of the current Belle II dataset brought to this analysis has seen throughout both Chapter 4 and this Chapter. Statistical uncertainty was seen to be one of the largest error contributions to $N^{\rm data}_{\rm SR,bkg}$ in Table 5.17. With this error contribution being used to create the toy MC necessary for confidence limit setting, it is thus evident how a low statistics sample can

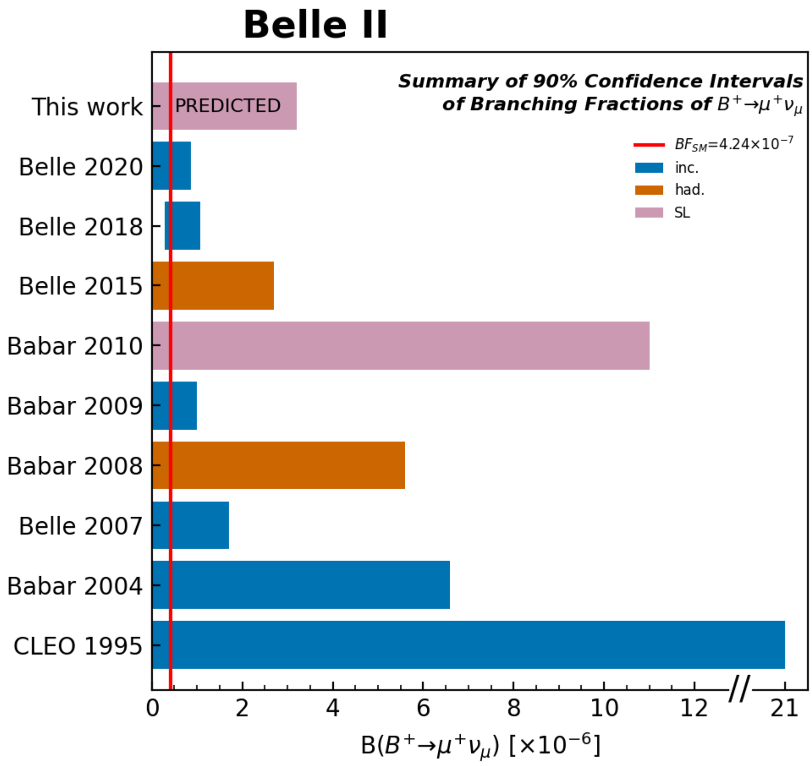


Figure 5.11. 90% upper limit on $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ derived from this analysis, in comparison to those from previous searches.

yield the larger upper limit of this rare decay search. The exclusion to this trend is in the Belle 2007 result, which used approximately 100 million less events than this current search. While repeating this analysis on a future larger Belle II dataset would improve the statistical uncertainty and thus lower the expected 90% upper limit, it remains to be seen whether this change alone could yield a limit comparable or better than the previous Belle results. If it does not, this could point towards other features of the analysis preventing a world leading limit on $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$, such as the tagging or selection efficiencies.

Given the differences in expected selection efficiency on the basis of inclusive and exclusive tagging, it could be considered more appropriate to compare the predicted upper limit of this thesis with those from previous exclusive tagged results. This work provides a more constrained limit than both the hadronic and semileptonic-tagged results from BaBar. The limits on this analysis fairs best against the BaBar 2010 result, given the similar choice of B-tagging mode. Here, this thesis predicts a 4$\times$ smaller limit with an equivalent-sized dataset. It is difficult to say which element of this Belle II search has resulted in an improved limit, given the multiple differences in the signal strategies. For example, the BaBar analysis effort reconstructs the same B_{tag} decay modes manually, while this work uses a machine-learning based method to build the B_{tag}. BaBar also used the Diamond Frame method to estimate the B-frame momentum of the signal muon, while this work used the lower resolution $\Upsilon(4S)$ frame muon momentum. Additionally, signal optimisation in this work is performed via sequential selections that optimise $\frac{S}{\sqrt{S+B}}$, while the BaBar search makes selections on constructed likelihood ratios of discriminating variables. A critical investigation of these differences and how they may affect the final Belle II $B^+ \to \mu^+\nu_\mu$ limit after unblinding are left as future work.

This Belle II search also improves the hadronic-tagged BaBar 2008 result, though underperforms with respect to that of Belle 2015 hadronic-tagged search. Notably, the Belle search implemented the Full Reconstruction algorithm, a predecessor to the FEI, whose large number of reconstruction modes in the full 711 fb^{-1} Belle dataset has allowed it to perform better than this search, despite the overall lower branching fraction of hadronic-tagged modes. This points to potential improvements that can be obtained in future work with a hadronic FEI tagged approach, which was originally considered in Chapter 3.

Using the systematic and statistical uncertainties evaluated throughout this Chapter, the method of Felman and Cousins has now completed the determination of the confidence intervals on potential signal yields and $B^+ \to \mu^+\nu_\mu$ branching fractions for this analysis. Without being able to unblind the number of data events in the signal region, a single interval cannot be quoted as the final result for this thesis. The predicted branching fraction limit for this analysis was instead obtained by choosing the interval compatible with the scaled signal yield, $N^{\text{data}}_{\text{SR,sig(est)}}$=1.47, which resulted in $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29\times10^{-6}$. With potential confidence intervals establish, the following section will evaluate the absolute branching fractions for the same range of potential $N^{\text{data}}_{\text{SR,all}}$.

5.4. Evaluation of Predicted $B^+ \to \mu^+\nu_\mu$ Branching Fractions

Following from the previous section, where the upper limits on the $B^+ \to \mu^+\nu_\mu$ branching fraction for this analysis were established for a range of potential yields in lieu of unblinding, in this section the corresponding absolute branching fractions will be evaluated. This will be performed by again using the background data yield estimated as $N^{\text{data}}_{\text{SR,bkg}}$=9.01

Table 5.22. Table of experimentally derived $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$, for varying $N^{\text{data}}_{\text{SR,all}}$. Evaluated with $N^{\text{data}}_{\text{SR,bkg}}$=9.01 and $N^{\text{data}}_{\text{SR,sig}}$=$N^{\text{data}}_{\text{SR,all}}$ - $N^{\text{data}}_{\text{SR,bkg}}$. $\mathcal{B}(B^+ \to \mu^+\nu_\mu)_{\text{SM}}$=$(4.24 \pm 0.24) \times 10^{-7}$, as per its evaluation in Eq. (1.7).

$N^{\text{data}}_{\text{SR,all}}$	$N^{\text{data}}_{\text{SR,sig}}$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu)$	$\Delta\mathcal{B}_{\text{stat}}$	$\Delta\mathcal{B}_{\text{sys}}$
2	-7.01	-2.25$\times10^{-6}$	1.22$\times10^{-6}$	9.59$\times10^{-7}$
3	-6.01	-1.93$\times10^{-6}$	1.26$\times10^{-6}$	8.26$\times10^{-7}$
4	-5.01	-1.61$\times10^{-6}$	1.30$\times10^{-6}$	6.94$\times10^{-7}$
5	-4.01	-1.29$\times10^{-6}$	1.34$\times10^{-6}$	5.63$\times10^{-7}$
6	-3.01	-9.67$\times10^{-7}$	1.37$\times10^{-6}$	4.34$\times10^{-7}$
7	-2.01	-6.46$\times10^{-7}$	1.41$\times10^{-6}$	3.11$\times10^{-7}$
8	-1.01	-3.25$\times10^{-7}$	1.45$\times10^{-6}$	2.04$\times10^{-7}$
9	-0.01	-4.32$\times10^{-9}$	1.48$\times10^{-6}$	1.51$\times10^{-7}$
10	0.99	3.17$\times10^{-7}$	1.52$\times10^{-6}$	2.01$\times10^{-7}$
11	1.99	6.37$\times10^{-7}$	1.55$\times10^{-6}$	3.08$\times10^{-7}$
12	2.99	9.58$\times10^{-7}$	1.58$\times10^{-6}$	4.31$\times10^{-7}$
13	3.99	1.28$\times10^{-6}$	1.62$\times10^{-6}$	5.59$\times10^{-7}$
14	4.99	1.60$\times10^{-6}$	1.65$\times10^{-6}$	6.90$\times10^{-7}$
15	5.99	1.92$\times10^{-6}$	1.68$\times10^{-6}$	8.23$\times10^{-7}$
16	6.99	2.24$\times10^{-6}$	1.71$\times10^{-6}$	9.56$\times10^{-7}$

with its previously derived statistical and systematic uncertainties. The same range of potential unblinded yields will be used, with integer values of $N^{\text{data}}_{\text{SR,all}}$ ∈[2,16] allowing for both null and excess $B^+ \to \mu^+\nu_\mu$ yields. These will be substituted into Eq. (5.4), alongside the data-calibrated $\epsilon_{\text{sig+tag}}$, to obtain the absolute experimental branching fractions. Additionally, the uncertainties on each input will also be considered, with the varying $N^{\text{data}}_{\text{SR,all}}$ assigned standard Poisson uncertainty. Examining each source of uncertainty in the evaluation of the branching fraction will give an indication of the leading error contributions in this calculation.

Table 5.22 contains the branching fractions evaluated for potential $N^{\text{data}}_{\text{SR,bkg}}$ ∈[2,16], with associated uncertainties in statistical and combined systematic uncertainties. For $N^{\text{data}}_{\text{SR,bkg}} \in [2, 9]$, all derived branching fractions are negative, as these are cases where the observed data yield is less than the estimated background yield in the signal region. For $N^{\text{data}}_{\text{SR,bkg}} \in [11, 16]$, the branching fractions exceed the Standard Model result of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = 4.24{\times}10^{-7}$. When considering these values alongside their statistical

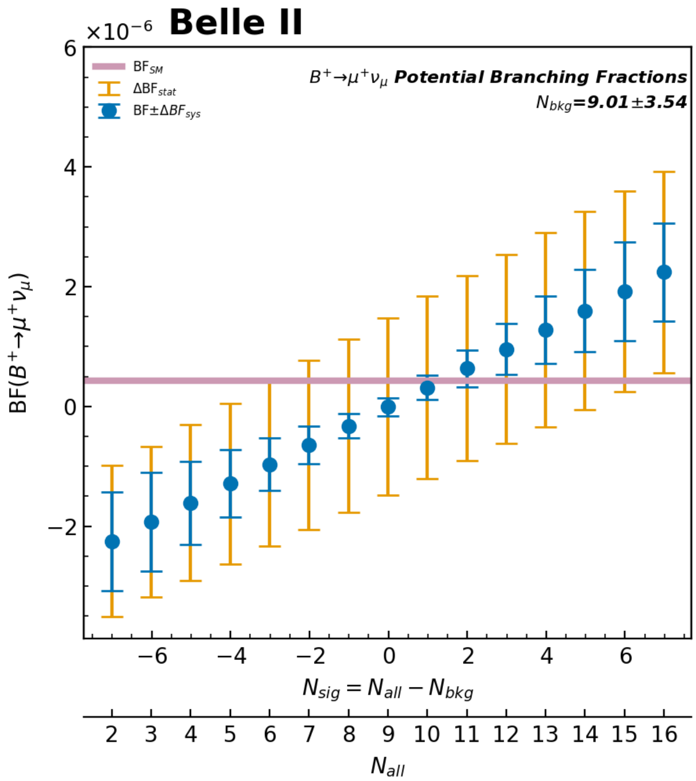


Figure 5.12. Potential $B^+ \to \mu^+\nu_\mu$ branching fractions for varying yields in data.

uncertainty, it is only $N^{\text{data}}_{\text{SR,bkg}}$=15 and 16 that exclude a null-result branching fraction within one unit of uncertainty. This requires a signal yield of 5.99 $B^+ \to \mu^+\nu_\mu$ events, around four times the expected signal yield. It again demonstrates that the main challenge of this search is its statistics, stemming from the rarity of the $B^+ \to \mu^+\nu_\mu$ process and the smaller $B\overline{B}$ dataset relative to previous searches. The statistical uncertainty is the largest source by an order magnitude, when compared to the systematic equivalent.Additionally, the statistical uncertainty increases with signal yield, a result of the Poisson error increasing with $N^{\text{data}}_{\text{SR,all}}$. This same trend is not seen in the other uncertainty sources, which all increase with the absolute value of the signal yield. Thus while an anomalously large yield may exclude the null result within statistical uncertainty, this will also come with larger systematic uncertainties. This highlights that systematic and statistical sources may not be independent, as is traditionally assumed when summing uncertainties in quadrature. The central values, their statistical and systematic uncertainties from Table 5.22 are plotted in Figure 5.12. The Figure emphasises the relationship between the yields and the magnitude of their error.

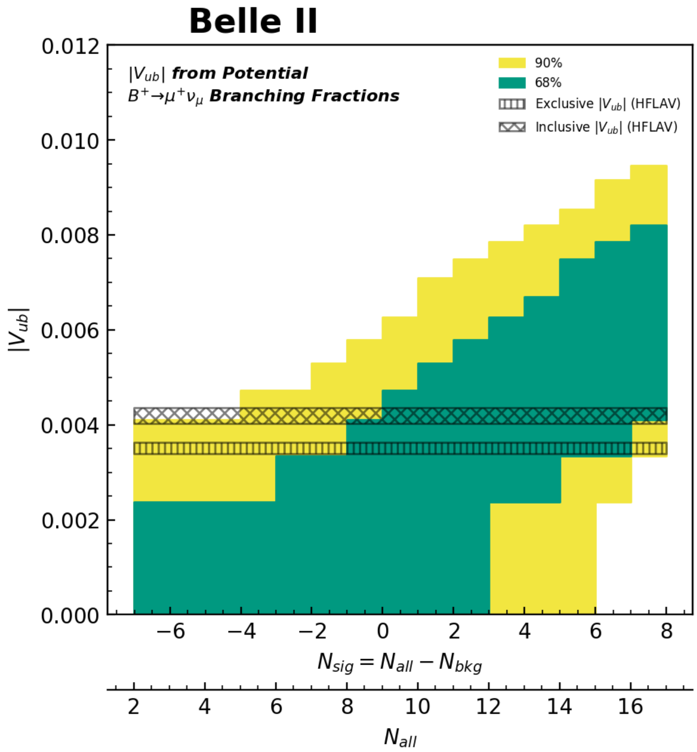
Similar to the previous section, the predicted branching fraction that will be the result for this thesis will be the one corresponding to the estimate of $N^{\text{data}}_{\text{SR,sig}}$=1.47. This is chosen to be result for $N^{\text{data}}_{\text{SR,sig}}$=1.99, which while larger than the estimate, matches the same $N^{\text{data}}_{\text{SR,all}}$=10 used for the upper limit. Thus for the predicted signal yield of $N^{\text{data}}_{\text{SR,sig}}$=1.99, the branching fraction is obtained.

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (6.4 \pm 15.5 \text{ (stat) } \pm 3.08 \text{ (sys) }) \times 10^{-7} \tag{5.15}$$

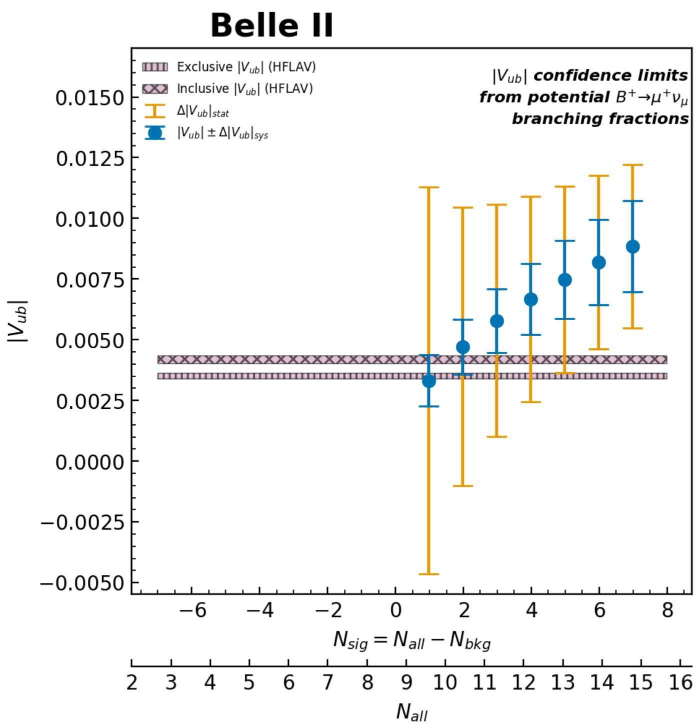
With this branching fraction dominated by statistical uncertainties, it must be quoted alongside the $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ upper limit obtained in the previous section.

5.5. Implications for $|V_{ub}|$

One of the first motivations discussed for measuring the $B^+ \to \mu^+\nu_\mu$ branching fraction in Section 1.2 was its dependence on CKM matrix element V_{ub}. V_{ub} was seen to be one of two least-well constrained parameter in the theoretical SM branching fraction equation presented in Eq. (1.6). The $B^+ \to \mu^+\nu_\mu$ branching fractions and upper limits produced in this Chapter are not world-leading, and thus it is not expected that the derivation of V_{ub} from the results of this thesis would produce an improvement on V_{ub}. However, for completions sake, the values of V_{ub} according to the expected experimentally derived $B^+ \to \mu^+\nu_\mu$ branching fraction is evaluated. Figure 5.13 presents the confidence limits and central values of $|V_{ub}|$ derived from the branching fractions that could be obtained when unblinding the signal region. These are based on the branching fractions from Table 5.22 and confidence limits from Table 5.20 and Table 5.21. While this analysis provides a measurement of $|V_{ub}|$ via an exclusive decay, it seen in Figure 5.13a, that most confidence limits agree with both the exclusive and inclusive determinations of $|V_{ub}|$. Therefore, the lack of precision in the analysis currently means it cannot be used to support one determination over the other with a 90% confidence. Similarly, for the absolute branching fractions demonstrated in Figure 5.13b, the large statistical uncertainties results in little discriminating power between the two HFLAV averages of $|V_{ub}|$. This plot also emphasises that alternate methods are better at constraining $|V_{ub}|$, given the relative magnitude of the uncertainties measured in this analysis and the narrow width of the shaded regions indicating much smaller uncertainty. This analysis, at least before the inclusion of future improvements discussed in Section 5.6, is not able to provide a competitive measurement of $|V_{ub}|$.



(a) 90% and 68% confidence bands on $|V_{ub}|$.



(b) $|V_{ub}|$ central values, with statistical and systematic uncertainties.

Figure 5.13. Potential derived values of $|V_{ub}|$ from the $B^+ \to \mu^+\nu_\mu$ branching fractions with varying yields in data. No branching fractions are obtained for negative yields, due to the need to square root the branching fraction to obtain $|V_{ub}|$.

To finish the discussion of $|V_{ub}|$, the measurements of $|V_{ub}|$ corresponding to the upper limit on the branching fraction from Eq. (5.15) and the corresponding absolute branching fraction from Eq. (5.14) will be stated. The following is the 90% upper limit on $|V_{ub}|$ from this analysis.

$$|V_{ub}| < 7.1 \times 10^{-3} \tag{5.16}$$

Additionally, the following is the value of $|V_{ub}|$ pertaining to the predicted signal yield of this analysis (alongside the statistical and systematic uncertainties respectively).

$$|V_{ub}| = (4.72 \pm 5.74 \pm 1.13) \times 10^{-3} \tag{5.17}$$

5.6. Future Improvements

In the previous two sections, a $B^+ \to \mu^+\nu_\mu$ branching fraction and corresponding upper limit on this branching fraction was produced for a variety of yields that could be seen when the data in the signal region is unblinded. For the upper limit derived from the predicted signal yield in data, it was demonstrated in Section 5.3.2 that such a limit would not improve the limit from previous $B^+ \to \mu^+\nu_\mu$ searches. Similarly, when looking at the absolute branching fraction in Section 5.4, the branching fraction for the same predicted signal yield was eclipsed by its large statistical uncertainty.

Future iterations of this work will have the capacity to improve the upper limit on the $B^+ \to \mu^+\nu_\mu$ branching fraction, as well as the uncertainties on the absolute branching fraction. Throughout this thesis, a number of factors have been identified which could strengthen these results. They are summarised here.

- **Unblinding**: Since incorporating collision data into the analysis in Chapter 4, the yields of all data surviving signal selection with $p^*_\mu \in$[2.5,2.85) GeV/c have not been examined. This blinding procedure is in place to avoid temptation to bias correction factors towards a non-null result. With uncertainties and confidence limits now derived in absence of data yield assigned to signal event, it is now appropriate to consider unblinding. Permission to unblind must be sought from the Belle II collaboration before doing so, and unfortunately such discussions did not finalise prior to this thesis. Thus, a first-order improvement to this analysis would be to proceed with unblinding, so that the quoted observables are reflective of the results in the signal region, rather than just predictions. Being able to unblind may be conditional on changes or further justifications of the signal strategy, some of which are discussed as additional facets of improvement below.

- **Run-dependent MC**: In Section 2.4.4, the attributes of the MC used in this analysis was introduced, particularly that run-independent MC would be used. Run-independent MC is generated using the same idealised detector and accelerator conditions, regardless of differences that may occur between events. Thus utilising run-dependent MC, as is now becoming the norm within Belle II analyses, simulates more specific conditions per event. Such changes can include beam energy deviations, variations in data calibration as well as subdetector component quality. Updating the analysis to use run-dependent MC will allow for better data-MC agreement, though given the current agreement achieved, it is not expected to dramatically impact the branching fraction measured.

- **Off-resonant data for continuum modelling**: Off-resonant data, (i.e. data recorded slightly below the $\Upsilon(4S)$ resonance) was first discussed in Section 2.2.4 to improve modelling of continuum events. It was posited in Section 4.3.2 when investigating data-MC agreement in p^*_μ sidebands that histogram bins with the worst agreement tended to be those that contained $c\bar{c}$ continuum as well. Improving data-MC agreement in p^*_μ sidebands will provide a more accurate estimate of signal yield and thus better $B^+ \to \mu^+\nu_\mu$ branching fraction. Using data with collision energy below that of the $\Upsilon(4S)$ data will be dominated by continuum events, and thus should provide a better representation of $c\bar{c}$ events, thus avoiding the difficulties in MC simulating the fragmentation process of continuum events. At the current luminosity of the Belle II dataset, there is unfortunately a much smaller off-resonant dataset than there is available continuum MC. However, as luminosity increases, this may become a more viable option.

- **Higher luminosity dataset**: This analysis was performed on $362\,\mathrm{fb}^{-1}$ of $e^+e^- \to \Upsilon(4S)$ collisions recorded by the Belle II detector between 2019 and 2022. Since completing maintenance during the first long-shutdown period in 2023, around $100\,\mathrm{fb}^{-1}$ more e^+e^- collisions have been recorded, with more to come. Having access to more data will assist in reducing the impact of statistical uncertainties, in the establishment of the branching fraction and upper limits. Figure 5.14 demonstrates how the statistical precision on the $B^+ \to \mu^+\nu_\mu$ branching fraction can increase with a larger dataset for this same analysis strategy. Using the $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ equation in Eq. (5.3), the expression for the statistical uncertainty $\Delta\mathcal{B}_{\mathrm{stat}}$ was derived by only considering contributions from $N^{\text{data}}_{362\text{fb}^{-1},B\overline{B}}$ and $N^{\text{data}}_{\text{SR,sig}}$ as Poisson quantities. Thus the expression below could be obtained from standard error propagation formulae (inclusive of correlations) of a quotient.

$$\frac{\Delta\mathcal{B}}{\mathcal{B}} = \frac{1}{\sqrt{\frac{1-\sqrt{2f^{+-}\epsilon_{\text{sig+tag}}\mathcal{B}_{SM}}}{N_{\text{sig}}} + \left(\frac{\Delta N_{362\text{fb}^{-1},B\overline{B}}}{N_{362\text{fb}^{-1}B\overline{B}}}\right)^2}} \tag{5.18}$$

Further simplifications include assuming $\epsilon_{\text{sig+tag}}$ is that of this analysis, as well as that $(\Delta N_{B\overline{B}}/N_{B\overline{B}})^2$ will remain unchanged from its value derived from Table 5.1 for higher luminosities. Thus, it is seen in Figure 5.14 that if the analysis strategy of this thesis were used at least $7\,\text{ab}^{-1}$ of $B\overline{B}$ events, the statistical uncertainty in $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ can reduced to a fifth of the branching fraction, where 25 events will be identified in almost 10 billion $B\overline{B}$ pairs. Such an integrated luminosity is well within the scope of the Belle II experiment, which aims to collect $50\,\text{ab}^{-1}$ over the next decade.

- **Larger MC dataset**: While the precision of the measured branching fraction can increase with a higher luminosity collision dataset, constraints can also be improved by increasing the amount of MC used. This becomes especially important given the $\frac{S}{\sqrt{S+B}}$ optimisation procedure in Section 4.2.3, where selections could be made on the basis of a single less background event in an already low yield signal region. The more underlying events that model the low statistics of the number of events in the signal region, the more stable a selection can be considered to be. This would also allow for a finer binning to be used in the signal region plots of Section 4.2.4. Here, a single underlying event was used to model the fractional yield of the $c\overline{c}$ component, which made judging if further $c\overline{c}$ rejection was required challenging. Thus more raw MC may also improve the modelling of the selection criteria boundaries. An additional consequence of the size of the MC dataset implemented was the lack of unscaled $u\overline{u}$, $d\overline{d}$ and $s\overline{s}$ events that remained after the final selections were decided. It is unlikely this is a reflection of there being zero possibility of the continuum events surviving the selection. Rather, it reflects that there was insufficient raw MC to precisely model the small, but non-zero probability of containing $u\overline{u}$, $d\overline{d}$ and $s\overline{s}$ events in the signal region. With zero raw MC events remaining, it was estimated that these components would not contribute to the overall statistical uncertainty on the estimated background yield in data. Given that a fractional, non-zero number of events should be expected, this uncertainty is potentially underestimated. In lieu of using a larger MC dataset, the fractional number of $u\overline{u}$, $d\overline{d}$ and $s\overline{s}$ events that would remain in the signal region and thus their statistical contributions could be extrapolated from sideband yields. Specifically, the *uds* contributions in the signal region could be estimated by relaxing the $E_{\text{ECL}} < 0.25$ GeV selection, obtaining a ratio of signal region to sideband yields, and then using this to scale the original sideband *uds* yield into a signal region estimate. However, with a large enough MC dataset, such estimations would not be necessary and the statistical uncertainty contribution from *uds* would be available directly.

- **Treatment of statistical uncertainties**: Throughout this thesis, statistical uncertainty has been treated as $\sqrt{N}$ by assuming Poisson statistics. However, when introduced at the end of Section 4.1.5, it was addressed that this is not the most accurate treatment of statistical uncertainty. Instead, an uncertainty that takes into account the number of unscaled MC events that represent the yields should be used, of the form $N/\sqrt{N_{raw}}$, for N_{raw} MC events that have been scaled to yield N events. Thus future iterations of this work will require a reconsideration of the yields of MC statistical uncertainties and how this affects the resultant upper bound. Using the unscaled-based statistical uncertainty also incorporates the expectation that using a larger underlying MC dataset should result in higher statistical precision of measured observables and thus smaller statistical uncertainty.

- **Alternate signal selection optimisation strategies**: To determine the optimal selections that reject background and prioritise signal retention, a procedure using the signal optimisation variable $\frac{S}{\sqrt{S+B}}$ is adopted. Here, multiple selection variables and their bounds were decided sequentially, where the resultant selection was chosen such that it had maximal $\frac{S}{\sqrt{S+B}}$. While this method was successful in determining a set of selection criteria that rejected background contributions in the signal region, some selections such as the tight constraint on E_{ECL} may have been too stringent. The efforts to prioritise the ratio of $B^+ \to \mu^+\nu_\mu$ events in the signal region required the complete removal of raw *uds* events, as well as many suitable $B^+ \to \mu^+\nu_\mu$ events. This may be addressed by revisiting the selections of $\frac{S}{\sqrt{S+B}}$ procedure via a $N-1$ analysis. Here, all but one of the decided selections are implemented and a new bound on the remaining selection is trialled in the same manner as previously. Alternatively, machine learning methods may offer a more sophisticated method of leveraging the same features of $B^+ \to \mu^+\nu_\mu$ decays to obtain more discernible signal events with well-modelled backgrounds. Boosted decision tree learning could similarly consider sequential selections that best divide the data into signal-like and background-like subsets. However, while this analysis only considers selections that maximises $\frac{S}{\sqrt{S+B}}$ in each step, the boosted decision tree algorithm can investigate a much larger area of the selection parameter space, by considering multiple sequences of initial and secondary selections. In this way, the resultant selections decided by the boosted decision tree algorithm could maximise the final $\frac{S}{\sqrt{S+B}}$ (or alternative performance metrics) over multiple sequences, which may be an improvement to the sequence of selections obtained by maximising at each step. Another suitable machine learning method is to use neural network learning. Neural network methods have an advantage over boosted decision tree methods as they can consider combinations of variables, whose weight contributions are adjusted simultaneously rather than sequentially. Such a method excels in exploiting non-linear correlations between variables, as opposed to traditional 'rectangular' selections. Both machine learning methods are available in `basf2` and could thus improve the signal retention of this analysis. The drawback of using such methods in the future would be the requirement for more signal and background MC to be created. This is needed so as to establish distinct training, testing and validation datasets required to minimise biases in machine learning methods, while maintaining (or improving) the current statistical precision of the analysis. While not a machine learning algorithm, the $\frac{S}{\sqrt{S+B}}$ optimisation procedure was performed in the same dataset as what was used to evaluate the yields used in the evaluation of the $B^+ \to \mu^+\nu_\mu$ branching fraction. Future analyses could validate the sequence of selections in an alternate MC dataset, to thus ensure that signal efficiency derived is reflective of $B^+ \to \mu^+\nu_\mu$ events generally and not the biases of the specific subset of MC used.

- ***B*-frame muon momentum and other alternate signal variables**: When deciding which signal variable to use in Section 4.2.2, a number of constructed kinematic variables were introduced as alternatives. This includes variables which combine the tag-side and signal-side $\cos\theta_{BY}$ information from previous Belle [113] and Babar [112] analyses. Additionally, using an estimated version of the B-frame muon momentum to obtain the B-frame muon momentum was discussed. One of the largest drawbacks of using a semileptonic B_{tag} is the lack of complete four-momentum information due to the unmeasured neutrino. Thus the singular momentum of the muon boosted into the parent B-frame, first derived in Section 1.2.1, cannot be used to identify signal events in this analysis. This is a clear advantage of hadron and inclusive tagged modes, which was hinted at given the use of B-frame muon momentum in these analyses and their better-constrained $B^+ \to \mu^+\nu_\mu$ branching fractions demonstrated in Figure 5.11. Preliminary investigations were done to emulate the Diamond Frame method of the 2010 Babar $B^+ \to \mu^+\nu_\mu$ search to estimate the four-momentum of the FEI B_{tag}. This was seen to have poor agreement with truth B-frame muon momentum distributions, and discontinued in favour of the simpler $\Upsilon(4S)$-frame muon momentum. Renewed efforts in improving the semileptonic FEI B_{tag} with the Diamond Frame method can allow the use of the B-frame muon momentum, allowing for a tighter signal region that further rejects background contributions and improves signal-to-background ratios. Such constraints could also function as additional selection criteria, by considering the characteristics of the two missing neutrinos that characterise semileptonic-tagged $B^+ \to \mu^+\nu_\mu$ events. The six degrees of freedom produced by the two neutrinos three-momenta can be reduced using the known energy and geometry of the $\Upsilon(4S)$ event, where the solution space of remaining unknowns can be sampled to potentially produce new signal discriminating variables. This includes methods that combine tag-side and signal-side $\cos\theta_{BY}$ information, which may provide additional powerful background rejection.

- **Hadronic FEI, Inclusive B-tagging and GraFEI**: The primary advantage of using a hadronically decaying B_{tag}, as opposed to the semileptonic B_{tag} used in this analysis, is in the purity of events selected. This is namely due to the complete B_{tag} four-momentum, which has already been discussed in this section. The hadronic FEI was explored in Section 3.4.3 as a potential tagging mode for this analysis, where it was not pursued in favour of the higher retention of the semileptonic FEI. Once sufficient integrated luminosity (i.e. multiple attobarns) is achieved by Belle II, the purity of hadronic decay modes can lead to higher precision constrains on the $B^+ \to \mu^+\nu_\mu$ branching fractions. With the current sub-attobarn luminosity available, pursuing the hadronic-tagged $B^+ \to \mu^+\nu_\mu$ search alongside the semileptonic-tagged one will allow for a larger dataset of $\Upsilon(4S) \to B\overline{B}$ decays in which to search for the rare leptonic decay. Results of a combined search could therefore improve the current estimate of the $B^+ \to \mu^+\nu_\mu$ branching fraction. The use of a hadronic B_{tag} the FEI would also allow for comparison with the Belle 2015 result, where the improvement of the FEI on the Full-Reconstruction algorithm could be examined. Additionally, the lower luminosity could be mitigated by using an inclusive B_{tag} in a future study. Inclusive B-tagging, in principle, has higher retention that exclusive tagging, with no assumptions made on the specific B-decay mode. The drawback of this is the need for sophisticated selection criteria, to be able to reject background contributions without access to decay-specific variables. A solution to this would could be to use the graph-based neural network version of the

FEI, GraFEI, discussed in Section 3.4.6. The GraFEI inclusively identifies final state particles most likely to belong to a B_{tag} decay tree, without reconstructing a limited set of predetermined B decays. Thus, a new $B^+ \to \mu^+\nu_\mu$ search using the GraFEI could increase efficiency by identifying B_{tag} candidates beyond the 8 used in the semileptonic FEI. The method could also have increased purity, with the potential for neural network-based learning to identify subtle features of suitable inclusive B_{tag} candidates, which may not be available when building inclusive B_{tag} candidates using traditional selection criteria. Using GraFEI-based tagging for a $B^+ \to \mu^+\nu_\mu$ search would also allow comparison with the inclusive B_{tag} search performed by Belle in 2020. While there are currently no searches within the collaboration using the GraFEI with a $B^+ \to \mu^+\nu_\mu$ search, there are efforts to extend the Belle 2020 analysis to Belle II data [126]. This result aims to be a combined Belle and Belle II result, which will be discussed in the next improvement.

- **Combined Belle and Belle II analysis**: Many discussions of uncertainties throughout this Chapter have centered on the small expected signal yield, compared to the original Belle dataset. While additional data is being recorded, it may still be some time before it is calibrated and skimmed. To address this discrepancy in a more timely manner, it is possible to use the $711\,\text{fb}^{-1}$ of $e^+e^- \to \Upsilon(4S)$ dataset recorded by the original Belle experiment. Repeating this analysis using the exact same dataset as previous Belle $B^+ \to \mu^+\nu_\mu$ searches would allow easier comparison of the efficacy of different B-tagging analysis strategies. Additionally, with the ability of `basf2` to analyse original Belle events via the `B2BII` library, it is possible to produce a future measurement with a combined Belle and Belle II dataset. Such a combined study would be the largest set of $B\overline{B}$ events used in a $B^+ \to \mu^+\nu_\mu$ search, with around 1 billion $B\overline{B}$ pairs, producing approximately 500 $B^+ \to \mu^+\nu_\mu$ events. Assuming the $\epsilon_{\text{sig+tag}}$ used in this analysis, this corresponds to 4.03 signal events, around three times what is predicted at the current luminosity.

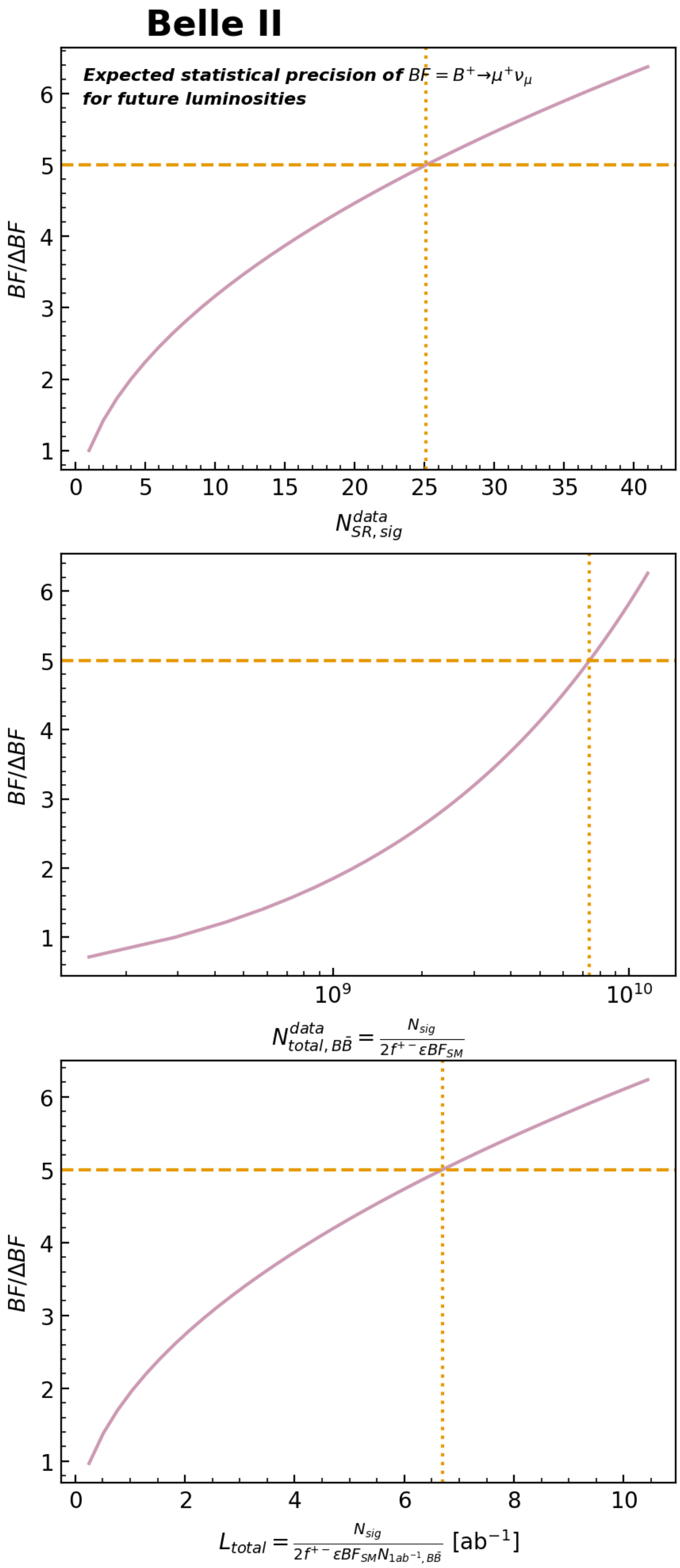


Figure 5.14. Predicted proportional statistical uncertainty in $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ given the measured signal efficiency and assuming the $B^+ \to \mu^+\nu_\mu$ SM branching fraction. The proportional uncertainty is presented to vary with the number of $B^+ \to \mu^+\nu_\mu$ events measured, number of $B\overline{B}$ events used and their corresponding integrated luminosity.

5.7. Summary

In this Chapter, the goal of this thesis, that is to measure the branching fraction of $B^+ \to \mu^+\nu_\mu$ events measured in $362\,\mathrm{fb}^{-1}$ of data recorded at the Belle II experiment, was partially completed.

The Chapter began with a derivation of how the $B^+ \to \mu^+\nu_\mu$ branching fraction could be constructed via a ‘cut-and-count’ procedure using the data and MC yields constrained in Chapter 4. The key parameters to the branching fraction derivation were identified to be the estimate of data yield due to background events ($N^{\mathrm{data}}_{\mathrm{SR,bkg}}$) and the combined signal and tagging efficiency ($\epsilon_{\mathrm{sig+tag}}$).

Before any evaluation could take place, a rigorous survey of the sources of systematic uncertainty in the analysis was presented. Several sources originated from corrections made to improve MC modelling of collision data, or to improve how well the recorded subdetector signals reflect the measured physics phenomena. These included systematic uncertainties from correction factors in particle ID, track momentum scaling and momentum range.

As the signal efficiency was derived using purely MC, the signal strategy was applied to an alternate B decay, $B^+ \to \bar{D}^0\pi^+$. This decay was chosen as it matched the two-body kinematics of $B^+ \to \mu^+\nu_\mu$, where the pion could play the role of the signal muon. The control mode analysis was used to assess if the signal strategy showed preference for $B^+ \to \mu^+\nu_\mu$ events in MC over those in data. This was especially important as no data-MC corrections had been made to account for the FEI tagging, where the dataset used to train the algorithm contains only MC decays. A bias towards MC decays would result in an underestimation of the number of $B^+ \to \mu^+\nu_\mu$ events in the Belle II collision dataset, so a correction factor on the signal efficiency was derived. This was taken to be the ratio of data-driven and MC-based efficiencies, where the former worked backwards from the expected $B^+ \to \bar{D}^0\pi^+$ branching fraction to obtain the efficiency. This correction factor was very close to 1, indicating that very little bias was shown to either data type. With this factor and its associated uncertainties, the estimated background data yield in the signal region and the now calibrated signal efficiency were complete.

A confidence limit on the ultimate branching fraction was then created via the method of Feldman and Cousins. This involved the creation of ‘toy’ MC, which emulated the measurement of the signal yield as the difference between the amount of data in the signal region and background estimate. The resulting confidence limit would consist of the set of 'true' signal yields that could be taken to be consistent with a measured signal yield. The convention set by previous $B^+ \to \mu^+\nu_\mu$ searches was to find the subset of possible signal yields that had 90% confidence of originating from the measured result.

The total yield of the data was not examined in this experiment, so as to avoid temptation to bias corrections and confidence intervals towards a significant branching fraction measurement. This yield remains unmeasured, as discussions within the Belle II review committee have not yet converged. As such, a single confidence limit cannot be quoted on the number of $B^+ \to \mu^+\nu_\mu$ collision events recorded. However, with the understanding that the following is not necessary what will be observed when measuring the real signal yield, the confidence limit on the signal yield predicted by the analysis can be quoted. In a similar method to how the background data yield was estimated, the signal yield can be estimated to be $N^{\text{data}}_{\text{SR,sig}}$=1.47 $\pm$ 1.31(stat) $\pm$ 0.08(sys). From the Feldman-Cousins method, the 90% upper limit on this potential yield can be taken to be $N^{\text{data}}_{\text{SR,sig}} < 10$. This corresponds to the branching fraction 90% upper limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29 \times 10^{-6}$, which is an order of magnitude larger than the SM prediction. This upper limit is not the current best limit on the $B^+ \to \mu^+\nu_\mu$ branching fraction, thought it has improved on the previous semileptonic-tagged result obtained by BaBar in 2010 [43]. Additionally, the central value of the branching fraction for this potential yield, which is inclusive of the null result, was found to be $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ $=6.37\times10^{-7} \pm 1.55\times10^{-6}$(stat) $\pm$ 3.08×10^{-7}(sys). Again it is emphasised that in absence of measuring the full yield in collision data, these are predicted (and not guaranteed) measurements for when the data yield is properly examined.

A number of improvements to the signal strategy. While unblinding the signal yield will be the focus of future iterations, there also other factors that can be addressed. This includes minimising the effect of large statistical uncertainty by using newly recorded Belle II data or original Belle data, as well as looking in to alternate B-tagging methods to complement the semileptonic dataset. The proportional uncertainty of the branching fraction was extrapolated to determine what luminosity of $e^+e^- \to B\overline{B}$ events would be required to obtain precision of one-fifth of the branching fraction. This was found to be $7\,\text{ab}^{-1}$, just less than 20 times the dataset used for this analysis, but still within Belle II's goal of recording $50\,\text{ab}^{-1}$ over the next decade.

To conclude the section, measurements of the CKM matrix parameter, $|V_{ub}|$ were derived. These were obtained using the predicted signal yield and its corresponding measured branching fraction. The bounds placed on $|V_{ub}|$ were not competitive with precision of the current HFLAV averages. They also did not support either of the inclusive of exclusively determined results over one another. This demonstrated that the analysis, at its current luminosity, is currently unable to resolve the current tension in the two results.

This concludes the work of this thesis. The next section will place this result in the context of the previous four Chapters, as well as summarise the outlook for this $B^+ \to \mu^+\nu_\mu$ analysis.

$\bowtie \bowtie \bowtie$

Chapter 6.

Conclusion

This thesis has presented a search for the rare leptonic B decay, $B^+ \to \mu^+\nu_\mu$, using data recorded with the Belle II detector. This was performed using $362\,\mathrm{fb}^{-1}$ of $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ processes, occurring from 2019 to 2022. The exclusive-tagging algorithm, the Full Event Interpretation (FEI), was used for the first time to tag events containing semileptonic B decays that may occur alongside the B decay of interest. This was performed with the goal of improving previous limits on the $B^+ \to \mu^+\nu_\mu$ branching fraction from the previous 20 years of particle physics searches. A summary of the content of this thesis will be presented below, followed by the key results and an outlook on the future of this analysis.

6.1. Summary

In Chapter 1, the $B^+ \to \mu^+\nu_\mu$ process was introduced as a rare leptonic decay theorised by the Standard Model of Particle Physics. Specifically, $B^+ \to \mu^+\nu_\mu$ was introduced as a flavour-physics process, as it involves the annihilation of two quarks of different generations at tree level. This invokes the quark flavour mixing of the CKM mechanism, which has parameters that could indicate the presence of charge parity violation when they are measured. This led to the expression for the SM $B^+ \to \mu^+\nu_\mu$ branching fraction, which is proportional to CKM parameter V_{ub} as $|V_{ub}|^2$. Key SM properties of the $B^+ \to \mu^+\nu_\mu$ decay that could be leveraged in the $B^+ \to \mu^+\nu_\mu$ search were discussed, such as the signal muon kinematics within the two body B-decay, the 'missing energy' due to the undetectable neutrino and the low-QCD background expected in a leptonic B-decay. Next, the 20 years of previous $B^+ \to \mu^+\nu_\mu$ searches by electron-positron collision experiments such as CLEO, BaBar and Belle were reviewed, with emphasis on the current best result from Belle in 2020 and the last semileptonic-tagged result in 2010. The Chapter concluded with a discussion of what aspects of the previous analyses would be useful in the Belle II study related to this thesis.

Chapter 2 focused on how the $e^+e^- \to B\overline{B}$ processes used in this $B^+ \to \mu^+\nu_\mu$ search were created, using the SuperKEKB accelerator that operates for the Belle II experiment. The origins and goals of the Belle II physics program were introduced, as a flavour factory experiment capable of constraining BSM phenomena at the luminosity and precision frontier. The function of the SuperKEKB accelerator was described, from individual electron sources to electron-positron collisions at energy equivalent to the $\Upsilon(4S)$, and consequently coalescing as $B\overline{B}$ pairs. The kinematics of the $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ process were presented, with emphasis on the heavy momentum and energy constrains are on B mesons produced using SuperKEKB. Cases where the e^+e^- collision at $\sqrt{s} = 10.58\,\text{GeV}$ produce background continuum events, or when the collision takes place at an alternate energy are discussed. The Belle II detector, which surrounds the e^+e^- of the collision, was then described as a series of subdetectors that each provide unique information about particles resulting from the collision event. How the raw electronic information recorded by the subdetectors become reconstructed into particle observables was then presented, which relies heavily on high-performance computing and data processing workflows. The role of MC simulation in modelling Belle II data and how it is analysed using the Belle II Analysis Software Framework (`basf2`) was introduced. Finally, the attributes of the $362\,\text{fb}^{-1}$ of $B\overline{B}$ events at Belle II were outlined, in addition to MC dataset compatible with the collision data.

Chapter 3 presented the algorithm that would be used to partially reconstruct $B\overline{B}$ events suitable for the $B^+ \to \mu^+\nu_\mu$ search, the Full Event Interpretation (FEI). To begin, the advantages and disadvantages of different B-tagging modes were discussed, with emphasis on which modes excel signal purity and which excel in signal efficiency. The FEI was then introduced, as an exclusive-tagging machine learning algorithm which selects the best final state particles to create the most probable B_{tag} candidates. The specifics of the algorithm were discussed before presenting the choice of operating modes of the FEI, namely that of using hadronic B decays against semileptonic B decays. Before investigating the observables of each mode, a best candidate procedure was determined so as to choose the most suitable B_{tag} produced by the FEI per event. Key distributions of the hadronic FEI, as applied to $200\,\text{fb}^{-1}$ of charged MC, were presented next. Attributes such as the specific B_{tag} reconstruction, the signal probability output from the algorithm and the B_{tag} energy and mass were investigated. Where relevant, the same plots for the semileptonic FEI were presented, with additional semileptonic-specific variables like $\cos\theta_{BY}$ and the D/D^* mass splitting were shown. By consulting these distributions, as well as the efficiency and purity statistics of each mode in MC fixed to produce $B^+ \to \mu^+\nu_\mu$ decays, it was decided that the semileptonic FEI would be more appropriate for this rare decay search, as a higher efficiency tagging mode. At this stage it was necessary to only consider B_{tag}'s with tag-side lepton momentum less than 2.5 GeV/c, so as to avoid the high momentum of the signal muon being mistaken for a tag-side one. The same semileptonic distributions created in charged MC could now be replicated using a dedicated B^+B^- sample where one B meson was fixed to proceed as $B^+ \to \mu^+\nu_\mu$ (or its charged conjugate), so as to investigate the behaviour of those B_{tag}'s decaying alongside $B^+ \to \mu^+\nu_\mu$. Finally, the distributions were again reproduced, but this time incorporating background MC scaled according to the magnitude expected in the 362 fb^{-1}. These distributions demonstrated the vast difference between the number of $B^+ \to \mu^+\nu_\mu$ signal events expected and the much larger number of background events expected. This emphasised the need for signal-optimising selection criteria, so as to allow for the best signal yield extraction.

After demonstrating the large difference between signal and background events expected in the B-tagged 362 fb^{-1} dataset, Chapter 4 explored how to best identify events that contain $B^+ \to \mu^+\nu_\mu$ processes from non-$B\overline{B}$ and non-$B^+ \to \mu^+\nu_\mu$ events. With the tag-side reconstruction performed via the FEI, the signal-side reconstruction began by identifying the muons most likely to have originated from the $B^+ \to \mu^+\nu_\mu$ decay. The best candidate signal muon criteria was decided using $B^+ \to \mu^+\nu_\mu$ MC, where the choosing the muon with highest $\Upsilon(4S)$-frame momentum of the event performed better than the detector-based muon identification probability `muonID`. After this the muons that remain after B-tagging were analysed, where it was decided that requiring the tag-side muon momentum be lesser than the signal side muon momentum was a more powerful constraint in preventing signal-side muons being used in B_{tag} reconstruction. Key distributions were presented again, for the B_{tag} and signal muon before and after reconstruction was performed. Doing so presented how the rejection of background events improved after constructing the $\Upsilon(4S)$ candidate, while also allowing for a preliminary look at what selections could be made in signal optimisation. The distributions now also contained signal muon variables such as the $\Upsilon(4S)$-frame momentum and the opening angle between the signal-side muon and the tag-side lepton. The strategy to systematically choose selection criteria that optimise the significance estimator $\frac{S}{\sqrt{S+B}}$ was then introduced. This estimator was first used to determine the signal variable that best discriminates signal and background within a signal region, where $p^*_\mu \in[2.45,2.85)$ GeV/c was chosen. Next, the estimator was used to determine the successive selections in this region, where more than 100 variables were trialled. This resulted in cuts based on unused energy in the detector, the thrust axes of the event, $\cos\theta_{BY}$, missing momentum of the event as well as tightening the signal region to $p^*_\mu \in[2.5,2.85)$ GeV/c. With the set of selections chosen, they were applied to the full-set of MC and data, to assess the remaining yields in background and signal. Distributions were again produced to understand how well the MC-simulated data matched collision data. Collision data was only inspected outside of the signal region, so as to proceed with a blinded analysis and avoid biases towards a desired result.

In Chapter 5, the yields from the previous Chapter were used to quantify sources of systematic uncertainty, to understand how they could impact the calculation of the $B^+ \to \mu^+\nu_\mu$ branching fraction in this analysis. A method of evaluating an experimental $B^+ \to \mu^+\nu_\mu$ branching fraction was put forward, where the signal yield in data would be obtained as the difference between the number of observed events in data and events estimated to be from background events, according to a 'cut-and-count' analysis approach. Before substituting the yields to obtain the final branching fraction, systematic uncertainties were derived for each input yield required for the calculation. The assigned systematic uncertainties accounted for differences in lepton identification between data and MC, uncertainties in magnetic field extrapolation of track momentum during calibration in data and data-MC differences in tracking efficiency. To account for potential differences in signal efficiency in data and MC, the analysis was repeated for $B^+ \to \bar{D}^0\pi^+$ signal. As this decay is more precisely known than $B^+ \to \mu^+\nu_\mu$, the $B^+ \to \bar{D}^0\pi^+$ efficiency and yields derived from the signal strategy could be compared to what is expected from the $B^+ \to \bar{D}^0\pi^+$ branching fraction, from which a calibration factor for the $B^+ \to \mu^+\nu_\mu$ efficiency is derived. With all corrections and uncertainties understood, they were used alongside the background yields in the method of Feldman and Cousins, a procedure used to obtain 90% confidence limits on potential signal yields. As unblinding the signal yield requires the permission of a internal Belle II review committee and such discussions have not converged in time for this thesis, the yield in the signal region is not measured. To proceed, an estimate of the signal yield in data derived from MC is constructed, which is then used to predict and evaluate the 90% upper limit on the branching fraction, in addition to the absolute branching fraction with propagated uncertainties. A prediction of $|V_{ub}|$ was also derived, using the theoretical branching fraction expression and the experimental branching fraction pertaining to the predicted signal yield.

The results from this analysis are summarised in the next section.

6.2. Results

The analysis pertaining to this thesis was performed as a blinded experiment, where the subset of data used to produce the final branching fraction was kept concealed, to avoid biasing the calculation towards a non-null result. Future efforts are required within the Belle II collaboration before the $B^+ \to \mu^+\nu_\mu$ branching fraction can be evaluated using this subset. However, after a rigorous systematic analysis, an expected branching fraction can be stated with well-defined uncertainty and confidence limits. Thus the results of this thesis are summarised below.

- From the Belle II dataset of $362\,\text{fb}^{-1}$ of $B\overline{B}$ processes, or 387.1 million $B\overline{B}$ pairs, 169 of the charged B mesons are expected to proceed as $B^+ \to \mu^+\nu_\mu$ (and its charged conjugate), assuming the SM branching fraction of 4.24×10^{-7}.
- The signal efficiency of measuring the $B^+ \to \mu^+\nu_\mu$ processes was obtained to be $(7.83 \pm 0.03 \pm 3.30){\times}10^{-3}$, where the uncertainties are statistical and systematic respectively.
- Number of events in data expected to be from background events in signal region (muon $\Upsilon(4S)$-frame momentum as $p^*_\mu \in$[2.5,2.85) GeV/c) was estimated to be $9.01 \pm 3.51 \pm 0.47$.
- In absence of unblinding, a predicted number of signal events expected in the signal region can be evaluated, as the product of the MC yield and that data-MC ratio in the sidebands. This value is $1.47 \pm 1.31 \pm 0.08$. This agrees within uncertainties of the estimate from the SM branching fraction.
- The 90% confidence interval pertaining to this predicted yield is $N^{\text{data}}_{\text{SR,sig}} < 10$. This corresponded to an upper limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 3.29\ {\times}10^{-6}$.
- This did not improve the previous 90% limit by the Belle inclusively-tagged limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 0.86\ {\times}10^{-6}$ [27]. It was, however, better constrained than the prior semileptonically-tagged limit by BaBar of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 11\ {\times}10^{-6}$ [43]. This has taken place, even though a larger $B\overline{B}$ dataset was used compared to the dataset analysed in this thesis.

- The nominal branching fraction evaluated from the predicted yield is $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (6.4\pm 15.5 \pm 3.1) \times 10^{-7}$. The branching fraction matches the same order of magnitude as the SM prediction, though the propagated uncertainties result in an overlap with zero.
- The current size of the dataset has lead to a larger source uncertainty in this measurement. If this analysis were to be repeated (with the same signal efficiency) in a higher luminosity dataset, more than 6500 fb^{-1} of $B\overline{B}$ processes (i.e. 7 billion) would be required to achieve a statistical uncertainty one-fifth of the $B^+ \to \mu^+\nu_\mu$ branching fraction.
- As this analysis did not improve the current limit on the $B^+ \to \mu^+\nu_\mu$ branching fraction, it also will not improve limits on the relevant CKM matrix element $|V_{ub}|$. For completions sake, $|V_{ub}| < 7.1\times 10^{-3}$ was derived as the 90% confidence interval and $|V_{ub}|$=4.72 $\pm$ 5.7 4$\pm$ 1.13 as the nominal value of $|V_{ub}|$. This supports both the exclusive $(3.51 \pm 0.12)\times 10^{-3}$ and inclusive $(4.19 \pm 0.17)\times 10^{-3}$ HFLAV averages, with insufficient precision to distinguish between them.

6.3. Outlook

With the rarity expected with the $B^+ \to \mu^+\nu_\mu$ SM decay, combined with the limited size of the initial dataset recorded by the Belle II experiment, future iterations of the measurement of the $B^+ \to \mu^+\nu_\mu$ branching fraction will continue to be challenged by statistical uncertainties until a larger number of B decays is obtained. At these early stages, using semileptonic B-tagging remains useful in maximising the efficiency of the signal strategy, where its been seen that the use of an exclusive-tagging algorithm like the FEI greatly improves the number of tagging modes implemented, as well as the ease with which suitable $B\overline{B}$ events are provided. Provided unblinding proceeds as has been predicted, this search stands to be the first $B^+ \to \mu^+\nu_\mu$ result published by Belle II, as well as the first $B^+ \to \mu^+\nu_\mu$ search using the FEI. In the next ten years, once Belle II enters the attobarn era, it would be expected that the purity of the hadronic modes of the FEI will become more competitive than the semileptonic case in providing narrow limits on the $B^+ \to \mu^+\nu_\mu$ branching fraction. The latest limits on $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ have challenged the latest BSM models, such as contributions from a charged Higgs interaction. However, to exclude beyond Standard Model phenomena from being the major contribution to the $B^+ \to \mu^+\nu_\mu$ process, a statistically significant

$B^+ \to \mu^+\nu_\mu$ branching fraction consistent $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ =4.24×10^{-7} is required. In this pursuit, using semileptonic-tagged modes in $B^+ \to \mu^+\nu_\mu$ searches can continue to provide a complementary dataset to hadronic-tagged modes, to sooner achieve the goal of a precision $B^+ \to \mu^+\nu_\mu$ measurement with the Belle II experiment.

⋈ ⋈ ⋈

Appendix A.

The effect of variations in the selection criteria on the final $B^+ \to \mu^+\nu_\mu$ branching fraction

Throughout Chapter 5, significant efforts were dedicated to understanding how aspects of the signal procedure could impact the final $B^+ \to \mu^+\nu_\mu$ branching fraction. This included considerations from the Belle II detector, data-MC corrections and the $B^+ \to \bar{D}^0\pi^+$ control mode. The resulting systematic uncertainties were then propagated to the final result.

In considering which aspects of the signal procedure may invite systematic uncertainty, an attempt was made to account for the finite step width used with optimising $\frac{S}{\sqrt{S+B}}$ in selection criteria boundaries. In Section 4.2.3, an algorithm was defined to systematically adopt selection criteria that would maximise the significance estimator $\frac{S}{\sqrt{S+B}}$, derived from the resulting signal and background yields. Evaluating a systematic uncertainty on this procedure was done with the goal of addressing the scenario where the 'true' maximum $\frac{S}{\sqrt{S+B}}$ occurs at a different selection boundary than what was chosen in Section 4.2.3. Thus, varying the used selection boundaries would assist in understanding how the measured yields would be impacted if an unknown experimental factor shifted the chosen selection from the 'true' maximum.

The uncertainty was estimated by varying selection boundaries one at a time, by a single bin higher or lower than what was chosen in the $\frac{S}{\sqrt{S+B}}$ optimisation. The difference between the resulting yields were halved and then treated as the uncertainty one the original yields. However, the magnitude of the quantity treated as the uncertainty ended

Table A.1. Table of selection variations. Cut variation magnitude is chosen to be the same as the step width used when maximising $\frac{S}{\sqrt{S+B}}$. Percentage cut variation indicates what proportion of the cut boundary the cut variation represents.

Variable	Cut boundary	Cut variation	Cut variation [%]
E_{ECL}	0.25	0.05	20.00
$B_{\mathrm{tag}}\ \cos\theta_{T_B T_{\mathrm{ROE}}}$	0.75	0.05	6.67
$\cos\theta_{BY}$ (upper)	1.00	0.10	10 .00
$\cos\theta_{BY}$ (lower)	-2.00	0.10	5.00
$\cos\theta_{\vec{p}_{\mathrm{miss}}}$(upper)	0.90	0.05	5.56
$\cos\theta_{\vec{p}_{\mathrm{miss}}}$(lower)	-0.90	0.05	5.56
B_{tag} decay mode	4.00	1.00	-

up being much larger than other contributions, bringing into question the reliability of the uncertainty procedure. It may also be the case that such a method is inappropriate for uncertainty estimation, instead better suited to assessing the stability of the determined selection criteria. The consideration of uncertainties in the selection procedure was thus removed from the main body of this work. It is instead reproduced here for completeness, where future efforts may be needed to consider its place within a $B^+ \to \mu^+\nu_\mu$ search.

A.1. Method

Table A.1 contains the selection variables of the $B^+ \to \mu^+\nu_\mu$ analysis in addition to the selection boundaries, the magnitude each selection is varied by, and the variation from the central value expressed as a percentage. A systematic uncertainty will only be estimated in the selection variables. Thus, the boundaries of the signal variables are excluded from the treatment (i.e. no variation in p^*_μ), as well as the pre-selection variables outside of the signal optimisation (i.e. no variation of the number of remaining tracks in the ROE after $\Upsilon(4S)$ reconstruction). Additionally, the decay mode number is treated ordinally, despite being a mostly categorical variable that refers to the included B-tagging decay modes. Originally, selection on the decay mode was allowed to be treated ordinally, as omitting the final four modes was found to be consistent with placing a limit on a continuous variable. This variable was the energy of the hadronic system of the B_{tag}, where the selection limited it to the energy of a D meson alone (i.e. excluding the four decay modes which contained a pion) in Section 4.2.3. Thus, to vary the included B_{tag} decay modes, the yield will be measured after the inclusion of a single pionic mode ($B^+ \to D^-\pi^+e^+\nu_e$) and the removal of one of the non-pionic modes ($B^+ \to \bar{D}^*\mu^+\nu_\mu$).

To estimate the systematic uncertainty on each yield measured, one selection boundary from Table A.1 is varied at a time, with all other selections implemented in the sample. The number of events retained after increasing and decreasing the selection boundary is recorded. This is done both in the muon momentum signal region and sideband as required. Half of the absolute difference in events is treated as the uncertainty due to the given selection. The total systematic uncertainty predicted to be from the selection procedure is taken to be the half differences derived from each selection added in quadrature.

A.2. Results

A.2.1. $B^+ \to \mu^+\nu_\mu$

Table A.2 contains the systematic uncertainties due to the selection procedure, alongside their corresponding yields in data and MC.

Table A.2. Systematic uncertainty due to signal selection, for yields in p^*_μ signal (SR) and sideband (SB) regions. Yields are formatted as $N \pm N_{\rm stat} \pm N_{\rm sys}$, where statistical error is $\sqrt{N}$.

Sample	Reg.	Yield	$\%\Delta N_{\rm stat}$	$\%\Delta N_{\rm sys}$	Reg.	Yield	$\%\Delta N_{\rm stat}$	$\%\Delta N_{\rm sys}$
$B^+ \to \mu^+\nu_\mu$	SR	1.29±1.14±0.18	87.92	14.26	SB	0.11±0.33±0.02	301.70	15.07
B^+B^-	SR	5.96±2.44±1.49	40.97	25.03	SB	38.17±6.18±9.8	16.18	25.66
$B^0\overline{B}^0$	SR	1.74±1.32±0.38	75.73	21.53	SB	32.33±5.69±7.6	17.59	23.52
$c\overline{c}$	SR	0.24±0.49±0.41	203.91	170.41	SB	1.37±1.17±0.62	85.33	45.21
$u\overline{u}$	SR	0.00±0.00±0.14	0.00	-	SB	0.26±0.51±0.19	194.99	70.71
$d\overline{d}$	SR	0.00±0.00±0.00	0.00	-	SB	0.0±0.0±0.13	0.00	-
$s\overline{s}$	SR	0.00±0.00±0.29	0.00	-	SB	0.0±0.0±0.18	0.00	-
tot. bkg.	SR	7.94±2.82±2.34	29.58	29.45	SB	72.14±8.49±17.76	11.36	24.62
MC	SR	9.24±3.04±2.52	32.01	27.29	SB	72.25±8.50±17.78	11.37	24.60
data	SR	0.00±0.00±0.00	0.00	-	SB	82.00±9.06±21.14	11.04	25.78

Table A.3. Inputs to $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$ affected by systematic uncertainty in signal selection.

Observable	$N \pm \Delta N_{\text{sys}}$	$\%\Delta N_{\text{sys}}$
$N^{\text{data}}_{\text{SB,all}}$	82.00±21.14	25.78
$N^{\text{MC}}_{\text{SR,bkg}}$	7.94±2.34	29.45
$N^{\text{MC}}_{\text{SB,all}}$	72.25±17.78	24.60
$N^{\text{MC}}_{\text{SR(gen)},B^+ \to \mu^+\nu_\mu}$	$(5.96{\pm}0.21){\times}10^4$	14.26
$\epsilon_{\text{sig+tag}} \left(= \frac{N^{\text{MC}}_{\text{SR(gen)},B^+ \to \mu^+\nu_\mu}}{N^{\text{MC}}_{\text{gen},B^+ \to \mu^+\nu_\mu}}\right)$	$(7.45{\pm}1.27){\times}10^{-3}$	14.26
$N^{\text{data}}_{\text{SR,bkg}} \left(= N^{\text{MC}}_{\text{SR,bkg}} \times \frac{N^{\text{data}}_{\text{SB,all}}}{N^{\text{MC}}_{\text{SB,all}}}\right)$	9.01±4.23	46.92
$N^{\text{data}}_{\text{SR,all}}$	×	×

The estimated systematic uncertainty due to the selection criteria chosen has contributed large percentage uncertainty, when compared to their counterparts derived from the Belle II detector conditions previously in Table 5.3, Table 5.6 and Table 5.8. This is especially true for the sideband quantities of Table A.1, where the data and MC yields attract a larger percentage systematic uncertainty in the selection procedure than in statistical case. This impacts the calculation of the data yield attributed to background processes in the signal region ($N^{\text{data}}_{\text{SR,bkg}}$), where the uncertainties from the underlying yields are propagated to give 46.92% of the nominal value, as presented in Table A.3.

To understand why this method of estimating systematic uncertainty results in such large values, the calculated uncertainty contribution from each selection boundary is presented. This is done specifically for the data and MC yields that will be used in the branching fraction evaluation, summarised in Table A.4. Each contribution is expressed as the decimal ratio of the variable-specific selection uncertainty to the total selection uncertainty on the yield. From the table, it is seen that variations in the $E_{\text{ECL}} < 0.25$ GeV boundary represents the largest fraction of the total selection uncertainty for all three quantities, more than double that of the next highest contribution from the $\cos\theta_{T_B T_{\text{ROE}}}$ bound. As the selection on the energy of unused clusters in the ECL was the first determined in the $\frac{S}{\sqrt{S+B}}$ optimisation strategy, it is understandable that variations in this criterion lead to the largest changes in yields. From this, it is evident that the $B^+ \to \mu^+\nu_\mu$ analysis is incredibly sensitive to changes in E_{ECL}.

Notably, the method that has been used to assign selection-based uncertainties is not the most stringent approach. With the variations per selection summed in quadrature,

Table A.4. Table of systematic uncertainties per selection criterion, expressed as decimal ratios of the total selection uncertainty for each yield.

Sample	E_{ECL}	$\cos\theta_{T_B T_{\mathrm{ROE}}}$	$\cos\theta_{BY}$ (hi)	$\cos\theta_{BY}$ (lo)	$\cos\theta_{\vec{p}_{\mathrm{miss}}}$(lo)	$\cos\theta_{\vec{p}_{\mathrm{miss}}}$(hi)	B_{tag} mode
$N^{\mathrm{data}}_{\mathrm{SB,all}}$	0.82	0.44	0.09	0.09	0	0.24	0.24
$N^{\mathrm{MC}}_{\mathrm{SB,all}}$	0.84	0.34	0.12	0.06	0.05	0.29	0.27
$N^{\mathrm{MC}}_{\mathrm{SR,bkg}}$	0.80	0.45	0.12	0.07	0.08	0.27	0.24

any correlations between the choice of boundaries have been ignored. It is expected that there is a non-trivial relationship between successive selections. For example, if the E_{ECL} boundary was restricted to $E_{\mathrm{ECL}} < 0.2$ GeV as the first selection, the selections that follow via the $\frac{S}{\sqrt{S+B}}$ optimisation procedure may also differ. An alternative method would be to repeat the optimisation procedure in a large number of new, independent MC datasets. This could then capture the underlying correlations between variables, where the standard deviation of the resulting yields would be treated as the systematic uncertainty. A drawback to such a method is the large amount of additional MC that would be needed to ensure an accurate result, beyond what is currently made available by the Belle II Data Production group.

An alternative to constructing several new datasets is to use bootstrap resampling on the current dataset. Bootstrap resampling creates a 'new' dataset by randomly sampling events (with replacement) from the given dataset, until the number of events in the sample matches that of the original dataset. This is analogous to recording new data, provided that the distribution of variables in the original dataset covers that of the theoretical distribution. Repeating this method a large number of times establishes an ensemble of datasets, which can be used to repeat the procedure and evaluate the standard deviation of yields. While these are not statistically independent datasets, it has been shown that for a large enough ensemble that attributes calculated via this method generally converge to that across newly sampled datasets [127]. Such a method, while sufficient for estimate uncertainties, is not appropriate to use as a replacement for recording new data when seeking increased statistical precision.

Thus, with the knowledge that the systematic uncertainty due to the selection strategy requires more careful consideration, the uncertainties derived in Chapter 5 will be not be used in any of the limit-setting procedures. They will not be combined with the other detector-based systematic uncertainties. For completeness, they will be propagated and quoted separately for the $B^+ \to \mu^+\nu_\mu$ branching fraction, where they can be treated as a preliminary estimate of the impact of the selection procedure. Future efforts must address this considerations, so that it can be combined with the other systematics. As an input to the Feldman-Cousins method, this may in turn increase the estimated upper limits.

A.2.2. $B^+ \to \bar{D}^0\pi^+$

Signal-selection based uncertainties can also be estimated in the $B^+ \to \bar{D}^0\pi^+$ control mode, with the same technique of varying the boundaries matches what was originally described in Section A.1. This is presented in Table A.5, as an extension to the original table in Table 5.14. The Table shows that signal-based uncertainties are much worse than the next largest, in the statistical uncertainty in $B^+ \to \bar{D}^0\pi^+$. As was discussed in its implementation in the $B^+ \to \mu^+\nu_\mu$ sample, future efforts are needed to better account for correlations between changing selection boundaries, which could also reduce the systematic assigned by this feature. For this reason, similar to the signal analysis, the uncertainties derived due to selections in the control mode will not be considered alongside the standard systematic uncertainties, propagated separately for qualitative consideration.

A.3. Propagation of uncertainty in intermediate observables

The final quantities and $B^+ \to \bar{D}^0\pi^+$ yields used to evaluate the data/MC efficiency ratio $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$ are presented in Table A.6. This table is extended from Table 5.15 to include selection-based uncertainty. Thus, the uncertainties are separated into three groups: statistical uncertainty, detector-related systematic uncertainties and the selection systematic uncertainty. The final component is kept separate from the other systematic uncertainties given its need for future efforts to better evaluate the selection uncertainty.

The impact of each systematic uncertainty contribution propagated and expressed as a percentage of $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$ is presented in Table A.7. This is extended from Table 5.16. Again, it is seen that the tagging and selection criteria dominates in its current method of estimation, with the next largest being the contribution from the hadron-ID correction.

Table A.5. Extended summary of uncertainties in yields from $B^+ \to \bar{D}^0\pi^+$ control mode reconstruction. Statistical uncertainty is denoted as ΔN_{stat}. Systematic uncertainty due to hadron-ID is denoted as ΔN_{HID}, due to uncertainty to track momentum scaling and magnetic field modelling denoted as $\Delta N_{\text{mom.scale}}$, due to mid-to-high momentum tracking efficiency as $\Delta N_{\text{mom.eff.}}$ and due to implemented selection criteria as $\Delta N_{\text{sel.}}$.

	N	ΔN_{stat}	$\%\Delta N_{\text{stat}}$	ΔN_{HID}	$\%\Delta N_{\text{HID}}$
$N^{\text{data}}_{\text{SR,all}}$	156.00	12.49	8.01	-	-
$N^{\text{MC}}_{\text{SR,bkg}}$	88.44	9.40	10.63	2.06	2.33
$N^{\text{data}}_{\text{SB,all}}$	67.00	8.19	12.22	-	-
$N^{\text{MC}}_{\text{SB,all}}$	62.55	7.91	12.64	6.77	10.82
$N^{\text{data}}_{\text{SR,bkg}}*$	94.73	19.46	20.55	10.48	11.07
$N^{\text{MC}}_{\text{SR(gen),sig}}$	466.02	21.59	4.63	2.06	0.44
$\epsilon^{\text{MC}}*$	3.18×10^{-5}	0.15×10^{-5}	4.63	0.01×10^{-5}	0.44

	$\Delta N_{\text{mom.scale}}$	$\%\Delta N_{\text{mom.scale}}$	$\Delta N_{\text{mom.eff.}}$	$\%\Delta N_{\text{mom.eff.}}$	$\Delta N_{\text{sel.}}$	$\%\Delta N_{\text{sel.}}$
$N^{\text{data}}_{\text{SR,all}}$	1	0.64	-	-	39.63	25.40
$N^{\text{MC}}_{\text{SR,bkg}}$	-	-	0.21	0.24	11.50	19.37
$N^{\text{data}}_{\text{SB,all}}$	0.5	0.75	-	-	23.92	35.70
$N^{\text{MC}}_{\text{SB,all}}$	-	-	0.15	0.24	19.69	31.48
$N^{\text{data}}_{\text{SR,bkg}}*$	0.71	0.75	0.32	0.34	52.87	55.81
$N^{\text{MC}}_{\text{SR(gen),sig}}$	-	-	1.12	0.24	90.26	19.37
$\epsilon^{\text{MC}}*$	-	-	0.01×10^{-5}	0.24	0.62×10^{-5}	19.37

Table A.6. Extended final measurements for $\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$. The selection-based uncertainties are kept separate as they require a more sophisticated method of estimation.

Measurement	$N \pm \Delta N_{\text{stat}} \pm \Delta N_{\text{sys}}$	ΔN_{sel}
$N^{\text{data}}_{\text{SR,all}}$	$156.00\pm 12.49 \pm 1.00$	39.63
$N^{\text{data}}_{\text{SR,bkg}}*$	$94.73\pm 19.46 \pm 10.51$	52.87
$\epsilon^{\text{MC}}*$	$(3.18 \pm 0.15 \pm 0.02)\times10^{-5}$	0.62×10^{-5}
$\epsilon^{\text{data}}*$	$(3.34 \pm 1.26 \pm 0.58)\times10^{-5}$	3.60×10^{-5}
$\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}*$	$1.05 \pm 0.40 \pm 0.18$	1.15

Table A.7. Extended summary of systematic uncertainties and sources in evaluation of the data-MC efficiency ratio derived in $B^+ \to \bar{D}^0\pi^+$ processes. Tagging and selection criteria uncertainty is considered preliminary given the need for a more sophisticated estimation procedure.

Source	$\frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$ [%]
`hadronID` correction	17.11
Track momentum scale	2.00
Tracking efficiency	0.57
$\mathcal{B}(B^+ \to \bar{D}^0\pi^+)$	2.17
f^{+-}	1.17
$N_{B\bar{B}}$	1.45

Table A.8. Extended summary of statistical and systematic uncertainties and sources in evaluation of the cumulative signal efficiency ($\epsilon_{\text{sig+tag}}$) and the number of signal region events in data estimated to be from background events ($N^{\text{data}}_{\text{SR,bkg}}$). Selection based uncertainties are quoted for completeness, though they require evaluation through a more sophisticated procedure.

Source	$\epsilon_{\text{sig+tag}}$ [%]	$N^{\text{data}}_{\text{SR,bkg}}$ [%]
Statistical	4.10	38.98
Particle-ID correction (`muonID`)	5.02	4.88
Track momentum scale	-	1.83
Tracking efficiency	0.24	0.34
Control mode efficiency correction (sys.)	17.47	-
Control mode efficiency correction (stat.)	38.02	-
Tagging and selection criteria	14.26	46.22
Control mode efficiency correction (sel.)	109.55	-

In Eq. (5.12), a data-sensitive $B^+ \to \mu^+\nu_\mu$ signal efficiency was obtained, by applying the efficiency correction factor derived from the $B^+ \to \bar{D}^0\pi^+$ control mode. This is quoted to be $(7.83 \pm 2.98(\text{stat}) \pm 1.37(\text{sys})) \times 10^{-3}$. An additional selection-based uncertainty of 8.58×10^{-3} could be considered alongside the corrected signal efficiency. This again makes the selection-based uncertainty the dominant uncertainty, where its magnitude is larger than even the central values.

Table A.9. Extended final measurements for $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$. Selection based uncertainties are quoted for completeness, though they require evaluation through a more sophisticated procedure.

Measurement	$N \pm \Delta N_{\rm stat} \pm \Delta N_{\rm sys}$	$\Delta N_{sel.}$	$\Delta N_{contr.sel.}$
$\epsilon_{\rm sig+tag}$	$(7.83 \pm 0.03 \pm 3.30) \times 10^{-3}$	1.12×10^{-3}	8.58×10^{-3}
$N^{\rm data}_{\rm SR,bkg}$	9.01±3.51±0.47	4.17	0.00

Table A.10. Extended estimate of data yield in signal region attributed to $B^+ \to \mu^+\nu_\mu$ events. Selection based uncertainties are quoted for completeness, though they require evaluation through a more sophisticated procedure.

Measurement	$N \pm \Delta N_{\rm stat} \pm \Delta N_{\rm sys}$	ΔN_{sel}	$\Delta N_{contr.,sel}$
$N^{\rm data}_{\rm SR,sig(est)} = N^{\rm MC}_{\rm SR,sig} \times \frac{N^{\rm data}_{\rm SB,all}}{N^{\rm MC}_{\rm SB,all}}$	1.47 ± 1.31 ± 0.08	0.56	0.00
$N^{\rm data}_{\rm SR,sig(est)} = 2f^{+-} N^{\rm data}_{362{\rm fb}^{-1},B\bar{B}} \times \epsilon^{\rm data}_{\rm sig+tag} \times \mathcal{B}(B^+ \to \mu^+\nu_\mu)_{\rm SM}$	1.32± 0.01 ±0.56	0.19	1.45

In Table A.8, the systematic uncertainty is expressed as a percentage of the observable the factor effects, presented according source and observable. This is an extension of Table 5.17 to include the selection-based uncertainties in both $B^+ \to \mu^+\nu_\mu$ and the $B^+ \to \bar{D}^0\pi^+$ control mode. While originally the control mode efficiency corrections were the highest proportional uncertainty, this is exceeded by the selection criteria-based uncertainties. This becomes especially large in the effect of varying the boundaries for the efficiency correction from the $B^+ \to \bar{D}^0\pi^+$ control mode.

Table A.10 presents estimates of the $B^+ \to \mu^+\nu_\mu$ yield in data within the signal region, evaluated in two methods with both $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \bar{D}^0\pi^+$ selection based uncertainties. This Table is extended from Table 5.19. It is notable now that the selection systematic for $\epsilon_{sig+tag}$ from $B^+ \to \mu^+\nu_\mu$ does not exceed the total systematic uncertainty of other factors added in quadrature. However, the selection systematic from the sideband does, while also exceeding the central value of the efficiency. As the control mode is not used in the estimate of the number of events in data assumed to originate from background events ($N^{data}_{\rm SR,bkg}$), no $B^+ \to \bar{D}^0\pi^+$-based uncertainty is propagated to this quantity.

Table A.9 presents the final values, with sources combined in quadrature to obtain total uncertainties. The selection-based uncertainties from the $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \bar{D}^0\pi^+$ samples have again been separated from the others, so as to make a comparison of magnitude. The statistical and systematic uncertainties in the final parameters were used in Section 5.3 to determine an upper limit on the $B^+ \to \mu^+\nu_\mu$ branching fraction. Given the previous discussion on how the method of estimating the systematic assigned to the selection boundaries could be improved, they were omitted from the Feldman-Cousins process. Both the signal and control mode based selection uncertainties are instead propagated to the final branching fraction calculation for completeness, where they can be qualitatively interpreted separately from the statistical and systematic counterparts.

A.4. Impact on final results

Table A.11 contains the branching fractions evaluated for potential $N^{\text{data}}_{\text{SR,bkg}} \in$[2,16], extended from Table 5.22 to contain both the propagated signal and control mode selection based uncertainties. The statistical uncertainty is the largest source for $N^{\text{data}}_{\text{SR,all}} <$ 4 and $N^{\text{data}}_{\text{SR,all}} > 14$, while the selection based uncertainties are the largest otherwise. Thus from Table A.11, the predicted branching fraction corresponding to the estimate of $N^{\text{data}}_{\text{SR,sig}}$=1.47, with selection based uncertainties considered, is the following.

$$\begin{aligned}\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (&6.4 \pm 15.5(\text{stat.}) \ \pm 3.08(\text{sys.}) \ ) \times 10^{-7} \\ &\pm 13.4 \times 10^{-7}(\text{sel.}) \pm 6.98 \times 10^{-7}(\text{contr.sel.})\end{aligned} \tag{A.1}$$

This $B^+ \to \mu^+\nu_\mu$ branching fraction is dominated by statistical uncertainty, followed by the signal selection-based uncertainty.

Using this branching fraction to place a constraint on V_{ub}, as was done in Section 5.5, the following expression (inclusive of the selection-based uncertainties) is obtained.

$$\begin{aligned}|V_{ub}| = (&4.72 \pm 5.74(\text{stat.}) \ \pm 1.13(\text{sys.})) \ \times 10^{-3} \\ &\pm 4.96(\text{sel.}) \ \pm 2.58(\text{contr.sel})\end{aligned} \tag{A.2}$$

Table A.11. Extended table of experimentally derived $\mathcal{B}(B^+ \to \mu^+\nu_\mu)$, for varying $N^{\text{data}}_{\text{SR,all}}$. Evaluated with $N^{\text{data}}_{\text{SR,bkg}}$=9.01 and $N^{\text{data}}_{\text{SR,sig}}$=$N^{\text{data}}_{\text{SR,all}}$ - $N^{\text{data}}_{\text{SR,bkg}}$. $\mathcal{B}(B^+ \to \mu^+\nu_\mu)_{\text{SM}}$=$(4.24 \pm 0.24) \times 10^{-7}$, as per its evaluation in Eq. (1.7).

$N^{\text{data}}_{\text{SR,all}}$	$N^{\text{data}}_{\text{SR,sig}}$	$\mathcal{B}(B^+ \to \mu^+\nu_\mu)$	$\Delta\mathcal{B}_{\text{stat}}$	$\Delta\mathcal{B}_{\text{sys}}$	$\Delta\mathcal{B}_{\text{sel}}$	$\Delta\mathcal{B}_{\text{contr.sel.}}$
2	-7.01	-2.25×10^{-6}	1.22×10^{-6}	9.59×10^{-7}	1.38×10^{-6}	2.47×10^{-6}
3	-6.01	-1.93×10^{-6}	1.26×10^{-6}	8.26×10^{-7}	1.37×10^{-6}	2.11×10^{-6}
4	-5.01	-1.61×10^{-6}	1.30×10^{-6}	6.94×10^{-7}	1.36×10^{-6}	1.76×10^{-6}
5	-4.01	-1.29×10^{-6}	1.34×10^{-6}	5.63×10^{-7}	1.35×10^{-6}	1.41×10^{-6}
6	-3.01	-9.67×10^{-7}	1.37×10^{-6}	4.34×10^{-7}	1.34×10^{-6}	1.06×10^{-6}
7	-2.01	-6.46×10^{-7}	1.41×10^{-6}	3.11×10^{-7}	1.34×10^{-6}	7.08×10^{-7}
8	-1.01	-3.25×10^{-7}	1.45×10^{-6}	2.04×10^{-7}	1.34×10^{-6}	3.56×10^{-7}
9	-0.01	-4.32×10^{-9}	1.48×10^{-6}	1.51×10^{-7}	1.34×10^{-6}	4.73×10^{-9}
10	0.99	3.17×10^{-7}	1.52×10^{-6}	2.01×10^{-7}	1.34×10^{-6}	3.47×10^{-7}
11	1.99	6.37×10^{-7}	1.55×10^{-6}	3.08×10^{-7}	1.34×10^{-6}	6.98×10^{-7}
12	2.99	9.58×10^{-7}	1.58×10^{-6}	4.31×10^{-7}	1.34×10^{-6}	1.05×10^{-6}
13	3.99	1.28×10^{-6}	1.62×10^{-6}	5.59×10^{-7}	1.35×10^{-6}	1.40×10^{-6}
14	4.99	1.60×10^{-6}	1.65×10^{-6}	6.90×10^{-7}	1.36×10^{-6}	1.75×10^{-6}
15	5.99	1.92×10^{-6}	1.68×10^{-6}	8.23×10^{-7}	1.36×10^{-6}	2.10×10^{-6}
16	6.99	2.24×10^{-6}	1.71×10^{-6}	9.56×10^{-7}	1.37×10^{-6}	2.24×10^{-6}

Bibliography

[1] E. Tiesinga, P. J. Mohr, D. B. Newell and B. N. Taylor, *CODATA recommended values of the fundamental physical constants: 2018*, Reviews of Modern Physics **93** (2021) 025010. 1, 1.2

[2] Wikimedia Commons, *Standard Model of Elementary Particles*, 2024. 1.2

[3] S. Cecchini and M. Spurio, *Atmospheric muons: experimental aspects*, Geoscientific Instrumentation, Methods and Data Systems **1** (2012) 185–196. 1.1.1

[4] Particle Data Group collaboration, S. Navas et al., *Review of Particle Physics*, Physical Review D **110** (2024) 030001. 1.1, 1.1.2, 1.2, 1.10, 2.2.2, 2.2.4, 4.1.1, 5.2.2, 5.2.2

[5] J.-M. Richard, *Exotic hadrons: Review and perspectives*, Few-Body Systems **57** (Oct., 2016) 11851212. 1.1.1

[6] Y. Aoki, T. Blum, G. Colangelo, S. Collins, M. D. Morte, P. Dimopoulos et al., *FLAG Review 2021*, The European Physical Journal C **82** (Oct., 2022) . 1.1.1, 1.2, 1.2.5, 1.2.6

[7] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Physical Review **105** (Feb, 1957) 1413–1415. 1.1.1

[8] A. Bettini, *Introduction to Elementary Particle Physics*. Cambridge University Press, 2014. 1.1.2, 1.2.2

[9] N. Cabibbo, *Unitary symmetry and leptonic decays*, Physical Review Letters **10** (Jun, 1963) 531–533. 1.1.2

[10] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (02, 1973) 652–657, [https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-

652.pdf]. 1.1.2, 2.1.1

[11] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary particles*, Progress of Theoretical Physics **28** (11, 1962) 870–880, [https://academic.oup.com/ptp/article-pdf/28/5/870/5258750/28-5-870.pdf]. 1.1.2

[12] Y. Amhis, S. Banerjee, E. Ben-Haim, E. Bertholet, F. Bernlochner, M. Bona et al., *Averages of b-hadron, c-hadron and tau-lepton properties as of 2021*, Physical Review D **107** (Mar., 2023) . 1.2

[13] L. Wan, K. Abe, C. Bronner, Y. Hayato, M. Ikeda, K. Iyogi et al., *Measurement of the neutrino-oxygen neutral-current quasielastic cross section using atmospheric neutrinos at Super-Kamiokande*, Physical Review D **99** (Feb., 2019) . 1.2.3

[14] Hyper-Kamiokande Proto- collaboration, K. Abe, K. Abe, H. Aihara, A. Aimi, R. Akutsu, C. Andreopoulos et al., *Hyper-Kamiokande Design Report*, arXiv:1805.04163. 1.2.3

[15] ATLAS collaboration, G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (Sept., 2012) 129. 1.3.1

[16] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (Sept., 2012) 3061. 1.3.1

[17] T. D. Lee, *A Theory of Spontaneous T Violation*, Physical Review D **8** (1973) 1226–1239. 1.3.1

[18] W.-S. Hou, *Enhanced charged Higgs boson effects in $B^- \to \tau\overline{\nu}$, $\mu\overline{\nu}$ and $b \to \tau\overline{\nu} + X$*, Physical Review D **48** (Sep, 1993) 2342–2344. 1.3.1

[19] J. C. Pati and A. Salam, *Lepton number as the fourth "color"*, Physical Review D **10** (Jul, 1974) 275–289. 1.3.2

[20] G. Valencia and S. Willenbrock, *Quark-lepton unification and rare meson decays*, Physical Review D **50** (Dec., 1994) 68436848. 1.3.2

[21] J. Heeck and D. Teresi, *Pati-Salam explanations of the B-meson anomalies*, Journal of High Energy Physics **2018** (Dec., 2018) . 1.3.2

[22] J. Wess and B. Zumino, *Supergauge transformations in four dimensions*, *Nuclear Physics B* **70** (1974) 39–50. 1.3.3

[23] S. Baek and Y. G. Kim, *Constraints on R-parity violating couplings from $B^{\pm} \to l^{\pm}\nu$ decays*, *Physical Review D* **60** (Sep, 1999) 077701. 1.3.3

[24] Q.-Y. Hu, Y.-D. Yang and M.-D. Zheng, *Revisiting the B-physics anomalies in R-parity violating MSSM*, *The European Physical Journal C* **80** (May, 2020) . 1.3.3

[25] T. Asaka, S. Blanchet and M. Shaposhnikov, *The νMSM, dark matter and neutrino masses*, *Physics Letters B* **631** (2005) 151–156. 1.3.4

[26] T. Asaka and M. Shaposhnikov, *The νMSM, dark matter and baryon asymmetry of the universe*, *Physics Letters B* **620** (2005) 17–26. 1.3.4

[27] Belle collaboration, M. T. Prim et al., *Search for $B^+ \to \mu^+ \nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging*, *Physical Review D* **101** (2020) 032007, [arXiv:1911.03186]. 1.3.4, 1.4.3, 6.2

[28] Belle collaboration, C. S. Park et al., *Search for a massive invisible particle X^0 in $B^+ \to e^+ X^0$ and $B^+ \to \mu^+ X^0$ decays*, *Physical Review D* **94** (2016) 012003, [arXiv:1605.04430]. 1.3.4

[29] D. Andrews, P. Avery, K. Berkelman, R. Cabenda, D. Cassel, J. DeWire et al., *The CLEO detector*, *Nuclear Instruments and Methods in Physics Research* **211** (1983) 47–71. 1.4

[30] D. Atwood, *Technical Design Report for the Babar Detector*, tech. rep., 3, 1995. https://www.osti.gov/biblio/1454130. 1.4, 2.1.1

[31] A. Abashian, K. Gotow, N. Morgan, L. Piilonen, S. Schrenk, K. Abe et al., *The Belle detector*, *Nuclear Instruments and Methods in Physics Research Section A* **479** (2002) 117–232. 1.4

[32] CLEO Collaboration collaboration, M. Artuso, M. Gao, M. Goldberg, D. He, N. Horwitz, G. C. Moneti et al., *Search for $B \to \ell\overline{\nu}_\ell$*, *Physical Review Letters* **75** (Jul, 1995) 785–789. 1.4.1

[33] CLEO collaboration, C. Bebek et al., *Evidence for New Flavor Production at the $\Upsilon(4S)$*, *Physical Review Letters* **46** (1981) 84. 1.4.1, 2.1.1

[34] D. Buskulic, D. Casper, I. De Bonis, D. Decamp, P. Ghez, C. Goy et al., *Measurement of the $b \to \tau^- \bar{\nu}_\tau$ branching ratio and an upper limit on $B^- \to \tau^- \bar{\nu}_\tau$*, *Physics Letters B* **343** (1995) 444–452. 1.4.1

[35] BaBar collaboration, B. Aubert et al., *Search for the rare leptonic decay $B^+ \to \mu^+ \nu_\mu$*, *Physical Review Letters* **92** (2004) 221803, [hep-ex/0401002]. 1.4.1

[36] BaBar collaboration, B. Aubert, D. Boutigny, I. De Bonis, J.-M. Gaillard, A. Jeremie, Y. Karyotakis et al., *Measurement of CP-Violating Asymmetries in B^0 Decays to CP Eigenstates*, *Physical Review Letters* **86** (Mar, 2001) 2515–2522. 1.4.1, 2.1.1

[37] Belle collaboration, A. Abashian, K. Abe, K. Abe, I. Adachi, B. S. Ahn, H. Aihara et al., *Measurement of the CP Violation Parameter $\sin 2\varphi_1$ in B_d^0 Meson Decays*, *Physical Review Letters* **86** (Mar, 2001) 2509–2514. 1.4.1, 2.1.1

[38] ARGUS collaboration, H. Albrecht, T. Hamacher, R. Hofmann, T. Kirchhoff, R. Mankel, A. Nau et al., *Measurement of the polarization in the decay $B \to J/\psi K^*$*, *Physics Letters B* **340** (1994) 217–220. 1.4.1

[39] Belle collaboration, N. Satoyama et al., *A Search for the rare leptonic decays B+ -> mu+ nu(mu) and B+ -> e+ nu*, *Physical Letters B* **647** (2007) 67–73, [hep-ex/0611045]. 1.4.1, 1.11

[40] Belle Collaboration collaboration, K. Ikado, K. Abe, K. Abe, I. Adachi, H. Aihara, K. Akai et al., *Evidence of the purely leptonic decay $B^- \to \tau^- \overline{\nu}_\tau$*, *Physical Review Letters* **97** (Dec, 2006) 251802. 1.4.1

[41] BaBar collaboration, B. Aubert et al., *Searches for the decays $B^0 \to \ell^\pm \tau^\mp$ and $B^+ \to \ell^+ \nu$ (l=e, μ) using hadronic tag reconstruction*, *Physical Review D* **77** (2008) 091104, [arXiv:0801.0697]. 1.4.2, 1.12

[42] BaBar collaboration, B. Aubert et al., *Search for the Rare Leptonic Decays B+ —> l+ nu(l) (l=e,mu)*, *Physical Review D* **79** (2009) 091101, [arXiv:0903.1220]. 1.4.2

[43] BaBar collaboration, B. Aubert et al., *A Search for $B^+ \to \ell^+ \nu_\ell$ Recoiling Against $B^- \to D^0 \ell^- \bar{\nu} X$*, *Physical Review D* **81** (2010) 051101, [arXiv:0912.2453]. 1.4.2, 1.13, 5.7, 6.2

[44] Belle collaboration, Y. Yook et al., *Search for $B^+ \to e^+ \nu$ and $B^+ \to \mu^+ \nu$ decays*

using hadronic tagging, *Physical Review D* **91** (2015) 052016, [arXiv:1406.6356]. 1.4.3, 1.14

[45] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander et al., *A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories*, *Nuclear Instruments and Methods in Physics Research Section A* **654** (2011) 432–440. 1.4.3, 3.2

[46] Belle collaboration, A. Sibidanov et al., *Search for $B^- \to \mu^- \bar{\nu}_\mu$ Decays at the Belle Experiment*, *Physical Review Letters* **121** (2018) 031801, [arXiv:1712.04123]. 1.4.3, 1.15

[47] T. Keck et al., *The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment*, *Computing and Software for Big Science* **3** (2019) 6, [arXiv:1807.08680]. 1.5, 3.2

[48] L. Aggarwal, S. Banerjee, S. Bansal, F. Bernlochner, M. Bertemes, V. Bhardwaj et al., *Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond*, arXiv:2207.06307. 2.1.1

[49] A. J. Bevan, B. Golob, T. Mannel, S. Prell, B. D. Yabsley, H. Aihara et al., *The Physics of the B Factories*, *The European Physical Journal C* **74** (Nov., 2014) . 2.1.1

[50] I. Adachi, K. Adamczyk, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov et al., *First Measurement of $R(X_{\tau/\ell})$ as an Inclusive Test of $b \to c\tau\nu$ Anomaly*, *Physical Review Letters* **132** (May, 2024) . 2.1.2, 3.2

[51] I. Adachi, K. Adamczyk, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov et al., *Evidence for $B^+ \to K^+ \nu\bar{\nu}$ decays*, *Physical Review D* **109** (June, 2024) . 2.1.2, 3.2

[52] A. Filimonova, R. Schäfer and S. Westhoff, *Probing dark sectors with long-lived particles at Belle II*, *Physical Review D* **101** (May, 2020) 095006. 2.1.2

[53] F. Abudinén, N. Akopov, A. Aloisio, V. Babu, S. Banerjee, M. Bauer et al., *B-flavor tagging at Belle II*, *The European Physical Journal C* **82** (Apr., 2022) . 2.1.2

[54] Belle II collaboration, I. Adachi, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov, A. Aloisio et al., *A new graph-neural-network flavor tagger for Belle II and measurement of* $\sin 2\phi_1$ *in $B^0 \to J/\psi K^0_S$ decays*, arXiv:2402.17260. 2.1.2

[55] Belle II collaboration, I. Adachi, L. Aggarwal, H. Aihara, N. Akopov, A. Aloisio, N. A. Ky et al., *Measurement of the branching fraction of the decay* $B^- \to D^0\rho(770)^-$ *at Belle II*, arXiv:2404.10874. 2.1.2

[56] Belle II collaboration, I. Adachi, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov, A. Aloisio et al., *Measurement of branching fractions and direct CP asymmetries for* $B \to K\pi$ *and* $B \to \pi\pi$ *decays at Belle II*, *Physical Review D* **109** (Jan., 2024) . 2.1.2

[57] Belle and Belle II collaboration, I. Adachi, L. Aggarwal, H. Aihara, N. Akopov, A. Aloisio, N. Althubiti et al., *Measurements of the branching fractions of* $\Xi_c^0 \to \Xi^0\pi^0$, $\Xi_c^0 \to \Xi^0\eta$, *and* $\Xi_c^0 \to \Xi^0\eta'$ *and asymmetry parameter of* $\Xi_c^0 \to \Xi^0\pi^0$, arXiv:2406.04642. 2.1.2

[58] Belle II collaboration, I. Adachi, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov, A. Aloisio et al., *Search for the* $e^+e^- \to \eta_b(1s)\omega$ *and* $e^+e^- \to \chi_{b0}(1p)\omega$ *processes at* $\sqrt{s} = 10.745\,\mathrm{GeV}$, arXiv:2312.13043. 2.1.2

[59] F. Abudinén, I. Adachi, P. Ahlburg, H. Aihara, N. Akopov, A. Aloisio et al., *Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment*, *Chinese Physics C* **44** (Feb., 2020) 021001. 2.1.2, 2.2.3

[60] I. Adachi, P. Ahlburg, H. Aihara, N. Akopov, A. Aloisio, N. Anh Ky et al., *Search for an Invisibly Decaying* Z' *Boson at Belle II in* $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ *Plus Missing Energy Final States*, *Physical Review Letters* **124** (Apr., 2020) . 2.1.2

[61] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*, *Nuclear Instruments and Methods in Physics Research Section A* **907** (Nov., 2018) 188199. 2.3, 2.2.1

[62] H. Aihara, A. Aloisio, D. P. Auguste, M. Aversano, M. Babeluk, S. Bahinipati et al., *The Belle II Detector Upgrades Framework Conceptual Design Report*, arXiv:2406.19421. 2.2.1, 2.2.1

[63] M. H. Ahn, E. Aliu, S. Andringa, S. Aoki, Y. Aoyama, J. Argyriades et al., *Measurement of neutrino oscillation by the K2K experiment*, *Physical Review D* **74** (Oct., 2006) . 2.2.1

[64] K. F. Y. Funakoshi, *SuperKEKB Technical Design Report*, tech. rep., SuperKEKB, 2020. https://www-linac.kek.jp/linac-com/report/skb-tdr/11_Linac-201023.pdf. 2.2.1

[65] I. Abe, N. Akasaka, M. Akemoto, S. Anami, A. Enomoto, J. Flanagan et al., *The KEKB injector linac*, *Nuclear Instruments and Methods in Physics Research Section A* **499** (2003) 167–190. 2.2.1

[66] T. Abe, K. Akai, N. Akasaka, M. Akemoto, A. Akiyama and M. Arinaga, *Achievements of KEKB*, *Progress of Theoretical and Experimental Physics* **2013** (03, 2013) 03A001, [https://academic.oup.com/ptep/article-pdf/2013/3/03A001/4440618/pts102.pdf]. 2.2.1

[67] E. Kou, P. Urquijo, W. Altmannshofer, F. Beaujean, G. Bell, M. Beneke et al., *The Belle II Physics Book*, *Progress of Theoretical and Experimental Physics* **2019** (Dec., 2019) . 2.2.1, 2.3.2, 2.6

[68] G. C. Fox and S. Wolfram, *Observables for the analysis of event shapes in e^+e^- annihilation and other processes*, *Physical Review Letters* **41** (Dec, 1978) 1581–1585. 2.2.3, 4.1.5

[69] C.-Z. Yuan and S. L. Olsen, *The BESIII physics programme*, *Nature Reviews Physics* **1** (July, 2019) 480494. 2.2.3

[70] S. Dobbs, Z. Metreveli, K. K. Seth, A. Tomaradze and T. Xiao, *Observation of $\eta_b(2S)$ in $\Upsilon(2S) \to \gamma\eta_b(2S)$, $\eta_b(2S) \to$ hadrons, and Confirmation of $\eta_b(1S)$*, *Physical Review Letters* **109** (2012) 082001, [arXiv:1204.4205]. 2.2.4

[71] Y. Sato, H. Yamamoto, H. Aihara, D. M. Asner, V. Aulchenko, T. Aushev et al., *Measurement of the CP-Violation Parameter sin2ϕ_1 with a New Tagging Method at the $\Upsilon(5S)$ Resonance*, *Physical Review Letters* **108** (Apr., 2012) . 2.2.4

[72] C.-C. Peng, P. Chang, I. Adachi, H. Aihara, T. Aushev, T. Aziz et al., *Search for $B_s^0 \to hh$ decays at the $\Upsilon(5S)$ resonance*, *Physical Review D* **82** (Oct., 2010) . 2.2.4

[73] K.-F. Chen, W.-S. Hou, M. Shapkin, A. Sokolov, I. Adachi, H. Aihara et al., *Observation of Anomalous $\Upsilon(1S)\pi^+\pi^-$ and $\Upsilon(2S)\pi^+\pi^-$ Production near the $\Upsilon(5S)$ Resonance*, *Physical Review Letters* **100** (Mar., 2008) . 2.2.4

[74] Belle II collaboration, I. Adachi, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov, A. Aloisio et al., *Observation of $e^+e^- \to \omega\chi_{bJ}(1P)$ and Search for $X_b \to \omega\Upsilon(1S)$ at $\sqrt{s}$ near 10.75 GeV*, *Physical Review Letters* **130** (Mar, 2023) 091902. 2.2.4

[75] D. Matvienko, *The Belle II experiment: status and physics program*, *EPJ Web of*

Conferences **191** (01, 2018) 02010. 2.7

[76] Belle II Neutrals Performance Group, *Photon and neutral pion performance in early Belle II data*, Tech. Rep. BELLE2-NOTE-PH-2022-017, Belle II, Apr, 2022. `https://docs.belle2.org/record/2981`. 2.3.2, 2.4.1

[77] Belle II Lepton ID Group, *Muon and electron identification performance with* 189 fb^{-1} *of Belle II data*, Tech. Rep. BELLE2-NOTE-TE-2021-011, Belle II, Apr, 2021. `https://docs.belle2.org/record/2340`. 2.3.2, 5.2.1

[78] J. Bennett, *Calibration data flow summary*, Tech. Rep. BELLE2-NOTE-PH-2016-003, Belle II, Sep, 2016. `https://docs.belle2.org/record/415`. 2.4.1

[79] T. Hara, *Computing at the Belle II experiment*, *Journal of Physics: Conference Series* **664** (dec, 2015) 012002. 2.4.1

[80] D. J. Lange, *The EvtGen particle decay simulation package*, *Nuclear Instruments and Methods in Physics Research Section A* **462** (2001) 152–155. 2.4.2

[81] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Computer Physics Communications* **191** (June, 2015) 159177. 2.4.2

[82] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce et al., *Geant4 a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A* **506** (2003) 250–303. 2.4.2

[83] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth and N. Braun, *The Belle II Core Software*, *Computing and Software for Big Science* **3** (2019) . 2.4.3

[84] R. Brun, F. Rademakers, P. Canal, A. Naumann, O. Couet, L. Moneta et al., “root-project/root: v6.18/02.” Zenodo, June, 2020. `https://doi.org/10.5281/zenodo.3895860`. 2.4.3

[85] J. Pivarski, P. Das, C. Burr, D. Smirnov, M. Feickert, T. Gal et al., “scikit-hep/uproot: 3.12.0.” Zenodo, July, 2020. `https://doi.org/10.5281/zenodo.3952728`. 2.4.3

[86] J. Pivarski, C. Escott, N. Smith, M. Hedges, M. Proffitt and C. Escott, “scikit-hep/awkward-array: 0.13.0.” Zenodo, July, 2020. `https://doi.org/10.5281/zenodo.3952674`. 2.4.3

[87] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau et al., *Array programming with NumPy*, *Nature* **585** (Sept., 2020) 357–362. 2.4.3

[88] J. D. Hunter, *Matplotlib: A 2d graphics environment*, *Computing in Science & Engineering* **9** (2007) 90–95. 2.4.3

[89] Y. Sato, S. Cunliffe, F. Meier and A. Zupanc, *Monte Carlo matching in the Belle II software*, *EPJ Web of Conferences* **251** (2021) 03021. 2.4.3

[90] T. Keck, *Machine learning algorithms for the Belle II experiment and their validation on Belle data.* PhD thesis, Karlsruher Institut für Technologie (KIT), 2017. `https://doi.org/10.5445/IR/1000078149`. 3.1

[91] I. Adachi, K. Adamczyk, L. Aggarwal, H. Aihara, N. Akopov, A. Aloisio et al., *Tests of light-lepton universality in angular asymmetries $B^0 \to D^{*-}\ell\nu$ decays*, *Physical Review Letters* **131** (Oct., 2023) . 3.2

[92] Belle II collaboration, I. Adachi, K. Adamczyk, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov et al., *A test of lepton flavor universality with a measurement of $R(D^*)$ using :hadronic b tagging at the Belle II experiment*, `arXiv:2401.02840`. 3.2

[93] L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov, A. Aloisio, N. Anh Ky et al., *A test of light-lepton universality in the rates of inclusive semileptonic B-mesons decays at Belle ii*, *Physical Review Letters* **131** (Aug., 2023) . 3.2

[94] Belle II Collaboration collaboration, M. Welsch, *Measurement of Lepton Mass Squared Moments in $B \to X_c\ell\bar{\nu}_\ell$ Decays with the Belle II Experiment*, `arXiv:2205.06372`. 3.2

[95] W. Sutcliffe, *Early results from Full Event Interpretation at Belle II*, in *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)*, vol. 390, p. 787, 2021. DOI. 3.2

[96] Belle II collaboration, "Belle II Analysis Software Framework (`basf2`)." Nuclear Instruments and Methods in Physics Research Section A. `https://doi.org/10.5281/zenodo.5574115`. 3.4

[97] Belle collaboration, M. Hohmann, P. Urquijo, I. Adachi, H. Aihara, D. M. Asner, T. Aushev et al., *Measurement of the Ratio of Partial Branching Fractions of*

Inclusive $\bar{B} \to X_u \ell \bar{\nu}$ to $\bar{B} \to X_c \ell \bar{\nu}$ and the Ratio of their Spectra with Hadronic Tagging, arXiv:2311.00458. 3.4

[98] Belle II collaboration, M. Liu, T. Luo and K. Trabelsi, "Search for lepton flavor-violating decay modes $B^0 \to K_s^0 \tau^\pm \ell^\mp (\ell = \mu, e)$ with hadronic B-tagging at Belle and Belle II." Belle II Document Server, Jul, 2023. https://docs.belle2.org/record/3754. 3.4

[99] A. Tomaradze, S. Dobbs, T. Xiao and K. K. Seth, *Precision Measurement of the Mass of the D^{*0} Meson and the Binding Energy of the $X(3872)$ Meson as a $D^0 \overline{D^{*0}}$ Molecule*, Physical Review D **91** (2015) 011102, [arXiv:1501.01658]. 3.4.3

[100] Belle II Collaboration collaboration, F. Abudinén, I. Adachi, R. Adak, K. Adamczyk, P. Ahlburg, J. K. Ahn et al., *A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \to X \ell \nu$ decays*, arXiv:2008.06096. 3.4.5

[101] M. Gelb, *Search for the rare decay of $B^+ \to \ell \nu \gamma$ with improved hadronic tagging*, Physical Review D **98** (Dec., 2018) . 3.4.6

[102] N. Toutounji, *Reconstruction Methods for Semi-leptonic Decays of B-mesons with the Belle II Experiment*, Master's thesis, University of Sydney, 2018. https://ses.library.usyd.edu.au/handle/2123/19818. 3.4.6

[103] A. Huang and K. Varvell, *Calibration of the SL FE using exclusive $D^* \ell \nu$ decays*, Tech. Rep. BELLE2-NOTE-PH-2023-022, The University of Sydney, May, 2023. https://docs.belle2.org/record/3568. 3.4.6

[104] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nuclear Instruments and Methods in Physics Research Section A **555** (2005) 356–369, [physics/0402083]. 3.4.6

[105] J. Kahn, I. Tsaklidis, O. Taubert, L. Reuter, G. Dujany, T. Boeckh et al., *Learning tree structures from leaves for particle decay reconstruction*, Machine Learning: Science and Technology **3** (sep, 2022) 035012. 3.4.6

[106] J. Cerasoli, F. Bernlochner, T. Boeckh, G. Dujany, P. Goldenzweig, M. Gotz et al., *A Graph Neural Network for B decay reconstruction at Belle II*, 2022. https://indico.cern.ch/event/1106990/papers/4996235/files/12252-ACAT_2022_proceedings.pdf. 3.4.6

[107] CLEO collaboration, J. E. Bartelt et al., *Measurement of the $B \to D \ell \nu$ branching*

fractions and form-factor, Physical Review Letters **82** (1999) 3746, [hep-ex/9811042]. 3.5.2

[108] O. Behnke, K. Kroninger, G. Schott and T. Schorner-Sadenius, *Data Analysis In High Energy Physics: A Practical Guide to Statistical Methods.* Wiley-VCH, 2013. 4.2.1

[109] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Physical Review D **57** (Apr., 1998) 38733889. 4.2.1, 5.3

[110] G. Punzi, *Sensitivity of searches for new signals and its optimization*, *eConf* **C030908** (2003) MODT002, [physics/0308063]. 4.2.1

[111] Belle II collaboration, F. Abudinén, I. Adachi, K. Adamczyk, L. Aggarwal, P. Ahlburg, H. Ahmed et al., *Determination of $|V_{cb}|$ from $B \to D\ell\nu$ decays using 2019-2021 Belle II data*, arXiv:2210.13143. 4.2.2

[112] B. Aubert, R. Barate, M. Bona, D. Boutigny, F. Couderc, Y. Karyotakis et al., *Measurement of the $B \to \pi\ell\nu$ Branching Fraction and Determination of $|V_{ub}|$ with Tagged B Mesons*, Physical Review Letters **97** (Nov., 2006) . 4.2.2, 5.6

[113] T. Hokuue, K. Abe, K. Abe, I. Adachi, H. Aihara, Y. Asano et al., *Measurements of branching fractions and q^2 distributions for $B \to \pi\ell\nu$ and $B \to \rho\ell\nu$ decays with $B \to D^{(*)}\ell\nu$ decay tagging*, Physics Letters B **648** (May, 2007) 139–148. 4.2.2, 5.6

[114] Y. Amhis, S. Banerjee, E. Ben-Haim, F. U. Bernlochner, M. Bona, A. Bozek et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018: Heavy Flavor Averaging Group (HFLAV)*, The European Physical Journal C **81** (Mar., 2021) . 5.1, 5.1

[115] C. Cecchi, G. D. Nardo, E. Manoni and M. Merola, *B counting measurement of the Run 1 Belle II data sample*, Tech. Rep. BELLE2-NOTE-PH-2023-011, Mar, 2023. https://docs.belle2.org/record/3468. 5.1

[116] W. Sutcliffe, C. Lyu, H. Junkerkalefeld, M. Eliaschevitch and S. Duell, *PIDvar*, 2024. https://gitlab.desy.de/william.sutcliffe/pidvar. 5.2.1, 5.2.2

[117] Q.-D. Zhou, *Correction for tracking momentum bias based on invariant mass peak studies*, Tech. Rep. BELLE2-NOTE-PH-2020-030, Belle II, Jun, 2020. https://docs.belle2.org/record/1947. 5.2.1

[118] Q.-D. Zhou, T. Iijima, K. Kojima, K. Nakamura, K. Matsuoka and K. Hara,

"Tracking momentum scale for Proc13 and Prompt Moriond23." https://indico.belle2.org/event/8171/contributions/49600/attachments/19684/29190/Tracking_p_correction_QidongZhou_20221130.pdf, 2022. 5.2.1

[119] A. Glazov, P. Rados, A. Rostomyan, E. Paoloni and L. Zani, *Measurement of the tracking efficiency in Phase 3 data using tau-pair events.*, Tech. Rep. BELLE2-NOTE-PH-2020-006, Belle II, Feb, 2020. https://docs.belle2.org/record/1867. 5.2.1

[120] T. Allmendinger, B. Bhuyan, D. Brown, H. Choi, S. Christ, R. Covarelli et al., *Track finding efficiency in BaBar*, *Nuclear Instruments and Methods in Physics Research Section A* **704** (Mar., 2013) 4459. 5.2.1

[121] G. Inguglia, P. Rados, L. Zani, A. Glazov and A. Rostomyan, *Tracking Efficiency with Taus*, 2023. https://indico.belle2.org/event/8043/contributions/51113/attachments/20577/30471/tau_eff_f2f_31jan23.pdf. 5.2.1

[122] S. Sandilya and A. Schwartz, *Study of Kaon and Pion Identification Performances in Phase III data with D^*+ sample*, Tech. Rep. BELLE2-NOTE-PH-2019-048, Jul, 2019. https://docs.belle2.org/record/1558. 5.2.2

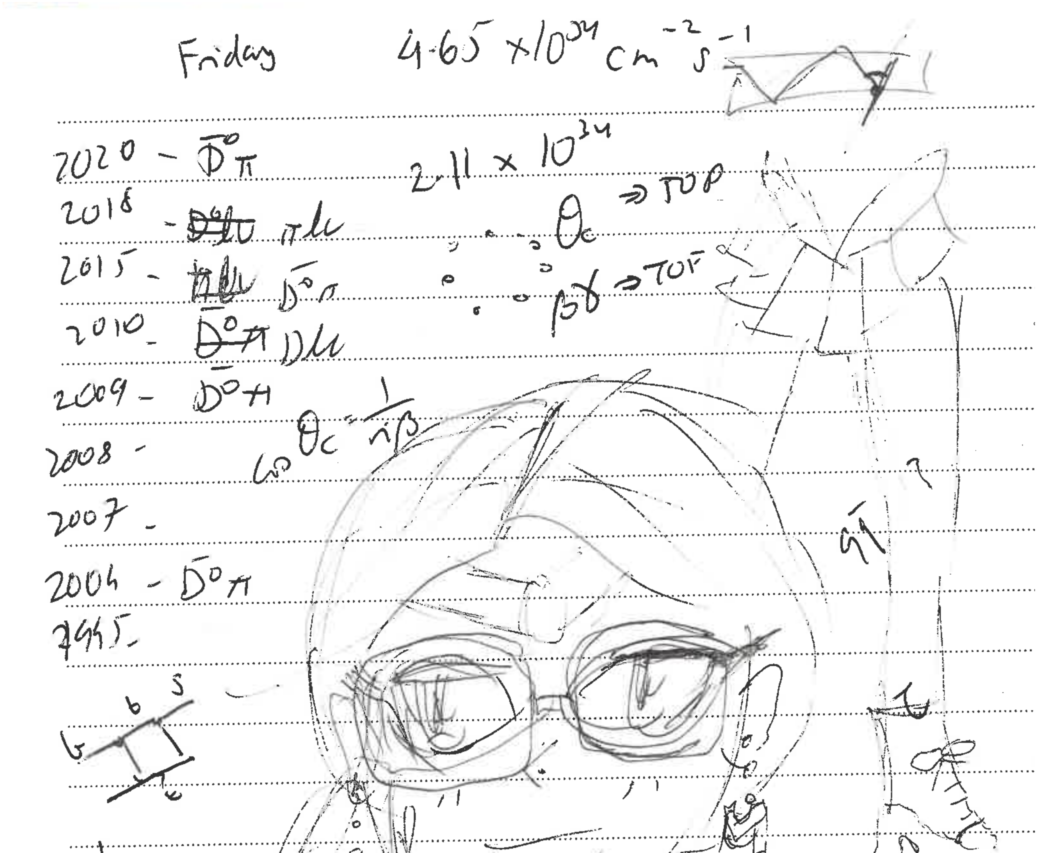
[123] U. Tamponi, *Study of hadron-ID using untagged K_s*, Tech. Rep. BELLE2-NOTE-TE-2019-024, Aug, 2019. https://docs.belle2.org/record/1652. 5.2.2

[124] S. Bilokin, "SysCorrFW." Gitlab, 2024. https://gitlab.desy.de/belle2/performance/systematic_corrections_framework. 5.2.2

[125] L. A. Corwin, *Leptonic Decays of the Charged B Meson.* PhD thesis, Ohio State University, 1, 2008. https://doi.org/10.2172/981681. 5.3

[126] D. Jacobi, J. Dingfelder, P. Lewis and S. K. Granderath, *Investigation of $B^+ \to \mu^+\nu_\mu$ with Inclusive Tagging at Belle II*, Master's thesis, Bonn, Physikalisches Institut, Bonn, 2021. https://docs.belle2.org/record/2635. 5.6

[127] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, *The Annals of Statistics* **7** (1979) 1 – 26. A.2.1



"Sailor Muon: Attending meetings by moonlight, fighting **ROOT** *by daylight."*