A Study of the Decay $B \to \pi\ell\nu_\ell$ and Determination of the Standard Model parameter $|V_{\mathrm{ub}}|$ with the Belle II Experiment

Nadia Ejlal Toutounji

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

Faculty of Science,
School of Physics,
University of Sydney

December 24$^{\text{th}}$, 2023

Abstract

Precision measurements of the magnitudes of the Cabibbo-Kobayashi-Maskawa (CKM)-matrix elements, which characterise the transitions between quarks of different flavours, are essential for testing the unitarity of the Unitarity Triangle (UT) described by the Standard Model. Of particular interest is the measurement of the magnitude of the matrix element $|V_{ub}|$, which forms a significant uncertainty in global UT fits. $|V_{ub}|$ can be measured via an exclusive approach from the differential decay rate of the semi-leptonic B-meson decay $B \to \pi \ell \nu_\ell$, where ℓ refers to a light lepton, either an electron e or muon μ. Using early data collected by the Belle II Experiment located in Tsukuba, Japan, an experiment based on the second-generation B-meson factory SuperKEKB which produces millions of pairs of B-mesons from electron-positron collisions, this study investigates the reconstruction of the signal $B \to \pi \ell \nu_\ell$ decay using a hadronic tagged approach. Using a state-of-the-art tagging algorithm developed for Belle II known as the Full Event Interpretation (FEI), $e^+ e^- \to \Upsilon(4S) \to B\bar{B}$ events are reconstructed with the other, non-signal B-meson in the event explicitly reconstructed in one of a number of specified hadronic decay modes. The signal yields in a preliminary subset of Belle II data equivalent to an integrated luminosity of $362~\mathrm{fb}^{-1}$ are estimated via a binned maximum likelihood fitting procedure using predictions from simulation. The partial branching fractions for the decays $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ are extracted in bins of the square of the momentum transfer to the leptonic system, q^2, from which the total branching fractions are determined. Finally, measurements of $|V_{ub}|$ are extracted using the distributions of the measured partial branching fractions alongside recent form factor constraints provided by lattice quantum chromodynamics (LQCD).

Statement of Originality

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

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

....

Nadia Ejlal Toutounji

Acknowledgments

Firstly, to my supervisor, Prof. Kevin Varvell, to whom I will forever owe my thanks for the constant support you have shown me over many years now. Thank you for your tremendous help and advice, and for the many hours you have spent reading drafts, discussing ideas and even responding to late-night emails when necessary. And to my co-supervisor, A/Prof. Bruce Yabsley, whose advice has been invaluable over the course of this project. Thank you for your brilliant suggestions and thought-provoking comments. I couldn't have asked for a better supervisory team.

Thanks also to Dr. Chia-Ling Hsu, who has been a constant source of help and guidance over the last few years, always willing to drop everything and help whenever I had a question to raise or something to discuss. To everyone in the Belle II Adelaide-Sydney group, it has been a pleasure meeting with you every week to talk through our analyses - thank you for all of your feedback and suggestions. Thank you to my parents, Akram and Sue, who have always encouraged me to pursue my best and to keep pushing forward through the difficult times. And finally, to my most dear friends, Asmaa, Fatima, Sarah and Tuba, whose support I have relied on heavily over the past few years - thank you for being there for me always.

Statement of Student Contribution

The analysis presented throughout the following chapters was conducted in part via the use of the Belle II Analysis Software Framework (basf2), the comprehensive suite of software developed by the Belle II Collaboration [57], of which I am a member. Most notably, the Full Event Interpretation (FEI) tagging algorithm implemented within the framework [13], developed primarily by Dr. Thomas Keck of Karlsruhe Institute of Technology and Dr. William Sutcliffe of the University of Bonn alongside the contributions of others including me, was used for the reconstruction of hadronic B-meson tags. In practice, the FEI was trained and applied centrally within the collaboration to produce a set of official samples known as skims, developed and implemented by a dedicated skimming sub-group, with analysts including me given the output of the skims in the form of lists of reconstructed tags. Being a large international collaboration of over 1000 members from more than 100 institutions, data processing and validation is conducted at Belle II by many individual collaborators divided across multiple working groups, with several analysis inputs and recommendations provided in an official capacity to all collaborators for use. One such example is the various correction factors applied to real and simulated Monte Carlo (MC) data to correct for known data-MC discrepancies, including lepton and pion identification corrections, π^0 reconstruction efficiency corrections, photon energy bias corrections and more. All such corrections applied to my analysis were developed as part of these independent studies performed by other collaborators.

All script writing, production of plots and analysis work was conducted by me with the exception of the following external sources, each referenced where appropriate in the text:

- A private software package developed by Dr. William Sutcliffe and Henrik Junkerkalefeld of the University of Bonn for providing plotting templates consistent with Belle

II style requirements was used to produce the majority of the plots included in this thesis [105].

- The eFFORT software package [72] developed by Dr. Markus Prim of the University of Bonn (and modified by Dr. Lu Cao of DESY) was used to generate the form factor weights described in Sections 4.1.3, 4.1.6, 5.2 and 7.1.3, and to perform the hybrid MC reweighting described in Section 5.1.1, with some independent code written by Prof. Florian Bernlochner of the University of Bonn used for the calculation of the relevant decay angles.

- Whilst the analysis selections and methodology were ultimately optimised and chosen by me, some limited inspiration was taken from prior studies including a hadronic FEI calibration study in [70] and a 2013 Belle study on hadronically-tagged $B \to \pi \ell \nu_\ell$ decays [32].

- Python scripts for the training and application of the continuum suppression boosted-decision-tree (BDT) algorithm in Section 5.6 were adapted from code written by Moritz Bauer of Karlsruhe Institute of Technology.

- The lepton and charged pion identification corrections in Sections 4.1.3, 4.1.6, 5.2 and 7.1.3 were applied using a private software package developed by Dr. William Suctliffe [80] and code written by Moritz Bauer.

- A Python interface developed by Dr. Lu Cao for the external RooUnfold software [95] was used to construct the q^2 migration matrices and perform the q^2 unfolding in Section 7.1.1.

- A private software package originally developed by Prof. Florian Bernlochner was used to generate the predicted $B \to \pi \ell \nu_\ell$ partial branching fractions and perform the χ^2 minimisation fits for the $|V_{\text{ub}}|$ extraction in Section 7.2 [99]. I adapted the software package substantially to include the functionality of performing simultaneous fits to two sets of partial branching fractions.

The content in the introductory Chapters 2 and 3 was largely adapted from my Master of Philosophy thesis completed in January 2019 [1]. Rather than just citing these chapters in the text, they were included again with light editing in the main matter for ease of communication. My supervisor, Prof. Kevin Varvell, read the contents of this thesis in its entirety and offered feedback and suggestions which have been implemented in the final text. In addition, at the time of writing this thesis, the presented $B \to \pi \ell \nu_\ell$ analysis was concurrently undergoing a process of internal peer review within the Belle II Collaboration for the purpose of later publishing a journal paper. As such, several suggestions and changes to the analysis methodology were implemented during this process based on the received feedback.

Throughout the course of this analysis, I wrote a large number of Python scripts and Jupyter notebooks for several purposes including the implementation of scaling and correction factors, the making of plots, various fitting procedures including Asimov, toy MC

and real data fits, the evaluation of systematic uncertainties and the determination of the $B \rightarrow \pi\ell\nu_\ell$ partial and total branching fractions. To construct the MC and data samples required for my analysis, I also wrote a set of Python scripts to access the sizeable samples available on the Belle II distributed computing system, reconstruct my signal-side decays, apply analysis pre-selections, and write out a set of kinematic variables to be used for plotting and further offline selection. I conducted optimisation studies to decide on the best analysis selections to implement, and validation studies for various aspects of the given analysis. Earlier iterations of the analysis presented in this thesis were written up independently and released publicly as conference papers, each available on the open-access arXiv website [2], [3] and [41].

....

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

Table of Contents

Chapter 1

Introduction

The Standard Model (SM) of Particle Physics is an extensive theory built upon the work of many scientists in the 1970s and earlier, and describes the fundamental particles making up the universe and the interactions between them. Since its conception, the SM has amassed a number of great successes, with experimental evidence confirming the existence of multiple particles described by the theory, often many years after they were first postulated. One such example includes the the discovery of the top quark in 1995 [4], some twenty years after the existence of a third generation of quarks was first predicted by Makoto Kobayashi and Toshihide Maskawa in 1973 [5]. Most notably in recent years was the groundbreaking discovery of the Higgs boson in 2012 [6], confirming a 50-year long hypothesis on the existence of a mechanism responsible for the origin of mass in fundamental particles.

With all of its successes, however, the Standard Model is unable to explain the existence of several observed phenomena. For instance, it contains no description of dark matter and dark energy, nor any mechanism for the gravitational interaction. The SM also does not provide answers to many questions that arise as to the nature of the fundamental particles it describes. One longstanding mystery is the existence of three separate generations of quarks and leptons, each seemingly identical to one another, distinguished only by their masses and the strength of their couplings under the weak interaction. Furthermore, a significant matter-antimatter asymmetry is observed in the current universe, the presence of such only being possible given certain conditions, as postulated by Andrei Sakharov in 1967 [7]. One of these conditions requires the violation of charge-parity (CP-)symmetry, the principle by which the laws of physics remain unchanged under the replacement of a particle with its anti-particle and an inversion of spatial coordinates. Though the SM does provide a mechanism through which CP-violation can occur [5], the magnitude of this effect is far too small to explain the large excess of matter observed.

It is for these reasons, along with several others, that the particle physics community has in recent years switched focus to theorising and searching for evidence of 'new physics', or physics Beyond the Standard Model (BSM). A multitude of new theories have thus far been postulated, some extensions to the SM, such as supersymmetric theories that introduce a new supersymmetric particle for each existing fundamental degree of freedom, significantly increasing the number of potential processes [8]. Other BSM theories introduce entirely new

concepts, such as the existence of extra dimensions beyond the four-dimensional spacetime we currently hold to be true [9]. Some BSM models also put forward various candidates for dark matter, with one example being the axion, a particle first hypothesised in order to solve the 'strong CP problem', a puzzling discovery whereby the strong nuclear force has been found to respect CP-symmetry unlike its weak counterpart [10].

Particle collider experiments around the world form one of the ways through which physicists aim to detect signatures of these new particles postulated by BSM theories. One such experiment is the Belle II Experiment, located in Tsukuba, Japan, which encompasses a broad physics program with a primary focus on searches for new physics. The experiment builds upon the work of its predecessor, the Belle Experiment, which achieved monumental success through its discovery of CP-violation in systems of neutral B-mesons [11], together with the BaBar Experiment situated in Stanford, California [12]. By its completion, Belle II aims to reach a total target data-set 50 times larger in size than that of Belle, which will greatly enhance its ability to study rare decays of B-mesons and potentially detect signatures of new physics that were inaccessible at earlier experiments with smaller data samples. Belle II commenced its first physics run in 2019, and has thus far collected over $400\,\mathrm{fb}^{-1}$ of data, a number roughly equivalent to the decays of 400 million pairs of B-mesons.

The following study utilises data from the Belle II Experiment, particularly from the semi-leptonic B-meson decay, $B \to \pi \ell \nu_\ell$, to extract the magnitude of the SM parameter $|V_{\mathrm{ub}}|$, one of the nine elements that comprise the Cabibbo-Kobayashi-Maskawa (CKM)-matrix, which characterises the weak transitions between quarks of different flavours. The matrix element $|V_{\mathrm{ub}}|$ forms one of the key components in testing a core axiom of the SM, namely the unitarity of the Unitarity Triangle (UT). The remainder of this chapter provides a summary of the Standard Model including the physics involved in semi-leptonic decays of B-mesons such as $B \to \pi \ell \nu_\ell$, and gives a detailed account of the current status of measurements of $|V_{\mathrm{ub}}|$. In Chapter 2, an in-depth description of the Belle II Experiment will be presented, including a breakdown of the various detector components and an overview of the software framework used for analysis within the collaboration. Chapter 3 will introduce the Full Event Interpretation (FEI) algorithm [13], the main analysis tool used for tagged analyses at Belle II, with an illustrative study on the calibration of the algorithm provided in Chapter 4. The subsequent chapters will collectively summarise the various components of the $B \to \pi \ell \nu_\ell$ analysis. After a comprehensive selection optimisation study detailed in Chapter 5, Chapter 6 will demonstrate the chosen signal extraction procedure used to estimate the number of signal events in the current Belle II data sample. Finally, measurements of the $B \to \pi \ell \nu_\ell$ branching fractions and $|V_{\mathrm{ub}}|$ will be extracted in Chapter 7, followed by a concluding discussion in Chapter 8.

Some final points should also be noted pertaining to the following sections. All equations, variables and constants listed in this report are expressed using natural units, as per the convention used within the field of particle physics. Both the speed of light c and the reduced Planck constant $\hbar$ are set to 1, and are thus henceforth excluded. In addition, a substantial portion of the introductory text contained in Chapters 2 and 3 was taken from previous work submitted as part of a Masters thesis undertaken prior to the current Doctor

of Philosophy.

1.1   The Standard Model of Particle Physics

The Standard Model (SM), established in its current form in the mid 1970s, details the fundamental particles that exist in nature, their properties, and the interactions that occur between them. It classifies these particles into two categories according to their spin, namely bosons, possessing integer values of spin, and fermions, with half-integer spin values. Fermions are further categorised into two groups, quarks and leptons, based on additional quantum properties. Figure 1.1 depicts the full range of fundamental particles included within the SM.
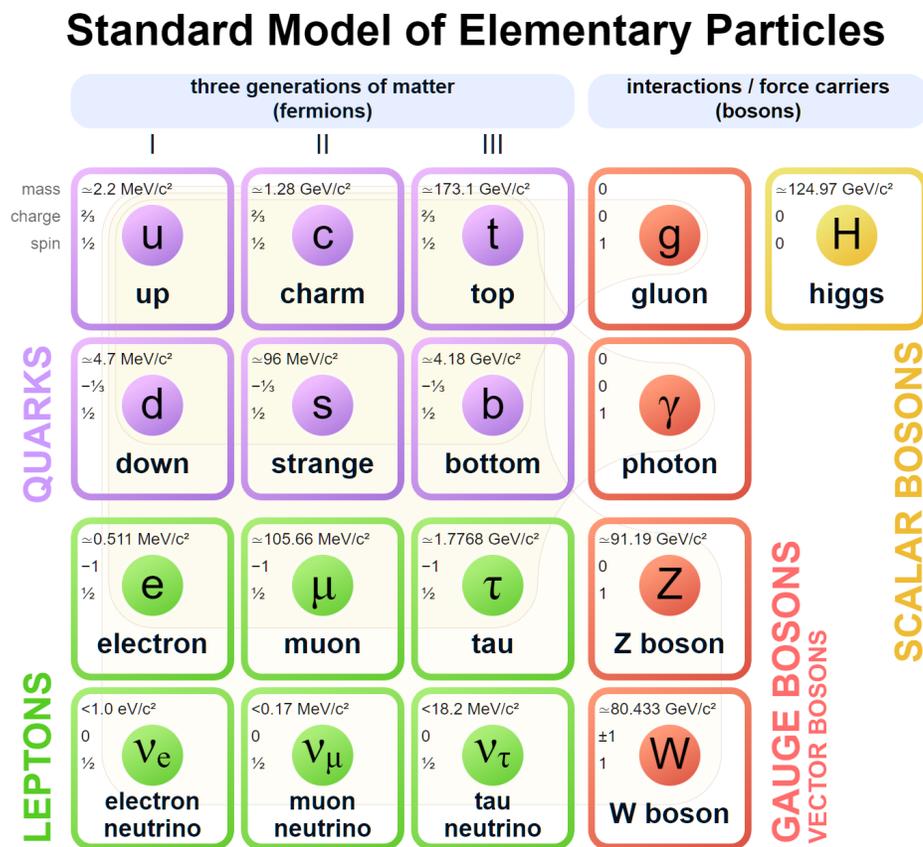
Figure 1.1: The fundamental particles of the Standard Model, including the values of their elementary properties, mass, charge and spin [14].

The bosons of the SM are further classified into gauge bosons, each with a spin of value 1, and a scalar boson named the Higgs boson, possessing a spin of 0. The gauge bosons are the force-carrying particles of the SM, mediating the interactions between the fundamental particles. The photon γ is a massless particles that mediates the electromagnetic force between particles that possess electric charge. The strong interaction is mediated by the gluon,

and is only experienced by those particles that possess a colour charge, namely quarks and gluons themselves. Colour charge is an additional elementary property with three different representations, red, green and blue, with the combination of all three colours into a colour singlet resulting in a net colour charge of 0, in a manner that resembles the combination of a positive and negative electric charge resulting in a state that is electrically neutral. An anti-colour also exists for each colour, anti-red, anti-green and anti-blue, such that the combination of a colour and its corresponding anti-colour also results in a colourless state. Gluons carry net colour, as they possess both a unit of colour and a contrasting anti-colour, such that eight effective gluon states exist. Gluons can thus not only mediate the strong interaction between quarks, but can be active participants, with gluon-gluon interactions possible under the SM formalism. This phenomenon significantly increases the complexity of the strong interaction when compared with the electromagnetic interaction, as whilst photons mediate the latter, they themselves are electrically neutral and no photon-photon interactions are possible to first-order [1]. The strong force is the strongest of the fundamental interactions, and is responsible for the strong nuclear force that binds protons and neutrons, each composite particles comprised of quarks, tightly together within the nuclei of atoms.

The weak force, named for being several orders of magnitude weaker at low energy than both the strong and electromagnetic forces, is mediated by both the electrically neutral Z-boson, and the positively and negatively charged W-bosons, W^+ and W^-. The Z- and W- bosons have the unique property of being the only gauge bosons that are not massless. All fermions participate in the weak interaction, as does the Higgs boson. It is the only fundamental interaction capable of changing quarks from one flavour to another, and is the only interaction known to violate charge-parity (CP)-symmetry, the mechanisms for which are described in detail in Section 1.1.1. The gravitational force is not incorporated within the SM, and has no gauge boson associated with its mediation under the SM formalism. In BSM theories that attempt to provide a quantum mechanical description for gravity ('quantum gravity'), such a mediating gauge boson does exist, namely the 'graviton' [15]. The gravitational interaction is the weakest of all interactions between fundamental particles at low energies. The last remaining boson of the SM is the scalar Higgs boson, a fundamental particle arising from the excitation of the Higgs field. The Higgs field was first postulated in 1964 [16] as a field that permeates all matter and space in the universe, providing a mechanism through which fundamental particles, by their interactions with the field, acquire mass. The Higgs field is responsible for the breaking of the electroweak symmetry arising from the unification of the weak and electromagnetic forces, a process necessary under the SM for force-carrying particles such as the Z- and W-bosons to be able to be massive.

The unification of the weak and electromagnetic forces under the SM considers these forces as two aspects of the same fundamental 'electroweak' force, with the distinct behaviour observed in their respective strengths and interaction distances being a direct consequence of the low energies characteristic of the current universe. At energies of ~ 246 GeV and above, corresponding to the so-called Fermi or electroweak scale, the two forces become largely

---

[1]At higher order, photon-photon interactions are possible under the SM via the exchange of virtual charged particles.

indistinguishable, forming a single electroweak force [17]. Whilst the SM incorporates this unification of the weak and electromagnetic forces, it does not possess any mechanism for the unification of the electroweak force with the strong force. However, many BSM theories postulate that an energy threshold exists for which all three of the fundamental forces described by the SM can be unified, with these collectively known as Grand Unified Theories (GUTs). Whilst individual theories disagree on the exact energy scale of this unification, many GUTs including those derived from supersymmetry suggest that grand unification would occur above energies of the order of 10^{16} GeV, suggesting that the three fundamental forces were indistinguishable in the very early universe [18]. Whilst it is impossible to probe such high energies at collider experiments, indirect searches for grand unification aim to find evidence for predictions made within the various GUTs, such as the existence of proton decay. After grand unification, higher level theories have been postulated that also incorporate the gravitational force, referred to as Theories of Everything (TOE) [19].

Quarks are SM fermions that experience all of the fundamental forces. In addition to the weak force affecting all fermions, quarks also participate in both the electromagnetic and strong interactions as they possess both electric and colour charge. There are six different types, or flavours, of quarks existing in three generations, with the generations identical to one another for all elementary properties except for their masses and the strength of their respective weak couplings. In order of increasing mass, these generations include the up u and down d quarks, the charm c and strange s quarks, and the top t and bottom b quarks. For each flavour of quark, a corresponding anti-quark exists with opposite electric charge and equivalent anti-colour, these being $\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$ and $\bar{b}$. Due to a property of the strong force known as colour confinement, quarks and anti-quarks are unable to exist freely at low energy, and can only be found bound together in composite particles called hadrons with no net colour charge. This is due to the fact that, for any attempt to separate these bound quarks, it becomes far more energetically favourable to produce a new quark-anti-quark pair instead. Hadrons can further be classified into baryons, mesons and exotic particles. Baryons are colourless bound states of three valence quarks, the proton (uud) and neutron (udd) being notable examples, with anti-baryons similarly comprised of three valence anti-quarks. Mesons are colourless bound states of a quark and an anti-quark, such as the pion π^+ ($u\bar{d}$) and its corresponding anti-particle π^- ($\bar{u}d$). Any hadrons consisting of more than three valence quarks or anti-quarks, or those states containing gluons, are referred to as exotic particles, and include both baryon-like hadrons such as the pentaquark with five valence quarks, and meson-like hadrons such as the tetraquark with four valence quarks, with experimental evidence suggesting the existence of both particles having been collected in recent years [20].

Finally, the remaining fermions of the SM are the leptons, distinguished from quarks in that they do not carry colour charge. Thus, while both quarks and leptons are fermions and participate in the weak interaction, leptons do not experience the strong force. Leptons are similarly categorised into three separate generations according to their masses, with the electron e^- and electron neutrino ν_e, muon μ and muon neutrino ν_μ, and tau τ and tau neutrino ν_τ. The electron, muon and tau leptons all possess electric charge, and thus also interact via the electromagnetic force. Their corresponding neutrinos, however, are electri-

cally neutral and only participate in weak (and gravitational) interactions. For each lepton, an equivalent anti-particle exists, including the positron e^+, anti-muon μ^+ and anti-tau τ^+, which contain the opposite electric charge to their matter counterparts. Anti-neutrinos $\bar{\nu}_e$, $\bar{\nu}_\mu$ and $\bar{\nu}_\tau$ also exist, but being neutral particles, they cannot be distinguished from neutrinos on the basis of electric charge. All quarks and charged leptons are considered to be Dirac fermions under the SM, that is, that a distinct anti-particle exists for each fermion. It is yet unclear, however, whether neutrinos are Dirac fermions, with an alternative representation postulated for the neutrino in 1937, the Majorana fermion [21]. Neutrinos were originally assumed to be massless in the SM, with experimental evidence regarding the neutrino helicity seeming to support the Dirac formalism. For massless neutrinos, the only known fundamental difference between neutrinos and anti-neutrinos is due to a property known as helicity, or handedness, whereby a particle is considered to be right-handed (or have a positive helicity) if the direction of the particle's spin is the same as its motion, or left-handed (with a negative helicity), if the spin and motion are in opposite directions. To date, all detected neutrinos have been found to be left-handed, whilst all anti-neutrinos possess a right-handed helicity. However, evidence for neutrino oscillation, a quantum mechanical process by which the measured flavour of neutrinos can change over time, was first established in 1998 [22], implying the existence of a small but non-zero neutrino mass. Massive neutrinos are allowed for in the Majorana representation, which states that Majorana particles are their own anti-particles, and thus, in the case of the neutrino, no distinct anti-neutrino would exist, and both terms would simply describe a single particle that can be either left- or right-handed. To date, there has been no definitive evidence as to whether neutrinos are truly Dirac or Majorana fermions.

1.1.1 The Cabibbo-Kobayashi-Maskawa (CKM) Matrix and the Unitarity Triangle (UT)

The Cabibbo-Kobayashi-Maskawa (CKM) matrix, also referred to as the quark mixing matrix, is a 3×3 matrix specified within the SM which characterises the strength of weak interactions involving quark flavour transitions. As the W-bosons that mediate the weak interaction are not electrically neutral, any quark emitting or absorbing a W-boson must change its flavour in order to conserve electric charge. Given the charges of the quarks listed in Figure 1.1, one can infer that the only potential quark mixing interactions that conserve this charge involve the former quark of each generation undergoing a transition to either of the three latter quarks, with a total of nine possible quark transitions. This gives rise to the following matrix,

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

where each matrix element of V_{CKM} is labelled according to the quark transition it specifies. The matrix elements are directly related to the strength of the given quark mixing interactions, with $|V_{\mathrm{ij}}|^2$ related to the probability of the transition between quarks of flavour

i and *j*.

The CKM matrix can further be parameterised a number of ways. In order to conserve probability, the CKM-matrix must be unitary under the SM, a property which results in only four independent real parameters being required to completely describe all of the nine complex elements. One such popular choice, the Wolfenstein parameterisation [23], defines four real parameters, A, ρ, η and $\lambda \approx |V_{\mathrm{us}}|$, and expands each element via a power series in λ, with the parameterised CKM-matrix written as,

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

where $\mathcal{O}(\lambda^4)$ refers to additional terms given by the higher order expansion at order λ^4. The parameter η represents a phase which has the consequences that, if non-zero, the CKM-matrix does not remain the same under complex conjugation, $V_{\mathrm{CKM}} \neq V_{\mathrm{CKM}}^*$, thus introducing CP-violation into the quark mixing matrix. Were $\eta = 0$, the CKM-matrix would no longer be complex in nature, and no CP-violation would be possible.

Under the premise that the CKM-matrix must be unitary, a total of six equivalent unitarity conditions can be derived from the matrix elements. One example is the relation,

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

where V_{ub}^*, V_{cb}^* and V_{tb}^* are the complex conjugates of their respective elements. This condition can be represented graphically as a triangle in the complex plane, termed a unitarity triangle (UT), as illustrated in Figure 1.2. Similar triangles can also be constructed for the five other unitarity conditions, though the UT depicted is, by convention, the most often referenced as it is the one for which the lengths of the sides are most similar. The lengths of each side of the triangle are determined from the magnitudes of the CKM-matrix elements, and the vertices are situated at (0,0), (1,0) and $(\bar{\rho}, \bar{\eta})$, where $\bar{\rho}$ and $\bar{\eta}$ are related to ρ and η via the following relation,

$$\rho + i\eta = (1 + \lambda^2/2)(\bar{\rho} + i\bar{\eta}) + \mathcal{O}(\lambda^4).$$

Finally, the angles of the UT are likewise a function of the individual matrix elements, and are defined using two notations that are often interchanged within the literature (α, β, γ or ϕ_2, ϕ_1, ϕ_3),

$$\alpha \equiv \phi_2 = \arg\left(-\frac{V_{\mathrm{td}}V_{\mathrm{tb}}^*}{V_{\mathrm{ud}}V_{\mathrm{ub}}^*}\right),$$

7
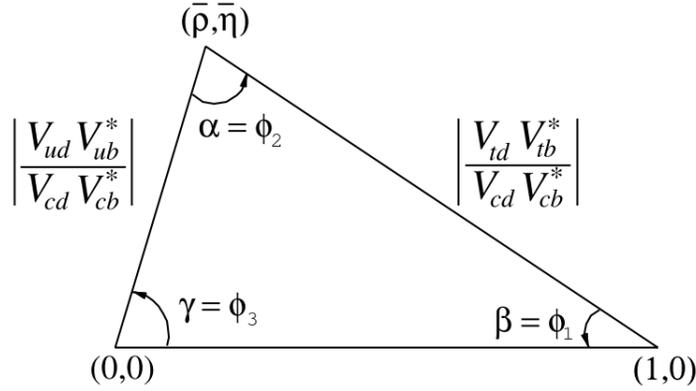
Figure 1.2: The Unitarity Triangle [20].

$$\beta \equiv \phi_1 = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right),$$

$$\gamma \equiv \phi_3 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right).$$

Testing the unitarity of the CKM-matrix is a fundamental test of the SM and an essential part of the search for BSM physics. The unitarity conditions can be tested directly through experimental determinations of the matrix elements, with any deviation from zero indicative of potential new physics. Alternatively, the angles of the UT can also be measured experimentally, where a result in which $\alpha + \beta + \gamma \neq 180°$ further suggests a departure from the SM formalism. It is largely desirable, therefore, to produce experimental measurements of these quantities with as high a precision as possible. Multiple collaborations exist, including the CKMfitter [24] and UTfit [25] groups, that produce global fits to the UT using a range of current experimental results. Several parameters are input into the global fits, including the measured values of the CKM-matrix elements and a number of key observables relating to CP-violation, with a full list detailed here [26]. Figure 1.3 depicts an example of the latest fit performed by CKMfitter [24] using experimental results available mid-2021. The measurements are plotted on the $\bar{\rho} - \bar{\eta}$ plane with their associated uncertainty bands, and the fitted apex of the UT is situated at the overlap of each band. To date, no significant deviation from unitarity has been observed [20].

The importance of $|V_{ub}|$

The individual elements of the CKM-matrix are currently measured experimentally to varied levels of precision. As of mid-2021, the world averages for the magnitudes of the nine elements as determined from the global fits described above are [20],

$$|V_{CKM}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix},$$
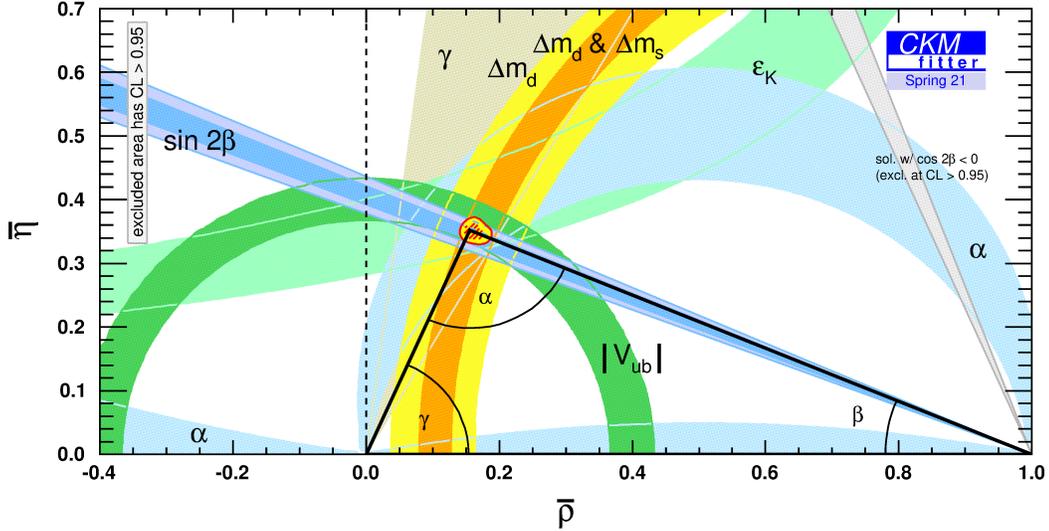
Figure 1.3: Global fit to the Unitarity Triangle, provided by the CKMfitter group [24].

where the quoted uncertainties on each value include contributions from both statistical and systematic sources, and are listed as a single value for symmetric errors or as separate positive and negative uncertainties in the case that the errors are asymmetric. The element V_{ub}, possessing the smallest magnitude of each of the nine elements, also suffers from the highest relative percentage uncertainty. This in turn results in $|V_{ub}|$ contributing a significant uncertainty in the global UT fits, as can be seen in Figure 1.3. Improving the precision on measurements of $|V_{ub}|$ is thus highly sought after, a goal shared by the Belle II Experiment, which aims to contribute to decreasing the uncertainty on $|V_{ub}|$ with its large projected data-set.

Further interest in the determination of $|V_{ub}|$ arises from the fact that historically, experimentally measured values of $|V_{ub}|$ have been known to differ from each other by as many as 3 standard deviations, or 3σ, depending on the method of extraction. Namely, these methods include exclusive extractions, in which $|V_{ub}|$ is determined via measuring the branching fraction of the decay of a B-meson to a specified final state by way of a $b \to u$ quark transition, such as the decay to a pion, electron and electron neutrino, $B \to \pi e \nu_e$, a process described in depth in Section 1.2.3. Alternatively, inclusive measurements of $|V_{ub}|$ are made via B-meson decays to non-specified final states, such as $B \to X_u e \nu_e$, where X_u refers to any hadronic system containing a u quark. Experimental methods for exclusive and inclusive extractions differ substantially, and require separate theoretical inputs. The current world average measurements for $|V_{ub}|$ are quoted as $|V_{ub}|(\text{exclusive}) = (3.70 \pm 0.10 \pm 0.12) \times 10^{-3}$ for exclusive determinations, and $|V_{ub}|(\text{inclusive}) = (4.25 \pm 0.12^{+0.15}_{-0.14} \pm 0.23) \times 10^{-3}$ for inclusive determinations [20], where the first uncertainty is experimental and the second theoretical. The third uncertainty assigned to the inclusive measurement is due to an observed model dependence, whereby an additional spread in the inclusive results is obtained when using three different theoretical models. The observed discrepancy between current $|V_{ub}|$ measurements from

both approaches further motivates the need for newer, more precise measurements of $|V_{\mathrm{ub}}|$ alongside improvements to theoretical models, in hopes of resolving this tension. The study detailed throughout the coming chapters focuses on an exclusive approach to extracting $|V_{\mathrm{ub}}|$.

1.2 B-meson Decay

The single most important particle at the heart of many electron-positron collider experiments including Belle II is the B-meson, a meson characterised as possessing either a b quark or anti-quark. B-mesons come in several different types, or flavours, including both charged and neutral B-mesons, B^+ and B^0, and their associated anti-particles, B^- and $\bar{B}^0$. B^+ and B^0 mesons both contain a bottom anti-quark, pairing this with an up quark in the case of B^+ ($u\bar{b}$), or a down quark for B^0 ($d\bar{b}$). The corresponding anti-particles thus contain a bottom quark together with either an anti-up, as for B^- ($\bar{u}b$) or anti-down quark, for $\bar{B}^0$ ($\bar{d}b$). Strange and charmed B-mesons also exist, which pair the bottom (anti-)quark with either a strange or charm (anti-)quark instead, resulting in the particles B_s^0 ($s\bar{b}$) and B_c^+ ($c\bar{b}$), with respective anti-particles $\bar{B}_s^0$ ($\bar{s}b$) and B_c^- ($\bar{c}b$).

The study of B-mesons has been the focus of the first-generation electron-positron collider experiments Belle and BaBar, and remains the focus of the second-generation Belle II Experiment, with the aims of the individual experiments evolving over time. Both Belle [11] and BaBar [12] were instrumental in discovering the existence of CP-violation in systems of neutral B-mesons, the purpose upon which both experiments were conceptualised some time after evidence of CP-violation was first observed in neutral kaons [27]. Moving forward, Belle II, with the ambitious goal of collecting 50 times the data-set of Belle, aims to use rare B-meson decays as a probe to search for new physics. The Belle II Experiment will be discussed at length in the following chapter.

B-mesons decay via hundreds of allowed channels [20], with the rarity of each decay quantified by its associated branching fraction. The branching fraction of a particle decay refers to the fraction of particles that decay according to the given mode with respect to the total number of decaying particles. The decays of B-mesons are typically classified into three separate groups according to their decay products.

1.2.1 Hadronic Decays

Hadronic B-meson decays refer to those in which the decay products are purely hadronic in nature. That is, no leptonic decay products are present amongst the immediate daughters of the decaying B-meson. One such example is the decay of a neutral B-meson to a pair of charged pions, $B^0 \to \pi^+\pi^-$, a rare decay with a branching fraction of $(5.12 \pm 0.19) \times 10^{-6}$ [20]. $B^0 \to \pi^+\pi^-$ is one of several hadronic decay modes in which CP-violation can be directly observed, and through which the angle α of the UT can be measured [28]. A Feynman diagram representing the decay is pictured in Figure 1.4(a).

(a) Hadronic decay $B^0 \to \pi^+\pi^-$ (b) Leptonic decay $B^+ \to \tau^+\nu_\tau$

(c) Semi-leptonic decay $B^+ \to \pi^0\ell^+\nu_\ell$
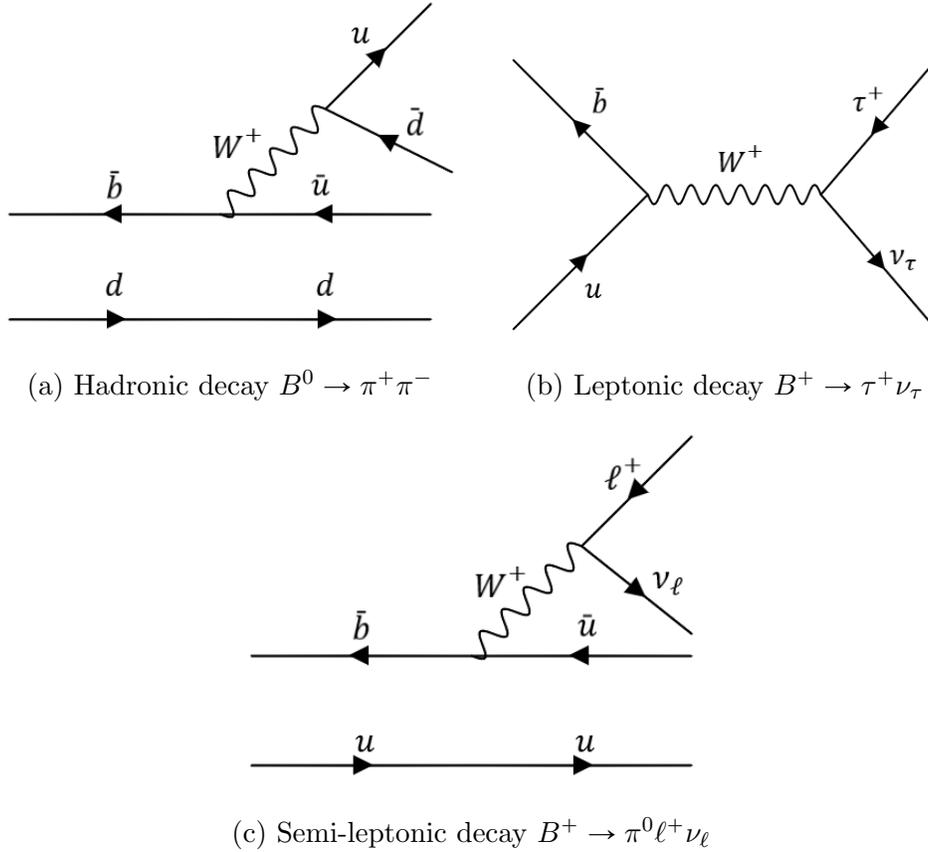
Figure 1.4: Feynman diagrams for example hadronic (a), leptonic (b), and semi-leptonic (c) B-meson decays.

Hadronic decays also play a significant role in the coming analysis, forming the basis of the tagging approach used to reconstruct the signal decay of interest. Hadronic tagging is explored in detail in Chapter 3.

1.2.2 Leptonic Decays

Conversely, B-mesons can also decay purely leptonically. Notable examples include the decays $B^+ \to e^+\nu_e$ and $B^+ \to \mu^+\nu_\mu$, both substantially rare decays with branching fractions predicted by the SM of $\mathcal{B}(B^+ \to e^+\nu_e) = (8.1 \pm 0.6) \times 10^{-12}$ and $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (3.5 \pm 0.3) \times 10^{-7}$, respectively [28]. Due to their rarity, only upper limits on these branching fractions have thus far been able to be determined experimentally, corresponding to world averages of $\mathcal{B}(B^+ \to e^+\nu_e) < 9.8 \times 10^{-7}$ and $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 8.6 \times 10^{-7}$ [20]. Whilst these decays both involve a $b \to u$ quark transition, $|V_{\mathrm{ub}}|$ is not typically measured via these modes due to their small branching fractions.

Of particular interest is the decay $B^+ \to \tau^+\nu_\tau$, which possesses a significantly higher branching fraction corresponding to a world average measurement of $(1.09 \pm 0.24) \times 10^{-4}$

11

[20], and is considered as a potential avenue through which signatures of BSM physics may be detected. The Two Higgs Doublet Model (2HDM) is a proposed extension to the SM, introducing an additional Higgs doublet that ultimately would result in the existence of two charged (H^+ and H^-) and three neutral (h, H and A) Higgs bosons [29]. According to the 2HDM, in an additional allowed process, a charged Higgs $H^\pm$ could replace the $W^\pm$ boson in the Feynman diagram for the $B^+ \to \tau^+\nu_\tau$ decay depicted in 1.4(b). Both processes would then interfere quantum mechanically, resulting in a measured branching fraction that departs from the SM prediction. Similar interference effects have also been proposed by other BSM theories including several GUTs that postulate the existence of an additional particle known as the leptoquark, a particle that decays into a quark and a lepton [30]. Coupling to both fermions, virtual leptoquarks would similarly be able to mediate leptonic decays such as $B^+ \to \tau^+\nu_\tau$, contributing to the magnitude of the total measured branching fraction. Precision measurements of the $B^+ \to \tau^+\nu_\tau$ branching fraction at Belle II would thus provide a direct test for any such deviations due to the existence of BSM particles such as the charged Higgs and the leptoquark.

1.2.3   Semi-leptonic Decays

Finally, B-mesons may decay to a mixture of hadronic and leptonic daughters in what is known as semi-leptonic decay. These decays are the standard processes through which the magnitudes of the CKM-matrix elements are measured, due to the simplicity introduced by the W-boson coupling to a leptonic pair which does not interact with the recoiling hadronic system via the strong force. One prominent example is the decay of a B-meson to a pion, an electron or muon, and its associated neutrino, $B \to \pi\ell\nu_\ell$, where ℓ refers to either of the two light leptons (e or μ). This decay forms the basis of the study detailed in the coming chapters, and is explored in depth in the following section.

Semi-leptonic decays involving τ leptons, often termed semi-tauonic decays, are presently the subject of great interest at collider experiments worldwide. This is largely due to discrepancies observed between the SM predictions and experimental results of the quantities $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$. These terms refer to the ratio between the branching fractions of semi-leptonic B-meson decays to τ leptons versus to electrons or muons, for D and D^* mesons, respectively, and are defined as,

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

where the denominator is given by the average of the electron and muon modes, $\mathcal{B}(\bar{B} \to D^{(*)}e^-\bar{\nu}_e)$ and $\mathcal{B}(\bar{B} \to D^{(*)}\mu^-\bar{\nu}_\mu)$. Current world average experimental results exceed the SM predictions by 2.0σ and 2.2σ for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$, respectively [31]. These discrepancies have been hypothesised to potentially be the result of interference from various BSM physics models, including those proposing the existence of charged Higgs bosons, such as the 2HDM [28], or the existence of leptoquarks [30]. Improving the precision on measure-

ments of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ as a probe for new physics is a key aim of the Belle II Experiment.

$B \to \pi \ell \nu_\ell$

The semi-leptonic decay of a B-meson to a pion, light lepton and corresponding neutrino, is largely considered the ideal decay for an exclusive extraction of $|V_{\mathrm{ub}}|$. Whilst several semi-leptonic B-meson decays involve a $b \to u$ quark transition, pions are scalar mesons with an integer spin of 0, differentiating them from other vector mesons such as the ρ-meson with a spin of 1. Strong interaction effects pertaining to scalar particles are the simplest to model, making the semi-leptonic $B \to \pi \ell \nu_\ell$ decay the most desirable choice for the exclusive extraction. Both decays of charged and neutral B-mesons are possible, written respectively as $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^0 \to \pi^- \ell^+ \nu_\ell$, where $\ell = e$ or μ. Neutral pions (π^0) are mesons containing either an up or down quark-antiquark pair[2], and exist as their own anti-particles. Conversely, the negatively charged pion (π^-) is comprised of an up anti-quark paired with a down quark ($\bar{u}d$), with associated anti-particle π^+ possessing a quark content of $u\bar{d}$. A Feynman diagram of the $B^+ \to \pi^0 \ell^+ \nu_\ell$ decay illustrating the $b \to u$ quark transition is shown in 1.4(c).

For any given particle decay, a decay rate can be defined which corresponds to the probability of the decay occurring per unit time. The differential decay rate for the $B \to \pi \ell \nu_\ell$ process to first order is evaluated as a function of q^2,

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} = \frac{G_F^2}{24\pi^3} |V_{\mathrm{ub}}|^2 |f_+(q^2)|^2 |\vec{p}_\pi|^3,$$

where q^2 is the square of the four-momentum of the W-boson mediating the decay, and characterises the four-momentum transferred to the leptonic decay products. Here, G_F is the Fermi coupling constant, $\vec{p}_\pi$ is the three-momentum of the daughter pion in the B-meson rest frame, and $f_+(q^2)$ is a vector form factor, an analytical function of q^2 derived from theoretical models. The form factor encodes the effects of the strong interactions between the quarks contained within the hadrons involved in the decay. Due to the theoretical simplicity of the scalar pion, these effects can be parameterised within a single form factor, unlike related semi-leptonic decays to vector mesons such as the ρ, which possesses a differential decay rate dependent on a total of three separate form factors [32]. The determination of the form factor $f_+(q^2)$ is non-trivial, and calculations made using lattice quantum chromodynamics (LQCD) only provide values of the form factor in the kinematic range of 17 GeV$^2 \leqslant q^2 \leqslant$ 26.4 GeV^2 [28]. At lower values of q^2, the coupling constant of the strong interaction becomes too large such that perturbative QCD approximations no longer hold, making these form factor calculations exceedingly difficult. The upper limit of 26.4 GeV^2 on the kinematic range corresponds to the q^2 kinematic endpoint, the maximum possible q^2 for the $B \to \pi \ell \nu_\ell$ decay given the respective masses of the parent B-meson (M_B) and daughter pion (M_π), where $q_{\mathrm{max}}^2 = (M_B - M_\pi)^2$. To provide results over the full kinematic range, 0 GeV$^2 \leqslant q^2 \leqslant$ 26.4 GeV^2, these values must then be extrapolated to low values of q^2, a

---

[2]More accurately, a neutral pion exists in a superposition of two quantum states, with quark content $\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$.

process that requires a suitable parameterisation of the form factor. Several options exist for this extrapolation, including the BGL [33] and BCL [34] parameterisations.

Decay Mode	Measured Branching Fraction $\times 10^4$	Measured $|V_{ub}| \times 10^3$	Experiment/ Year	Recon. Method	Ref.
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.38 \pm 0.10 \pm 0.16 \pm 0.08)$	$(3.82 \pm 0.14 \pm 0.22 \pm 0.11^{+0.88}_{-0.52})$	BaBar, 2005	Untagged	[35]
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.41 \pm 0.05 \pm 0.07)$	$(3.78 \pm 0.13^{+0.55}_{-0.40})$ (Calc.) (2.95 ± 0.31) (χ^2 min. fit)	BaBar, 2011	Untagged	[36]
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.45 \pm 0.04 \pm 0.06)$	(3.25 ± 0.31)	BaBar, 2012	Untagged	[37]
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.33 \pm 0.18 \pm 0.11 \pm 0.01 \pm 0.07)$	$(3.24 \pm 0.22 \pm 0.13^{+0.55}_{-0.39} \pm 0.09)$	CLEO, 2003	Untagged	[38]
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.37 \pm 0.15 \pm 0.11)$	$(3.6 \pm 0.4 \pm 0.2^{+0.6}_{-0.4})$	CLEO, 2007	Untagged	[39]
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.49 \pm 0.09 \pm 0.07)$	(3.52 ± 0.29)	Belle, 2013	Had. Tag	[32]
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(1.38 \pm 0.19 \pm 0.14 \pm 0.03)$	$(3.60 \pm 0.41 \pm 0.20^{+0.62}_{-0.41})$ (FNAL LQCD), $(4.03 \pm 0.46 \pm 0.22^{+0.59}_{-0.41})$ (HPQCD LQCD)	Belle, 2007	SL Tag	[40]
$B^0 \to \pi^- e^+ \nu_e$	$(1.43 \pm 0.27 \pm 0.07)$	(3.88 ± 0.45)	Belle II Preliminary *, 2022	Had. Tag	[41]
$B^0 \to \pi^- \ell^+ \nu_\ell$ World Avg.	(1.50 ± 0.06)	-	-	-	[20]
$B^+ \to \pi^0 \ell^+ \nu_\ell$	$(0.80 \pm 0.08 \pm 0.04)$	(3.52 ± 0.29)	Belle, 2013	Had. Tag	[32]
$B^+ \to \pi^0 \ell^+ \nu_\ell$	$(0.77 \pm 0.14 \pm 0.08 \pm 0.00)$	$(3.60 \pm 0.41 \pm 0.20^{+0.62}_{-0.41})$ (FNAL LQCD), $(4.03 \pm 0.46 \pm 0.22^{+0.59}_{-0.41})$ (HPQCD LQCD)	Belle, 2007	SL Tag	[40]
$B^+ \to \pi^0 e^+ \nu_e$	$(0.83 \pm 0.17 \pm 0.06)$	(3.88 ± 0.45)	Belle II Preliminary *, 2022	Had. Tag	[41]
$B^+ \to \pi^0 \ell^+ \nu_\ell$ World Avg.	(0.78 ± 0.03)	-	-	-	[20]
Excl. $|V_{ub}|$ World Avg.	-	$(3.70 \pm 0.10 \pm 0.12)$	-	-	[20]

Table 1.1: Prior measurements of the branching fractions of $B \to \pi \ell \nu_\ell$ decays, and the corresponding extracted values of $|V_{ub}|$. Entries are ordered chronologically and grouped by experiment and tagging method. The current world averages are also listed. (*These results refer to an earlier iteration of the analysis presented in this thesis.)

In the BCL parameterisation [34], q^2 is first mapped to a separate variable z with the following relation,

$$z(t, t_0) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}},$$

where $t^{\pm} = (M_B \pm M_\pi)^2$, $t_0 = (M_B + M_\pi)(\sqrt{M_B} - \sqrt{M_\pi})^2$, and M_B and M_π refer to the masses of the B- and π-mesons, respectively. The form factor can then be parameterised as a function of z, with,

$$f_+(z) = \frac{1}{1 - q^2(z)/M_{B*}^2} \sum_{k=0}^{K-1} b_k \left[z^k - (-1)^{k-K} \frac{k}{K} z^K \right],$$

where b_k are the coefficients of the parameterisation and M_{B*} refers to the mass of a pole in the $d\Gamma/dq^2$ distribution, $M_{B*} = 5.325$ GeV, defined as a large peak in the distribution at high q^2. In an exclusive approach using $B \to \pi \ell \nu_\ell$ decays, the differential decay rate can be measured experimentally, and together with theoretical input on the structure of the form factor and predictions based on LQCD, a measurement of $|V_{\mathrm{ub}}|$ can therefore be extracted.

Table 1.1 lists a number of recent results for the branching fractions of $B \to \pi \ell \nu_\ell$ decays, alongside the corresponding values of $|V_{\mathrm{ub}}|$ extracted from the measurements. These include results from several $e^+ e^-$ experiments, utilising a range of different analysis techniques and LQCD predictions made over the last two decades, such as those from the Femilab/MILC (FNAL LQCD) [42] and HPQCD [43] Collaborations. As one example of the differences in analysis procedure, the listed results include determinations via both tagged and untagged reconstruction methods, differing in their treatment of the other, non-signal B-meson in an $\Upsilon(4S) \to B\bar{B}$ decay where the signal B-meson decays as $B \to \pi \ell \nu_\ell$. A detailed description of tagged and untagged reconstruction is left to Chapter 3, though it is noted here that tagged methods, which explicitly require the reconstruction of the other B-meson in the event via a known decay mode, offer a higher purity at the cost of a lower efficiency than untagged methods, making the two reconstruction methods complementary in approach. The systematic uncertainties of each approach also vary, with the higher backgrounds typical of untagged reconstruction methods resulting in background-related uncertainties such as form factor and branching fraction contributions playing a more significant role than in tagged determinations [28]. Conversely, tagged methods are often associated with an additional source of uncertainty related to the calibration of the tagging algorithms used for the reconstruction. The CLEO and BaBar results in Table 1.1 were determined via untagged analyses [35] – [39], with the two Belle results listed originating from tagged approaches, where $B \to \pi \ell \nu_\ell$ was reconstructed against a semi-leptonic B-meson tag in Ref. [40] and a hadronic tag in Ref. [32]. To date, the highest precision on $|V_{\mathrm{ub}}|$ extracted using an exclusive tagged approach corresponds to the $\sim 8.2\%$ achieved by the 2013 Belle analysis [32], with the 2012 BaBar analysis [37] representing the highest precision thus far achieved via untagged reconstruction methods, a value of $\sim 9.5\%$.

The various results in Table 1.1 also reflect differences in the methods used for the extraction of $|V_{\mathrm{ub}}|$. In one approach, LQCD predictions for the form factor $f_+(q^2)$ were combined with experimental data on the measured $B \to \pi \ell \nu_\ell$ differential decay rate to extract a measurement of $|V_{\mathrm{ub}}|$ using the relevant formula for the differential decay rate above, an approach

detailed in Ref. [35], [39] and [40]. Other extractions favoured a more data-driven approach, using LQCD form factor determinations to generate a set of predictions for the differential decay rate assuming an initial input value of $|V_{ub}|$. These predictions were then fit to experimental data in a χ^2 minimisation procedure to determine the value of $|V_{ub}|$ consistent with the best agreement between the predictions and real data, as demonstrated in Ref. [37], [38] and [32]. In each case, the LQCD predictions were also used as constraints in the fits to avoid returning results inconsistent with the LQCD parameters within uncertainties. In the BaBar result detailed in Ref. [36], both the calculation and fitting based methods were used for the extraction, with two values of $|V_{ub}|$ quoted. The χ^2 minimisation was the chosen method for the current analysis, with more detail on the fitting procedure provided in Chapter 7.

Finally, an earlier iteration of the analysis presented in this thesis is also included in Table 1.1 as a preliminary Belle II result [41]. Using a hadronic tagged approach on a smaller subset of Belle II data corresponding to an integrated luminosity of 189.3 fb^{-1}, the branching fractions of the decays $B^0 \to \pi^- e^+ \nu_e$ and $B^+ \to \pi^0 e^+ \nu_e$ were determined, together with their respective differential decay rates, and a preliminary value of $|V_{ub}|$ was extracted. As will be discussed in the coming chapters, the current iteration of the analysis was performed on a data-set roughly two times the size used for this preliminary result, with branching fractions extracted also from the muon channels, $B^0 \to \pi^- \mu^+ \nu_\mu$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$.

With the importance of extracting $|V_{ub}|$ established and its exclusive determination via the semi-leptonic $B \to \pi \ell \nu_\ell$ decay physically motivated, the following chapters will present the results of a hadronic tagged reconstruction of $B \to \pi \ell \nu_\ell$. After an introduction to the Belle II Experiment in the next chapter followed by a description of the Full Event Interpretation tagging algorithm and its calibration in Chapters 3 and 4, a detailed account of the methodology used for the analysis will be explored, culminating in the determination of the $B \to \pi \ell \nu_\ell$ branching fractions and an extraction of $|V_{ub}|$.

Chapter 2

The Belle II Experiment

The High Energy Accelerator Research Organisation, abbreviated as KEK, was established in 1997 on the site of a particle physics lab located in Tsukuba, part of the Ibaraki prefecture in Japan, with the purpose of providing particle accelerators for study in fields such as high energy physics. Since its establishment, the highly successful Belle Experiment was conducted, in which an asymmetric electron-positron collider KEKB was operated during the years 1998 - 2010, and the Belle detector constructed to detect the products of these collisions. The experiment ceased data-taking in 2010 to make way for major upgrades to both the accelerator, now known as SuperKEKB, and the detector, now referred to as Belle II.

Belle II officially commenced data-taking with the first electron-positron collisions in March of 2019 [44], and temporarily ceased operation for its first long shutdown (LS1) in June 2022, thus far recording a total of 428 fb^{-1} of data. Physics operation is scheduled to resume in 2024.

2.1 The Belle Experiment

The KEKB accelerator used in the Belle experiment consisted of two nearly circular rings of circumference 3 km, one used to accelerate electrons to energies of 8.0 GeV, and the other responsible for accelerating positrons to energies of 3.5 GeV [45]. The beams would then collide in the center of the Belle detector, with decay products propagating outwards throughout the various detector layers. The beam energies were designed such that at the time of collision, they would have a combined energy equivalent to the rest mass of an $\Upsilon(4S)$ meson, 10.58 GeV, when calculated in its centre-of-mass frame. That is, these electron-positron collisions would result in the production of an $\Upsilon(4S)$, a meson consisting of a bottom b quark and $\bar{b}$ antiquark, which would be produced approximately at rest in its centre of mass frame, but boosted along the electron flight direction as seen in the laboratory frame of reference. The $\Upsilon(4S)$ decays almost instantaneously to a pair of B-mesons 96% of the time, with $\sim 51\%$ of these being decays to charged B-mesons, B^+B^-, and $\sim 49\%$ to neutral B-mesons, $B^0\bar{B}^0$ [20]. As the B-mesons have a very short lifetime of the order of 10^{-12} s [20], the observed boost in the laboratory frame was extremely useful in separating the decay vertices of the two decaying B-mesons, which recoiled against each other [46]. Despite the KEKB design

energies of the electron and positron beams, collisions did not always result in the production of the $\Upsilon(4S)$ $b\bar{b}$ bound state, and events in which $e^+e^- \to c\bar{c}$, $u\bar{u}$, $d\bar{d}$ or $s\bar{s}$ were also frequently occurring. These events, termed the continuum background, constituted a significant proportion of the backgrounds associated with the B-meson studies performed at Belle. Figure 2.1 illustrates the cross-section of the continuum background processes as a function of the center-of-mass energy of the colliding e^+e^- beams. Several Υ resonances, corresponding to different excitations of the $b\bar{b}$ bound state, manifest as peaks above the underlying continuum for certain values of the collision energy.
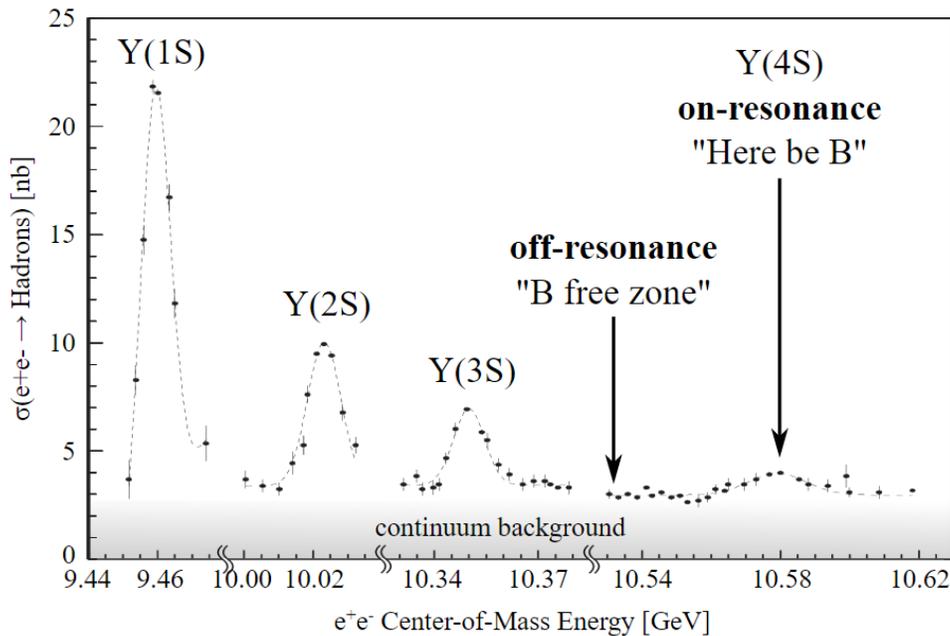
Figure 2.1: Cross-sections of processes involving the production of hadrons from e^+e^- collisions, as a function of the center-of-mass energy. Various Υ resonances ($e^+e^- \to b\bar{b}$) are shown as peaks above the continuum background ($e^+e^- \to c\bar{c}, u\bar{u}, s\bar{s}$ or $d\bar{d}$) [47].

The primary purpose of the Belle experiment was to search for CP-violation in neutral B-meson systems, after it was first experimentally observed in neutral kaons in 1964 [27]. This goal was realised in 2001 when both the Belle and BaBar collaborations independently reported results indicating the existence of CP-violation in neutral B-mesons [11, 12]. Alongside the success of this discovery, the data collected by Belle during its operation has been extremely valuable in updating various Standard Model parameters including the CKM-matrix elements, together with the branching fractions of a large number of B-meson decay modes. Many B-meson decays were also observed for the first time in Belle data, including the leptonic decay $B^+ \to \tau^+\nu_\tau$ introduced in Section 1.2.2 [48].

2.2 SuperKEKB

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

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

The Belle II Experiment aims to achieve a 40-fold increase in the maximum luminosity reached by the Belle Experiment. In order to accomplish this, the KEKB accelerator has had to undergo necessary upgrades to form the accelerator now referred to as SuperKEKB. One of the most significant changes to the technical design is the implementation of the "Nano-Beam" scheme, which aims to reduce the overlap of the electron and positron beams at their interaction point [49]. Figure 2.2 is a representation of the beam collisions, with d representing the size of the region of overlap. Minimising the value of d results in the squeezing of the vertical beta functions (β_y^*) of each beam at the interaction point, which has a significant effect on the resultant luminosity L given the following relationship,

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

Here, the $\pm$ suffix refers to each beam, electron (-) and positron (+), and $\gamma_\pm$ is the Lorentz factor, e the elementary electric charge, and r_e the classical electron radius. The luminosity depends also on the beam currents $I_\pm$ and the vertical beam-beam parameters $\xi_{y\pm}$. Finally, R_L and $R_{\xi y}$ refer to reduction factors for the luminosity and vertical beam-beam parameters respectively, and these are determined using quantities including the crossing angle ϕ, with their ratio varying only slightly from unity. The Nano-Beam scheme boasts a reduction in the vertical beta functions $(\beta_{y\pm}^*)$ by a factor of 20 when compared to those present at KEKB. Combined with a doubling of the beam currents $I_\pm$, and assuming a similar vertical beam-beam parameter $\xi_{y\pm}$, the design luminosity of the SuperKEKB accelerator is $8\times10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, 40 times larger than the luminosity achieved by KEKB [49]. Working towards this target, the highest luminosity achieved by SuperKEKB so far is $4.7\times10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, a new world record luminosity more than double that of KEKB. A comparison of the machine parameters achieved by Belle and the design values for Belle II are presented in Table 2.1.

However, such a substantial gain in the luminosity of SuperKEKB comes at the expense of a significant increase in the number of background processes arising from interactions within the individual beams and between the beams and foreign particles such as the surrounding residual gas in the evacuated beam pipe. These effects are collectively referred to as beam background, and encompass the following processes [49]:

- Synchrotron radiation: The electrons and positrons moving at relativistic speeds within the beams emit photons as they are accelerated radially. These photons can produce false hits in the detector and contribute quite extensively to the observed backgrounds.
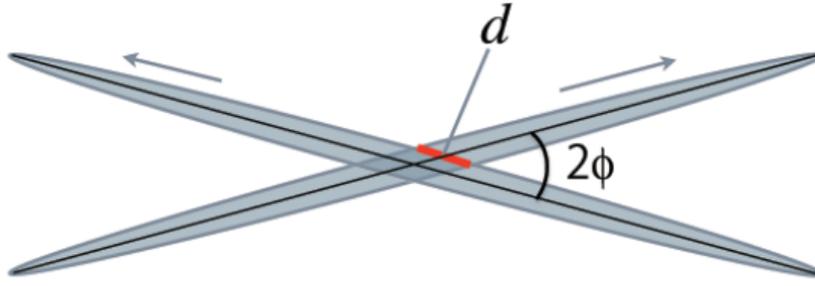
19

Figure 2.2: A depiction of the interaction between an electron and positron beam at Su-perKEKB. The grey areas represent the bunches of colliding particles, with the size of the beam overlap d and the crossing angle ϕ minimised with the Nano-Beam design [49].

- Beam-gas scattering: Whilst the beams are contained within vacuum chambers, in-evitably a small amount of gas particles remain with which the electrons and positrons can interact. These interactions include both Coulomb scattering, referring to elas-tic collisions that occur between charged particles, and bremsstrahlung scattering, whereby one of the electrons or positrons loses kinetic energy via the emission of a photon after hitting a gas particle. In each case, when a particle within a beam is scattered, its momentum is altered leading it to collide with the chamber walls or surrounding magnets, producing a shower of particles that contribute to the beam background.

- Touschek scattering: Particles within the beams can also scatter off one another, mod-ifying their momentum and leading to shower particles when the electrons or positrons hit the surrounding chamber walls. The rate of Touschek scattering is inversely pro-portional to the beam size, and thus the background effects from Touschek scattering can be quite severe with the greatly reduced beam sizes at Belle II.

- Radiative Bhabha scattering: The collision of the electron and positron beams does not always result in their annihilation, but rather the electrons and positrons can also scatter off of one another in a process known as Bhabha scattering. Bremsstrahlung emission can also occur whereby one of the scattered particles releases a photon, a phenomenon known as radiative Bhabha, with $e^+e^- \rightarrow e^+e^-\gamma$. The resultant photons can travel along the beam-pipe and interact with the iron detector magnets, producing significant amounts of background neutrons. Being neutral hadrons, these neutrons form a major background for the detection of K_L^0 mesons in the K-Long and Muon (KLM) Belle II sub-detector, introduced in Section 2.3.6.

- Pair production: Via the electromagnetic interaction, an electron-positron pair can be created from a single photon in a process known as pair production. Similarly, the collision of an electron and positron can result in their annihilation to a pair of photons, as per the same fundamental interaction. The combination of these two effects forms an additional beam background termed pair production via the two-photon process,

whereby colliding electrons and positrons annihilate to form two photons, both of which produce an electron and positron pair, $e^+e^- \rightarrow e^+e^-e^+e^-$. These background electrons and positrons enter the detector and contribute particularly to the background levels of the first sub-detector component, the Pixel Detector (PXD).

An additional change to the accelerator design for the Belle II Experiment that should be noted is a shift in the energies of the electron and positron beams. These have been changed from the Belle values, 8.0 GeV and 3.5 GeV, to 7.0 GeV and 4.0 GeV, for the electron and positron beams respectively. The new energy scheme still results in a centre-of-mass energy at the $\Upsilon(4S)$ resonance, but with a decreased boost in the laboratory frame along the electron flight direction. The separation between the decay vertices of both B-mesons is thus also reduced, but this effect is largely counteracted by the smaller beam sizes and consequently the narrower beam-pipe [49].

Machine Parameters	KEKB achieved	SuperKEKB target	SuperKEKB pre-LS1
Energy (GeV) (e^+/e^-)	3.5/8.0	4.0/7.0	4.0/7.0
ξ_y	0.129/0.090	0.090/0.088	0.041/0.028
β_y^* (mm)	5.9/5.9	0.27/0.41	1.0/1.0
I (A)	1.6/1.2	3.6/2.6	1.5/1.1
Luminosity (10^{34}cm^{-2}s^{-1})	2.1	80	4.7

Table 2.1: A comparison of the key machine parameters achieved by the KEKB accelerator, predicted for the SuperKEKB accelerator at its target luminosity, and achieved by SuperKEKB prior to the current shutdown. Where relevant, values for the positron and electron beams are quoted as e^+/e^- [49].

2.3   The Belle II Detector

The Belle II detector consists of a number of individual sub-detector layers, each with defined roles relating to the measurement of kinematic quantities and particle detection, identification and discrimination. The following sections present an overview of the structure and functionality of each detector component, beginning with those closest to the electron-positron beam line and continuing through to the outer regions of the detector. A schematic of the Belle II detector illustrating the individual layers is displayed in Figure 2.3.

2.3.1   Pixel Detector (PXD)

The Pixel Detector is the innermost component situated close to the beam-pipe housing the incoming electron and positron beams. Due to the Nano-Beam scheme chosen for Belle II, the radius of the beam-pipe in the interaction region of the beams is only approximately 10 mm, with the pixel detector designed to consist of two layers of sensors positioned at radii of 14 mm and 22 mm [49]. The PXD was only partially installed for Belle II data-taking prior to LS1, excluding the majority of the outermost layer [50]. A significant aim of the
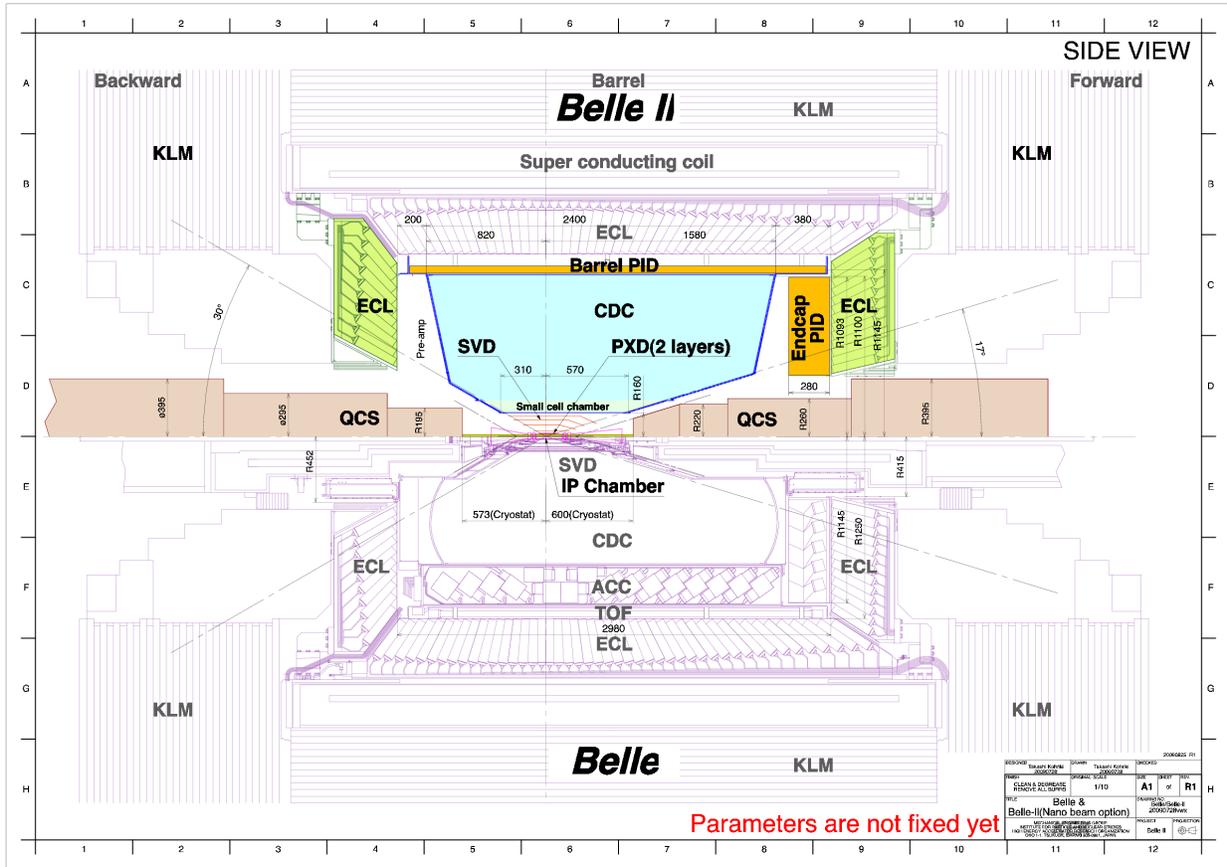
Figure 2.3: A side view of the Belle II detector (top), compared with its predecessor Belle (bottom). The various sub-detectors are labeled [51].

LS1 period has been to upgrade the PXD to the full two-layer installation in preparation for the next run period.

The PXD together with the Silicon Vertex Detector (SVD) and Central Drift Chamber (CDC) introduced in the following sections form the region of the detector responsible for tracking the trajectories of charged particles as they curve within the supplied solenoidal 1.5 T magnetic field, allowing for the determination of their momentum. Another main goal of these detector regions is to use this information for the precise determination of the decay vertices of particles, including those of the parent B-mesons.

The sensors used within the PXD are comprised of DEPleted Field Effect Transistor (DEPFET) pixels, with 8 million pixels in total divided between the two layers. The DEPFET technology is semi-conductor based, with hits recorded by collecting electrons that have been freed by the incidence of charged particles colliding with the transistors. The nature of the DEPFET structure is such that the sensors consume very little power, requiring only air cooling, and can be built to be quite thin, with a thickness of only 50 microns.
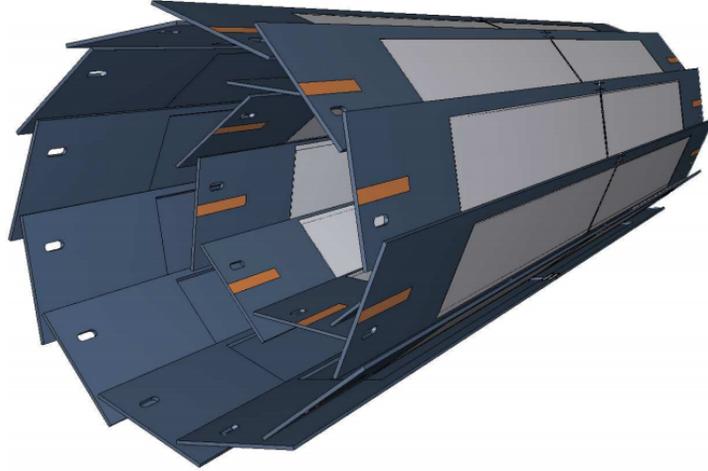
Figure 2.4: A design schematic of the PXD, corresponding to a total length of 174 mm and diameters of 22 mm and 44 mm for the inner and outer pixel layers, respectively [49].

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

2.3.2 Silicon Vertex Detector (SVD)

The Silicon Vertex Detector, as mentioned above, is an additional detector component designed for charged particle tracking and decay vertexing. It consists of four layers of strip sensors made of silicon, with the innermost layer positioned 16 mm from the outer layer of the PXD [49]. Given the high luminosity and expected levels of beam background, this is the safest distance to introduce strip sensors. It is not feasible to use pixel layers throughout the entire inner region as both the readout from the large number of channels and the estimated cost would be too large. Complementing the PXD, the SVD also allows for the measurement of decay vertices displaced further from the interaction point.

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

Figure 2.5: The Belle II SVD, containing four layers of silicon strip sensors with a distance of 20.4 cm separating the inner and outermost layers [49].

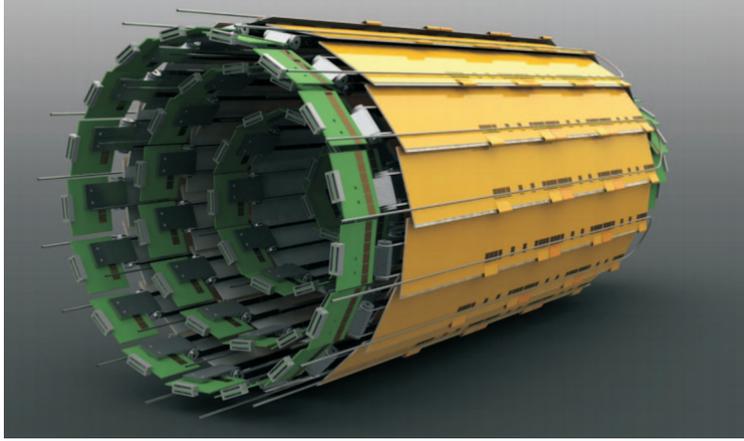### 2.3.3   Central Drift Chamber (CDC)

The Central Drift Chamber, located at a radial distance of 160 mm, 20 mm out from the fourth layer of the SVD, is a large chamber designed for a number of key purposes. It reinforces the tracking capabilities of the PXD and SVD and forms a crucial component in the reconstruction of charged particle tracks and the precise determination of track momenta. However, it holds the additional benefit of also being able to provide information on the identity of the charged tracks passing through the chamber. The CDC extends outwards to a radial distance of 1130 mm around the beam-pipe, making it the largest of the inner detector layers responsible for charged particle tracking [49].

The Belle II CDC is modeled closely on the CDC incorporated within Belle, which was highly successful in its performance. It consists of a total of 14336 drift cells arranged in layers, an improvement on the 8400 employed at Belle [45], with each drift cell filled with a mixture of helium and ethane gases in equal proportions. Charged tracks ionise the gases as they pass through the drift cells, releasing electrons that then drift towards sensor wires that register and amplify the signals. The measurement of the drift time of the electrons together with the location of the sensor wire hits can be used for the reconstruction of charged particle tracks. Additionally, the level of kinetic energy lost by each track as it ionises the gas within a cell can be measured. It is here that the CDC is able to contribute to particle identification, with different charged particles possessing characteristic signatures in their distributions of $\frac{dE}{dx}$, the energy loss as a function of distance. In this way, tracks with particularly low momentum that are unable to reach the additional detector components designed for particle identification (described in the upcoming section) can often be identified using the CDC alone. The arrangement of wires within the CDC is illustrated in Figure 2.6, with a higher density observed in the inner layers to accommodate the increased levels of beam background.
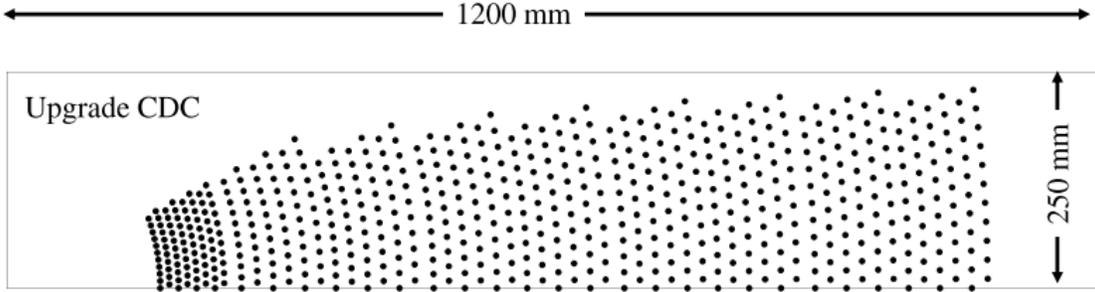
Figure 2.6: The sensor wire configuration of the Belle II CDC [49].

2.3.4   Particle Identification Components

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

The beam-pipe housing the two incoming electron and positron beams must pass through the detector, and thus when referring to Belle II, it is common to divide the detector into two sections, namely the barrel region and the endcaps. The barrel refers to the region of the detector located at radial distances from the beam-pipe, with the two endcaps referring to the left- and right-most detector walls surrounding the beam-pipe at each end. Additionally, due to the beam asymmetry and the consequent lab-frame boost, the two endcaps are termed forward and backward, with the forward endcap defined as the one facing the incoming electron beam and resultant boost. The structure of the barrel and endcap regions can be clearly distinguished in Figure 2.3.
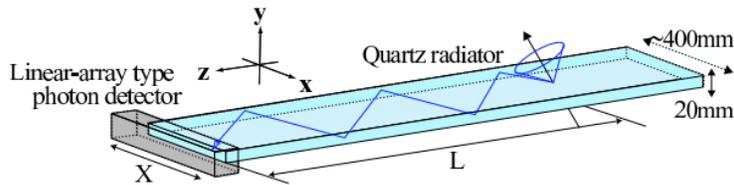
Figure 2.7: An illustration of the Belle II TOP counter depicting the total internal reflection of a Cherenkov photon through the quartz radiator [49].

The TOP counter lies in the barrel region of the detector, and consists of 16 2.7 m long modules distributed azimuthally around the beam-line [52]. Each module is comprised of a length of quartz acting as a radiator, with a spherical mirror attached to one end, and a prism wedge at the other. When a charged particle passes through the quartz bars with a velocity higher than the associated speed of light for the medium, Cherenkov radiation is produced, with the resulting photons traversing the length of the bar via total internal reflection, arriving at a series of photo-multiplier tubes attached to the prism. A 3-dimensional

Cherenkov image can then be simulated using the spatial coordinates and time of arrival of the Cherenkov photons, and charged particles can be discriminated based on these generated images. For Belle II, the TOP counter has replaced the equivalent Belle detector component, the Time-Of-Flight (TOF) detector, with the TOP designed particularly to improve the kaon-pion (K/π) separation capabilities.
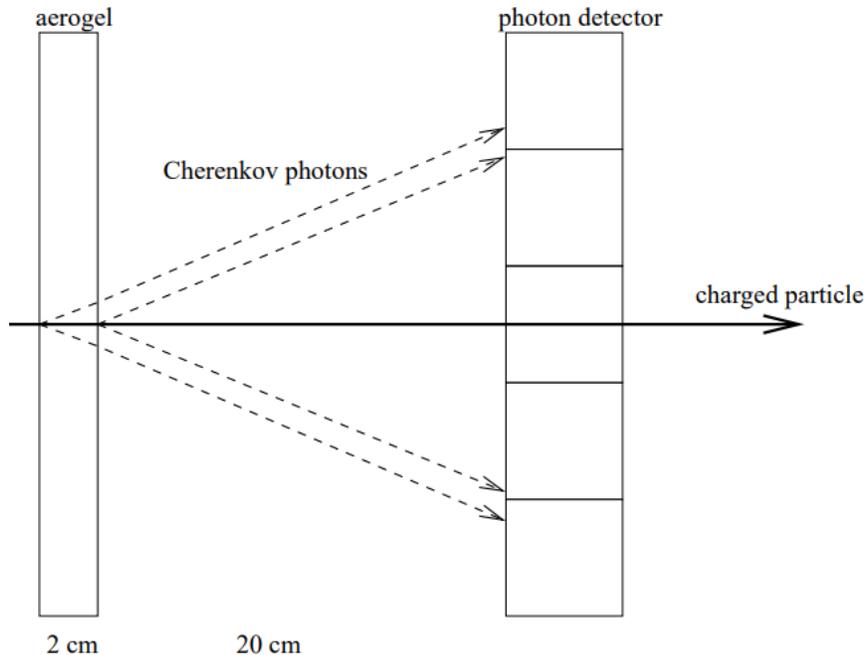
Figure 2.8: A representation of the Cherenkov ring-imaging technique employed by the ARICH detector [49].

Conversely, the ARICH detector is located at the forward endcap, and whilst differing in structure to the TOP counter, it also relies on Cherenkov radiation imaging for charged particle identification. It spans a 3.5 m^2 area of the endcap in two layers, with a radiator layer of aerogel tiles separated from a plane of photon detectors by a distance of 200 mm [53]. Planar mirrors surround the edges of the entire system to redirect the Cherenkov photons produced in the aerogel to the photon detector region. The large volume between the two layers allows the trajectories of the emerging photons to diverge enough that the incident photons form a ring on the detector plane, resulting in the Cherenkov image used for particle discrimination. The ARICH has similarly replaced the Aerogel Cherenkov Counter (ACC) configuration used at Belle, with improved kaon-pion separation over a large momentum range, and reasonable separation between pions, muons and electrons at momenta under 1 GeV [49].

2.3.5 Electromagnetic Calorimeter (ECL)

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

As for the structure of the ECL, a total of 8736 scintillator crystals cover a polar angle range of $\sim 12° < \theta < 155°$, with 6624 and 2112 divided between the barrel and endcaps, respectively [49]. When ionising particles enter the scintillator crystals, their energy is absorbed and re-emitted in the form of photons, which are then detected via photo-diodes glued to each crystal. The same crystals have been inherited from Belle for the Belle II ECL, due to only minimal damage received during the run-time of KEKB. At present, the crystals are made of Thallium-activated Cesium Iodide, CsI(Tl), a scintillating material with a considerably large light output. However, discussion and research are currently ongoing to upgrade the crystals to pure Cesium Iodide (CsI), at least in the end-cap region [50]. Although these produce a considerably lower light output, the decay time between the energy deposits and the emittance of photons is substantially shorter. The ECL electronics have also been significantly upgraded since Belle, with major improvements to the readout time achieved.

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

Finally, the K_L^0 and μ detector, the outermost layer of the detector similarly sectioned into the barrel and endcap regions, is designed for the detection and identification of muons and neutral K_L^0 mesons. It is located outside of the super-conducting solenoid used to provide the 1.5 T magnetic field to the detector, which occupies the space between the ECL and the KLM, as can be seen in Figure 2.3. The angular coverage of the KLM is comparable to that of the ECL with a $20° < \theta < 155°$ range in polar angle, and polyethylene sheets shield the KLM from multiple backgrounds including neutrons produced from beam background processes. Resembling closely the successful Belle KLM, detector elements known as resistive plate chambers (RPCs) are wedged between 4.7 cm thick iron plates in the barrel. When hadrons deposit energy into the ECL or collide with the iron plates, they interact strongly such that a characteristic shower of secondary hadrons is produced. These traverse the iron plates and are detected by the RPCs via the charged hadrons present in the shower. Each RPC consists of two parallel electrode sheets made from an insulating substance known as float glass across which a high voltage is applied. Charged particles can ionise the gas between the electrodes, with electrons and ions accelerated to the anode and cathode, respectively. At the end-caps, however, RPCs are no longer optimal when contending with

the higher luminosity beam background effects, and polystyrene scintillator strips are used in lieu of the RPCs within both the forward and backward endcaps [54].

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

The nature of the momentum of the muons produced at Belle and Belle II is such that the muons are most often low ionising, and can traverse the ECL and even the iron plates with minimal energy loss. As such, muons produce non-shower hits in the KLM that can be matched to tracks in the CDC [55]. This provides a useful feature for the discrimination between muons and charged hadrons, particularly pions, that do generate hadronic showers within the KLM.

2.3.7 Triggering and Data Acquisition

As explained, the sub-detectors described above are designed for particle tracking and identification based on certain signatures or hits occurring within the component geometry. This functionality is achieved through an efficient data acquisition system including a number of triggers associated with certain criteria by which the various physics processes occurring within the Belle II environment, including beam background, continuum and $\Upsilon(4S)$ events can be detected. The triggering system developed for Belle II was similarly modeled after the Belle system with advancements in the technology used, and consists of five sub-triggers activated by hits in the various detector components. These sub-trigger signals are then passed to the Global Decision Logic (GDL), which combines the information received from each sub-trigger system to make a final decision regarding the nature of the detected event, as per a number of secondary criteria [56]. This triggering system is depicted in Figure 2.9, with the sub-triggers shown associated with the CDC tracking, ECL and KLM clustering and the particle identification capabilities of the ECL (EPID) and TOP and ARICH detectors located in the barrel (BPID).

2.4 The Belle II Analysis Software Framework (basf2)

The Belle II Analysis Software Framework, abbreviated to basf2, is the term used to refer to the entire suite of software utilised throughout the Belle II experiment [57]. This extends not only to software used for physics analysis, but software related to the simulation and monitoring of the detector. The framework is primarily programmed using C++, but Python is the main language chosen for writing scripts that get executed by the framework. This
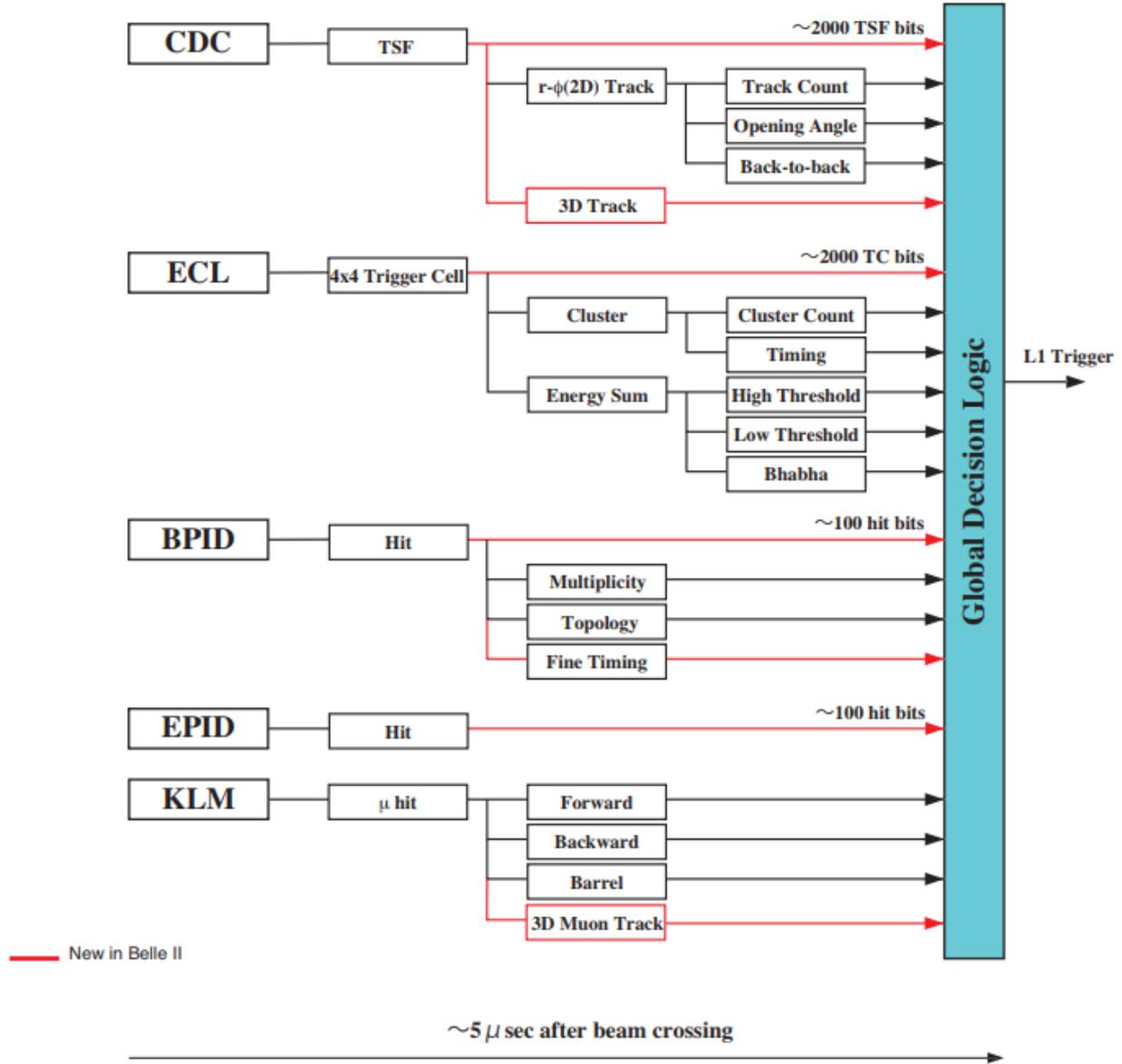
Figure 2.9: The structure of the Belle II triggering system, with the quoted information from each sub-trigger passed to the Global Decision Logic (GDL) [49].

is a significant shift from the Belle Analysis Software Framework (BASF) used in the Belle Experiment, which was programmed using a mixture of C, C++ and Fortran, with C++ being the primary language used for writing analysis scripts [58]. Basf2 also incorporates multiple independent libraries into the framework, with the most notable one being ROOT [59], a scientific software framework used for physics analysis developed at the European Organization for Nuclear Research, CERN, in Geneva. Other relevant external packages include EvtGen [60], used for the simulation of particle decays, particularly those of B and D-mesons, and Geant4 [61], responsible for simulating the interaction and consequent detection of particles by the various sub-detectors.

2.4.1 Framework Structure

The processing of data within basf2 relies on the use of modules, which are linearly arranged in what is termed a path [57]. The modules are self-contained processing blocks responsible for particular tasks which span a vast range of applications, such as reading information from and writing information to files, simulating particle decay and detection, and multivariate classification. When processing a path, the individual modules are read and implemented in order.
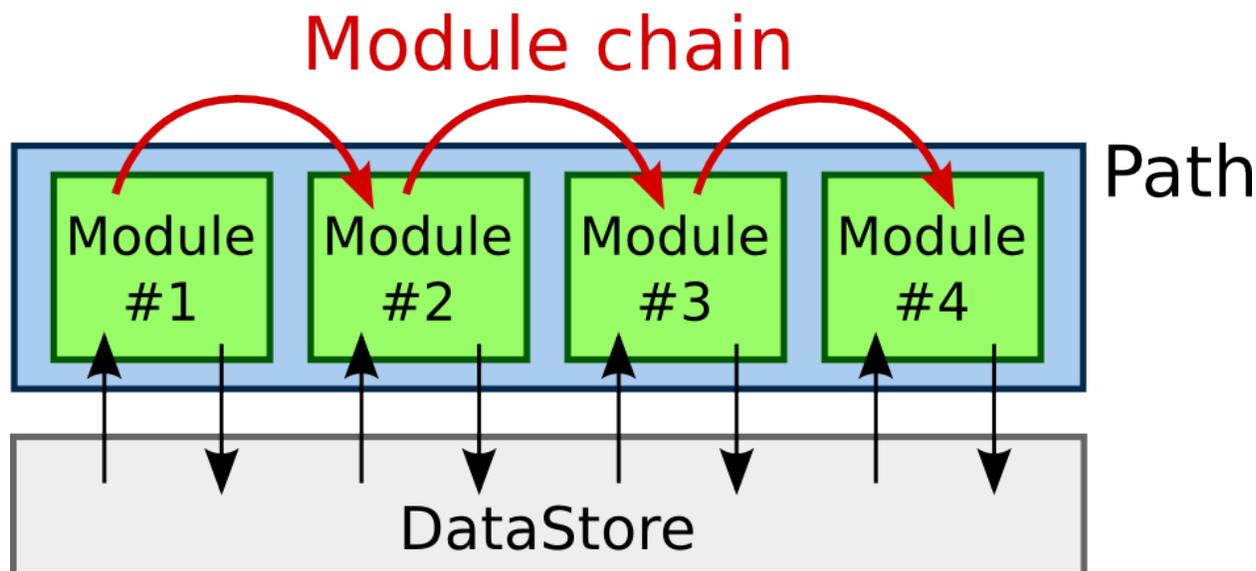
Figure 2.10: The basf2 data processing chain. Modules are arranged and executed linearly in a path, exchanging information using the DataStore [57].

Upon executing a path with multiple modules, data often needs to be shared between them. A primary example would be an analysis script that generates a number of $\Upsilon(4S)$ particles, and simulates their decay into subsequent particles which propagate through and interact with the detector. Such a script is known as a steering file, written in Python, and first creates and then executes a path containing the relevant modules. In this example, data including the list of particles involved in the $\Upsilon(4S)$ decays, ECL and KLM cluster information, and particle track information needs to be accessed by multiple modules. This is implemented within basf2 using the DataStore, which is a common storage available within the framework. Relevant data is stored in the DataStore as data-objects, with the aforementioned examples being `ParticleLists`, `ECLClusters`, `KLMClusters` and `Tracks`. All modules are able to read from and write to the DataStore, forming an efficient method of data processing. A schematic of the basf2 data processing chain is depicted in Figure 2.10.

2.4.2 Monte Carlo (MC)

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

1. Event Generation: Using the EvtGen package [60], full particle decay events are simulated. Users can select the required number of decay events, and can choose to specify any desired decay modes and their amplitudes, or default to current theoretical and experimental values. The particle mother may decay directly to the final particles in the chain, or to intermediate particles that further decay to produce these final-state particles. This step is entirely detector-independent, and is referred to as generator-level MC.

2. Event Simulation: The generated particles are then propagated throughout the detector. Final-state particles interact with the various sub-detectors including the ECL and KLM, the simulation of which is performed using the Geant4 package [61]. This results in simulated clusters in the ECL and KLM, as well as simulated particle tracks that can be matched to these clusters.

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

Unlike real data, MC provides the unique opportunity to access not only the result of the reconstruction from detector information, but also the input generated particles. As such, the generator-level MC is often referred to as the MC truth. basf2 implements a process by which reconstructed-level and generator-level information can be compared, or matched, in order to determine whether a certain particle has been reconstructed correctly. This process is invaluable in providing estimates of the quality of reconstruction that can be extrapolated to real data, for which no truth information is known.

2.4.3 Multivariate Classification

Multivariate classification is an extremely valuable tool used in a variety of different applications, with language recognition technology, internet search engines and email spam filters being some of a multitude of examples. It refers to the classification of data elements into several given categories using a number of input variables, and is often performed via the use of machine learning techniques. In these cases, in order for the classification to be meaningful, these techniques must be trained on large samples of data that have been categorised correctly.

Multivariate classification can be performed in several ways, including categorical methods (sorting data into independent categories such as languages) and ordinal methods (sorting data into ordered categories such as whether a duration is 'short', 'medium', or 'long'). However, most relevant for physics analyses, and the classification method that is described in detail below, is a numerical-based method. A typical use for multivariate classification in high energy physics analyses such as those performed at Belle II is the classification of particles or decay events as being either signal or background. When studying a particle decay of interest in particle collider experiments, these decay events, termed signal events, are produced alongside a multitude of other particle decays forming various types of backgrounds. Multivariate classification methods, after being trained on large samples of Monte Carlo containing a mixture of known signal and background events, aim to classify unseen data into one of these two categories. The Full Event Interpretation algorithm discussed at length in the coming chapter is a key example of the use of multivariate classification for this purpose [13].

In practice, the numerical method of multivariate classification employed in the techniques discussed throughout this thesis involves defining a multi-dimensional vector $\mathbf{x}$ for each reconstructed particle candidate involved in a decay [62]. The elements of the vector, termed features, can include variables such as particle momentum and decay vertex information among many other quantities. This multi-dimensional feature vector $\mathbf{x}$ can then be mapped onto a one-dimensional number known as a test-statistic T. Upon training the classification on large numbers of simulated Monte Carlo containing both signal and background decay events, the distribution of the test-statistic provides a meaningful way in which to discriminate between the two categories.

Figure 2.11 shows an example of a test-statistic distribution for a particular feature vector $\mathbf{x}$, in which two distinct peaks are observed. Given a particular null hypothesis H_0, such as that corresponding to the default assumption that $\mathbf{x}$ is signal, an alternate hypothesis H_1 exists capable of rejecting the null hypothesis, in this case referring to the determination of $\mathbf{x}$ as background. These conditions can be associated with probability density functions (PDFs) $T(\mathbf{x}|H_0)$ and $T(\mathbf{x}|H_1)$ for signal and background respectively, where

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

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

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

After training, a critical value of the test-statistic T_c can be chosen above which particle candidates can be considered signal. This naturally results in errors whereby candidates are misidentified. As seen in Figure 2.11,

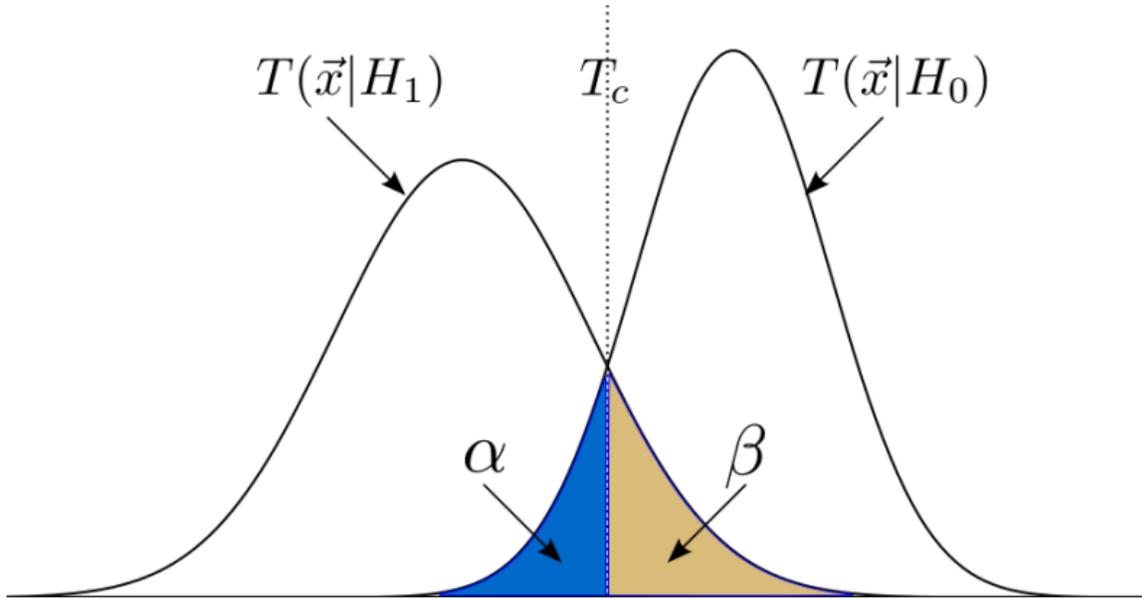$$\alpha = \int_{-\infty}^{T_c} T(\mathbf{x}|H_0)$$

Figure 2.11: An example distribution of the test-statistic T for signal $T(\mathbf{x}|H_0)$ and background $T(\mathbf{x}|H_1)$ PDFs. Above a critical value T_c, all candidates are considered signal. α and β represent the errors associated with candidates misidentified as background and signal, respectively [62].

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

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

represents the probability of errors in which background candidates are misidentified as signal. A number of multivariate classification techniques are supported by basf2.

Artificial Neural Networks

Artificial neural networks are computing applications that consist of multiple different interconnected nodes, or neurons, present in different layers. Each neuron in a layer is connected to each neuron in the adjacent layer, forming a complex network resembling the structure of the biological brain, after which the technique is named. A signal can be transferred along the connections between neurons, with each connection given an associated weight, w, that represents the strength of the connection. These weights are modified throughout the training process in what is known as machine learning. An example of the structure of an artificial neural network is depicted in Figure 2.12. In practice, the neurons of the first layer correspond to the elements of the input feature vector described above, with the final layer being synonymous with the output test-statistic. The NeuroBayes package [63] was the chosen software for artificial neural network analysis in the Belle Experiment. Basf2, how-

ever, now employs the use of the external ROOT-based Toolkit for Multivariate Analysis, or TMVA package [64], which is tailored for high energy physics applications and provides multiple different techniques for multivariate analysis.
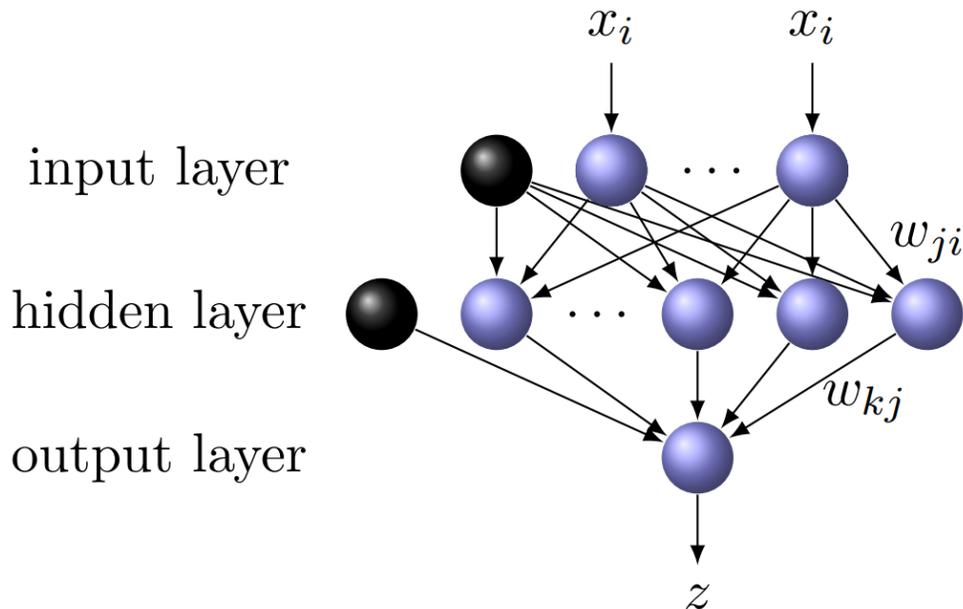
Figure 2.12: A schematic of an artificial neural network, with neurons interconnected between layers [65]. The subscripts i, j and k refer to the indices of successive layers.

Boosted Decision Trees (BDTs)

An alternative method for the execution of multi-variate analysis available in basf2 involves the use of decision tree classification. In this method, input features are evaluated against particular criteria by which they are then split into categories. This classification then propagates down a structure referred to as a decision tree, whereby categories at each node are further split based on relevant variables, and a multi-variate classifier is trained. The typical structure of a decision tree is illustrated in Figure 2.13. Individual decision trees, however, are particularly sensitive to over-training [62], and an extended method known as boosted decision tree learning exists with the purpose of counteracting this effect. In this technique, multiple decision trees are used, each with a minimal number of nodes to prevent over-training. Each tree is assigned a weight, w_i, and the final classifier output is obtained through the weighted sum of the classifiers resulting from these weaker trees in order to produce a stronger classifier output.

The `FastBDT` software [66] is the primary BDT-learning technique incorporated within basf2, and is more specifically an example of a stochastic gradient-boosted decision tree method. The gradient boost refers to one of multiple potential algorithms existing for the

determination of the weights described in detail here [67], with the stochastic descriptor referring to the fact that only a random subset of the input statistics are chosen for the classifier training. This in turn allows the resultant classifiers to better discriminate between signal and background when applied to unseen data [62]. The Full Event Interpretation, being the focus of this study and described in detail in the following chapter, utilises `FastBDT` as its machine learning algorithm of choice.
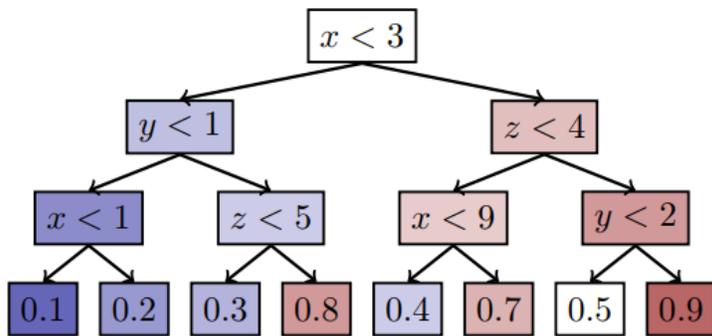
Figure 2.13: The structure of a single decision tree, depicting the categorisation of data at each node and the resultant trained classifiers [66]. The variables x, y and z are generic variables representing key selection criteria that can be used to discriminate between signal- and background-like events, such as particle momenta and identification information reconstructed from data collected by the various Belle II sub-detectors.

Chapter 3

The Full Event Interpretation

In the context of electron-positron collider experiments such as Belle and Belle II, the reconstruction of full $\Upsilon(4S)$ events containing B-meson decay modes of interest is achieved through two main reconstruction methods, namely tagged and untagged analyses. In such events, the decays of the daughter B-mesons can be precisely established, inferring the existence of the parent $\Upsilon(4S)$ meson with an initial state that can be deduced from the finely tuned energies of the incident electron and positron beams. Both reconstruction methods benefit from this knowledge, with kinematic information corresponding to the desired signal decay, or B_{sig}, able to be determined from 4-momentum conservation of the entire event. In the following sections, the two reconstruction methods will be explained in detail, followed by an in-depth description of the machine learning algorithm used for tagged analyses at Belle II, the Full Event Interpretation (FEI) [13].

3.1 Reconstruction Methods

3.1.1 Tagged Analysis

In tagged analyses involving the decay of an $\Upsilon(4S)$ to two B-meson daughters, in addition to the reconstruction of the desired B_{sig} decay mode, the second B-meson associated with the event, the B_{tag}, is also reconstructed using a number of exclusive decay channels. In this way, the reconstruction of the entire $\Upsilon(4S)$ decay chain is achieved. The B_{tag} decay modes can be either hadronic or semi-leptonic in nature, with examples of these presented in Figures 3.1(a) and 3.1(b), respectively, for a B_{sig} decay mode of particular relevance to this study, $B_{sig}^+ \to \pi^0 e^+ \nu_e$.

Hadronic and semi-leptonic tags differ significantly in their treatment of leptonic and semi-leptonic B_{sig} decay modes, due to the presence of neutrinos. As described in Chapter 1, neutrinos are the lone fundamental particles known to interact only via the weak force. As such, they possess a significantly low interaction cross-section, and are unable to be detected in collider experiments such as Belle II, subsequently carrying energy and momentum away from systems in which they are present in the decay products. For (semi-)leptonic B_{sig} decay modes, $\Upsilon(4S)$ events consisting of hadronic B_{tag} daughters therefore possess a single source of missing 4-momentum assigned to the signal side. Thus, taking into account the initial

state of the $\Upsilon(4S)$, the 4-momentum vector of the signal-side neutrino can be determined via 4-momentum conservation. This is not the case, however, for $\Upsilon(4S)$ decays to semi-leptonic B_{tag} daughters, in which neutrinos can be present on both the tag and signal sides of the event. Under such circumstances, the 4-momentum of the signal-side neutrino can no longer be exactly determined via 4-momentum conservation.
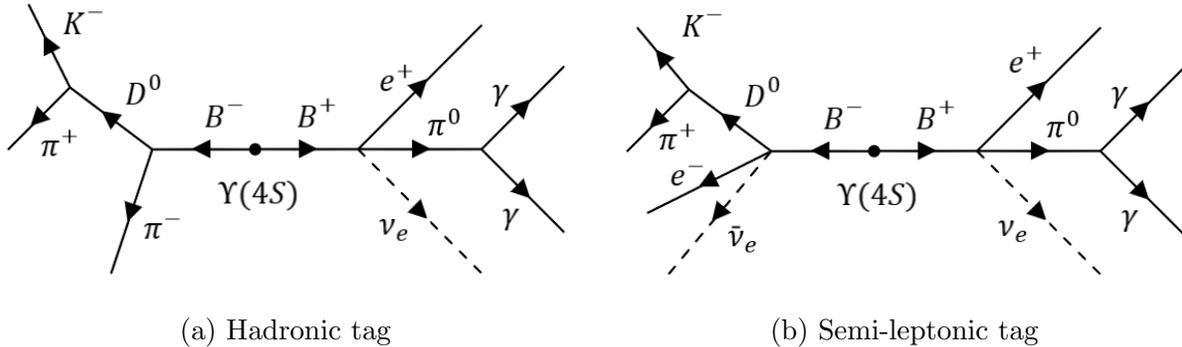

(a) Hadronic tag   (b) Semi-leptonic tag

Figure 3.1: Schematics giving examples of $\Upsilon(4S)$ events containing a B_{tag} and B_{sig} decaying as $B^+_{sig} \to \pi^0 e^+ \nu_e$, with $\pi^0 \to \gamma\gamma$. Dashed lines represent undetected neutrinos.

3.1.2 Untagged Analysis

In contrast to tagged analyses, untagged analyses require only the reconstruction of the B_{sig} daughter from an $\Upsilon(4S)$ decay, with all remaining particles in the event built from detector information assumed to have arisen from the second B-meson. Since no explicit B_{tag} is reconstructed, in the combination of the 4-momentum of these particles suitable selections must be implemented to retain only those events consistent with the nature of B-meson decays. Given that no neutrinos interact with the detector, it is likewise assumed that the only neutrinos present in an $\Upsilon(4S)$ event are associated with the B_{sig} decay. In this way, a value for the 4-momentum of the signal-side neutrino can be determined, though the precision of such a measurement is inferior to the hadronic tagged analysis able to reconstruct the entirety of the $\Upsilon(4S)$ decay chain excepting the signal neutrino. A representation of a typical untagged analysis in depicted in Figure 3.2, for the semi-leptonic signal decay $B^+_{sig} \to \pi^0 e^+ \nu_e$.

3.2 Overview of Algorithm

The studies performed and detailed throughout the remainder of this thesis take the tagged approach, utilising the hadronic tagging capability of the comprehensive FEI algorithm [13]. The FEI is the signature analysis tool available in basf2 for conducting tagged analyses and employs the use of multi-variate classification to effectively discriminate between signal and background decay events. It follows the same general approach as the equivalent technique that was used for the Belle Experiment, termed Full Reconstruction (FR) [32], but with
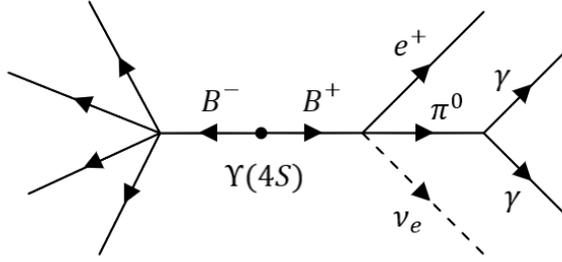
Figure 3.2: Schematic showing an untagged analysis of an $\Upsilon(4S)$ event with daughter B_{sig} decaying as $B_{sig}^{+} \to \pi^0 e^+ \nu_e$, with $\pi^0 \to \gamma\gamma$. No second B-meson is explicitly reconstructed.

several improvements and added features that are discussed in detail in the coming sections. As mentioned in Section 2.4.3, the FEI is a boosted decision tree-based algorithm developed using the `FastBDT` software [66].

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

The FEI tagging algorithm is trained in stages using a hierarchical approach that first uses detector information including tracks, ECL and KLM clusters to build final-state particle candidates [68]. A multi-variate classifier (MVC) is trained for each final-state particle using a number of relevant input variables detailed below, the result of which is summarised in a variable named the `SignalProbability`. This variable is related to the probability that a certain candidate is correctly reconstructed, and thus acts as a useful, tunable discriminant between signal and background events. The final-state particle candidates are then reconstructed into intermediate particle candidates within a variety of allowed decay channels, and a multi-variate classifier is trained for each individual channel that uses the final-state particle MVCs as input amongst other relevant variables. Finally, these intermediate candidates are used to build the mother B-mesons, with yet another MVC being trained for each reconstructed decay channel. In this way, the algorithm is trained stage by stage, with the `SignalProbability` of the candidates reconstructed at each level forming part of the input for the training of the MVCs performed in the following stages. This hierarchical approach is depicted in Figure 3.3.
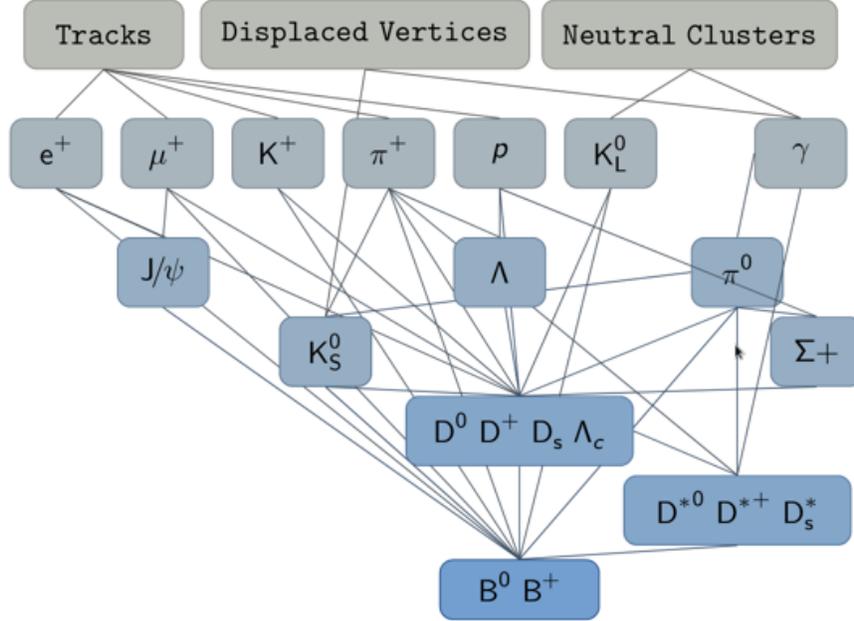
Figure 3.3: The hierarchical structure of the FEI tagging algorithm. Detector information including tracks and clusters is first used to reconstruct and train a classifier for each final-state particle. Intermediate particles are then reconstructed in stages, with the classifier output of each stage taken as input for the classifier training of the successive stage [13].

3.3 Input Variables Used for Classifier Training

The following variables are used as input into the multi-variate classifiers to determine the resulting `SignalProbability` of each candidate [69]:

3.3.1 Final-state Particles (FSPs)

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

- **Particle identification (ID) variables:** In basf2, a particle ID exists for each charged FSP which represents the likelihood ratio that a particular charged track is in fact the named particle as opposed to any other charged FSP. Each particle ID varies between 0 and 1, with higher values indicative of a greater likelihood that the charged track was created by the named FSP. The particle ID variables used in the multivariate classifier training are namely the `electronID`, `muonID`, `protonID`, and `kaonID`. A particle ID variable also exists for pions, `pionID`, but is not included in the training variables as any charged track with low values for the remaining four particle IDs can safely be assumed to be a pion. The deuteron is an additional charged final-state particle with an associated particle ID (`deuteronID`), but is not incorporated within the FEI.

- **Track momentum variables:** These include the magnitude of the total momentum p, as well as the individual transverse and z components of the momentum, p_t and p_z

respectively. These variables are useful for discriminating between tracks originating from different FSPs with characteristic momentum distributions.

- **Track impact-parameters:** The interaction point is defined as the point at which the electron-positron collision occurs. For any given track, the distance between the interaction point and the point closest to this on the track can be calculated. The radial, d_r, and z-components, d_z, of this distance are referred to as impact-parameters and they provide a way to discriminate well between particles produced at earlier stages in the decay chain, and those produced in subsequent decays with tracks originating further from the interaction point.

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

- **Particle rank:** When reconstructing particles at any point in a decay chain, there are often many candidates in the output that fit the reconstruction criteria, though only one of these candidates actually corresponds to the true particle. As such, in order to reduce combinatorics and unnecessary computing time, the FEI utilises a basf2 feature known as the best candidate selection, whereby a list of candidates can be ranked in ascending or descending order based on the values they possess for a particular variable. The FEI employs this technique for most particles to select only a subset of candidates most likely to be correct.

  For the charged final-state particles reconstructed by the FEI, an initial pre-classifier selection, also commonly referred to as a 'cut', is made on the impact-parameters of all charged tracks. A discussion on the pre- and post-classifier cuts made during the FEI is left to Section 3.4, with this simply noted here. All candidates surviving this selection are then ranked from highest to lowest according to the relevant particle ID for each FSP (`electronID`, `muonID`, `pionID`, `protonID` and `kaonID`), and only the best 20 candidates are kept for the pion, proton and kaon cases, and only 10 for the electron and muon cases. The rank of each candidate in the list is stored and this rank is also used as an input variable for training. Illustrating this with an example, for the training of the classifier of an electron, all tracks surviving the above pre-cut are then ranked in order of `electronID`, with the most electron-like candidate given the rank of first. Only the 10 most electron-like candidates are kept, and their associated rank in the list is taken into account by the classifier training.

Photons: γ

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

- **ECL cluster region:** The `clusterReg` variable simply returns which of the three regions of the ECL the cluster was found in, that is, the forward, backward and barrel

regions. This is valuable information as the geometry of the ECL crystals differs between these three regions.

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

- **ECL cluster shape:** Another significant aspect of cluster discrimination is the cluster shape, which is primarily described by the variable `E9E25`. This represents the ratio of the energy deposited in the inner 3×3 crystals of the cluster and the energy deposited in the outer 5×5 crystals. The value of `E9E25` can be used to distinguish between particles with different characteristic deposits in the calorimeter.

- **Energy and momentum:** The energy E and transverse and z-components of the momentum, p_t and p_z of the photon candidates are also passed to the classifier training.

- **Particle rank:** The appropriate pre-cut is applied to all photon candidates and they are then ranked from highest to lowest according to their energy. The best 40 candidates are kept and the ranks of each candidate are used as input for the MVC training.

K^0_L mesons

- **Energy:** The energy E of each K^0_L candidate is passed to the classifier training.

- **KLM cluster timing:** The time of detection of each cluster in the KLM is also used, analogous to the use of ECL cluster timing for training photon classifiers.

3.3.2 Intermediate Particles

π^0 mesons

- **Invariant mass:** The invariant or rest mass of π^0 candidates is an important input variable due to the fact that the π^0 mass is well-known and experimentally established to be 135 MeV, as recorded by the Particle Data Group (PDG) [20]. Candidates with an invariant mass close to this value are more likely to correspond to true π^0 mesons. Two variables are passed to the classifier training, namely, the invariant mass M itself, and the absolute value of the variable $|dM|$, which is defined as the difference between the mass of the candidate and the nominal mass as given by the PDG.

- **Angle between photon daughters:** π^0 mesons decay to a pair of photons approximately 99% of the time [20]. The angle between the flight directions of both daughter photons is a useful quantity containing kinematic information about this decay.

- **Daughter `SignalProbability`:** As the π^0 mesons are intermediate particles built from final-state neutral photons, the output from the photon classifier training, that is, the photon `SignalProbability`, is used as part of the input to the π^0 classifier. This reflects how the hierarchical structure of the FEI is accomplished in practice.

- **Energy and momentum:** The energy E and transverse and z-components of the momentum, p_t and p_z are also included for π^0 mesons.

- **Particle rank:** After the relevant pre-cut for π^0 mesons is applied, the candidates are then ranked from those with the lowest values of $|dM|$ to those with higher values. The best 20 candidates are then kept and the ranks of each are also passed to the classifier training.

K_S^0 mesons, Λ^0 and $\Sigma^\pm$ baryons

- **Distance variables:** For charged FSPs that produce tracks, the impact-parameters are measured between the interaction point and the closest point of the track, as mentioned above. In the case of intermediate particles that decay within the geometry of the detector, the equivalent distance is that between the interaction point and the decay vertex of the particle reconstructed from its daughters. The magnitude of the total displacement vector, given the variable name of `distance` in basf2, as well as the radial and z-components of this distance, dr and dz, are included as input variables. The significance of the distance forms another variable passed to the classifier, and this is defined as the ratio between the measured distance and the uncertainty of this measurement. This indicates that the more precise our determination of the particle vertex position, the more meaningful distance variables will be when considered by the MVC training.

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

- **Angle between vertex position and particle momentum:** This refers to the angle between the intermediate particle momentum vector and the distance vector formed between the interaction point and the particle decay vertex, and is given the name
`cosAngleBetweenMomentumAndVertexVector` in basf2. This variable contains relevant information related to the particle flight direction and the kinematics of its decay.

- **Invariant mass:** Much like for π^0 mesons, the invariant mass M of the reconstructed candidates as well as the absolute value of the difference $|dM|$ between these values and the nominal masses as quoted by the PDG [20], form input variables to the classifier.

- **Energy:** The energies E of the candidates calculated in their rest frames are also included.

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

- **Daughter momenta and distance variables:** The magnitude of the momentum p of each daughter candidate calculated in their rest frames is also passed as kinematic information to the classifier. Additionally, the radial and z- distance variables d_r and

d_z of each daughter are used, and these are crucial to the determination of the decay vertex.

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

- **Particle rank:** Finally, the particle rank is included in a similar way to the treatment of π^0 mesons. After the relevant pre-classifier cut, K_S^0, Λ^0 and $\Sigma^\pm$ candidates are ranked from lowest to highest according to their values of $|dM|$. The best 20 candidates are kept for each particle, and their ranks are passed to the classifier training.

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

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

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

- **Invariant mass:** Both the difference between the invariant mass of the candidates and the established nominal mass dM as well as the absolute value of this difference $|dM|$ are used as classifier input for these intermediate particles. The invariant mass of each potential combination of the candidate daughters is also passed to the training.

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

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

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

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

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

- **Daughter `decayModeID`:** Particles that decay within the geometry of the detector often have multiple possible decay modes, with these incorporated within the FEI. The FEI framework thus assigns an index to the decay channels of each of these particles, named the `decayModeID`, ranging from 0 to the number of supported channels. The `decayModeID`s of the daughters of the $\Lambda_c^\pm$ baryons, J/Ψ and various D-mesons all form part of the input to the classifier training.

3.3.3 B-mesons: B^0, $B^\pm$

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

- **Daughter `SignalProbability`:** The `SignalProbability` of each daughter, and the product of the `SignalProbability` of all daughters.

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

- **Daughter momenta and distance variables:** The magnitudes of the rest-frame momentum p and distance vectors for each daughter.

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

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

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

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

- **Daughter `decayModeID`:** The `decayModeID`s of each daughter of the B-mesons are also included.

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

3.4 Cuts Made Within FEI Algorithm

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

3.4.1 Pre-classifier Cuts

One of the most useful ways of reducing combinatorics is to apply appropriate selections to the particle candidates before any classifiers are trained. These cuts should be reasonably loose and aim to remove only candidates that are very unlikely to correspond to true particles, such that computing time is not wasted passing variables from these candidates to the classifier training. These pre-classifier cuts were alluded to in Section 3.3, and are quoted below for each particle type as defined by the FEI algorithm.

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

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

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

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

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

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

K_L^0 **mesons:** No pre-classifier cut is applied to the K_L^0 mesons. Only the 20 K_L^0 candidates with the lowest values of $|dM|$ are retained, with $|dM|$ calculated as the difference between the reconstructed K_L^0 mass and the nominal neutral kaon mass as recorded by the PDG, $M_{K^0} = 498$ MeV [20].

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

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

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

The pre-classifier cuts applied to the remainder of the intermediate particle lists correspond to restricting the invariant mass to a suitable mass window, and are summarised in Table 3.1.

Intermediate particle	Invariant mass cut (GeV)	Number of candidates retained with lowest values of $|dM|$
π^0	$0.08 < M < 0.18$	20
K_S^0	$0.4 < M < 0.6$	20
D^0, $D^\pm$	$1.7 < M < 1.95$	20
D_s^+	$1.68 < M < 2.1$	20
J/Ψ	$2.6 < M < 3.7$	20
$\Lambda_c^\pm$	$2.2 < M < 2.4$	20
Λ^0	$0.9 < M < 1.3$	20
$\Sigma^\pm$	$1.0 < M < 1.4$	20

Table 3.1: The invariant mass pre-cuts applied to a subset of the intermediate particles reconstructed by the FEI.

$B^\pm$, B^0 **mesons:** For hadronic channels, cuts are applied to two meaningful kinematic variables, the first of which being the energy difference ΔE introduced in Section 3.3.3. As mentioned, true B-meson candidates are expected to have a value of ΔE close to 0, and this is reflected in an appropriate pre-classifier cut:

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

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

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

where E_{beam} is the aforementioned beam energy defined in the $\Upsilon(4S)$ rest frame, and $\vec{p}$ corresponds to the reconstructed 3-momentum of the B-meson candidate. In essence, M_{bc} is a quantity that is analogous to the invariant mass, but with the beam energy replacing the B-meson energy such that the calculation does not rely on the B-meson mass determined from its daughters. For this reason, selections are typically made on M_{bc} rather than the invariant mass in order to improve resolution. True B-meson candidates should have M_{bc} values close to the nominal B-meson mass of 5.28 GeV [20], and a loose cut is therefore employed to reduce combinatorics before classifier training:

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

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

It should be noted that in all of the above cases, the number of candidates retained corresponds to the maximum possible number, that is, a channel with a number of candidates less than or equal to the defined cut-off will retain all candidates after such a pre-cut is applied. It is therefore pertinent that the cut-off values are reasonably defined for each particle type to remain considerably loose whilst also avoiding such cases in which all candidates are able to pass the implemented cut. All values quoted represent the default configuration of the FEI given the suggested order of input training events, with the opportunity available for individual users to override these cut-off values with their desired alternatives. No such changes were made for the analysis presented throughout this thesis, with the FEI implemented using the default configuration.

3.4.2 Post-classifier Cuts

After selecting a number of candidates to proceed to the classifier training, additional cuts can then be made on the list of candidates once the classifiers have been trained. Each candidate has a `SignalProbability` now associated with it as a result of the training, and this can be used as an additional variable for selection, given that it is expected that candidates with a relatively low `SignalProbability` are less likely to correspond to the true particles. Post-classifier cuts on this `SignalProbability` are enforced to further reduce combinatorics by decreasing the number of candidates proceeding to the subsequent reconstruction stage. The default cuts as implemented within the FEI algorithm are depicted in Table 3.2 for each particle type, and these can be overridden by the user as desired in the same way as for the pre-classifier cuts described above. No post-classifier cuts are applied to the B-mesons, as this is the final stage of the FEI reconstruction.

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

Table 3.2: The post-classifier cuts applied to the particles reconstructed by the FEI.

3.5   Comparison with Full Reconstruction

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

The FR algorithm involved the use of the external NeuroBayes package [63], with the multivariate analysis technique chosen for the FEI being that of the stochastic boosted decision tree method, FastBDT [66]. The training of the FR algorithm was slow, taking of the order of weeks. In contrast, the FEI is much faster, taking only of the order of days for the training process to be completed. The FEI is also significantly more user-friendly, with the entire algorithm automated such that users need only to specify the initial configuration they would like for the training [62]. The algorithm is then efficiently trained as per the specified configuration with no further intervention required by the user.

Another major difference between the two methods is that FR only employed the use of hadronic tags, with the algorithm designed to train only hadronic B-meson decay modes. The FEI incorporates both hadronic and semi-leptonic decay modes, allowing for tagged analyses using both types of tags [68]. The algorithm additionally affords users the option of turning either the hadronic or semi-leptonic modes off during the training in order to save computing time in the case that the user is interested in only one of the two tagging methods. One further advantage of the FEI is that it supports more B-meson hadronic decay channels than were supported by the FR. At the commencement of this study, the FEI incorporated a total of 36 hadronic charged and 31 hadronic neutral B-meson decay modes, a significant increase from the equivalent 17 and 15 respective modes characteristic of the FR [69]. With the inclusion of the 42 charged and 39 neutral semi-leptonic modes, the FEI is able to reconstruct a total of over 4000 exclusive decay channels compared with the roughly 1000 achieved by the FR. The analysis in the coming chapters considers only hadronic tagging via the FEI, and a full list of each hadronic B-meson decay channel supported by the FEI, together with the associated `decayModeID`s, is given in Table 3.3. The decay modes of the intermediate

Charged B_{tag} Modes		Neutral B_{tag} Modes	
decayModeID	Channel	decayModeID	Channel
0	$B^+ \to \bar{D}^0 \pi^+$	0	$B^0 \to D^- \pi^+$
1	$B^+ \to \bar{D}^0 \pi^+ \pi^0$	1	$B^0 \to D^- \pi^+ \pi^0$
2	$B^+ \to \bar{D}^0 \pi^+ \pi^0 \pi^0$	2	$B^0 \to D^- \pi^+ \pi^0 \pi^0$
3	$B^+ \to \bar{D}^0 \pi^+ \pi^+ \pi^-$	3	$B^0 \to D^- \pi^+ \pi^+ \pi^-$
4	$B^+ \to \bar{D}^0 \pi^+ \pi^+ \pi^- \pi^0$	4	$B^0 \to D^- \pi^+ \pi^+ \pi^- \pi^0$
5	$B^+ \to \bar{D}^0 D^+$	5	$B^0 \to \bar{D}^0 \pi^+ \pi^-$
6	$B^+ \to \bar{D}^0 D^+ K_S^0$	6	$B^0 \to D^- D^0 K^+$
7	$B^+ \to \bar{D}^{*0} D^+ K_S^0$	7	$B^0 \to D^- D^{*0} K^+$
8	$B^+ \to \bar{D}^0 D^{*+} K_S^0$	8	$B^0 \to D^{*-} D^0 K^+$
9	$B^+ \to \bar{D}^{*0} D^{*+} K_S^0$	9	$B^0 \to D^{*-} D^{*0} K^+$
10	$B^+ \to \bar{D}^0 D^0 K^+$	10	$B^0 \to D^- D^+ K_S^0$
11	$B^+ \to \bar{D}^{*0} D^0 K^+$	11	$B^0 \to D^{*-} D^+ K_S^0$
12	$B^+ \to \bar{D}^0 D^{*0} K^+$	12	$B^0 \to D^- D^{*+} K_S^0$
13	$B^+ \to \bar{D}^{*0} D^{*0} K^+$	13	$B^0 \to D^{*-} D^{*+} K_S^0$
14	$B^+ \to D_S^+ \bar{D}^0$	14	$B^0 \to D_S^+ D^-$
15	$B^+ \to \bar{D}^{*0} \pi^+$	15	$B^0 \to D^{*-} \pi^+$
16	$B^+ \to \bar{D}^{*0} \pi^+ \pi^0$	16	$B^0 \to D^{*-} \pi^+ \pi^0$
17	$B^+ \to \bar{D}^{*0} \pi^+ \pi^0 \pi^0$	17	$B^0 \to D^{*-} \pi^+ \pi^0 \pi^0$
18	$B^+ \to \bar{D}^{*0} \pi^+ \pi^+ \pi^-$	18	$B^0 \to D^{*-} \pi^+ \pi^+ \pi^-$
19	$B^+ \to \bar{D}^{*0} \pi^+ \pi^+ \pi^- \pi^0$	19	$B^0 \to D^{*-} \pi^+ \pi^+ \pi^- \pi^0$
20	$B^+ \to D_S^{*+} \bar{D}^0$	20	$B^0 \to D_S^{*+} D^-$
21	$B^+ \to D_S^+ \bar{D}^{*0}$	21	$B^0 \to D_S^+ D^{*-}$
22	$B^+ \to \bar{D}^0 K^+$	22	$B^0 \to D_S^{*+} D^{*-}$
23	$B^+ \to D^- \pi^+ \pi^+$	23	$B^0 \to J/\psi K_S^0$
24	$B^+ \to D^- \pi^+ \pi^+ \pi^0$	24	$B^0 \to J/\psi K^+ \pi^-$
25	$B^+ \to J/\psi K^+$	25	$B^0 \to J/\psi K_S^0 \pi^+ \pi^-$
26	$B^+ \to J/\psi K^+ \pi^+ \pi^-$	26	$B^0 \to \bar{\Lambda}_c^- p^+ \pi^+ \pi^-$
27	$B^+ \to J/\psi K^+ \pi^0$	27	$B^0 \to \bar{D}^0 p^+ \bar{p}^-$
28	$B^+ \to J/\psi K_S^0 \pi^+$	28	$B^0 \to D^- p^+ \bar{p}^- \pi^+$
29	$B^+ \to \bar{\Lambda}_c^- p^+ \pi^+ \pi^0$	29	$B^0 \to D^{*-} p^+ \bar{p}^- \pi^+$
30	$B^+ \to \bar{\Lambda}_c^- p^+ \pi^+ \pi^- \pi^+$	30	$B^0 \to \bar{D}^0 p^+ \bar{p}^- \pi^+ \pi^-$
31	$B^+ \to \bar{D}^0 p^+ \bar{p}^- \pi^+$	31	$B^0 \to \bar{D}^{*0} p^+ \bar{p}^- \pi^+ \pi^-$
32	$B^+ \to \bar{D}^{*0} p^+ \bar{p}^- \pi^+$		
33	$B^+ \to D^+ p^+ \bar{p}^- \pi^+ \pi^-$		
34	$B^+ \to D^{*+} p^+ \bar{p}^- \pi^+ \pi^-$		
35	$B^+ \to \bar{\Lambda}_c^- p^+ \pi^+$		

Table 3.3: The hadronic B-meson decay channels supported by the FEI algorithm, along with their `decayModeID` flags.

particles in these hadronic B-meson decay chains are left to Table A.1 in Appendix A.

Related in part to the increase in the number of supported decay channels, one of the most notable improvements of the FEI over the FR is the increase in the maximum possible tag-side efficiency, defined as the fraction of $\Upsilon(4\mathrm{S}) \to B\bar{B}$ events for which a B_{tag} is able to be correctly reconstructed [13]. Unlike untagged analyses which effectively correspond to a tag-side efficiency of 100% due to no event explicitly requiring the reconstruction of the non-signal B-meson via a specified decay mode, tagged analyses are associated with much lower tag-side efficiencies, typically of the order of 0.1% for hadronic tagging and 1% for semi-leptonic tagging. Studies on the performance of the FEI and FR have demonstrated over a two-fold increase in the tag-side efficiency achieved by the FEI [13], corresponding to an increase from 0.28% (FR) to 0.76% (FEI) for reconstructed hadronic B^+ tags, and from 0.18% (FR) to 0.46% (FEI) for hadronic B^0 tags. In these same studies, the maximum tag-side efficiencies achieved for semi-leptonic tagging via the FEI were 1.80% and 2.04% for charged and neutral tags, respectively.

3.6   FEI in Practice

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

3.6.1   Training

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

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

The training of the FEI algorithm is carried out centrally within the Belle II collaboration, with no individual analysts required to perform a training of the algorithm themselves.

---

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

This training is also updated periodically as the analysis software and MC generation evolves. The version of the FEI training relevant to the analyses presented in this study was executed on a sample of 108×10^6 charged (B^+B^-) and 102×10^6 mixed ($B^0\bar{B}^0$) MC events [2].

3.6.2 Application

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

Users wanting to perform a tagged analysis for a B_{sig} decay mode of choice can apply the FEI to a data or MC sample to produce a list of tags alongside building B_{sig} candidates, and can then combine these in order to reconstruct the entire $\Upsilon(4S)$ decay event. In this case, both the B_{sig} and B_{tag} candidates are built within the same path, with the basf2 functions used for reconstruction ensuring that no B_{sig} and B_{tag} candidates associated with the same daughters are combined together into an $\Upsilon(4S)$. The application of the FEI by individual users in this way is not typically encouraged within the collaboration, however, where users are instead recommended to utilise official FEI skims for their analyses, as described below.

3.6.3 FEI Skims

In addition to the training, the application of the FEI is also performed centrally within the collaboration via a process termed as skimming. Here, an FEI skim refers to the application of the FEI together with a number of very loose selections, producing an output list (a basf2 `ParticleList`) of reconstructed B_{tag} candidates. Analysts intending to utilise the FEI for tagging in their analyses can then simply access these B_{tag} lists directly and search for events in which the recoiling B-meson decays via their desired signal mode. Skimming is highly desirable due to the dramatic decrease in file size, as $\Upsilon(4S)$ events in which no B_{tag} candidate can be built are thrown away immediately, saving analysts from wasting time and computing resources on the processing of these events. Both hadronic and semi-leptonic FEI skims on MC and data are available for all analysts to access.

The selections imposed in the reconstruction of B_{tag} candidates by the FEI skim are designed to be very loose, rejecting only those events with a substantially high likelihood of possessing incorrectly reconstructed tags. Further background rejection is left to individual users to decide what is appropriate for any given analysis. Firstly, the minimum number of charged particle tracks in an event meeting a set of loose criteria is set to three. The vast majority of B-meson decay chains corresponding to the hadronic FEI channels include at least three charged particles, and such a cut is useful at suppressing background from events with only a small number of tracks. The criteria chosen for the track selection include cuts

---

[2]Specifically, this refers to the version with prefix 'FEIv4_2022_MC15_light-2205-abys', trained on Monte Carlo from the 'MC15' Belle II MC campaign.

on the z-coordinate $z0$ and the distance from the origin of the detector geometry $d0$ of the point on the track closest to the origin, $|z0| < 2.0$ cm, $|d0| < 0.5$ cm, to ensure proximity to the interaction point of the e^+e^- beams. A minimum threshold is also set for the transverse momentum of the track, $p_t > 0.1$ GeV. Similar restriction is applied to the ECL clusters in the event, with at least three clusters being required within the polar angle acceptance of the CDC satisfying a minimum energy threshold of E > 0.1 GeV. The total visible energy per event is required to be at least 4 GeV, defined as the sum of the energies of all reconstructed tracks and clusters in the CMS frame.

The number of B_{tag} candidates per event is similarly reduced with a tightening of the M_{bc} and $|\Delta E|$ selections enforced during the training of the FEI algorithm, with $M_{bc} > 5.2$ GeV and $|\Delta E| < 0.3$ GeV. A loose cut on the FEI classifier output, `SignalProbability` > 0.001, provides further background rejection at little signal loss, and is applied to all B_{tag} modes with the exception of $B^0 \to J/\psi K_S^0$ and $B^+ \to J/\psi K^+$. During a prior training of the FEI, the `SignalProbability` classifier was unsuccessfully trained for these two modes due to their high purity, with not enough background events present in the training sample. All event-level and B_{tag} candidate-based selections applied during the hadronic FEI skims are summarised in Table 3.4.

Event	visibleEnergyOfEventCMS > 4 GeV
	nCleanedTracks($|z0| < 2.0$ cm, $|d0| < 0.5$ cm, $p_t > 0.1$ GeV) $\geqslant 3$
	nCleanedECLClusters($0.2967 < \theta < 2.618$, E > 0.1 GeV) $\geqslant 3$
B_{tag}	`SignalProbability` > 0.001 or `decayModeID` = 23(25) for $B^0(B^+)$
	$M_{bc} > 5.2$ GeV
	$|\Delta E| < 0.3$ GeV

Table 3.4: Event- and candidate-based selections applied in the hadronic FEI skims of data and MC.

Chapter 4

Calibrating the FEI Algorithm

Monte Carlo (MC) used to simulate real data is constantly updated and refined over time as our understanding of the underlying physics behind particle decay is expanded. However, MC can only ever represent a best guess at the nature of data based on current knowledge, and any imperfections in this understanding are manifested as observed differences between measurements made using MC versus real data. One key example of this is the efficiency with which the FEI algorithm reconstructs B-meson tags, which differs by a non-negligible amount between simulated and real data. In order to correct for this difference in efficiency, the algorithm must be calibrated.

Calibrating a tagging algorithm typically involves reconstructing a signal B-meson decay with a high branching fraction recoiling against a B-meson tag in an $\Upsilon(4S) \to B\bar{B}$ event. A set of calibration factors can then be determined by evaluating the ratio between the number of signal events reconstructed from data versus MC. These factors can then be applied to weight the MC used in independent analyses to correct the FEI efficiency, given several conditions are met. Of these, the most critical is that any selection applied on the B-meson tag `SignalProbability` must be consistent with the selection applied in the calibration procedure, as the calibration is highly dependent on the chosen value. As will be discussed at length in the following chapter, the optimal selection on the `SignalProbability` found for the hadronically tagged $B \to \pi\ell\nu_\ell$ analysis was `SignalProbability` > 0.001, and thus the calibration detailed below was performed for this value of the selection. Some time after commencing this study, an FEI task force was established within the collaboration which has since provided results for the calibration at several values of the `SignalProbability` relevant for other analyses.

The Belle Full Reconstruction (FR) algorithm was calibrated using the full Belle dataset, equivalent to an integrated luminosity of 711 fb^{-1} [32]. The signal decays chosen for the calibration were the semi-leptonic decays $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$, where ℓ refers to an electron or muon, with a combined total branching fraction used of 6.26% for $B^- \to D^{(*)0}\ell^-\bar{\nu}_\ell$ and 5.31% for $\bar{B}^0 \to D^{(*)+}\ell^-\bar{\nu}_\ell$ [32] [1]. Given these high branching fractions and the large

---

[1]The total branching fractions for these decays according to the Particle Data Group [20] at the time were 7.91% for $B^- \to D^{(*)0}\ell^-\bar{\nu}_\ell$ and 7.22% for $\bar{B}^0 \to D^{(*)+}\ell^-\bar{\nu}_\ell$, but only a subset of the subsequent $D^{(*)0}$ and $D^{(*)+}$ decays were used in the signal reconstruction.

data-set size, the calibration was able to be performed separately for each individual hadronic tag decay channel. For this analysis, in order to utilise as many events as possible in the 362 fb^{-1} LS1 Belle II data-set, the total inclusive semi-leptonic decays $B^+ \to X^0 \ell^+ \nu_\ell$ and $B^0 \to X^- \ell^+ \nu_\ell$ were chosen as the signal decay modes used for the hadronic FEI calibration, where X^0 and X^- refer to any neutral or charged hadronic system, respectively. The total branching fractions for these inclusive decays as listed by the Particle Data Group (PDG) are $\mathcal{B}(B^+ \to X^0 \ell^+ \nu_\ell) = (10.99 \pm 0.28)\%$ and $\mathcal{B}(B^0 \to X^- \ell^+ \nu_\ell) = (10.33 \pm 0.28)\%$ [20]. In addition, the calibration was not performed for each tag decay channel individually, but rather a global calibration factor was determined for the hadronic tagging algorithm as a whole, in order to further maximise these statistics. After detailing the methodology of the chosen calibration procedure and the results obtained, these will be compared with an official Belle II calibration performed some time after the given analysis in Section 4.1.7.

4.1 Reconstructing $B \to X \ell \nu_\ell$

To reconstruct hadronically-tagged $B \to X \ell \nu_\ell$ decays in MC and data, lists of B_{tag} candidates provided by the hadronic FEI skims were combined with a single electron or muon to reconstruct the parent $\Upsilon(4S)$ meson, alongside several analysis selections. All remaining detected tracks and cluster deposits in the ECL and KLM were then grouped together into a basf2 object termed the rest-of-event (ROE). After some loose selections on the ROE to reject effects due to beam background, this ROE object was taken to represent the decay products of the hadronic system X. As described in Section 3.1.1, neutrinos are unable to be detected by Belle II, and the neutrinos in the $B \to X \ell \nu_\ell$ decays were thus treated as missing energy and momentum. Both charged $B^+ \to X^0 \ell^+ \nu_\ell$ and neutral $B^0 \to X^- \ell^+ \nu_\ell$ decays were studied.

4.1.1 Description of Monte Carlo

For this analysis, a number of MC samples were utilised in order to simulate real data. For generic $B\bar{B}$ decays, both charged $B^+ B^-$ and neutral $B^0 \bar{B}^0$ events equivalent to an integrated luminosity of 800 fb^{-1} were studied. These samples were used to simulate both signal events, where one B-meson in the pair decayed as $B \to X \ell \nu_\ell$, together with background events where neither B-meson decayed via this mode. Additional background samples were also studied, in which the initial electron-positron collision did not produce an $\Upsilon(4S)$ meson at all, but rather an alternative quark-antiquark pair, namely, $e^+ e^- \to q\bar{q}$, where, $q = u, d, s$ or c. These events, introduced in Section 2.1, are referred to as the continuum background, and a total of 1 ab^{-1} of $q\bar{q}$ events were used for the background simulation. Table 4.1 lists the number of events for each MC sample studied, before any selections including the application of the FEI skims.

MC sample	Integrated luminosity (ab^{-1})	N_{events} used ($\times 10^6$)
B^+B^-	0.8	432
$B^0\bar{B}^0$	0.8	408
$c\bar{c}$	1.0	1329
$u\bar{u}$	1.0	1605
$s\bar{s}$	1.0	383
$d\bar{d}$	1.0	401

Table 4.1: Integrated luminosity and corresponding number of events for each MC sample studied, before any selections including the application of the FEI skims.

4.1.2 Event Selection

In addition to the selections imposed by the hadronic FEI skim, as listed in Section 3.6.3, the following selections were applied in order to further reject background events, with limited inspiration taken from Ref. [70]:

- **Tag M_{bc}**: The selection on the beam-constrained mass of the hadronic B-meson tag was tightened substantially from the skim selection to $M_{\text{bc}} > 5.27$ GeV, a value close to the world average measurement of the B-meson invariant mass, 5.28 GeV [20], ensuring only the selection of tag candidates that were sufficiently B-meson-like.

- **Tag ΔE**: The selection on the difference between the centre-of-mass energy of the B_{tag} candidates and the beam energy was likewise tightened to $-0.15 < \Delta E < 0.1$ GeV, values of ΔE close to 0 being characteristic of real B-mesons as described in Section 3.3.3.

- **Continuum suppression**: Background events from continuum processes differ kinematically from $B\bar{B}$ decay events in a number of ways, and these differences can be utilised to suppress continuum backgrounds. One such example is the angle between the thrust vectors of a reconstructed B-meson candidate and the remaining particles in the event, where a thrust vector for a given particle or collection of particles refers to the direction in which the projection of the total momentum is highest. In a true $\Upsilon(4S)$ $\rightarrow B\bar{B}$ decay, the B-mesons are produced with little momentum, resulting in thrust vectors that are uniformly distributed on average, forming a spherical event shape. For continuum events, in which $e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}$ or $c\bar{c}$, the relatively light hadrons possess a significantly higher momentum and are largely produced back-to-back, resulting on average in a collimated, jet-like event shape. Continuum events misreconstructed as $B\bar{B}$ events thus tend towards having smaller angles between the thrust axes of the reconstructed B-meson candidate and the rest of the event.

  To suppress the continuum background in this analysis, after the reconstruction of a B_{tag} candidate, all remaining reconstructed particles in the event were grouped together into an ROE object. Several loose selections were first placed on the ROE in order to reject clear background events originating from beam background effects. Reconstructed tracks within the ROE were required to fall within the acceptance of

the CDC detector, possess a transverse momentum greater than 0.2 GeV, and satisfy impact-parameter selections placing them close to the interaction point, $dr < 2$ cm and $|dz| < 5$ cm, as first defined in Section 3.3.1. ECL clusters corresponding to particles in the ROE were similarly required to fall within the CDC acceptance, and clusters were required to be built from hits to more than one ECL crystal. The ECL cluster timing, defined as the time of the waveform of the crystal recording the highest energy in the cluster, was required to satisfy `clusterTiming` < 200 ns. This selection required that the particle associated with the given ECL cluster originated close to the interaction point, as, under this definition, a cluster produced at the interaction point would record a cluster timing consistent with 0 ns. Finally, loose minimum thresholds on the energies of the ECL clusters within the ROE were enforced for the three detector regions, $E > 0.08$ GeV for the forward endcap, $E > 0.03$ GeV for the barrel, and $E > 0.06$ GeV for the backward endcap.

After the cleaning of the ROE, the cosine of the angle between the thrust axes of the reconstructed B_{tag} candidate and the ROE, `cosTBTO`, was determined. In contrast to $B\bar{B}$ events with `cosTBTO` values distributed uniformly between 0 and 1, continuum backgrounds characteristic of `cosTBTO` values close to 1 were rejected via selecting only events satistfying `cosTBTO` < 0.9.

- **Tag** `SignalProbability`: As mentioned at the opening of this chapter, the cut on the B_{tag} `SignalProbability` imposed during the hadronic FEI skim was maintained, with no tightening of the selection, `SignalProbability` > 0.001. No `SignalProbability` cut was applied to the $B^0 \to J/\psi K_S^0$ and $B^+ \to J/\psi K^+$ tagging modes, as described in Section 3.6.3.

- **Best tag candidate selection**: After all analysis selections on the B_{tag} were applied, multiple B_{tag} candidates per event still remained in the analysis sample. To choose a single B_{tag} for each event, only the candidate with the highest value of the `SignalProbability` was retained.

- **Lepton track selections**: The reconstructed signal-side electron or muon tracks were also required to originate close to the interaction point through the impact-parameter selections $dr < 2$ cm and $|dz| < 5$ cm, and to fall within the polar angle acceptance of the CDC, $0.297 < \theta < 2.618$ radians.

- **Lepton particle identification**: To reject pions and kaons misreconstructed as electrons and muons, a minimum threshold was placed on the relevant lepton identification variables. The variables chosen differed from those introduced in Section 3.3.1 due to advances in the lepton identification techniques since the development of the FEI. Rather than the likelihood-based variables `electronID` and `muonID`, a new set of variables (`pidChargedBDTScore`) based on a boosted decision tree (BDT) classifier trained to better discriminate between charged particles was used. These BDT-based variables demonstrated lower rates of misreconstructed leptons than the likelihood-based variables for similar efficiencies. For this analysis, the relevant electron and muon BDT-based variables were required to satisfy `pidChargedBDTScore` > 0.9.

- **Lepton momentum cuts**: The momentum of the reconstructed signal-side lepton in the CMS frame, $p_{\ell,\mathrm{CMS}}$, was the variable chosen for the signal extraction in this analysis, with only leptons satisfying $p_{\ell,\mathrm{CMS}} > 1$ GeV considered. These leptons were also required to meet a minimum threshold on the transverse momentum, $p_t > 0.3$ GeV.

- **Best $\Upsilon(4S)$ candidate selection**: Finally, multiple lepton candidates per event were retained after the signal-side selections. Combining these lepton candidates with the single B_{tag} candidate chosen for each event likewise resulted in multiple $\Upsilon(4S)$ candidates. To select a single $\Upsilon(4S)$ candidate per event, only the one with the highest magnitude of the signal-side lepton CMS momentum was kept.

4.1.3 Data and MC Corrections

As described in the opening of this chapter, various differences between measurements extracted from MC and data arise due to our limited knowledge of the underlying nature of real data. The calibration of the FEI is indeed one such attempt at correcting for the observed discrepancy in the efficiency of reconstructing B-meson tags. In order to perform such an analysis, however, the data and MC used for the calibration must also be corrected for various other data-MC discrepancies, each described in detail in the following text.

$B \to X_u \ell \nu_\ell$ Branching Fraction Corrections

The branching fractions used for the official Belle II MC generation are updated periodically between individual campaigns to correspond to the latest values recorded by the PDG [20]. However, differences may occasionally arise due to the occurrence of any updates to the world average results after the conclusion of a particular MC generation campaign. For exclusive $B \to X_u \ell \nu_\ell$ decays, where the first daughter of the B-meson refers to a specified hadronic system containing an up-quark (X_u), such as π^0, ρ^0 or η, two differences were found between the MC generation and the latest results from the PDG. For $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, the PDG lists a branching fraction of $\mathcal{B}_{\mathrm{PDG}}(B^0 \to \pi^- \ell^+ \nu_\ell) = (1.50 \pm 0.06) \times 10^{-4}$, compared with the equivalent branching fraction used for the generation of the MC used in the analysis, $\mathcal{B}_{\mathrm{MC}}(B^0 \to \pi^- \ell^+ \nu_\ell) = 1.45 \times 10^{-4}$. For $B^+ \to \eta \ell^+ \nu_\ell$ decays, the branching fractions given by the PDG and the MC generation are $\mathcal{B}_{\mathrm{PDG}}(B^+ \to \eta \ell^+ \nu_\ell) = (3.9 \pm 0.5) \times 10^{-5}$ and $\mathcal{B}_{\mathrm{MC}}(B^+ \to \eta \ell^+ \nu_\ell) = 3.8 \times 10^{-5}$, respectively. Whilst the branching fractions corresponding to the world average measurements and the MC generation were consistent within uncertainties, making the effect of such corrections minimal, the MC was nonetheless updated to reflect the latest branching fraction measurements. For $\Upsilon(4S) \to B\bar{B}$ events in MC where one B-meson decayed via the $B^0 \to \pi^- \ell^+ \nu_\ell$ or $B^+ \to \eta \ell^+ \nu_\ell$ modes in the underlying MC truth, a weighting factor of $w_{\mathrm{BF}} = \mathcal{B}_{\mathrm{PDG}}/\mathcal{B}_{\mathrm{MC}}$ was applied. In the rare occurrence that both B-mesons decayed as $B^0 \to \pi^- \ell^+ \nu_\ell$ or $B^+ \to \eta \ell^+ \nu_\ell$, this weighting factor was applied twice to the MC event, once for each B-meson decay.

The total inclusive $B \to X_u \ell \nu_\ell$ branching fractions recorded by the PDG [20], where X_u refers to any hadronic system containing an up-quark, differ substantially from the sum

of the branching fractions of each exclusive $B \to X_u \ell \nu_\ell$ decay. In reality, the spectrum of $B \to X_u \ell \nu_\ell$ decays is complex, consisting of both individual resonances and a large underlying non-resonant contribution that is not well understood. For example, the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^0 \to \rho^- \ell^+ \nu_\ell$ decays present as resonances in the total spectrum of inclusive $B^0 \to X_u^- \ell^+ \nu_\ell$ decays, contributing a combined branching fraction of 0.444×10^{-3}, compared to the total inclusive $B^0 \to X_u^- \ell^+ \nu_\ell$ branching fraction of $(1.76 \pm 0.22) \times 10^{-3}$ [20]. To account for the gap between the sum of the resonant branching fractions and the total inclusive rate, the Belle II MC generation includes an additional $B \to X_u \ell \nu_\ell$ channel with a branching fraction equivalent to this difference, $\mathcal{B}(B \to X_{u,\mathrm{gap}} \ell \nu_\ell)$.

The total inclusive $B \to X_u \ell \nu_\ell$ branching fractions used to define this $B \to X_{u,\mathrm{gap}} \ell \nu_\ell$ gap contribution in the version of the MC used for this analysis are outdated with respect to the PDG. Updated values for these gap contributions were thus calculated from subtracting the total sum of all resonant $B \to X_u \ell \nu_\ell$ branching fractions from the up-to-date values of the total inclusive rates. Any $\Upsilon(4S) \to B\bar{B}$ events in the MC with B-mesons found to decay via the $B^0 \to X_{u,\mathrm{gap}}^- \ell^+ \nu_\ell$ or $B^+ \to X_{u,\mathrm{gap}}^0 \ell^+ \nu_\ell$ modes were similarly reweighted by a factor $w_{\mathrm{BF}} = \mathcal{B}_{\mathrm{PDG}} / \mathcal{B}_{\mathrm{MC}}$, applied once or twice per event depending on whether one or both of the B-mesons decayed via these channels. All $B \to X_u \ell \nu_\ell$ branching fractions used for the MC generation together with the updated values used for this analysis are listed in Table 4.2. More detail on the resonant and non-resonant contributions to the $B \to X_u \ell \nu_\ell$ total inclusive rate will be provided for the $B \to \pi \ell \nu_\ell$ analysis in Section 5.1.1.

Decay mode	MC branching fraction ($\times 10^{-4}$)	Updated branching fraction ($\times 10^{-4}$)
$B^0 \to X_u^- \ell^+ \nu_\ell$ modes		
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.45	1.50 ± 0.6
$B^0 \to \rho^- \ell^+ \nu_\ell$	2.94	2.94 ± 0.21
$B^0 \to X_{u,\mathrm{gap}}^- \ell^+ \nu_\ell$	17.01	13.16 ± 2.1
Total $B^0 \to X_u^- \ell^+ \nu_\ell$	21.4	17.6 ± 2.2
$B^+ \to X_u^0 \ell^+ \nu_\ell$ modes		
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.78	0.78 ± 0.03
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.58	1.58 ± 0.11
$B^+ \to \omega \ell^+ \nu_\ell$	1.19	1.19 ± 0.09
$B^+ \to \eta \ell^+ \nu_\ell$	0.38	0.39 ± 0.05
$B^+ \to \eta' \ell^+ \nu_\ell$	0.23	0.23 ± 0.08
$B^+ \to X_{u,\mathrm{gap}}^0 \ell^+ \nu_\ell$	17.24	15.04 ± 2.4
Total $B^+ \to X_u^0 \ell^+ \nu_\ell$	21.4	19.2 ± 2.4

Table 4.2: The branching fractions of all $B \to X_u \ell \nu_\ell$ contributions used in the MC generation, together with the updated values used for the analysis.

$B \to X_c \ell \nu_\ell$ Branching Fraction Corrections

The $B \to X_c \ell \nu_\ell$ decays present in the MC were also reweighted in a similar manner to the $B \to X_u \ell \nu_\ell$ decays above, in order to update the branching fractions assumed in the MC generation to more recent values. In this case, most of the relevant branching fractions used for the MC generation were already consistent with the PDG. However, alternative methods for calculating these branching fractions at higher precision exist, utilising a property of charged and neutral B-mesons known as strong isospin symmetry.

Isospin is a quantum number describing the up (u)- and down (d)-quark content of a particle, with the third component, I_3, defined as $I_3 = \frac{1}{2}(n_u - n_d)$, where n_u and n_d are the number of u- and d-quarks, respectively. Given the very similar masses of these quarks, isospin symmetry is a good approximation for the invariance of the strong interaction under the transformation between u- and d-quarks. One natural consequence of this symmetry is that the semi-leptonic $B \to X_c \ell \nu_\ell$ decays $B^0 \to D^- \ell^+ \nu_\ell$ and $B^+ \to D^0 \ell^+ \nu_\ell$ can be differentiated purely based on the ratio of the B^0 and B^+ lifetimes. The branching fractions of $B^0 \to D^- \ell^+ \nu_\ell$ and $B^+ \to D^0 \ell^+ \nu_\ell$ can be defined as:

$$\mathcal{B}(B^0 \to D^- \ell^+ \nu_\ell) = \frac{\Gamma(B \to D \ell \nu_\ell)}{\Gamma(B^0 \to \text{All})} = \Gamma(B \to D \ell \nu_\ell)\tau_{B^0},$$

$$\mathcal{B}(B^+ \to D^0 \ell^+ \nu_\ell) = \frac{\Gamma(B \to D \ell \nu_\ell)}{\Gamma(B^+ \to \text{All})} = \Gamma(B \to D \ell \nu_\ell)\tau_{B^+},$$

where $\Gamma(B \to D \ell \nu_\ell)$ is the total decay rate of all semi-leptonic decays of a B-meson to a D-meson and lepton-neutrino pair, $\Gamma(B^0 \to \text{All})$ and $\Gamma(B^+ \to \text{All})$ are the total decay rates of B^0 and B^+ mesons to all possible decay products, respectively, and τ_{B^0} and τ_{B^+} are the respective lifetimes of these B-mesons. It follows from these equations that:

$$\mathcal{B}(B^0 \to D^- \ell^+ \nu_\ell) \cdot \frac{\tau_{B^+}}{\tau_{B^0}} = \mathcal{B}(B^+ \to D^0 \ell^+ \nu_\ell),$$

and thus the branching fraction of the charged decay is simply the branching fraction of the neutral decay scaled by the ratio of the lifetimes, a value measured to be 1.076 ± 0.004 [31]. Assuming isospin symmetry, a weighted average of the measurements of these two branching fractions can be made, from which updated values of the individual charged and neutral branching fractions can be extracted with better precision. An equivalent set of equations can also be written for the semi-leptonic decays $B \to D^* \ell \nu_\ell$, with a similar outcome.

Decay mode	MC branching fraction (%)	Updated branching fraction (%)
$B^0 \to X_c^- \ell^+ \nu_\ell$ modes		
$B^0 \to D^- \ell^+ \nu_\ell$	2.140	2.2396 ± 0.0664
$B^0 \to D^{*-} \ell^+ \nu_\ell$	5.110	5.1137 ± 0.1082
$B^0 \to D_1^- \ell^+ \nu_\ell$	0.704	0.6164 ± 0.1013
$B^0 \to D_0^{*-} \ell^+ \nu_\ell$	0.362	0.3903 ± 0.0697
$B^0 \to D_1^{'-} \ell^+ \nu_\ell$	0.401	0.3903 ± 0.0837
$B^0 \to D_2^{*-} \ell^+ \nu_\ell$	0.347	0.2727 ± 0.0302
$B^0 \to D^- \pi^0 \ell^+ \nu_\ell$	0.046	0
$B^0 \to \bar{D}^0 \pi^- \ell^+ \nu_\ell$	0.092	0
$B^0 \to D^{*-} \pi^0 \ell^+ \nu_\ell$	0.046	0
$B^0 \to \bar{D}^{*0} \pi^- \ell^+ \nu_\ell$	0.092	0
$B^0 \to D \pi \pi \ell^+ \nu_\ell$	0.049	0.0579 ± 0.0823
$B^0 \to D^* \pi \pi \ell^+ \nu_\ell$	0.245	0.2007 ± 0.0952
$B^0 \to D^- \eta \ell^+ \nu_\ell$	0.217	0.4092 ± 0.4092
$B^0 \to D^{*-} \eta \ell^+ \nu_\ell$	0.217	0.4092 ± 0.4092
Total $B^0 \to X_c^- \ell^+ \nu_\ell$	10.1	10.1 ± 0.4
$B^+ \to X_c^0 \ell^+ \nu_\ell$ modes		
$B^+ \to D^0 \ell^+ \nu_\ell$	2.310	2.4098 ± 0.0709
$B^+ \to D^{*0} \ell^+ \nu_\ell$	5.490	5.5023 ± 0.1146
$B^+ \to D_1^0 \ell^+ \nu_\ell$	0.757	0.6632 ± 0.1089
$B^+ \to D_0^{*0} \ell^+ \nu_\ell$	0.389	0.4200 ± 0.0750
$B^+ \to D_1^{'0} \ell^+ \nu_\ell$	0.431	0.4200 ± 0.0900
$B^+ \to D_2^{*0} \ell^+ \nu_\ell$	0.373	0.2934 ± 0.0325
$B^+ \to D^- \pi^+ \ell^+ \nu_\ell$	0.100	0
$B^+ \to \bar{D}^0 \pi^0 \ell^+ \nu_\ell$	0.050	0
$B^+ \to D^{*-} \pi^+ \ell^+ \nu_\ell$	0.100	0
$B^+ \to \bar{D}^{*0} \pi^0 \ell^+ \nu_\ell$	0.050	0
$B^+ \to D \pi \pi \ell^+ \nu_\ell$	0.053	0.0623 ± 0.0886
$B^+ \to D^* \pi \pi \ell^+ \nu_\ell$	0.263	0.2160 ± 0.1025
$B^+ \to D_s^- K^+ \ell^+ \nu_\ell$	0.030	0.0300 ± 0.0142
$B^+ \to D_s^{*-} K^+ \ell^+ \nu_\ell$	0.030	0.0290 ± 0.0194
$B^+ \to \bar{D}^0 \eta \ell^+ \nu_\ell$	0.201	0.3770 ± 0.3770
$B^+ \to \bar{D}^{*0} \eta \ell^+ \nu_\ell$	0.201	0.3770 ± 0.3770
Total $B^+ \to X_c^0 \ell^+ \nu_\ell$	10.8	10.8 ± 0.4

Table 4.3: The branching fractions of all $B \to X_c \ell \nu_\ell$ contributions used in the MC generation, together with the updated values used for the analysis.

The remaining resonant contributions to the $B \to X_c\ell\nu_\ell$ branching fraction include semi-leptonic decays to heavier states ($X_c = D_1, D_1', D_0^*$ or D_2^*), often referred to collectively as $B \to D^{**}\ell\nu$ decays. Measuring the branching fractions for these decays is not as straight-forward due to the fact that their resonances are very broad, with many of them having overlapping decay products, and they are instead indirectly reconstructed from $B \to D\pi\pi\ell\nu_\ell$ and $B \to D^*\pi\pi\ell\nu_\ell$ final states. These measurements also directly depend on the assumption that isospin symmetry holds true for D-mesons.

Table 4.3 lists the branching fractions used in the MC generation for each of the $B \to X_c\ell\nu_\ell$ decays, together with the updated isospin-averaged branching fractions used for the analysis. A weighting factor of $w_{\mathrm{BF}} = \mathcal{B}_{\mathrm{NEW}}/\mathcal{B}_{\mathrm{MC}}$ was applied for each reconstructed B-meson found to decay via one of these modes.

In the MC, the $B \to D\pi\pi\ell\nu_\ell$ and $B \to D^*\pi\pi\ell\nu_\ell$ contributions are defined by subtracting the branching fractions of the resonant $B \to D^{**}\ell\nu$ modes from the measured total branching fractions $\mathcal{B}(B \to D\pi\pi\ell\nu_\ell)$ and $\mathcal{B}(B \to D^*\pi\pi\ell\nu_\ell)$. These contributions were corrected in the analysis using the updated values of the $B \to D^{**}\ell\nu$ branching fractions. The MC also includes contributions from $B \to D\pi\ell\nu_\ell$ and $B \to D^*\pi\ell\nu_\ell$ decays, but these are covered entirely within the branching fractions of the remaining resonant modes. As a result, these contributions were simply set to zero in the reweighting procedure.

In summing all of the resonant $B \to X_c\ell\nu_\ell$ contributions, there still exists a discrepancy between this sum and the measured total inclusive $B \to X_c\ell\nu_\ell$ branching fraction. This discrepancy is known as the $B \to X_c\ell\nu_\ell$ gap, analogous to the gap contribution defined for $B \to X_u\ell\nu_\ell$. For $B^+ \to X_c^0\ell^+\nu_\ell$, two components of this gap contribution have been measured by Belle [71], namely the $B^+ \to D_s^-K^+\ell^+\nu_\ell$ and $B^+ \to D_s^{*-}K^+\ell^+\nu_\ell$ decays, with the branching fractions of both modes measured to be relatively small with large systematic uncertainties. For the remainder of the $B^+ \to X_c^0\ell^+\nu_\ell$ gap and the $B^0 \to X_c^-\ell^+\nu_\ell$ gap, no information is known, and the contributions are estimated in the MC by assigning the remaining portion of the total inclusive $B \to X_c\ell\nu_\ell$ rates to $B \to \bar{D}\eta\ell\nu_\ell$ and $B \to \bar{D}^*\eta\ell\nu_\ell$ decays. Since this is no more than a guess at the composition of the $B \to X_c\ell\nu_\ell$ gap, these decays are assigned a 100% uncertainty. For the analysis, the branching fractions of the $B \to \bar{D}\eta\ell\nu_\ell$ and $B \to \bar{D}^*\eta\ell\nu_\ell$ decays were reweighted such that the sum of all $B \to X_c\ell\nu_\ell$ modes matched the current world average inclusive prediction [20].

$B \to X_u\ell\nu_\ell$ Form Factor Corrections

Section 1.2.3 introduced the differential decay rate for the $B \to \pi\ell\nu_\ell$ process as a function of the form factor $f_+(q^2)$, an analytical function calculated using lattice quantum chromo-dynamics (LQCD) alongside a suitable parameterisation. The differential decay rates for other exclusive $B \to X_u\ell\nu_\ell$ decays can likewise be defined, though these are substantially more complicated due to the existence of multiple form factors that depend not only on the square of the four-momentum transfer to the leptonic system q^2, but also on a number of kinematic angular variables. As theoretical predictions improve over time, new or updated

models for the parameterisation of these form factors become available.

For this analysis, in addition to the branching fraction corrections, the form factor models used in the generation of the MC were also updated to reflect the latest parameter calculations, or in some cases, replaced entirely where a newer model was available. This form factor reweighting was implemented via the eFFORT software package [72], and involved calculating the total and differential decay rates for each exclusive $B \to X_u \ell \nu_\ell$ process under both the decay model used for the MC generation and the updated model, such that a form factor correction weight w_{FF} could be defined per-event as

$$
w_{\text{FF}} = \frac{\Gamma_{\text{MC}}}{\Gamma_{\text{NEW}}} \cdot \frac{\dfrac{d\Gamma}{dq^2}^{\text{NEW}}}{\dfrac{d\Gamma}{dq^2}^{\text{MC}}},
$$

for the decays $B \to \rho \ell \nu_\ell$ and $B \to \omega \ell \nu_\ell$, and

$$
w_{\text{FF}} = \frac{\Gamma_{\text{MC}}}{\Gamma_{\text{NEW}}} \cdot \frac{\dfrac{d\Gamma}{dw}^{\text{NEW}}}{\dfrac{d\Gamma}{dw}^{\text{MC}}},
$$

for the decays $B \to \eta \ell \nu_\ell$ and $B \to \eta' \ell \nu_\ell$. Here, Γ_{MC} and Γ_{NEW} refer to the total decay rates, $\frac{d\Gamma}{dq^2}^{\text{MC}}$ and $\frac{d\Gamma}{dq^2}^{\text{NEW}}$ refer to the differential decay rates as a function of q^2, and $\frac{d\Gamma}{dw}^{\text{MC}}$ and $\frac{d\Gamma}{dw}^{\text{NEW}}$ refer to the differential decay rates as a function of the recoil parameter w, for the MC and updated models, respectively. The recoil parameter w is in turn defined here as the dot product of the velocities of the B-meson and the η or η' mesons, $v_B \cdot v_{\eta^{(\prime)}}$. These correction weights were subsequently applied per-event for each B-meson in the MC generation decaying via one of these modes.

The $B \to \rho \ell \nu_\ell$ and $B \to \omega \ell \nu_\ell$ decays present in the MC were generated using the BCL form factor parameterisation [34]. The same decay model was used for the form factor reweighting of these decays, but the values of the individual parameters were updated to more recent values [73]. For $B \to \eta \ell \nu_\ell$ and $B \to \eta' \ell \nu_\ell$, the decay model was updated completely from the ISGW2 model [74] to a new model proposed by physicists G. Duplancic and B. Melic described in detail in Ref. [75]. The ISGW2 model used in the MC generation is somewhat outdated and relies on the naive quark model, with large discrepancies observed between experimental data and theoretical predictions in the context of B-meson decays to η and η' mesons [75]. The later Duplancic-Melic model recalculates the relevant form factors, including several gluonic contributions and corrections related to the square of the $\eta^{(\prime)}$ mass. The values of the BCL parameters used in the MC generation of $B \to \pi \ell \nu_\ell$ decays were already up-to-date with respect to recent calculations [73], and thus these events were not reweighted in this manner.

$B \to X_c \ell \nu_\ell$ Form Factor Corrections

The semi-leptonic decays $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ were similarly reweighted in the MC to update the parameters used in the form factor calculations for these decays. Both modes were generated using the BGL parameterisation [76], with central values of the given parameters updated for the analysis to those given in Ref. [77] and [78] for $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ decays, respectively. For each B-meson in the MC decaying via one of these channels, a form factor correction weight was applied, defined as

$$ w_{\text{FF}} = \frac{\Gamma_{\text{MC}}}{\Gamma_{\text{NEW}}} \cdot \frac{\dfrac{d\Gamma}{dw}^{\text{NEW}}}{\dfrac{d\Gamma}{dw}^{\text{MC}}}, $$

for $B \to D\ell\nu_\ell$, and

$$ w_{\text{FF}} = \frac{\Gamma_{\text{MC}}}{\Gamma_{\text{NEW}}} \cdot \frac{\dfrac{d\Gamma}{dw d\cos(\theta_\ell) d\cos(\theta_V) d\chi}^{\text{NEW}}}{\dfrac{d\Gamma}{dw d\cos(\theta_\ell) d\cos(\theta_V) d\chi}^{\text{MC}}}, $$

for $B \to D^*\ell\nu_\ell$, where $\dfrac{d\Gamma}{dw d\cos(\theta_\ell) d\cos(\theta_V) d\chi}^{\text{MC}}$ and $\dfrac{d\Gamma}{dw d\cos(\theta_\ell) d\cos(\theta_V) d\chi}^{\text{NEW}}$ refer to the MC and updated partial decay rates as a function of the recoil parameter w, here defined as the the dot product of the velocities of the B- and D^* mesons, $v_B \cdot v_{D*}$, and a number of angular variables, namely $\cos(\theta_\ell)$, $\cos(\theta_V)$ and χ. The variable θ_ℓ refers to the polar angle of the lepton in the rest frame of the W-boson mediating the semi-leptonic decay, with respect to the flight direction of this W-boson in the B-meson rest frame. The D^* meson subsequently decays to a D-meson and either a pion or a photon, and θ_V is defined as the polar angle of the daughter D-meson in the rest frame of the D^* meson with respect to the D^* flight direction in the B-meson rest frame. Finally, the variable χ refers to the angle between the decay planes of the W-boson and the D^* meson. The definition of these angular variables for the semi-leptonic $B \to D^*\ell\nu_\ell$ decays is illustrated in Figure 4.1.

Bremsstrahlung Corrections

Whilst traveling through the Belle II detector, electrons and positrons can lose kinetic energy by radiating photons in a process known as bremsstrahlung radiation. When these leptons are detected and reconstructed, the magnitude of their measured energy and momentum will therefore be lower than the values at the time of their initial production. To combat the effects of this bremsstrahlung radiation on the analysis, the four-momentum of any photon below a given energy detected within a certain angular cone around an electron or positron was added to the lepton four-momentum. These photons were then excluded from the rest of the event to avoid any double-counting. In the case that multiple photons were found within this angular range, only the photon closest to the reconstructed lepton was treated
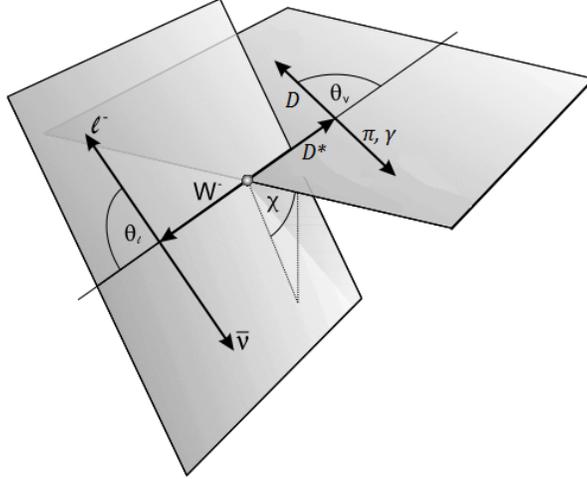
Figure 4.1: Definition of the angular variables θ_ℓ, θ_V and χ for the semi-leptonic $B \to D^* \ell \nu_\ell$ decays. For the neutral B-meson decay $B^0 \to D^{*-} \ell^+ \nu_\ell$, the D^{*-} meson subsequently decays as $D^{*-} \to D^- \pi^0, D^0 \pi^-$ or $D^- \gamma$. For the charged B-meson decay $B^+ \to D^{*0} \ell^+ \nu_\ell$, the daughter D^{*0} meson decays via $D^{*0} \to D^0 \pi^0$ or $D^0 \gamma$ [79].

as bremsstrahlung radiation.

The photon energy and angular thresholds used were based on standard recommendations within the collaboration, and were optimised for certain ranges of the reconstructed electron or positron lab-frame momentum, p_e. At low values of the lepton momentum, $p_e < 0.6$ GeV, only photons with energies satisfying $E_\gamma < 0.09$ GeV detected within an angle of 0.1368 radians of a given lepton were utilised for the bremsstrahlung correction. For electrons or positrons with a mid-range momentum $0.6 \leqslant p_e < 1.0$ GeV, the photon energy cut was loosened to allow for the emission of higher energy bremsstrahlung photons, with $E_\gamma < 0.9$ GeV, and the angular threshold was tightened to 0.0737 radians. Finally, for high energy leptons with $p_e \geqslant 1.0$ GeV, the photon energy threshold was loosened further to $E_\gamma < 1.2$ GeV, for photons within an angular distance of 0.0632 radians. Electrons and positrons reconstructed from both MC and data were corrected via this method.

Lepton Identification Efficiency and Fake-Rate Corrections

A further source of discrepancy between simulated and real data at Belle II is the efficiency with which electrons and muons satisfying particular selections on the lepton identification variables are reconstructed. This is true both for cases in which these leptons are reconstructed correctly, such as a real electron being identified as an electron in the reconstruction, and those cases in which a different charged final-state particle, either a pion or kaon, is mistakenly reconstructed as an electron or muon from detector information. In this latter case, the efficiencies with which these pions are kaons are reconstructed as leptons are often referred to as the pion and kaon fake-rates.

To account for the observed differences in the lepton reconstruction efficiencies and fake-rates between MC and data, a set of correction factors must be applied to the MC for any events reconstructed from decay products that include electrons or muons for which a particular selection on the lepton identification variables has been enforced. These corrections depend heavily on the value of the selection chosen and, for any given selection, are provided in bins of the magnitude of the lab-frame momentum p and the polar angle θ of the reconstructed electron or muon track.

The lepton identification correction factors are official Belle II recommendations, and have been derived from multiple independent studies by other collaboration members. For the lepton efficiency corrections, several decay channels were used in order to maximise the coverage over the lepton momentum and polar angle ranges, each involving the reconstruction of a decay from electron or muon daughters. Six channels in total were utilised in the derivation of these corrections, including the reconstruction of a J/ψ meson from a pair of lepton daughters, $J/\psi \to \ell^+\ell^-$, and the two-photon pair production ($e^+e^- \to e^+e^-\ell^+\ell^-$) and radiative Bhabha scattering ($e^+e^- \to e^+e^-\gamma$) processes described in Section 2.2, where $\ell = e$ or μ. For each of these studies, the efficiencies of various lepton identification selections were measured, and the final correction factors were determined via a statistical combination of the results from all channels.

Correction factors for the pion and kaon fake-rates were similarly determined via dedicated studies within the collaboration, in which several channels with pion or kaon decay products were reconstructed under electron or muon identification criteria. For the pion fake rates, two channels were used to gain sufficient p-θ coverage, namely the decay of a K_S^0 meson to a pair of pions, $K_S^0 \to \pi^+\pi^-$, and the production of a pair of τ leptons from an electron-positron pair, $e^+e^- \to \tau^\pm(1P)\tau^\mp(3P)$, where the suffixes (1P) and (3P) refer to those decays of a τ to one or three charged daughters, respectively, often termed 1- and 3-prong decays. Here, the 3-prong τ decay was chosen in order to provide a clean sample of pions, where $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$. The kaon fake-rate correction factors were determined through the reconstruction of charged D^* mesons from a charged pion and neutral D-meson, the latter of which subsequently decayed to a charged kaon-pion pair, $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$. For each of these channels, the numbers of pions or kaons remaining in the MC and data samples with and without the application of the lepton identification criteria were used to determine the fake-rates for the various thresholds. In the pion case, the final fake-rate corrections were derived by performing a statistical combination of the results of the two channels studied.

For this analysis, as detailed in Section 4.1.2, electrons and muons were selected based on a lepton identification threshold of `pidChargedBDTScore` > 0.9. The reconstructed MC events were divided into bins of the lepton momentum p and polar angle θ, and the correction weights derived for these thresholds were subsequently applied using a private software package [80]. Both the efficiency and fake-rate corrections were implemented, the former for events in which the reconstructed leptons were matched to real leptons in the MC truth, and the latter for those matched to pions or kaons. For those MC events with values of p and θ not covered by the available lepton identification corrections, no weighting was applied.

Data Track Momentum Scale Correction

The reconstruction of charged particle tracks at Belle II relies heavily on the assumption that the magnitude of the strength of the magnetic field spanning the detector volume is well measured. Whilst the superconducting solenoid encompassing the majority of the detector produces a stable 1.5T magnetic field, small, local fluctuations in the measured field strength occur between different regions. Using measurements of the magnetic field made in 2015, a 3-dimensional magnetic field map spanning the range of the Belle II detector was simulated and implemented within basf2 [81].

As the magnetic field determines the curvature of charged particles as they traverse the detector, measurements of the three momentum components p_x, p_y and p_z of charged tracks are particularly sensitive to any discrepancies or imperfections within the magnetic field map. These discrepancies would in turn affect the distributions of the invariant mass of these particles and subsequently any particles reconstructed from them. In an independent Belle II study, a distinct shift in the invariant mass distribution for reconstructed D^0 mesons was observed in data when compared to MC simulation, a direct result of imperfections in the simulated magnetic field map [82]. Using this shift, scaling factors were determined in order to correct all charged particle tracks reconstructed from data. These scaling factors differed only slightly from unity, ranging from $0.99918 - 0.99990$ for the entire recorded data-set, and were applied within this analysis to all reconstructed data tracks. Further measurements of the Belle II magnetic field and improvements to the field map simulation are planned for the future of the experiment [81].

Continuum Rescaling

Finally, an additional correction was applied solely to the events reconstructed from the continuum MC in order to rescale the contribution of this background component. As described in Section 2.1, at the $\Upsilon(4S)$ resonance corresponding to a center-of-mass energy of 10.58 GeV, e^+e^- collisions can not only produce the $b\bar{b}$ bound state of an $\Upsilon(4S)$ meson, but often result in the production of a $c\bar{c}, u\bar{u}, d\bar{d}$ or $s\bar{s}$ pair, collectively forming the continuum background. As illustrated in Figure 2.1, this continuum background spans a range of center-of-mass energies and does not exclusively lie under the $\Upsilon(4S)$ peak.

At Belle II, data collected at the $\Upsilon(4S)$ resonance is often referred to as on-resonance data, for which a total integrated luminosity of 362 fb^{-1} has thus far been collected. Belle II has also collected a small amount of data at a lower center-of-mass energy of 10.52 GeV, equivalent to an integrated luminosity of 42.3 fb^{-1}. This data is referred to as off-resonance data, and contains only continuum events with no B-meson content. One of the main purposes of collecting this off-resonance data is to improve the current modelling and understanding of the continuum MC.

Much like the discrepancies between the MC and the on-resonance data necessitating the calibration of the FEI, discrepancies between the measured off-resonance data and the continuum MC should also be corrected for. In this analysis, the number of events remaining
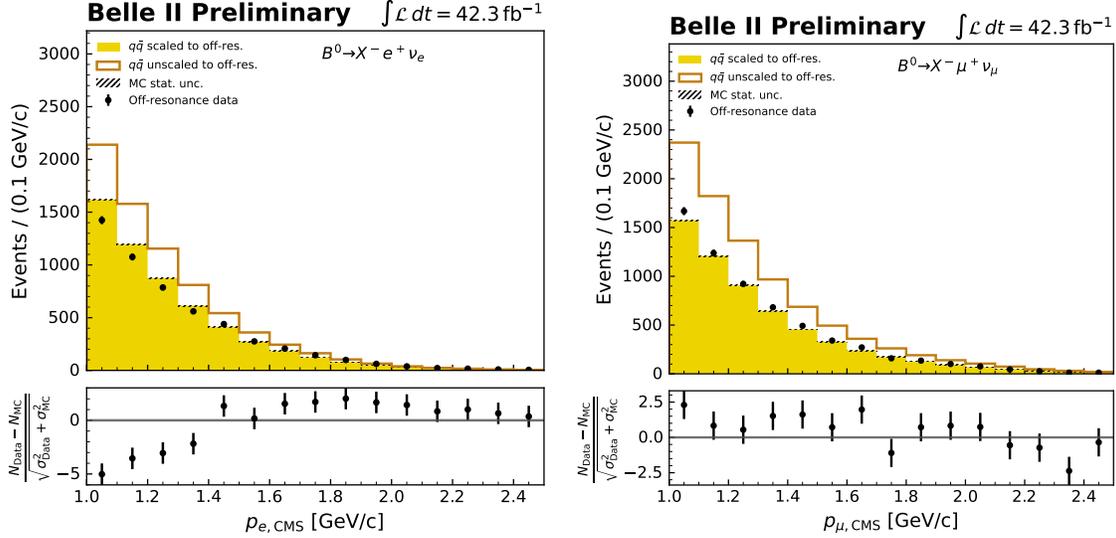
Figure 4.2: off-resonance data compared with continuum MC for $B^0 \to X^- e^+ \nu_e$ (left) and $B^0 \to X^- \mu^+ \nu_\mu$ (right) decays. The filled and unfilled histograms depict continuum MC that has and has not been rescaled to the off-resonance prediction, respectively.

in the off-resonance data passing all event selections was used to develop a prediction for the size of the continuum contribution. For this off-resonance sample, the relevant selection on the B_{tag} beam-constrained mass had to be rescaled due to the lower beam-energy. The rescaled beam-constrained mass, $M_{\text{bc,off}}$, can be derived from the on-resonance equivalent using the following relation,

$$M_{\text{bc,off}} = \sqrt{E_{\text{off}}^{*}{}^2 - \vec{p}_{\text{off}}^{\,2}} = \sqrt{E_{\text{on}}^{*}{}^2 - \left[\left(\frac{E_{\text{on}}^{*}}{E_{\text{off}}^{*}} \right)^2 \times \vec{p}_{\text{off}}^{\,2} \right]},$$

where $\vec{p}_{\text{off}}$ is the off-resonance B_{tag} momentum. The parameters E_{off}^{*} and E_{on}^{*} refer to the center-of-mass energies of the colliding electron–positron beams for off- and on-resonance data, 10.52 GeV and 10.58 GeV, respectively. Using this relation, the nominal event selection $M_{\text{bc}} > 5.27$ GeV was modified to $M_{\text{bc,off}} > 5.24$ GeV for the off-resonance sample. A similar rescaling can be performed for the selection on the tag-side energy difference, ΔE, but the resultant shift is negligible.

The continuum MC was subsequently rescaled to the off-resonance prediction by the application of a scaling factor w_{qq},

$$w_{qq} = \frac{N_{\text{off}}}{N_{\text{qq}}} \times \frac{\text{L}_{\text{qq}}^{\text{int}}}{\text{L}_{\text{off}}^{\text{int}}} \times \left(\frac{E_{\text{off}}^{*}}{E_{\text{on}}^{*}} \right)^2,$$

where N_{off} and N_{qq} refer to the number of events remaining in the off-resonance data and continuum MC after all selections, respectively, and $\text{L}_{\text{off}}^{\text{int}}$ and $\text{L}_{\text{qq}}^{\text{int}}$ are the respective integrated luminosities for each data-set, namely 42.3 fb^{-1} for the off-resonance data and 1 ab^{-1} for
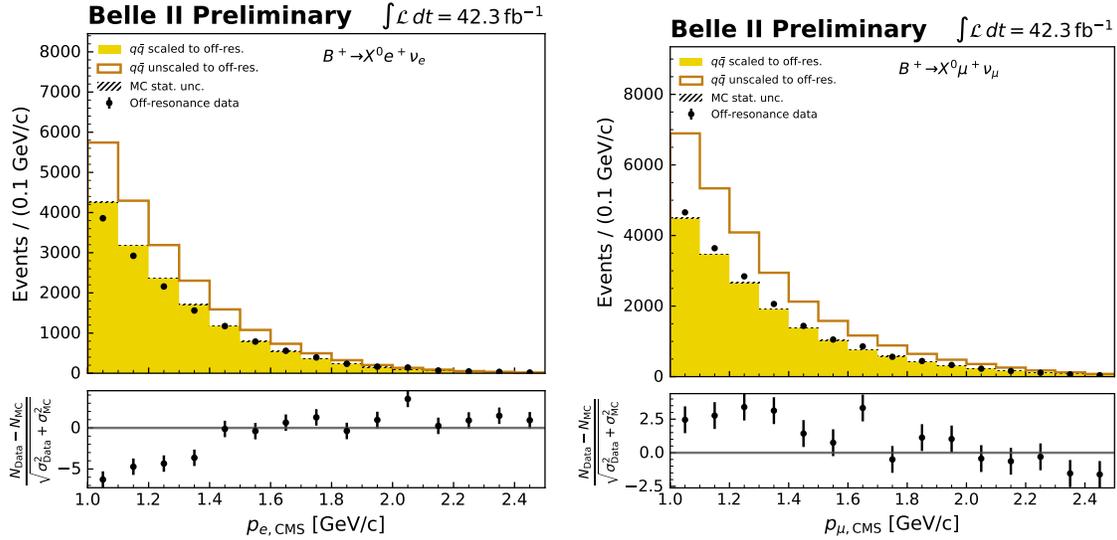
the continuum MC.

Figure 4.3: off-resonance data compared with continuum MC for $B^+ \rightarrow X^0 e^+ \nu_e$ (left) and $B^+ \rightarrow X^0 \mu^+ \nu_\mu$ (right) decays. The filled and unfilled histograms depict continuum MC that has and has not been rescaled to the off-resonance prediction, respectively.

Figures 4.2 and 4.3 depict a comparison between the off-resonance data and the continuum MC in the signal selection variable chosen for the analysis, $p_{\ell,\mathrm{CMS}}$, for the reconstruction of $B^0 \rightarrow X^- \ell^+ \nu_\ell$ and $B^+ \rightarrow X^0 \ell^+ \nu_\ell$, respectively. The continuum MC has been normalised to the integrated luminosity of the off-resonance data sample, and the distributions are shown separately for the two lepton flavours studied, e and μ. Both the continuum MC contributions with and without the application of the rescaling factors are plotted against the off-resonance data points. In each case, the rescaling significantly improves the agreement between MC and data when compared to the situation in which no rescaling is applied.

At present, the rescaling chosen only takes into account the total relative normalisation of the continuum MC with respect to the off-resonance data. As a result, the data-MC agreement fluctuates between individual $p_{\ell,\mathrm{CMS}}$ bins as can be seen in the given distributions. To better illustrate the agreement observed per-bin, a secondary plot is presented at the bottom of each distribution in Figures 4.2 and 4.3, depicting the data-MC residuals. These residuals represent how well described the data is by the MC, and are defined in each bin as the ratio between the difference in the number of data and MC events and the uncertainty on this difference, $(N_{\mathrm{Data}} - N_{\mathrm{MC}})/\sqrt{\sigma_{\mathrm{Data}}^2 + \sigma_{\mathrm{MC}}^2}$. Here, σ_{Data} and σ_{MC} refer to the statistical uncertainties in each bin for the data and MC, respectively, calculated according to Poisson statistics. The number of events in the MC or data sample passing all analysis selections can be estimated as the mean of a Poisson distribution μ, defined as $\mu = \sigma \mathcal{L}$, where σ is the standard deviation of the distribution and $\mathcal{L}$ is the integrated luminosity of the sample. In the Poissonian description, the uncertainty on the number of events N can be estimated as the square root of this value, $\sqrt{N}$ [83]. Each residual is assigned an uncertainty of ± 1 to

illustrate the bounds of a 1σ deviation from the central value.

4.1.4   Fitting and Results

In order to extract the number of signal $B \to X\ell\nu_\ell$ events in data, the distribution in the signal lepton CMS momentum of all data events passing the event selections was studied against the corresponding MC. The MC used for the analysis was separated into three main categories; signal, other $B\bar{B}$ and continuum events, the latter two categories representing the backgrounds for the analysis. Here, the signal contribution described cases in which the signal-side lepton was both correctly identified and originated from the semi-leptonic decay of a B-meson in the MC truth. The other $B\bar{B}$ contribution covered the cases in which the signal lepton reconstructed from the generic $B\bar{B}$ MC was either misidentified or produced from a secondary decay of the daughter of a B-meson, with the continuum contribution describing all candidates built from $q\bar{q}$ MC.
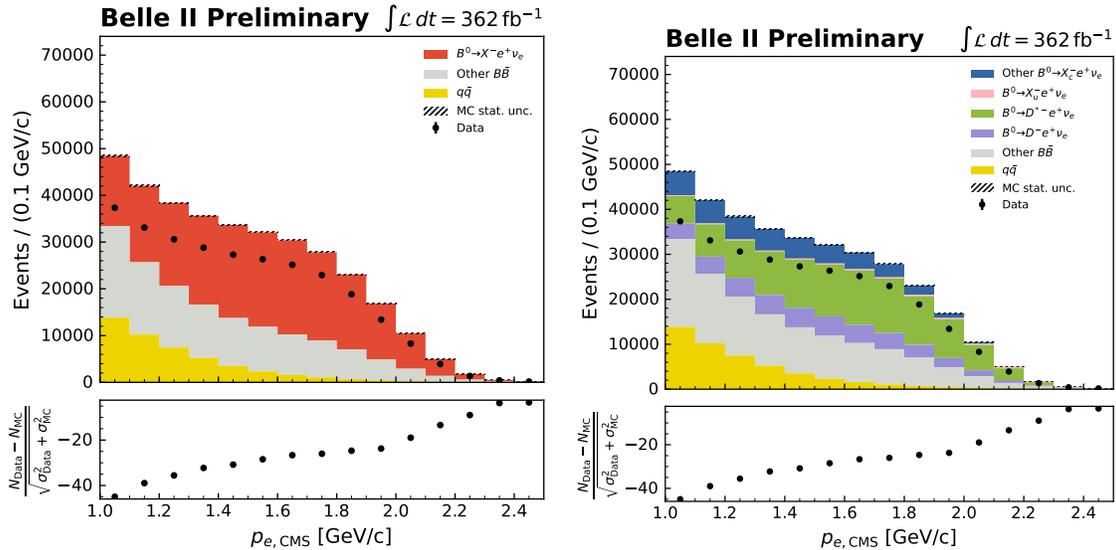
Figure 4.4: Distribution of the signal-side lepton CMS momentum $p_{\ell,\mathrm{CMS}}$ in data against normalised MC for $B^0 \to X^- e^+ \nu_e$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

Figures 4.4, 4.5, 4.6 and 4.7 depict the $p_{\ell,\mathrm{CMS}}$ distributions in data and MC for reconstructed $B^0 \to X^- e^+ \nu_e$, $B^0 \to X^- \mu^+ \nu_\mu$, $B^+ \to X^0 e^+ \nu_e$ and $B^+ \to X^0 \mu^+ \nu_\mu$, respectively, with the MC split into the three contributions described above. For each distribution, an additional plot is included showing the breakdown of the signal MC into four components, illustrating the relative contributions of various semi-leptonic B-meson decays to the total $B \to X\ell\nu_\ell$ signal MC. Contributions from $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ decays are included, together with the total contribution from all $B \to X_u\ell\nu_\ell$ decays. The remainder of the $B \to X_c\ell\nu_\ell$ contribution, containing both the resonant $B \to D^{**}\ell\nu_\ell$ decays and the $B \to X_c\ell\nu_\ell$ gap, form the fourth component of the plotted signal MC. The $B \to D^*\ell\nu_\ell$
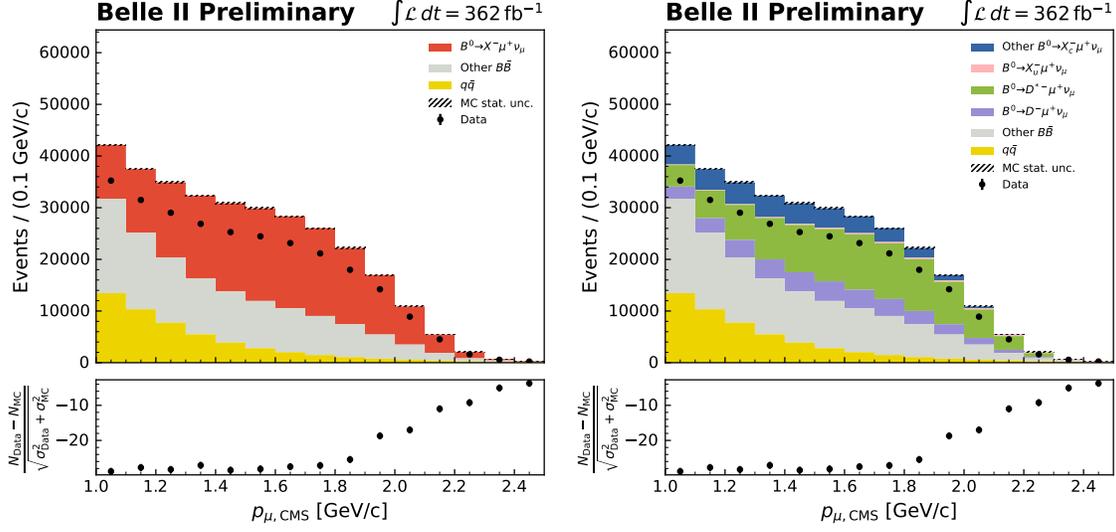
Figure 4.5: Distribution of the signal-side lepton CMS momentum $p_{\ell,\text{CMS}}$ in data against normalised MC for $B^0 \rightarrow X^- \mu^+ \nu_\mu$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).
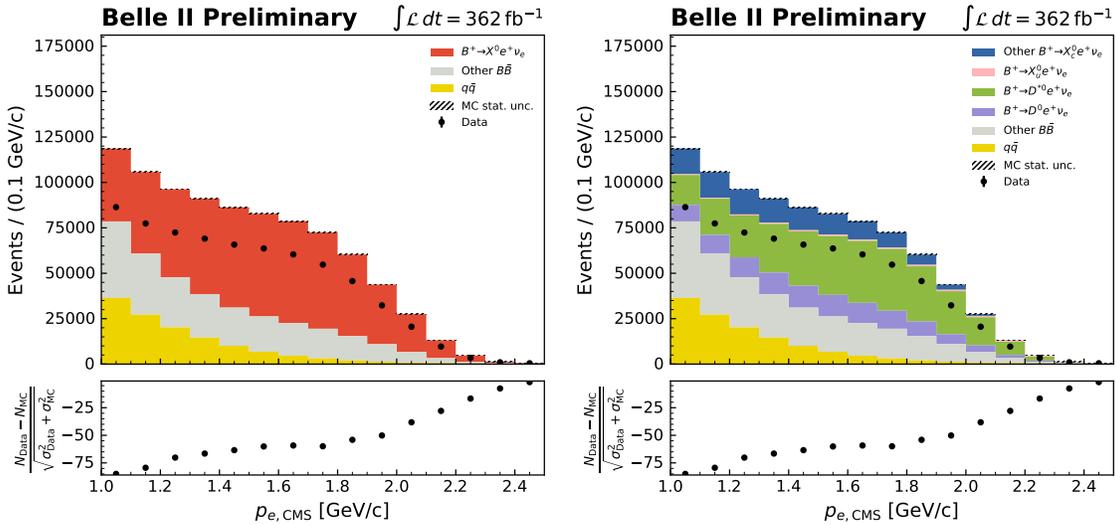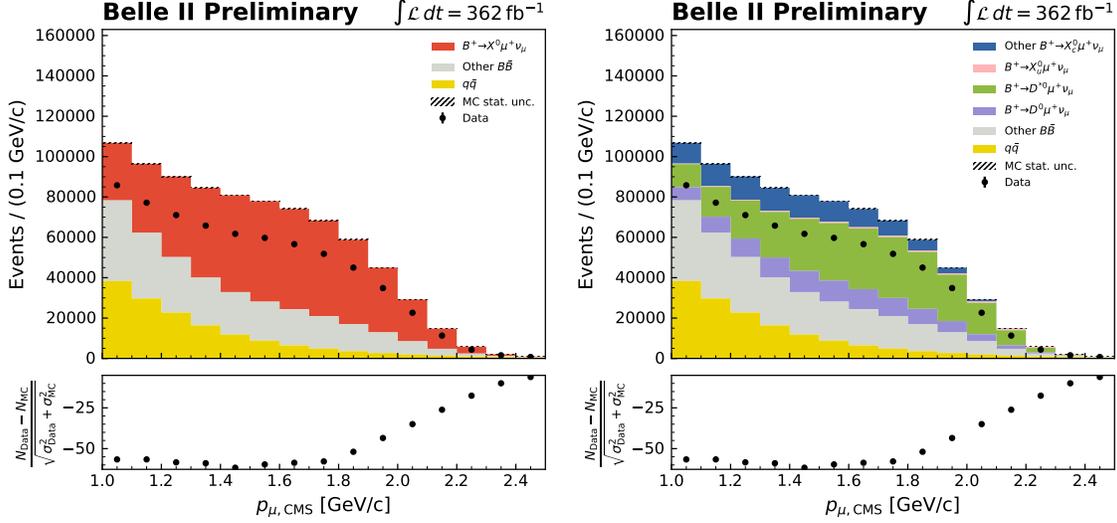

Figure 4.6: Distribution of the signal-side lepton CMS momentum $p_{\ell,\text{CMS}}$ in data against normalised MC for $B^+ \rightarrow X^0 e^+ \nu_e$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

decays constitute the dominant signal component. In each distribution, the MC has been normalised to the data integrated luminosity.

In each of the given distributions, a significant discrepancy between data and MC can be observed, a direct result of the difference in efficiency between the application of the FEI to real vs. simulated data. To perform the calibration, the MC was fitted to the given data and an estimate of the number of signal events in data was extracted via a strategy known

Figure 4.7: Distribution of the signal-side lepton CMS momentum $p_{\ell,\mathrm{CMS}}$ in data against normalised MC for $B^+ \to X^0\mu^+\nu_\mu$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

as a binned maximum likelihood template fit. In this strategy, the $p_{\ell,\mathrm{CMS}}$ distributions of the three MC components described above were used to generate probability-density-functions (pdfs) of the form $p(x|\theta)$, which describe the probability of the occurrence of a set of observations x given a set of parameters θ. In this case, the observed data x refers to the number of events measured in each bin of the $p_{\ell,\mathrm{CMS}}$ distribution, with the parameters θ including the relative yields of the individual components: signal, $B\bar{B}$ background, and continuum, and the relative fractions of the four components of the signal contribution: $f_{D*\ell\nu_\ell}$, $f_{D\ell\nu_\ell}$, $f_{X_u\ell\nu_\ell}$ and $f_{\mathrm{other}X_c\ell\nu_\ell}$. As the shapes of the $p_{\ell,\mathrm{CMS}}$ distributions in simulation are used to define the model that is fit to data, in contrast to a specified functional form such as a Gaussian, such a fit is classified as a template fit, with the histograms representing the shapes of the individual MC components referred to as templates.

In the maximum likelihood estimation, a likelihood function $L(\theta) = p(x|\theta)$ is defined, which, taking the observed data x as given, characterises how likely a given set of parameters θ represents the observed data. The given model is then fitted to the data by finding the values of the parameters that maximise the likelihood function [84]. For this analysis, the RooFit software [85] was utilised to perform the maximum likelihood fits to data, and the relative yields and signal component fractions were extracted. The maximum likelihood fitting algorithm provided by RooFit returns each fitted parameter with an associated uncertainty derived from the finite statistical size of the samples used in the fit.

Several constraints were also imposed during the fit, with each constraint implemented by taking the product of the template pdfs and a Gaussian distribution with mean and width defined accordingly. Firstly, the relative fractions of the signal component of the fit were constrained in order to maintain the composition of the total $B \to X\ell\nu_\ell$ branching fraction

73

observed in nature. To constrain the $B \to D\ell\nu_\ell$, $B \to D^*\ell\nu_\ell$ and $B \to X_u\ell\nu_\ell$ fractions to their expectations, the means of the Gaussian distributions were set to the nominal fractions derived from MC, with the widths set to the uncertainties on these fractions assuming the branching fraction uncertainties as listed in Tables 4.2 and 4.3. A constraint term for the remaining component of the signal, other $B \to X_c\ell\nu_\ell$ decays, was not implemented in the fit as this was implicitly achieved through constraining three out of the four fractions comprising the signal template.

Furthermore, in order to achieve good discrimination between the shapes of the generic $B\bar{B}$ background and continuum templates, the available off-resonance data-set was used to impose a Gaussian constraint on the continuum template. The off-resonance prediction for the number of continuum events in MC after all analysis selections, as introduced in the previous section, was taken as the mean of the Gaussian, with the width taken from the propagation of the Poisson uncertainty on the number of off-resonance data events post-selections.

The values of the parameters returned from the fits are listed in Tables 4.4 and 4.5 for reconstructed $B^0 \to X^-\ell^+\nu_\ell$ and $B^+ \to X^0\ell^+\nu_\ell$ decays, respectively. The signal, $B\bar{B}$ and continuum background yields are included, as well as the fitted fractions $f_{D^*\ell\nu_\ell}$, $f_{D\ell\nu_\ell}$ and $f_{X_u\ell\nu_\ell}$. The fourth fraction of the signal component, $f_{\mathrm{other}X_c\ell\nu_\ell}$ is also listed, calculated from the fitted values of the other fractions as $1 - f_{D^*\ell\nu_\ell} - f_{D\ell\nu_\ell} - f_{X_u\ell\nu_\ell}$. For each fitted parameter derived from data, the expectation based on MC is included for comparison. The fits confirm that the MC significantly overestimates the data, with larger expected than observed signal and $B\bar{B}$ yields. Given the constraints on the relative fractions of the signal component and the continuum contribution, little variation is observed between the MC expectation and the observed data for these two contributions.

The resultant fitted $p_{\ell,\mathrm{CMS}}$ distributions are given in Figures 4.8, 4.9, 4.10 and 4.11, for reconstructed $B^0 \to X^-e^+\nu_e$, $B^0 \to X^-\mu^+\nu_\mu$, $B^+ \to X^0e^+\nu_e$ and $B^+ \to X^0\mu^+\nu_\mu$, respectively. The fits significantly improve the agreement between data and simulation. The right-most plot of each Figure depicts the signal component broken up into the relative contributions of the various semi-leptonic decays, with the composition varied slightly from the pre-fit case as described.

4.1.5 Calibration Factors

The numbers of signal events in MC and data were subsequently used to compute the FEI calibration factors ϵ for each $B \to X\ell\nu_\ell$ mode studied using the following formula:

$$\epsilon = \frac{N_{\mathrm{sig}}^{\mathrm{Data}}}{N_{\mathrm{sig}}^{\mathrm{MC}}},$$

where $N_{\mathrm{sig}}^{\mathrm{Data}}$ refers to the fitted signal yield derived from data, and $N_{\mathrm{sig}}^{\mathrm{MC}}$ refers to the total number of signal events in the MC, weighted by all of the various analysis corrections

Parameter	MC prediction	Fitted data result
$B^0 \to X^- e^+ \nu_e$		
Signal yield	188170	180590 ± 2397.6
$f_{D*e\nu_e}$	0.564	0.531 ± 0.008
$f_{De\nu_e}$	0.208	0.202 ± 0.006
$f_{X_u e\nu_e}$	0.019	0.016 ± 0.001
$f_{\text{other}X_c e\nu_e}$	0.210	0.251 ± 0.010
Other $B\bar{B}$ yield	111310	53932 ± 2681.3
Continuum yield	43668	43540 ± 603.59
$B^0 \to X^- \mu^+ \nu_\mu$		
Signal yield	162020	138050 ± 1781.9
$f_{D*\mu\nu_\mu}$	0.570	0.578 ± 0.008
$f_{D\mu\nu_\mu}$	0.206	0.207 ± 0.006
$f_{X_u \mu\nu_\mu}$	0.019	0.015 ± 0.002
$f_{\text{other}X_c \mu\nu_\mu}$	0.205	0.200 ± 0.011
Other $B\bar{B}$ yield	107910	74355 ± 2143.7
Continuum yield	52298	52296 ± 662.25

Table 4.4: Yields and signal component fractions returned from fits to data for $B^0 \to X^- \ell^+ \nu_\ell$ decays, compared with predictions from MC.
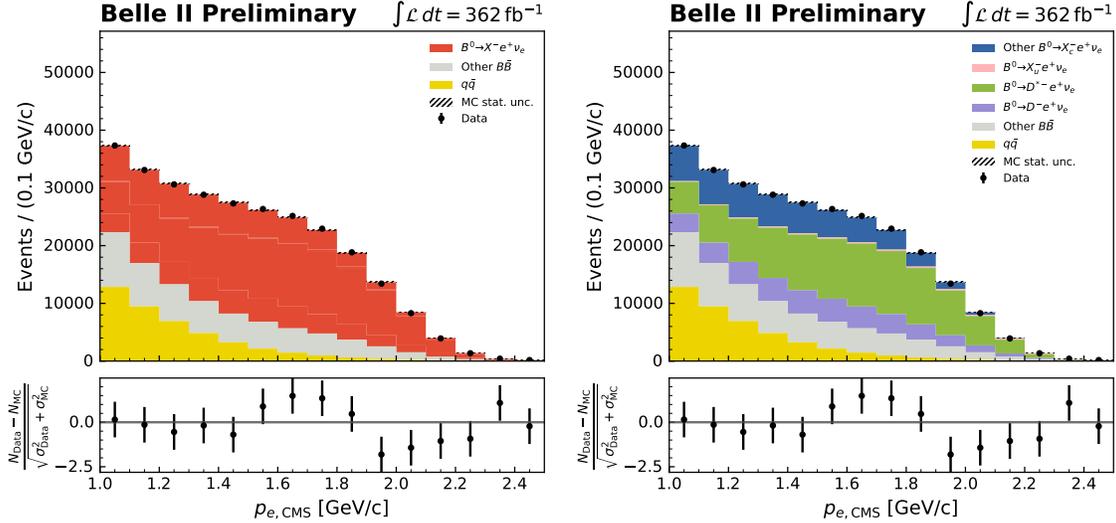
Figure 4.8: Distribution of the signal-side lepton CMS momentum $p_{\ell,\text{CMS}}$ with MC fitted to data, for $B^0 \to X^- e^+ \nu_e$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

Parameter	MC prediction	Fitted data result
$B^+ \to X^0 e^+ \nu_e$		
Signal yield	519770	460750 ± 9301.5
$f_{D^*e\nu_e}$	0.569	0.531 ± 0.007
$f_{De\nu_e}$	0.206	0.200 ± 0.006
$f_{X_u e\nu_e}$	0.018	0.019 ± 0.001
$f_{\mathrm{other}X_c e\nu_e}$	0.207	0.250 ± 0.009
Other $B\bar{B}$ yield	234790	82846 ± 11148
Continuum yield	119490	119620 ± 2398.4
$B^+ \to X^0 \mu^+ \nu_\mu$		
Signal yield	451080	350870 ± 2792.6
$f_{D^*\mu\nu_\mu}$	0.573	0.598 ± 0.006
$f_{D\mu\nu_\mu}$	0.206	0.208 ± 0.006
$f_{X_u \mu\nu_\mu}$	0.019	0.014 ± 0.001
$f_{\mathrm{other}X_c \mu\nu_\mu}$	0.203	0.181 ± 0.009
Other $B\bar{B}$ yield	233700	143030 ± 3371.6
Continuum yield	156390	156240 ± 1136.2

Table 4.5: Yields and signal component fractions returned from fits to data for $B^+ \to X^0 \ell^+ \nu_\ell$ decays, compared with predictions from MC.

Signal mode	Calibration factor
$B^0 \to X^- e^+ \nu_e$	$0.96 \pm 0.01(\mathrm{stat}) \pm 0.03(\mathrm{sys})$
$B^0 \to X^- \mu^+ \nu_\mu$	$0.85 \pm 0.01(\mathrm{stat}) \pm 0.03(\mathrm{sys})$
Average $B^0 \to X^- \ell^+ \nu_\ell$	$0.91 \pm 0.01(\mathrm{stat}) \pm 0.03(\mathrm{sys})$
$B^+ \to X^0 e^+ \nu_e$	$0.89 \pm 0.02(\mathrm{stat}) \pm 0.02(\mathrm{sys})$
$B^+ \to X^0 \mu^+ \nu_\mu$	$0.78 \pm 0.01(\mathrm{stat}) \pm 0.03(\mathrm{sys})$
Average $B^+ \to X^0 \ell^+ \nu_\ell$	$0.84 \pm 0.01(\mathrm{stat}) \pm 0.02(\mathrm{sys})$

Table 4.6: Calibration factors determined from the analysis of $B \to X \ell \nu_\ell$ decays.

described in Section 4.1.3 and normalised to the data integrated luminosity. The resultant calibration factors are listed in Table 4.6 for the reconstruction of $B^0 \to X^- e^+ \nu_e$, $B^0 \to X^- \mu^+ \nu_\mu$, $B^+ \to X^0 e^+ \nu_e$ and $B^+ \to X^0 \mu^+ \nu_\mu$ decays. Each calibration factor is assigned both a statistical and systematic uncertainty. The statistical uncertainties represent the errors on the central values as a result of the size of the given data sample, and were determined by replacing the numerator in the above equation with the uncertainty on the signal yield returned from the fit. Systematic uncertainties correspond to additional sources of uncertainty related to limitations in the detector resolution and apparatus, theoretical
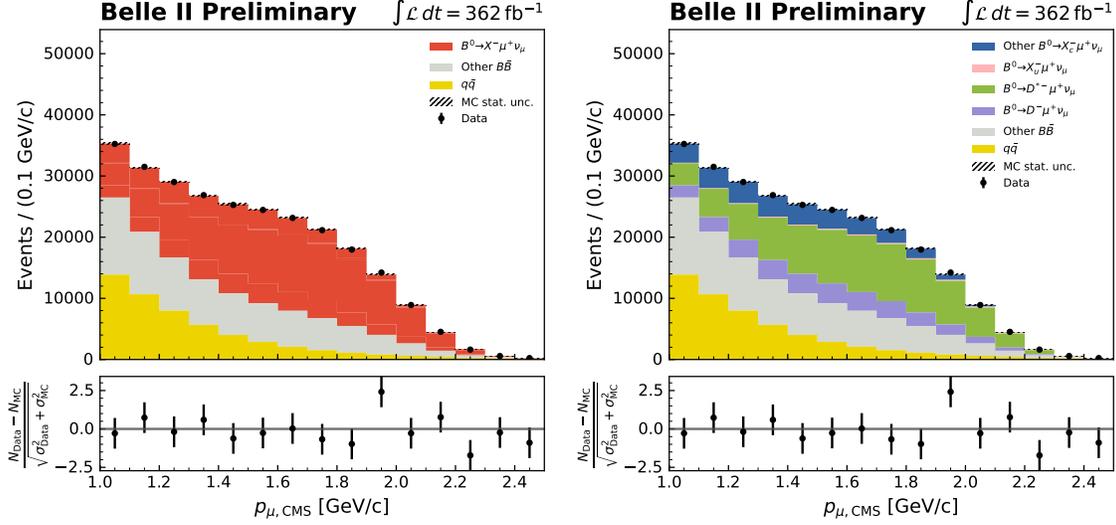
Figure 4.9: Distribution of the signal-side lepton CMS momentum $p_{\ell,\text{CMS}}$ with MC fitted to data, for $B^0 \to X^- \mu^+ \nu_\mu$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

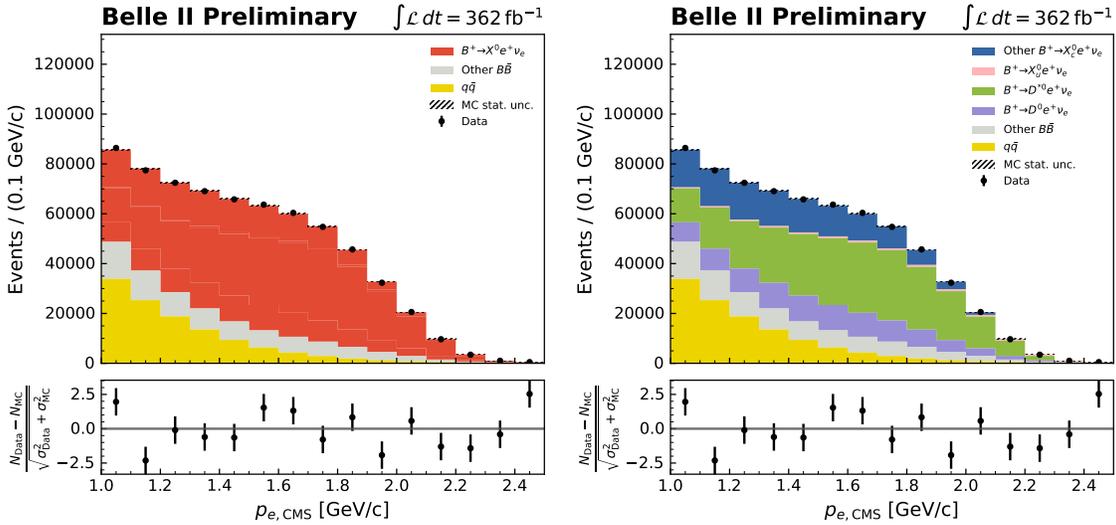Figure 4.10: Distribution of the signal-side lepton CMS momentum $p_{\ell,\text{CMS}}$ with MC fitted to data, for $B^+ \to X^0 e^+ \nu_e$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

models used for the MC and other external effects. A full list of the systematic uncertainties considered for this analysis is provided in the following section, together with a description of how each uncertainty was determined.

A sizeable difference between the factors determined from the electron and muon channels is observed for both charged and neutral $B \to X\ell\nu_\ell$ decays, with the electron channel exhibiting better agreement between the signal yields derived from data and MC. Whilst

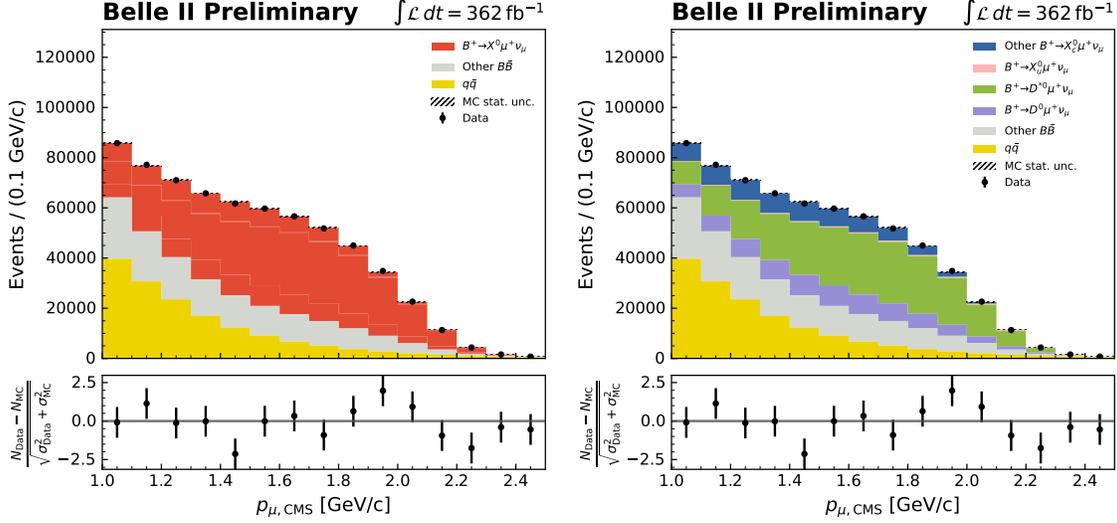Figure 4.11: Distribution of the signal-side lepton CMS momentum $p_{\ell,\text{CMS}}$ with MC fitted to data, for $B^+ \to X^0 \mu^+ \nu_\mu$ decays. The distribution is shown for the signal MC as a whole (left) and split into its four contributions (right).

no discrepancy between the lepton flavours is expected due to the lepton universality described in the Standard Model [28], several differences exist in the way in which decays involving electrons and muons are reconstructed in this analysis. As one example, the lepton identification corrections applied to the MC are evaluated for the two lepton flavours separately via different decay channels, offering a potential source of discrepancy. Additionally, the electron momentum, which is used for the signal extraction, is corrected to account for bremsstrahlung radiation. Finally, the muon channels correspond to substantially higher continuum backgrounds than the electron channels, with the relative proportion of these backgrounds constrained during the fitting procedure. A combination of one or more of these differences in the reconstruction could explain the observed discrepancies between lepton flavours, with further investigation forming the basis of a potential future study.

The individual calibration factors evaluated for the electron and muon channels were then averaged to obtain an estimate for the calibration factors from $B^0 \to X^- \ell^+ \nu_\ell$ and $B^+ \to X^0 \ell^+ \nu_\ell$ decays, with these values also listed in Table 4.6. In combining the electron and muon channels, the statistical uncertainties were combined in quadrature under the assumption that each data sample was statistically independent. The systematic uncertainties were similarly propagated for those sources deemed independent between the two channels, namely those due to the MC statistics and the lepton identification efficiency. All remaining systematic uncertainties were assumed to be completely correlated between the electron and muon channels, with the uncertainties on the average factors determined as $\sigma_{\text{avg}} = \sqrt{\sigma_e^2 + \sigma_\mu^2 + 2\sigma_e\sigma_\mu}/2$, where σ_e and σ_μ are the systematic uncertainties of the electron and muon channels, respectively.

4.1.6 Systematic Uncertainties

Table 4.7 lists the relative systematic uncertainties on the calibration factors for a number of independent sources, described in detail below:

- $B \to X\ell\nu_\ell$ **branching fraction:** The relative uncertainty on the total $B \to X\ell\nu_\ell$ branching fraction forms the dominant source of systematic uncertainty for the analysis. For $B^0 \to X^-\ell^+\nu_\ell$ and $B^+ \to X^0\ell^+\nu_\ell$, the world average measurements correspond to uncertainties of 2.71% and 2.55%, respectively, with $\mathcal{B}(B^0 \to X^-\ell^+\nu_\ell) = (10.33 \pm 0.28)\%$ and $\mathcal{B}(B^+ \to X^0\ell^+\nu_\ell) = (10.99 \pm 0.28)\%$ [20].

- **MC statistics:** An additional systematic uncertainty for this analysis arises from the finite sample sizes of the MC used. In order to estimate the magnitude of this uncertainty, the number of signal MC events post-selections was taken as the mean of a Gaussian distribution with a width corresponding to the Poissonian uncertainty on this given event yield. 200 variations on the number of signal events were then sampled from this Gaussian distribution and the calibration factors were recalculated for each of these variations. The magnitude of the systematic uncertainty was then determined from the ratio between the standard deviation of the resultant calibration factors and the nominal result.

- **Tracking:** The efficiency with which tracks are reconstructed at Belle II similarly differs between data and MC. To account for this effect, official recommendations specify that a systematic uncertainty of 0.24% should be assigned per reconstructed track for all Belle II analyses. The value of this uncertainty was determined via a dedicated study on tracks reconstructed from $e^+e^- \to \tau^+\tau^-$ decays.

- **Lepton identification:** The values of the correction factors determined for the lepton identification efficiency, evaluated in bins of the lepton momentum and polar angle as detailed in Section 4.1.3, each possess an associated uncertainty. In a similar manner to the systematic uncertainties derived from the MC statistics, using the software package in Ref. [80], 200 variations on the nominal correction weights were generated via sampling from a Gaussian distribution with width given by the uncertainties on these weights, with the total systematic uncertainty taken from the standard deviation of the varied event yields.

- $B \to D^{(*)}\ell\nu_\ell$ **form factors:** The values of the parameters used in the form factor parameterisation of the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ decays described in Section 4.1.3 are likewise associated with particular uncertainties. To determine the magnitude of the systematic uncertainty on the calibration factors as a result of these form factor uncertainties, the eFFORT package [72] was used to generate varied form factor correction weights. Each parameter was modified individually by either adding or subtracting its uncertainty from the central value, and the form factor weights were recalculated for each of these up- and down-variations. Under the BGL parameterisation used [76], the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ form factors are described by 4 and 6 parameters, respectively, resulting in a total of 8 and 12 varied form factor weights. The numbers

of weighted signal MC events were then determined for each of these varied form factor weights. To conservatively estimate the systematic uncertainties due to these form factor contributions, the maximum difference observed between the varied signal MC yields and the nominal yield was computed, with the ratio between this difference and the nominal yield taken as the systematic uncertainty. Similar systematic uncertainties were computed for the form factor variations of $B \to X_u \ell \nu_\ell$ modes, but these were found to be negligible.

The total systematic uncertainties for this analysis were subsequently determined by adding the above contributions in quadrature, under the assumption that each of the individual sources were completely independent from one another.

Source	Systematic uncertainty (%)			
	$B^0 \to X^- e^+ \nu_e$	$B^0 \to X^- \mu^+ \nu_\mu$	$B^+ \to X^0 e^+ \nu_e$	$B^+ \to X^0 \mu^+ \nu_\mu$
$B \to X\ell\nu_\ell$ branching fraction	2.71	2.71	2.55	2.55
MC statistics	0.16	0.18	0.10	0.11
Tracking	0.24	0.24	0.24	0.24
LeptonID	0.57	1.21	0.57	1.21
$B \to D\ell\nu_\ell$ form factors	0.04	0.04	0.04	0.05
$B \to D^*\ell\nu_\ell$ form factors	0.61	0.61	0.60	0.61
Total	2.85	3.04	2.69	2.90

Table 4.7: Systematic uncertainties for the measured calibration factors in % from a number of identified sources.

4.1.7 Belle II Official FEI Calibration

Some time after the commencement of this study, an official FEI task force was established within the Belle II collaboration in order to manage and coordinate all efforts related to the calibration and development of the FEI. One main goal of this task force was to provide official, standardised FEI calibration factors for all users to access and utilise in their individual analyses. As per the recommendation of the collaboration, these official calibration factors were used to correct the MC for the $B \to \pi\ell\nu_\ell$ analysis presented in the forthcoming chapters. The study detailed in this chapter provides an illustrative example of the method in which the FEI calibration can be performed. In the discussion below, the calibration factors produced from this study are compared to the values given by the official recommendations.

Within the FEI task force, two independent calibration studies were performed for the hadronic FEI, each adopting a similar method to the analysis described throughout this chapter. The first effort similarly involved the reconstruction of a $B \to X\ell\nu_\ell$ signal-side decay recoiling against a hadronic tag, utilising the large branching fraction of this decay in order to maximise the available statistics. Whilst the choice of signal decay mode is

common between this calibration study and the analysis forming the basis of this chapter, several differences exist regarding the event selections and the fitting procedure, making a direct comparison impractical.

The second calibration method opted for a much purer signal-side decay, with the signal B-meson decaying to a D or D^* meson and a pion, $B \to D^{(*)}\pi$. In this study, after hadronic B-tagging via the FEI, a single pion was reconstructed on the signal side. The momenta of this B_{tag} and pion were used to determine the mass of the recoiling particle (either the D or D^* meson), with this recoil mass, M_{recoil}, used for the signal extraction. In this way, the $D^{(*)}$ meson was not required to be explicitly reconstructed on the signal side.

Hadronic signal-side decays offer a strong advantage for calibration studies due to the low backgrounds associated with fully reconstructing the event. However, the high purity comes at the cost of a low reconstruction efficiency, which, together with the low branching fractions characteristic of individual hadronic B-meson decays, results in sample sizes that are quite statistically limited. The $B \to D^{(*)}\pi$ calibration study thus aimed for a middle approach, partially reconstructing the signal-side decay to increase the statistical power of the calibration samples whilst maintaining a high purity. As a result, the two independent calibration studies complemented one another well, representing signal-side decays with both a high efficiency ($B \to X\ell\nu_\ell$) and a high purity ($B \to D^{(*)}\pi$). The calibration factors resulting from both studies were then combined to produce a final set of official calibration factors for use in Belle II analyses.

Both studies performed the calibration for individual B_{tag} decay modes (or subsets of modes), with the integrated luminosity of the current data sample allowing for such an extraction for the first time since the commencement of Belle II data taking. A total of 10 neutral and 11 charged hadronic B_{tag} decay modes were individually calibrated, with a single factor determined as an average over all of the remaining modes. As a secondary result, both calibration studies also quoted a global calibration factor determined for all tagging modes, analogous to the study presented in this chapter. Table 4.8 lists the global calibration factors extracted from the official $B \to X\ell\nu_\ell$ and $B \to D^{(*)}\pi$ studies alongside those obtained from this analysis. Only the calibration factors split by B_{tag} decay mode were produced for the combination of the two official studies, and an estimate of the global calibration factor for this combination was not available for comparison.

In comparing the calibration factors derived from $B \to X\ell\nu_\ell$ in this analysis and the official calibration, significant discrepancies are observed, with the presented study measuring larger factors overall. Known differences in the analysis selections as well as the signal extraction strategy between the two methods are the likely cause for the observed disagreement. The official $B \to X\ell\nu_\ell$ calibration additionally does not observe a large discrepancy between the factors obtained from the electron and muon modes. The results of the analysis in this chapter are likewise inconsistent with the global calibration factors derived from the $B \to D^{(*)}\pi$ study, which agree relatively well with the official $B \to X\ell\nu_\ell$ factors with the exception of $B^+ \to \bar{D}^0\pi^+$, which exhibits some tension with the rest of the official measurements. As stated prior, the study presented in this chapter is illustrative of the methodology

and technique with which the FEI tagging algorithm can be calibrated, and the official recommendations, which have been substantially peer-reviewed and approved by the collaboration, will be utilised instead for the main $B \to \pi \ell \nu_\ell$ analysis detailed in the coming chapters. Potential further study on the results presented in this chapter would focus on adapting the analysis selections and signal extraction to better match the official $B \to X \ell \nu_\ell$ analysis, and investigating the source of the discrepancy observed between the electron and muon channels.

Signal mode	This analysis	Official calibration
$B^0 \to X^- e^+ \nu_e$	$0.96 \pm 0.01(\text{stat}) \pm 0.03(\text{sys})$	0.80 ± 0.02
$B^0 \to X^- \mu^+ \nu_\mu$	$0.85 \pm 0.01(\text{stat}) \pm 0.03(\text{sys})$	0.77 ± 0.02
$B^0 \to D^- \pi^+$	-	0.81 ± 0.03
$B^0 \to D^{*-} \pi^+$	-	0.75 ± 0.04
$B^+ \to X^0 e^+ \nu_e$	$0.89 \pm 0.02(\text{stat}) \pm 0.02(\text{sys})$	0.73 ± 0.02
$B^+ \to X^0 \mu^+ \nu_\mu$	$0.78 \pm 0.01(\text{stat}) \pm 0.03(\text{sys})$	0.72 ± 0.02
$B^+ \to \bar{D}^0 \pi^+$	-	0.62 ± 0.02
$B^+ \to \bar{D}^{*0} \pi^+$	-	0.72 ± 0.02

Table 4.8: Calibration factors determined from the presented study compared with those from the official Belle II FEI calibration.

Chapter 5

Event Selection and Optimisation

The two previous chapters aimed at providing a detailed description of the full-event-interpretation (FEI), the state-of-the-art algorithm developed for tagged analysis at Belle II. Using simulated and real data samples that have been skimmed with the FEI and armed with a set of correction factors used to calibrate the algorithm, the remaining chapters of this thesis will focus on the hadronically tagged analysis of semi-leptonic $B \to \pi \ell \nu_\ell$ decays.

With the ultimate aim of measuring the branching fractions of the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays and extracting the magnitude of the CKM matrix element $|V_{\mathrm{ub}}|$, the core foundations of the analysis must first be established. These, in essence, are the selections applied during the event reconstruction, which must be carefully chosen and optimised in order to maximise the discrimination between signal and background events whilst retaining a suitable reconstruction efficiency. After introducing the composition of the MC samples used for the analysis and any relevant data-MC corrections, the following sections will detail the chosen strategy for the optimisation of the event selections.

5.1   Description of Monte Carlo

As for the calibration study presented in the previous chapter, MC samples of generic mixed, $\Upsilon(4S) \to B^0 \bar{B}^0$, charged $\Upsilon(4S) \to B^+ B^-$, and continuum $e^+ e^- \to u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ events were utilised to simulate the data in the analysis. The samples chosen corresponded to a total integrated luminosity of 2.8 ab^{-1} of $B\bar{B}$ MC and 1 ab^{-1} of continuum MC. The same real data sample was used, corresponding to 362 fb^{-1}, the complete physics data-set recorded by Belle II before the first long shutdown of the experiment. All samples were skimmed using the hadronic FEI according to the procedure detailed in Section 3.6.3. Table 5.1 lists the integrated luminosity and equivalent number of events before all selections including the application of the FEI, for each MC sample used in the analysis.

To better simulate signal events as well as background events from other $B \to X_u \ell \nu_\ell$ decays, additional large MC samples of $B^0 \to X_u^- \ell \nu_\ell$ and $B^+ \to X_u^0 \ell \nu_\ell$ decays were used, each containing 100 million events. As these decays comprised a relatively small subset of the generic $B\bar{B}$ MC due to their low branching fractions, these large samples ensured sufficient

MC sample	Integrated luminosity (ab^{-1})	N_{events} used ($\times 10^6$)
B^+B^-	2.8	1512
$B^0\bar{B}^0$	2.8	1428
$c\bar{c}$	1.0	1329
$u\bar{u}$	1.0	1605
$s\bar{s}$	1.0	383
$d\bar{d}$	1.0	401
$B^0 \to X_u^- \ell\nu_\ell$ resonant	55.8	50
$B^0 \to X_u^- \ell\nu_\ell$ non-resonant	13.9	50
$B^+ \to X_u^0 \ell\nu_\ell$ resonant	55.6	50
$B^+ \to X_u^0 \ell\nu_\ell$ non-resonant	12.1	50

Table 5.1: Integrated luminosity and corresponding number of events for each MC sample studied, before any selections including the application of the FEI skims.

MC statistics were available for the study. The $B \to X_u\ell\nu_\ell$ events in the generic $B\bar{B}$ MC were removed and replaced with these samples, scaled down to the appropriate integrated luminosity.

The use of the dedicated $B \to X_u\ell\nu_\ell$ samples provided another main advantage related to the relative resonant and non-resonant contributions to the total $B \to X_u\ell\nu_\ell$ branching fractions. These samples allowed for the combination of the resonant and non-resonant events in a way motivated by theory to form what is known as the hybrid MC, described in detail below [86].

5.1.1 Hybrid MC

As discussed in Section 4.1.3 in the context of the FEI calibration study, the branching fractions of the non-resonant contributions to the total measured $B \to X_u\ell\nu_\ell$ inclusive branching fractions, $\mathcal{B}(B \to X_{u,\mathrm{gap}}\ell\nu_\ell)$, are substantially overestimated in the Belle II MC generation. To combat this in the calibration study, the non-resonant branching fractions were simply reweighted such that the sum of the resonant and non-resonant contributions matched the world average measurement of the total inclusive rate [20]. This reweighting was done via the application of a single correction factor equivalent to the ratio of the updated branching fractions and those used for the MC generation, $w_{\mathrm{BF}} = \mathcal{B}_{\mathrm{PDG}}/\mathcal{B}_{\mathrm{MC}}$.

Whilst this approach was more than suitable for the FEI calibration study given that the $B \to X_u\ell\nu_\ell$ decays contributed minimally to the signal-side reconstruction of $B \to X\ell\nu_\ell$, an alternative approach was implemented for the main $B \to \pi\ell\nu_\ell$ analysis where the modelling of the signal and background $B \to X_u\ell\nu_\ell$ decays was crucial. In this approach, the reweighting of the non-resonant contribution was not performed using a global scale factor, but rather, on an event-by-event basis according to the values of three key kinematic quanti-

ties: the lepton energy in the B-meson frame, E_l^B, the mass of the hadronic system M_{X_u} and the square of the four-momentum transfer to the leptonic system, q^2. Using these variables, a set of 3-dimensional bins was established, with theoretically motivated bin boundaries [86].
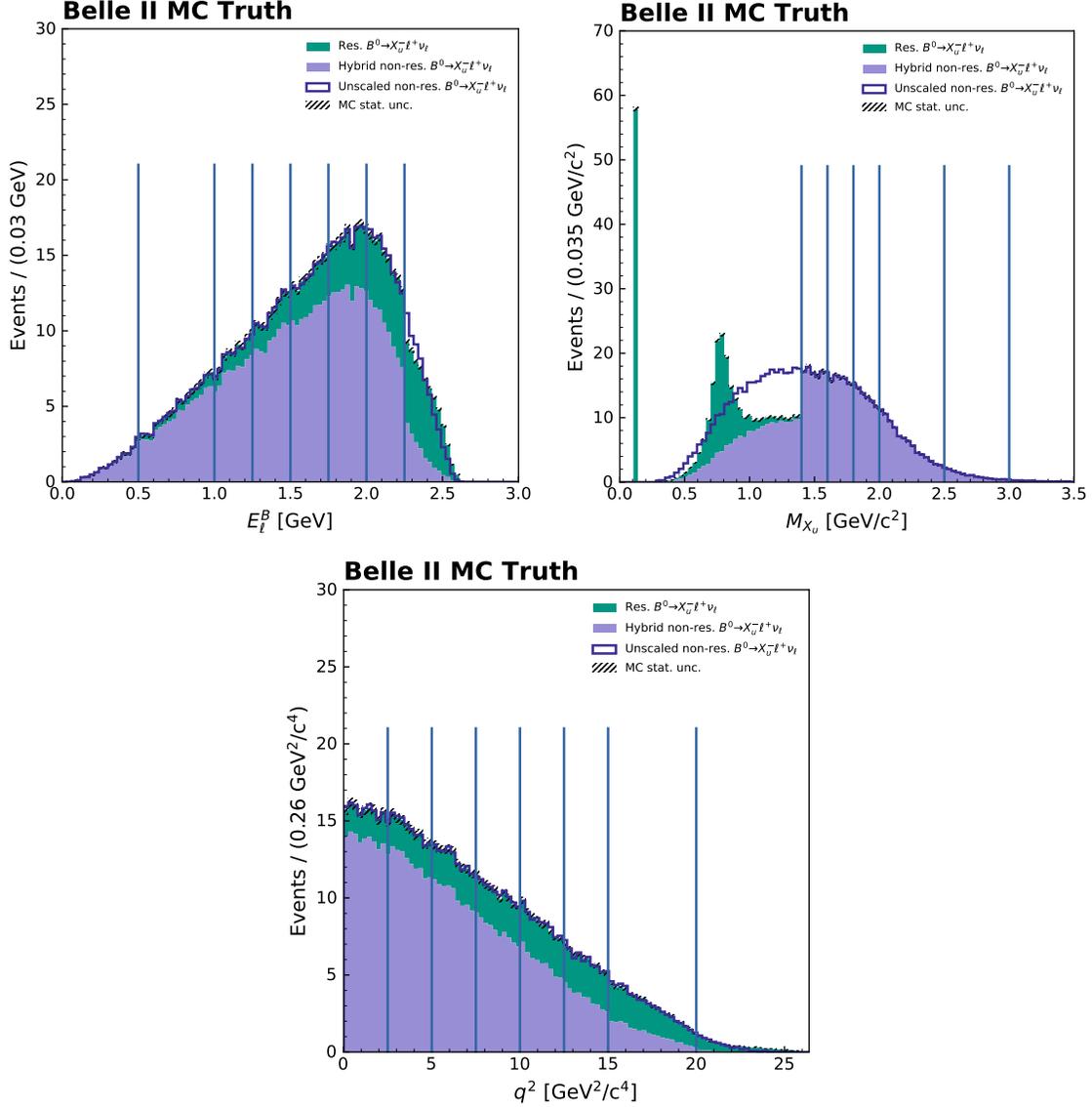

Figure 5.1: Distributions of the kinematic variables E_l^B, M_{X_u} and q^2 in MC truth for the hybrid $B^0 \to X_u^- \ell^+ \nu_\ell$ sample. The vertical lines represent the projection of the 3-dimensional bin boundaries.

To form the hybrid MC samples of $B^0 \to X_u^- \ell^+ \nu_\ell$ and $B^+ \to X_u^0 \ell^+ \nu_\ell$, 50 million $B\bar{B}$ events with at least one non-resonant B-meson decay and 50 million events with at least one B-meson decaying to a resonant final state were combined. The non-resonant events were generated using the BLNP model defined in Ref. [87], which defines the shape of the non-resonant contribution over the full kinematic region studied. Using the eFFORT software

package [72], the events in each sample were divided into 3-dimensional bins according to the values of E_l^B, M_{X_u} and q^2 given in the MC truth. A weight for each non-resonant event was then defined such that the sum of the resonant and non-resonant contributions in each bin matched the total inclusive rate according to the following relation:

$$H_i = R_i + w_i N_i,$$

where H_i refers to the number of total hybrid events per bin i, R_i and N_i are the number of resonant and non-resonant events per bin, and w_i is the derived per-bin weight. The non-resonant $B \to X_u \ell \nu_\ell$ events in the MC analysis sample were thus reweighted by applying the hybrid weight for each B-meson in the event decaying via the $\mathcal{B}(B \to X_{u,\text{gap}} \ell \nu_\ell)$ modes.

Figures 5.1 and 5.2 display the distributions in MC truth of E_l^B, M_{X_u} and q^2 for the hybrid $B^0 \to X_u^- \ell^+ \nu_\ell$ and $B^+ \to X_u^0 \ell^+ \nu_\ell$ samples, respectively. The filled histograms illustrate the resonant and non-resonant contributions of the hybrid, with the latter weighted down according to the above procedure. The effect of the hybrid reweighting can be clearly seen with the inclusion of the unscaled non-resonant component in each distribution. The projection of the 3-dimensional bin boundaries for each variable are also displayed in the plots as vertical lines.

Whilst the resonant and non-resonant contributions of $B \to X_u \ell \nu_\ell$ are largely kinematically similar in the distributions of E_l^B and q^2, they exhibit substantially different shapes in M_{X_u}. The individual resonant $B \to X_u \ell \nu_\ell$ decays can be clearly discerned from the non-resonant component, appearing as spikes in the distribution of varying widths. For $B^0 \to X_u^- \ell^+ \nu_\ell$ in Figure 5.1, these correspond to a sharp, narrow peak for the $B^0 \to \pi^- \ell^+ \nu_\ell$ decay to the scalar charged pion of mass 140 MeV [20], and the broad peak characteristic of the vector ρ-meson in the $B^0 \to \rho^- \ell^+ \nu_\ell$ decay, centered at 775 MeV [20]. The peaks corresponding to the $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \rho^0 \ell^+ \nu_\ell$ decays can be similarly observed in Figure 5.2, with the addition of the $B^+ \to \eta \ell^+ \nu_\ell$ resonance at 548 MeV, the $B^+ \to \eta' \ell^+ \nu_\ell$ resonance at 958 MeV, and the $B^+ \to \omega \ell^+ \nu_\ell$ resonance at 783 MeV, with the latter unable to be clearly discerned from the broad $B^+ \to \rho^0 \ell^+ \nu_\ell$ peak [20]. The reweighting using the given bin boundaries produces a sharp feature at $M_{X_u} = 1.5$ GeV, though this ultimately will be smeared out during the reconstruction due to the finite detector resolution. In general, this feature is unproblematic so long as no event selections are made at this boundary.

As mentioned earlier, all $B \to X_u \ell \nu_\ell$ decays in the generic $B\bar{B}$ MC were removed and replaced with the hybrid samples. In order to scale this hybrid MC down to match the integrated luminosity of the $B\bar{B}$ MC, additional scaling factors were applied, defined as:

$$\text{SF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \text{res}} = \frac{2\text{BF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \text{res}} N_{B^0 \bar{B}^0}}{N_{B^0 \to X_u^- \ell^+ \nu_\ell, \text{res}}},$$

$$\text{SF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \text{non-res}} = \frac{2\text{BF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \text{non-res}} N_{B^0 \bar{B}^0}}{N_{B^0 \to X_u^- \ell^+ \nu_\ell, \text{non-res}}},$$
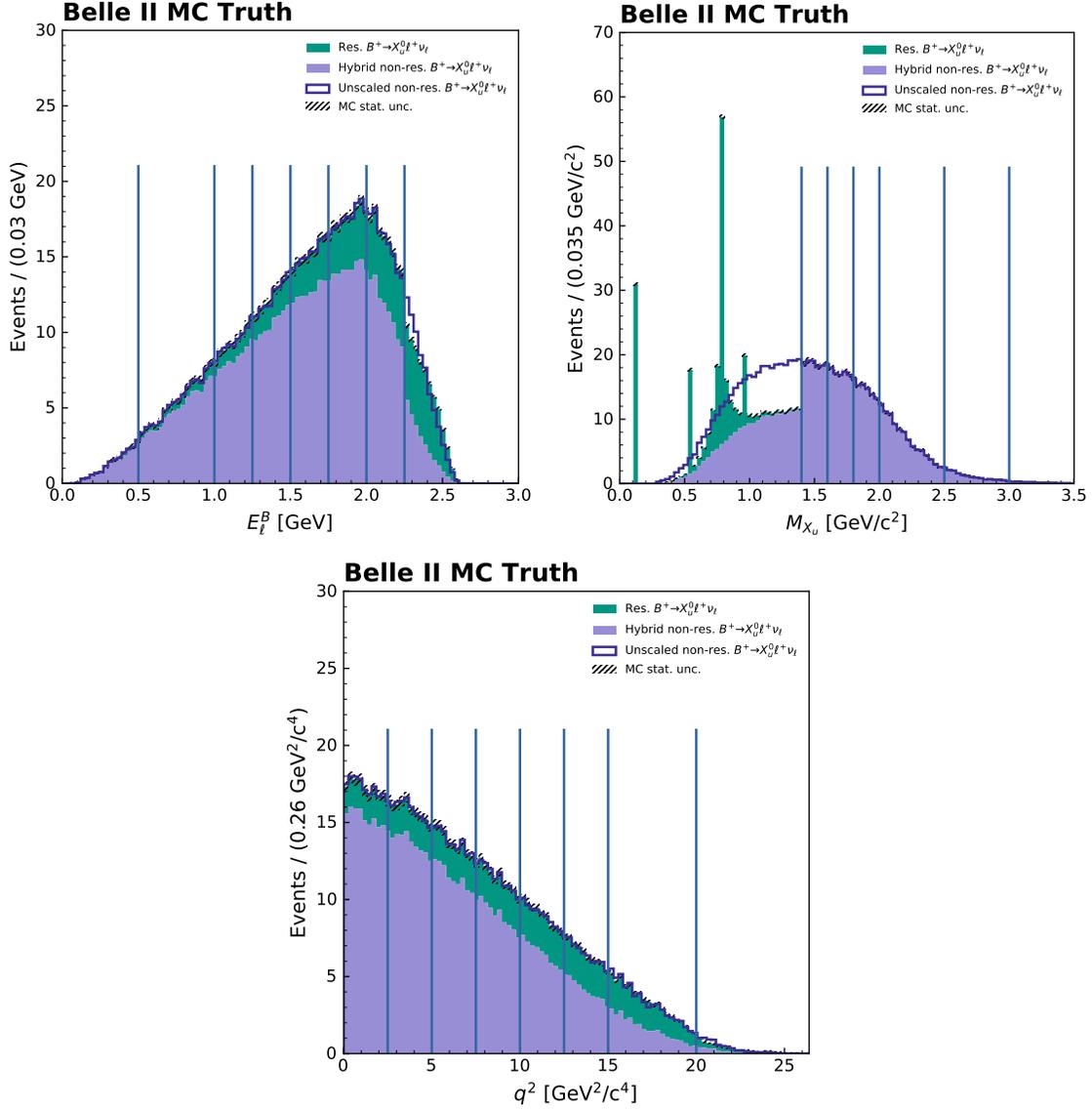
86

Figure 5.2: Distributions of the kinematic variables E_l^B, M_{X_u} and q^2 in MC truth for the hybrid $B^+ \to X_u^0 \ell^+ \nu_\ell$ sample. The vertical lines represent the projection of the 3-dimensional bin boundaries.

$$\mathrm{SF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{res}} = \frac{2\mathrm{BF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{res}} N_{B^+ B^-}}{N_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{res}}},$$

and

$$\mathrm{SF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{non\text{-}res}} = \frac{2\mathrm{BF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{non\text{-}res}} N_{B^+ B^-}}{N_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{non\text{-}res}}},$$

where $\mathrm{SF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \mathrm{(non\text{-})res}}$ and $\mathrm{SF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{(non\text{-})res}}$ are the scaling factors for the (non-)resonant hybrid $B^0 \to X_u^- \ell^+ \nu_\ell$ and $B^+ \to X_u^0 \ell^+ \nu_\ell$ samples, respectively, $\mathrm{BF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \mathrm{res}}$

87

and $\mathrm{BF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{res}}$ are the sum of the branching fractions for all resonant $B^0 \to X_u^- \ell^+ \nu_\ell$ and $B^+ \to X_u^0 \ell^+ \nu_\ell$ decays, respectively, $\mathrm{BF}_{B^0 \to X_u^- \ell^+ \nu_\ell, \mathrm{non\text{-}res}}$ and $\mathrm{BF}_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{non\text{-}res}}$ are the non-resonant $B^0 \to X_u^- \ell^+ \nu_\ell$ and $B^+ \to X_u^0 \ell^+ \nu_\ell$ branching fractions, $N_{B^0 \to X_u^- \ell^+ \nu_\ell, \mathrm{(non\text{-})res}}$ and $N_{B^+ \to X_u^0 \ell^+ \nu_\ell, \mathrm{(non\text{-})res}}$ are the number of events prior to selections and before the application of the FEI in the (non-)resonant hybrid $B^0 \to X_u^- \ell^+ \nu_\ell$ and $B^+ \to X_u^0 \ell^+ \nu_\ell$ samples, respectively, and $N_{B^0 \bar{B}^0}$ and $N_{B^+ B^-}$ are the number of events (also pre-selection and pre-FEI application) for the generic $B^0 \bar{B}^0$ and $B^+ B^-$ MC samples, respectively, as listed in Table 5.1. A factor of 2 is applied to the numerator of each equation to account for the two B-mesons in each event.

5.2 Data and MC Corrections

A number of corrections were also applied to all samples used for the $B \to \pi \ell \nu_\ell$ analysis, in order to compensate for known discrepancies between data and MC. Many of these corrections are the same as those implemented in the FEI calibration study as explained in detail in Section 4.1.3, and are thus only listed briefly below. For those corrections unique to this study, a full description is provided.

- **$B \to X_u \ell \nu_\ell$ branching fraction corrections:** The branching fractions of the various $B \to X_u \ell \nu_\ell$ contributions in MC were reweighted to the updated world average values [20].

- **$B \to X_c \ell \nu_\ell$ branching fraction corrections:** The branching fractions of the various $B \to X_c \ell \nu_\ell$ contributions in MC were reweighted to the isospin-averaged values as described in Section 4.1.3.

- **$B \to X_u \ell \nu_\ell$ and $B \to X_c \ell \nu_\ell$ form factor corrections:** The $B \to \rho \ell \nu_\ell$ [73], $B \to \omega \ell \nu_\ell$ [73], $B \to D \ell \nu_\ell$ [77] and $B \to D^* \ell \nu_\ell$ [78] decays in MC were reweighted to update the values of the form factor parameterisation. For $B \to \eta \ell \nu_\ell$ and $B \to \eta' \ell \nu_\ell$, the decay model was updated to the Duplancic-Melic model [75].

- **Bremsstrahlung corrections:** The momentum of reconstructed electrons in MC and data were corrected for bremsstrahlung radiation by adding the four-momentum of any photons within the angular cone of the electron defined in Section 4.1.3, removing the photons from the event. The corrections were applied in optimised bins of the angular distance, photon energy and electron momentum. With regards to the signal MC, the number of corrected events corresponded to roughly 18% of all $B^0 \to \pi^- e^+ \nu_e$ decays and 22% of all $B^+ \to \pi^0 e^+ \nu_e$ decays. In data, the number of events corrected for bremsstrahlung radiation corresponded to approximately 5% of the total number of events for the two decay modes, including both signal and background contributions.

- **Lepton identification corrections:** Correction factors to account for the data-MC discrepancies in the efficiency of reconstructing electrons and muons as well as the pion and kaon fake-rates were applied to all MC samples [80].

- **Data track momentum scale correction:** All charged particle tracks in data were scaled to correct for a momentum shift resulting from imperfections in the simulated magnetic field map [81].

- **Charged pion identification corrections:**

  Much like the differences observed between data and MC in the lepton reconstruction efficiencies and fake-rates, similar discrepancies also exist for reconstructed charged pions. Both the efficiency with which charged pions are reconstructed correctly and the efficiency with which kaons are misreconstructed as pions, the kaon fake-rate, must be corrected for in MC by the application of a set of scaling factors. Like the lepton identification corrections, the pion corrections were similarly determined via independent studies by other collaboration members, quoted in bins of the measured polar angle θ and magnitude of the momentum p of the reconstructed pion track, for various thresholds on the relevant pion identification variable, `pionID`. The same decay channels were also used to derive the corrections.

  The pion efficiency corrections were derived from the $K_S^0 \to \pi^+\pi^-$ channel, previously used for the determination of the pion fake-rate corrections in the lepton identification case. In this case, the pion daughters were reconstructed under the pion identification criteria, and the correction factors were determined by evaluating the relative efficiencies of various `pionID` selections in MC and data.

  A similar strategy was adopted for the kaon fake-rate corrections, where the relative efficiencies with which kaons originating from the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decay chain were reconstructed under the various `pionID` identification criteria were established, a method analogous to the derivation of the fake-rate corrections due to kaons faking leptons in the lepton identification studies. The pion identification corrections were applied to all $B^0 \to \pi^-\ell^+\nu_\ell$ events reconstructed from MC.

- **Neutral pion reconstruction efficiency corrections:**

  The efficiency with which neutral pions are reconstructed from two photon daughters constitutes another source of discrepancy observed between data and MC. To account for this, official studies within the collaboration have derived correction factors in bins of the π^0 momentum and evaluated for several sets of standardised event selections recommended for π^0 reconstruction at Belle II. As will be described in Section 5.5.1, a set of selections on the daughter photon energy and the π^0 invariant mass corresponding to a $\sim 40\%$ π^0 reconstruction efficiency was chosen for this analysis, and the relevant set of π^0 efficiency corrections were applied to all MC events from which $B^+ \to \pi^0\ell^+\nu_\ell$ decays were reconstructed.

  To determine the correction factors, one such independent study was performed involving the decay of a D^0 meson to a charged kaon, charged pion and neutral pion, $D^0 \to K^-\pi^+\pi^0$. The signal yields were extracted by fitting the D^0 invariant mass,

and the ratio between these yields in data and MC were determined for each π^0 momentum bin. An additional study was performed using 3-prong τ lepton decays of the form $\tau \to 3\pi\pi^0\nu_\tau$, in which the τ decayed to three charged pions, a neutral pion and a tau neutrino, with the data-MC efficiency ratios calculated from the signal yields extracted from fits to the π^0 invariant mass peak. As per official Belle II recommendations, the correction factors from the D^0 decay channel were used to weight the MC, and the difference between the factors derived from the two independent studies were incorporated as part of the systematic uncertainty on each individual correction factor.

- **Data photon energy bias correction:**

  In early Belle II data, a slight shift was observed between data and MC in the ΔE distributions of various signal B-meson decays involving neutral final-state particles such as $B^+ \to K^+\pi^0 (\pi^0 \to \gamma\gamma)$ and $B^+ \to \eta' K^+ (\eta' \to \rho^0\gamma)$. This shift was investigated by other Belle II analysts and was determined to be a result of a discrepancy in the energies of the reconstructed photons between data and MC. To evaluate the extent of this observed bias and correct for it, a dedicated, independent study was performed within the collaboration using the photon daughters originating from the π^0 in the decay chain $D^{*+} \to D^0\pi^+ (D^0 \to K^-\pi^+\pi^0)$. The π^0 invariant mass distributions in both MC and data were fitted using the combination of a Gaussian probability density function (pdf) to represent the signal decays, and an exponential pdf for the relevant backgrounds. The variation in the position of the mass peak and the width of the fitted Gaussian were then used to determine the magnitude of the observed bias, which was found to be of the order of 1%. In this way, a set of factors was evaluated to correct all photons reconstructed from data, with these determined with respect to the photon energy and the detector region within which the photons were detected.

  For this analysis, the energies of the π^0 daughter photons reconstructed from data as part of the signal $B^+ \to \pi^0\ell^+\nu_\ell$ decay were corrected for by applying the relevant photon energy bias corrections using a dedicated basf2 function (`correctEnergyBias`).

- **FEI calibration corrections:** The official Belle II FEI calibration factors described in Section 4.1.7, determined from the combined analysis of $B \to X\ell\nu_\ell$ and $B \to D^{(*)}\pi$, were applied to all reconstructed MC events. The correction factors determined in bins of the decay mode of the reconstructed B_{tag} were used for this analysis, as per official recommendations.

Finally, it should be noted that for the $B \to \pi\ell\nu_\ell$ analysis, unlike the calibration study performed in the previous chapter, the rescaling of the continuum MC to the prediction based on off-resonance data was not performed. Given the relatively low $B \to \pi\ell\nu_\ell$ branching fractions and the limited off-resonance data-set with only 42.3 fb^{-1} of total integrated luminosity available, too few events were reconstructed from off-resonance data to allow for a valid prediction. In lieu of this rescaling, a systematic uncertainty on the continuum MC normalisation will be included in the measurement of the $B \to \pi\ell\nu_\ell$ branching fractions, as will be discussed extensively in Section 7.1.3.

5.3 Reconstructing $B \to \pi \ell \nu_\ell$

Using the MC and data samples described in the previous sections, both the neutral, $B^0 \to \pi^- \ell^+ \nu_\ell$, and charged, $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays were reconstructed. The electron and muon channels were reconstructed separately, resulting in a total of four signal decay modes studied: $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$. As the hadronic FEI skims of the samples were used, each $\Upsilon(4S) \to B\bar{B}$ event retained after the skims contained at least one B_{tag} candidate, with these being combined with B_{sig} candidates reconstructed via either of the four given modes to build the full event.

To reconstruct the signal $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, final-state charged pion and electron or muon tracks were first reconstructed from detector information. The charge of the pion and lepton were explicitly required to be of opposite sign to ensure charge conservation. The signal B^0_{sig} meson was then reconstructed from this pion and lepton, with the undetected neutrino taken as missing energy and momentum in the event. For $B^+ \to \pi^0 \ell^+ \nu_\ell$, the lepton was similarly built from detector information. The neutral pion, however, is not a final-state particle, decaying rapidly to a pair of photons $\sim 99\%$ of the time [20]. Photons built from detector information were used to reconstruct the neutral π^0 mesons, and these were combined with the reconstructed leptons to form B^+_{sig} candidates. Figure 5.3 displays an example of the reconstruction of a full $\Upsilon(4S) \to B_{\mathrm{tag}}B_{\mathrm{sig}}$ event, where $B^-_{\mathrm{sig}} \to \pi^0 \ell^- \bar{\nu}_\ell$ and the hadronic B_{tag} decays via the mode $B^+_{\mathrm{tag}} \to \bar{D}^0 \pi^+$.
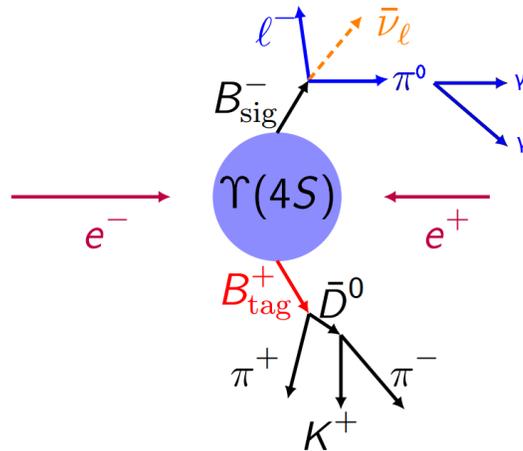

Figure 5.3: An example of an $\Upsilon(4S) \to B_{\mathrm{tag}}B_{\mathrm{sig}}$ decay, where $B^-_{\mathrm{sig}} \to \pi^0 \ell^- \bar{\nu}_\ell$ [88].

5.4 M^2_{miss}

Whilst a detailed description of the signal extraction strategy for the analysis is left to the following chapter, the variable chosen for this extraction is introduced here, as the optimisation of the analysis selections relies heavily upon this choice. As described in Section 3.1.1, a

hadronic tagged analysis of a semi-leptonic signal decay possesses a unique advantage related to the missing energy and momentum in the event. Since the B_{tag} decays hadronically and all particles on the signal side are able to be reconstructed with the exception of the neutrino, the neutrino four-momentum can be directly inferred using momentum conservation, as the initial state of the electron-positron collision is known to a high degree of precision. To utilise this advantage in the hadronic tagged analysis of $B \to \pi \ell \nu_\ell$, a variable known as the square of the missing mass in the event, M^2_{miss}, is defined. In the center-of-mass (CMS) frame, the four-momentum of the signal B-meson can be determined from that of the B_{tag}, given that the $\Upsilon(4S)$ meson is produced at rest with the two daughter B-mesons being released back-to-back. The B_{sig} four-momentum in the CMS frame, $p_{B_{\text{sig}}}$, is thus defined as:

$$p_{B_{\text{sig}}} \equiv (E_{B_{\text{sig}}}, \vec{p}_{B_{\text{sig}}}) = \left(\frac{m_{\Upsilon(4S)}}{2}, -\vec{p}_{B_{\text{tag}}} \right),$$

where $E_{B_{\text{sig}}}$ and $\vec{p}_{B_{\text{sig}}}$ are the CMS energy and momentum of the signal B-meson, respectively, $m_{\Upsilon(4S)}$ is the mass of the $\Upsilon(4S)$, and $\vec{p}_{B_{\text{tag}}}$ is the CMS momentum of the B_{tag}. The missing four-momentum in the event evaluated in the CMS frame, p_{miss}, can then be determined as:

$$p_{\text{miss}} \equiv (E_{\text{miss}}, \vec{p}_{\text{miss}}) = p_{B_{\text{sig}}} - p_Y,$$

where E_{miss} and $\vec{p}_{\text{miss}}$ are the missing CMS energy and momentum in the event, and p_Y is the four-momentum in the CMS frame of the combined pion-lepton system, Y. Finally, the M^2_{miss} is defined simply as the square of the missing four-momentum in the event:

$$M^2_{\text{miss}} \equiv p^2_{\text{miss}}.$$

In a correctly reconstructed $\Upsilon(4S) \to B_{\text{tag}} B_{\text{sig}}$ decay where $B_{\text{sig}} \to \pi \ell \nu_\ell$, the only missing particle expected in the event is the neutrino, resulting in a clear peak at the square of the neutrino mass, 0 GeV^2 in the M^2_{miss} distribution. For background events in which no true $B_{\text{sig}} \to \pi \ell \nu_\ell$ decay was present, no such behaviour is expected as the missing particles in the event would no longer be comprised of a single neutrino. The M^2_{miss} distribution thus provides a powerful mechanism for discriminating between signal and background events, and was the variable chosen for the signal extraction in this analysis. With this variable defined, the strategy for the selection optimisation can now be outlined in the following sections.

5.5   Choosing the Analysis Selections

A number of factors contributed to the determination of the selections chosen for the analysis. Official Belle II recommendations for selections were taken by default where applicable. Where no such recommendations were given, an initial set of baseline selections were developed, with heavy inspiration taken from a prior Belle study performed using the full Belle data-set, in which a measurement of $|V_{\text{ub}}|$ was extracted from a hadronic tagged analysis of $B \to \pi \ell \nu_\ell$ using the Full Reconstruction algorithm [32].

5.5.1 Default Selections

Several selections were chosen based on Belle II recommendations, particularly those concerning final-state particles and with consistency with the official FEI calibration in mind.

Standard impact-parameter cuts were chosen for reconstructed electrons, muons and charged pions, with $dr < 2$ cm and $|dz| < 5$ cm. Each of these final-state particles was also required to fall within the polar angle acceptance of the CDC, $0.297 < \theta < 2.618$ radians. Selections on the particle identification variables were chosen based on thresholds for which data-MC efficiency and fake-rate corrections were available. A high-purity threshold of `pidChargedBDTScore` > 0.9 was chosen for the reconstructed leptons. A looser threshold was selected for the charged pions in the reconstruction of $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, `pionID` > 0.6, as this variable was later used to select the best $\Upsilon(4S)$ candidate, as will be discussed in the following section. The reconstructed pion tracks were also required to correspond to greater than 20 distinct hits recorded by the CDC based on standard recommendations, `nCDCHits` > 20. Such a selection was effective at rejecting fake tracks reconstructed from a small number of isolated CDC hits, as opposed to real pions which interact frequently, producing several CDC hits along the entire trajectory of the track.

To reconstruct π^0 mesons from pairs of photons in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ analysis, a set of official, standard selections were applied. These were derived from an independent study of neutral pions reconstructed from $B\bar{B}$ MC, where optimal selections were chosen by finding those that resulted in the highest purity for a set of different efficiency thresholds. For this analysis, the standard selection list for reconstructing neutral pions with a 40% efficiency was chosen. This involved selecting photons within the CDC acceptance that were associated with clusters built from more than one crystal in the ECL. The energy of the photons was also required to be above certain thresholds dependent on the detector region from which the photon was reconstructed: $E > 0.080$ GeV for the forward end-cap, $E > 0.030$ GeV for the barrel, and $E > 0.060$ GeV for the backward end-cap. Under these standard recommendations, the invariant mass of the π^0 mesons reconstructed from the photon pairs was additionally required to fall within a window defined around the known π^0 mass [20], $0.120 < M < 0.145$ GeV.

Selections on the reconstructed hadronic B_{tag} candidates were chosen to align with those imposed during the official hadronic FEI calibration. These included a selection on the tag-side energy difference to favour B-meson-like tag candidates, $-0.15 < \Delta E < 0.1$ GeV, and a selection on the cosine of the angle between the thrust axes of the B_{tag} and its rest-of-event (ROE, as first defined in Section 4.1), `cosTBTO` < 0.90, in order to reject continuum events. The ROE was required to satisfy the same conditions as described in Section 4.1.2, where the `cosTBTO` variable was first introduced. For reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, this selection was tightened to `cosTBTO` < 0.85 due to higher levels of continuum background present.

Finally, after the reconstruction of the $\Upsilon(4S)$ meson from the tag and signal B-meson daughters, an ROE object was built for the $\Upsilon(4S)$ containing all remaining tracks and clusters in the event. Given that the only missing particle in a correctly reconstructed event

should be a single neutrino, an additional requirement was placed such that the number of tracks present in the ROE passing a set of specified selections must be zero. This selection is a common requirement enforced in hadronic tagged analyses, and is referred to as the completeness constraint in other literature [13]. Whilst sufficiently strict, this constraint is particularly powerful in terms of enhancing the purity of a given analysis. In this study, the threshold selections imposed upon the ROE tracks were relatively high to exclude only those events that were likely to be due to backgrounds. These included standard impact-parameter and CDC acceptance cuts, $dr < 2$ cm, $|dz| < 5$ cm, and $0.297 < \theta < 2.618$ radians, a moderate selection on the transverse momentum, $p_t > 0.2$ GeV, and energy thresholds for the associated ECL clusters: $E > 0.10$ GeV for the forward end-cap, $E > 0.09$ GeV for the barrel, and $E > 0.16$ GeV for the backward end-cap.

5.5.2 Best Candidate Selection (BCS)

During the reconstruction of each particle in the $\Upsilon(4S)$ decay chain, multiple candidates often arose due to the existence of more than one combination of particles (or detector information, in the case of final-state particles) that satisfied the given selections. Much like the case for the FEI calibration study in the previous chapter, a method for choosing a single, best candidate per event for each particle was thus implemented.

A single B_{tag} candidate was chosen first for each event, with the candidate with the highest value of the FEI `SignalProbability` selected. After the reconstruction of the lepton on the signal side, a single lepton per event was also chosen based on the highest value of the relevant lepton identification variable, the `pidChargedBDTScore`.

For the reconstructed pions, one candidate was not explicitly selected. Instead, the single B_{tag}, the single lepton, and several pion candidates were combined to produce multiple candidates for the $\Upsilon(4S)$ meson. In this way, the candidate multiplicity of the $\Upsilon(4S)$ was driven entirely by the multiple pion candidates. The enforcement of the completeness constraint detailed in the previous section was highly effective at reducing the number of events containing multiple $\Upsilon(4S)$ candidates, as most events with particles remaining in the ROE after the full reconstruction of the event were discarded. However, the nature of this constraint was such that only those events with ROE particles satisfying the selections described above were rejected, resulting in a small number of events in which multiple $\Upsilon(4S)$ candidates were retained even after the application of the completeness constraint. Thus, a single $\Upsilon(4S)$ candidate was also chosen for each event. For reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, the $\Upsilon(4S)$ candidate with a charged pion granddaughter corresponding to the highest value of the pion identification variable, `pionID`, was selected. For $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, the best $\Upsilon(4S)$ candidate was chosen based on the lowest opening angle of the photons originating from the decay of the signal neutral pion, $\psi_{\gamma\gamma}$. Here, the opening angle refers to the angle between the flight directions of the two photons, with lower values typically characteristic of photons arising from a real $\pi^0 \to \gamma\gamma$ decay. Fake π^0 candidates formed from photons released from interactions in the forward and backward endcap regions of the detector tend to possess

larger values of the opening angle, as the two photons cannot be traced back to a common decay vertex [32].

5.5.3 Selection Optimisation

For this analysis, in addition to the default selections given above, a set of further selections were taken from the prior Belle study [32] and optimised individually. The nominal values of these selections as found in Ref. [32] are listed in Table 5.2, with each defined and described in detail in the coming sections.

Variable	Selection, $B^0 \to \pi^- \ell^+ \nu_\ell$	Selection, $B^+ \to \pi^0 \ell^+ \nu_\ell$
E_{miss}	$E_{\mathrm{miss}} > 0.3$ GeV	
E_{residual}	$E_{\mathrm{residual}} < 1.0$ GeV	$E_{\mathrm{residual}} < 0.6$ GeV
$B_{\mathrm{tag}}\ M_{\mathrm{bc}}$	$B_{\mathrm{tag}}\ M_{\mathrm{bc}} > 5.27$ GeV	
$p_{\ell,\mathrm{lab}}$	$p_{e,\mathrm{lab}} > 0.3$ GeV	
	$p_{\mu,\mathrm{lab}} > 0.6$ GeV	
$\psi_{\gamma\gamma}$	–	$\psi_{\gamma\gamma} < 1.32$ rad.
$|z_\ell - z_\pi|$	$|z_\ell - z_\pi| < 0.10$ cm	–
$\cos\theta_{BY}$	$|\cos\theta_{BY}| \leqslant 3.0$	

Table 5.2: A number of analysis selections imposed during the Belle hadronic tagged $B \to \pi\ell\nu_\ell$ study detailed in Ref. [32].

The M^2_{miss} distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and the nominal selections given in Table 5.2 are illustrated in Figures 5.4 and 5.5 for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively. All MC components have been normalised to the integrated luminosity of the available continuum sample. The signal components of each distribution are given by the red filled histograms, with the total backgrounds broken down into several different components. These include background contributions from continuum events, crossfeed from other semi-leptonic $B \to X_u \ell \nu_\ell$ decays, and other $B\bar{B}$ backgrounds, such as those from $B \to X_c \ell \nu_\ell$ decays. For $B^0 \to \pi^- \ell^+ \nu_\ell$ in Figure 5.4, the background $B^0 \to \rho^- \ell^+ \nu_\ell$ contribution was extracted from the $B \to X_u \ell \nu_\ell$ component and plotted separately to illustrate its considerable relative size. In these figures, the majority of signal events can be clearly seen to fall within the region $-1.0 \leqslant M^2_{\mathrm{miss}} \leqslant 1.0$ GeV^2, with a clear peak observed at 0 GeV^2, as expected. This M^2_{miss} region is thus referred to as the signal region, with higher M^2_{miss} values defined within the range $1.0 < M^2_{\mathrm{miss}} \leqslant 3.0$ GeV^2 termed the M^2_{miss} sideband.

Whilst the discrimination between signal and background events was already evident using the nominal selections taken from the Belle study, this was sought to be further improved via the optimisation of each selection, a process aimed at finding the values of the selections that best reduced the backgrounds whilst maintaining as many signal events as possible. In
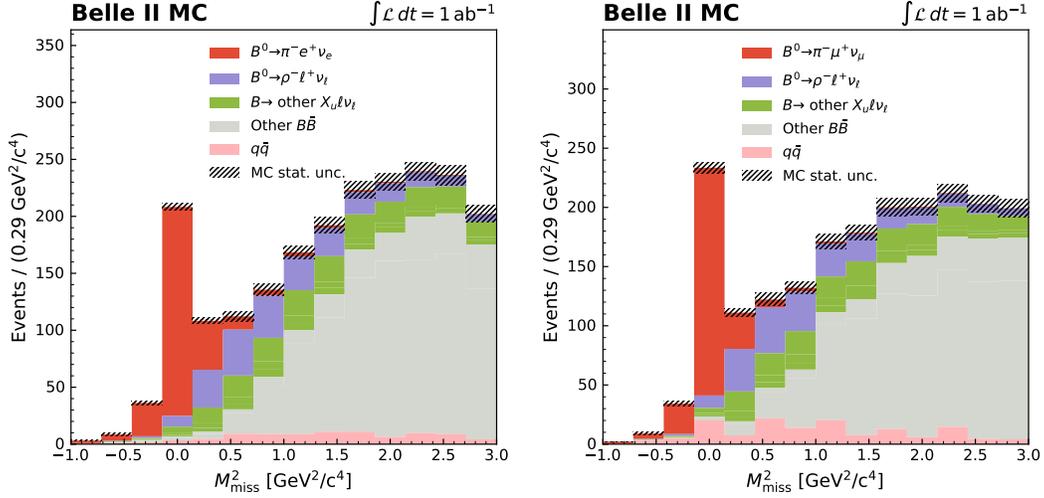
Figure 5.4: M_{miss}^2 distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and the nominal selections given in Table 5.2 [32], for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (right). All MC has been normalised to the integrated luminosity of the continuum sample, 1 ab^{-1}.
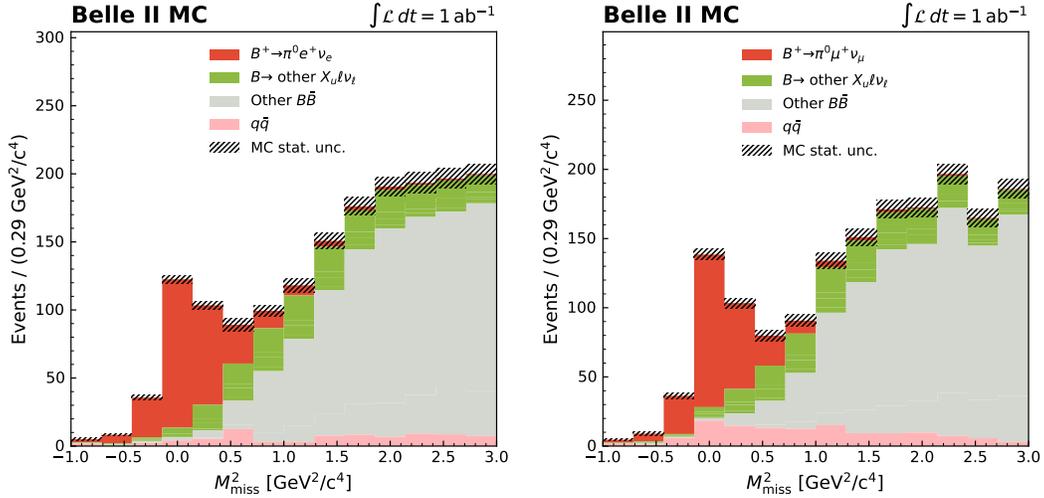

Figure 5.5: M_{miss}^2 distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and the nominal selections given in Table 5.2 [32], for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right). All MC has been normalised to the integrated luminosity of the continuum sample, 1 ab^{-1}.

order to achieve this, a common metric for estimating the statistical significance of the signal was used, $S/\sqrt{S+B}$, where S and B refer to the number of signal and background events remaining post-selection in a chosen region as described below. For each individual selection, the value of the cut was varied across a window defined around the nominal Belle selection, keeping all other selections constant at the values listed in Table 5.2. The $S/\sqrt{S+B}$ significance estimator was then calculated from the resultant M_{miss}^2 distribution for each cut value.

96

Other selection optimisation methods exist which take a more nuanced approach, optimising the full set of analysis selections concurrently in such a way that any correlations between selection variables are taken into account. However, given that the selection values chosen as starting points for the given analysis were taken from the peer-reviewed and highly acclaimed Belle study published in 2013 [32], with these selections already demonstrating a high degree of signal-to-background discrimination as seen in Figures 5.4 and 5.5, the simpler method of optimising the analysis selections individually was deemed sufficient for the present study.

To determine the number of signal and background events used in the calculation, the resultant M_{miss}^2 distributions in MC were used to generate what is known as an Asimov data-set. Asimov data has a variety of applications, and is defined as a set of data points that exactly matches the given expectation for a particular distribution - in essence, a 'perfect data-set' [89]. In this context, the Asimov data-set generated for each M_{miss}^2 distribution corresponded to the total number of weighted signal and background MC events in each bin. In a similar manner to those described in Section 4.1.4, maximum likelihood fits were then performed to the given Asimov data-sets to extract the yields for each cut value studied. Two MC templates were used in the fit, a signal and background template, with the latter containing all of the various background contributions merged together. The fits were performed over the full M_{miss}^2 range, but only the yields within the signal region were extracted by integrating the fitted pdfs over the $-1.0 \leqslant M_{\text{miss}}^2 \leqslant 1.0$ GeV^2 range. Optimising the selections in this way specifically targeted the reduction of background events in the M_{miss}^2 region from which the the majority of signal events will eventually be reconstructed in real data. Whilst the number of signal and background events remaining after each selection could have been determined simply by counting the relative event numbers in MC using the given truth information, the use of Asimov data allowed also for the validation of the fitting technique which will ultimately be used to extract these real data yields. Only the MC was used for the selection optimisation, in order to avoid any bias with respect to the real data. The following sections will detail the optimisation of each individual selection and the resultant distributions of the $S/\sqrt{S+B}$ significance estimators in MC.

E_{miss}

The missing energy in the event, defined in Section 5.4, was optimised by scanning cut values around the nominal Belle $E_{\text{miss}} > 0.3$ GeV selection, which was applied to all four reconstructed $B \to \pi \ell \nu_\ell$ modes. The $S/\sqrt{S+B}$ distributions corresponding to these cut values are displayed in Figure 5.6, with error bars calculated from the propagation of the errors on the signal and background yields returned from the maximum likelihood fits to the Asimov data. The equivalent E_{miss} distributions are also included in the figure, plotted for the range spanning the tested cut values.
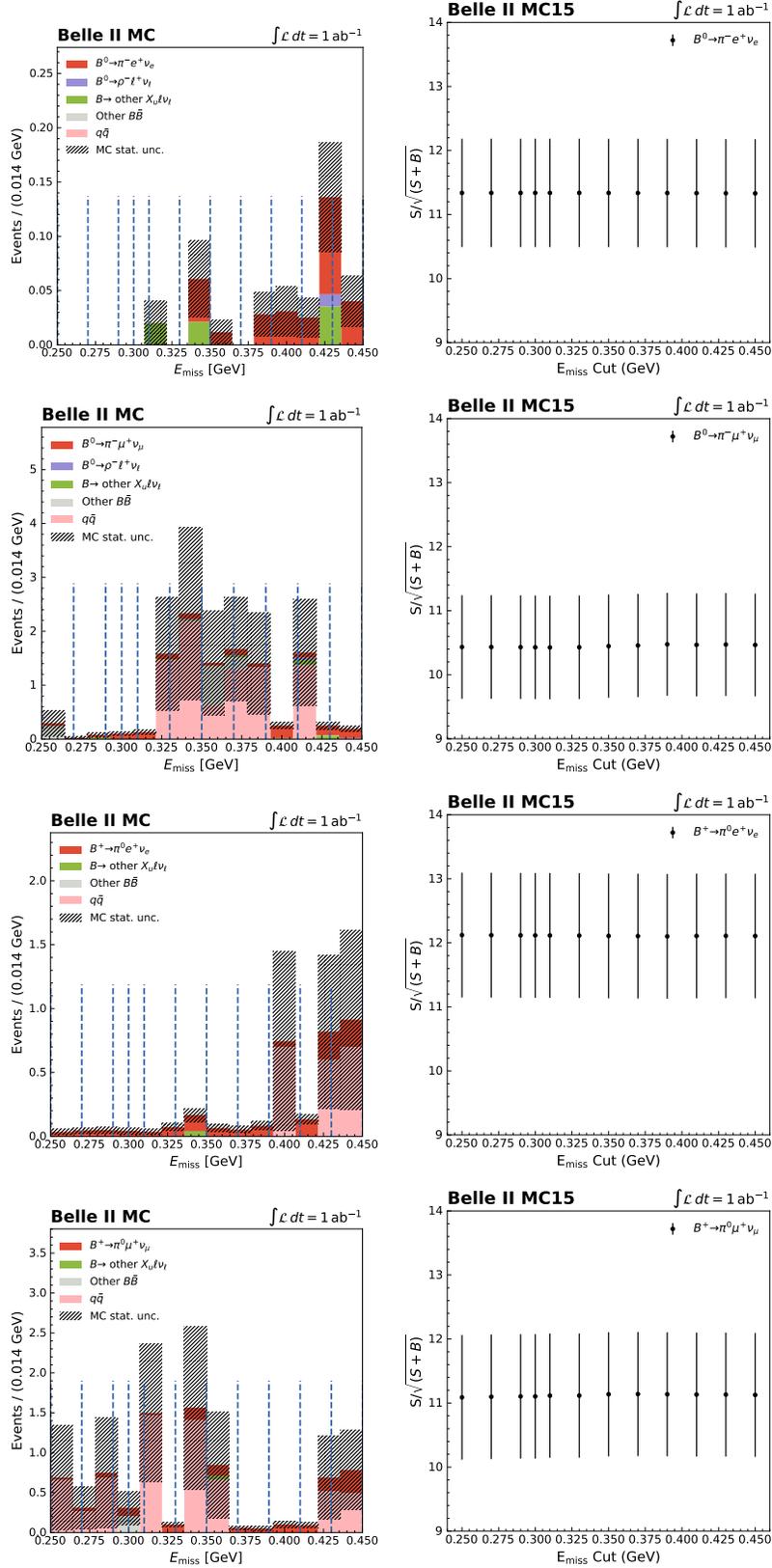
97

Figure 5.6: Distributions of E_{miss} (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

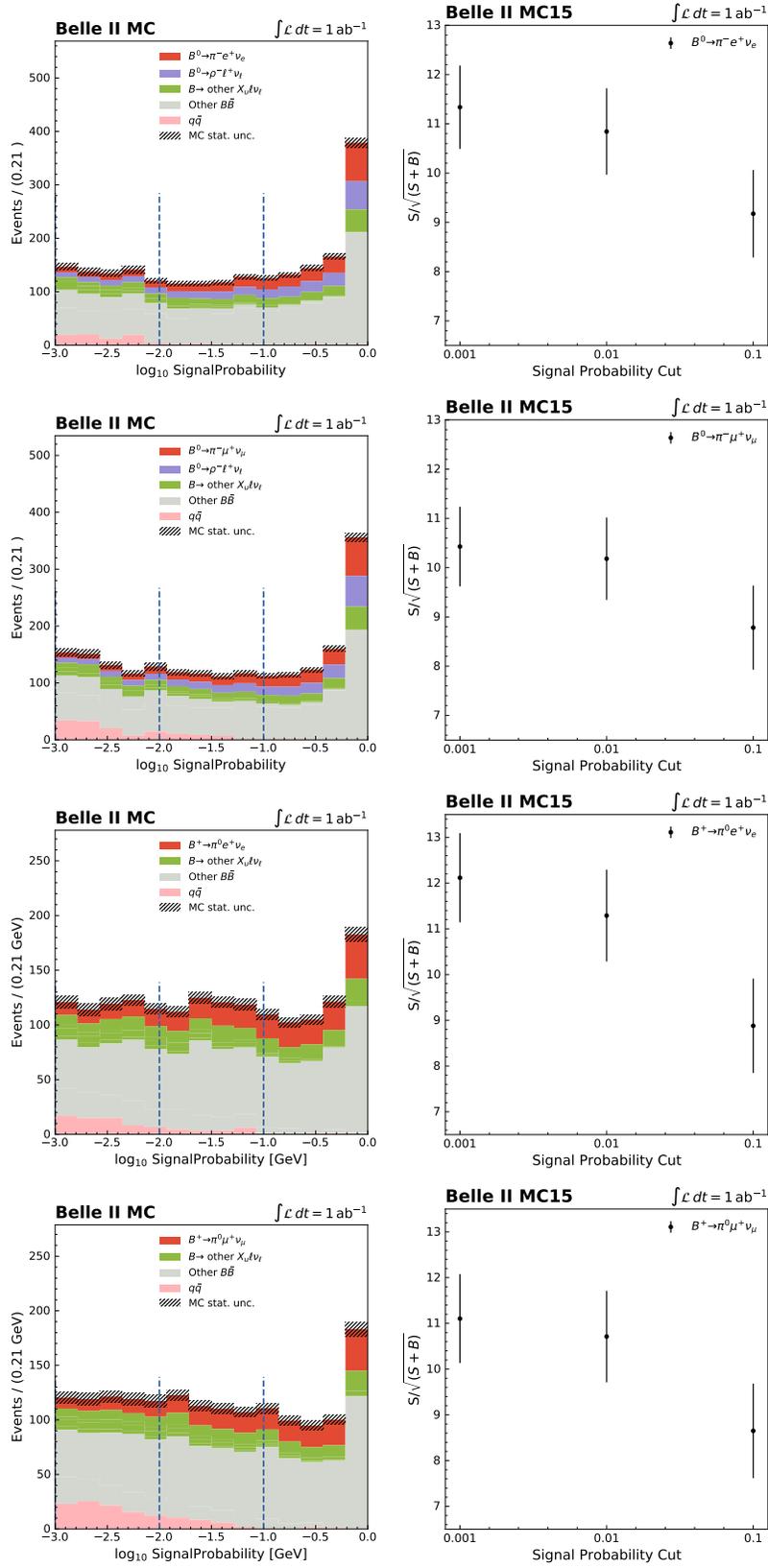Figure 5.7: Distributions of the FEI `SignalProbability` (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

For each decay mode, the distribution of the significance estimator is largely flat across all cut values studied, indicating that the chosen value for this selection has no substantial impact on the overall significance. This is unsurprising given the E_{miss} distributions provided, with only very few events falling within the plotted window. In light of this, the loosest selection was chosen for each mode to retain as many signal events as possible:

$$E_{\text{miss}} > 0.25 \text{ GeV}$$

FEI `SignalProbability`

The official hadronic FEI calibration factors are provided only for certain selections on the B_{tag} `SignalProbability`, starting from the loose selection imposed during the application of the hadronic FEI skims, `SignalProbability` > 0.001. The different thresholds were evaluated to determine which selection resulted in the best signal significance for the analysis, as can be observed in the $S/\sqrt{S+B}$ distributions shown in Figure 5.7. The equivalent `SignalProbability` distributions are also included in the figure, plotted on a logarithmic scale for better visualisation of the relative components. In considering all four decay modes, maintaining the skim selection was found to be the best option both for maximising the signal significance and efficiency:

$$B_{\text{tag}} \text{ `SignalProbability` } > 0.001$$

The chosen `SignalProbability` selection was applied to all B_{tag} decay modes, in line with official recommendations. Both the FEI skims and the personal calibration study detailed in Chapter 4 made an exception for the modes $B^0 \to J/\psi K_S^0$ and $B^+ \to J/\psi K^+$ for which no `SignalProbability` cut was applied. This decision arose due to issues resulting from the insufficient training of these modes during a prior instance of training the FEI algorithm. These issues have since been resolved and the exception was no longer required, and thus, was not applied.

B_{tag} M_{bc}

The beam-constrained mass, M_{bc}, introduced in Section 3.4.1, is a powerful variable used to distinguish between real and fake B-mesons misreconstructed from continuum events. Correctly reconstructed B-mesons are characterised by a peak in M_{bc} at the known B-meson mass, 5.28 GeV [20], clearly demonstrated in the M_{bc} distributions included in Figure 5.8 for the four reconstructed $B \to \pi \ell \nu_\ell$ modes.

Several cut values for the tag-side M_{bc} were tested within a window defined around the 2013 Belle selection, $M_{\text{bc}} > 5.27$ GeV, with the $S/\sqrt{S+B}$ distributions also given in Figure 5.8. A clear trend was observed with loose selections resulting in a poor signal significance due to contamination from continuum backgrounds. Additionally, selecting too tightly at the B-meson mass, $M_{\text{bc}} > 5.28$ dramatically decreased the significance due to the rejection of a high proportion of signal events. The optimal selection was found to align with that of the Belle study:

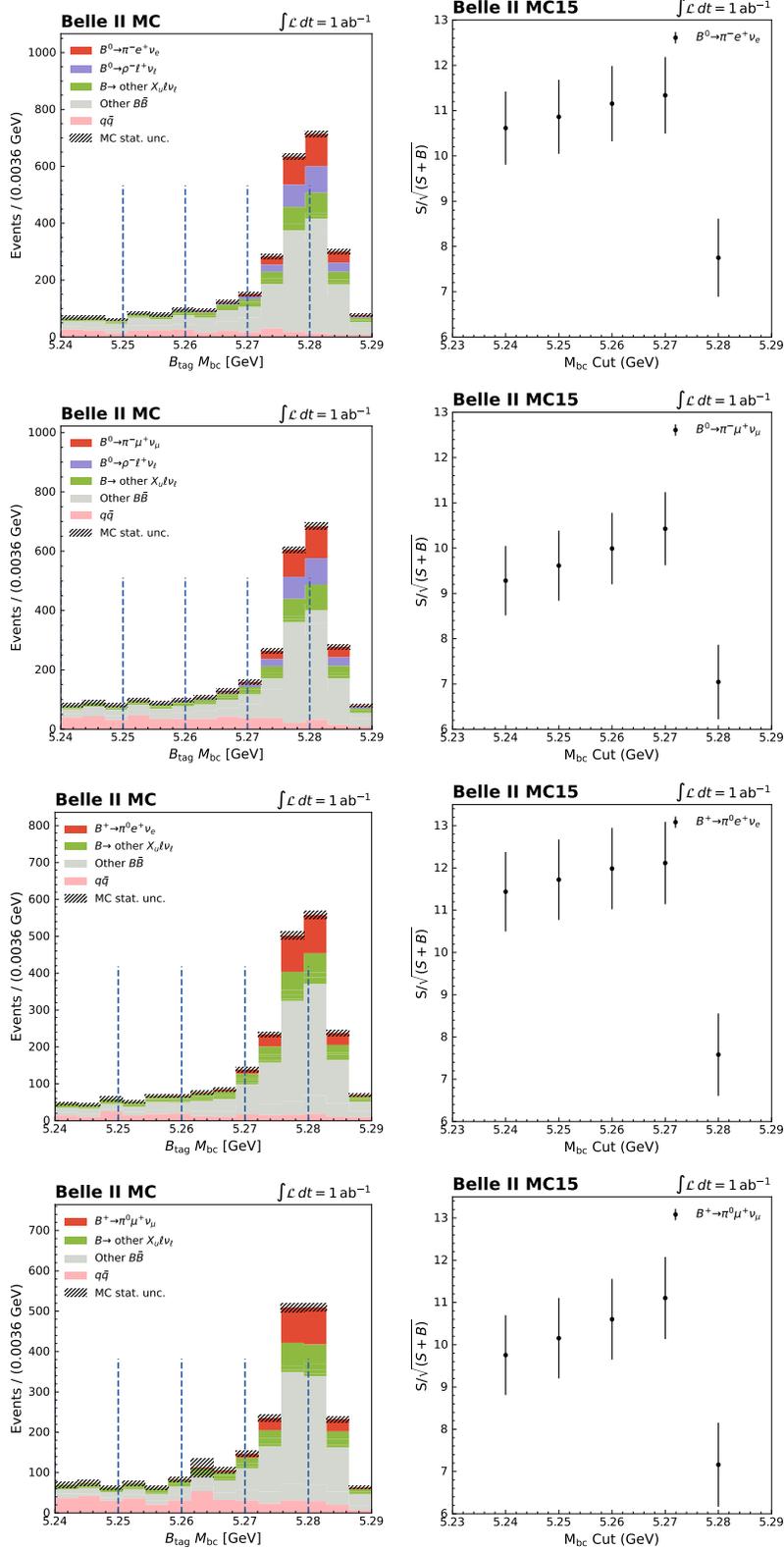$$B_{\text{tag}} \ M_{\text{bc}} > 5.27 \text{ GeV}$$

Figure 5.8: Distributions of the B_{tag} M_{bc} (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

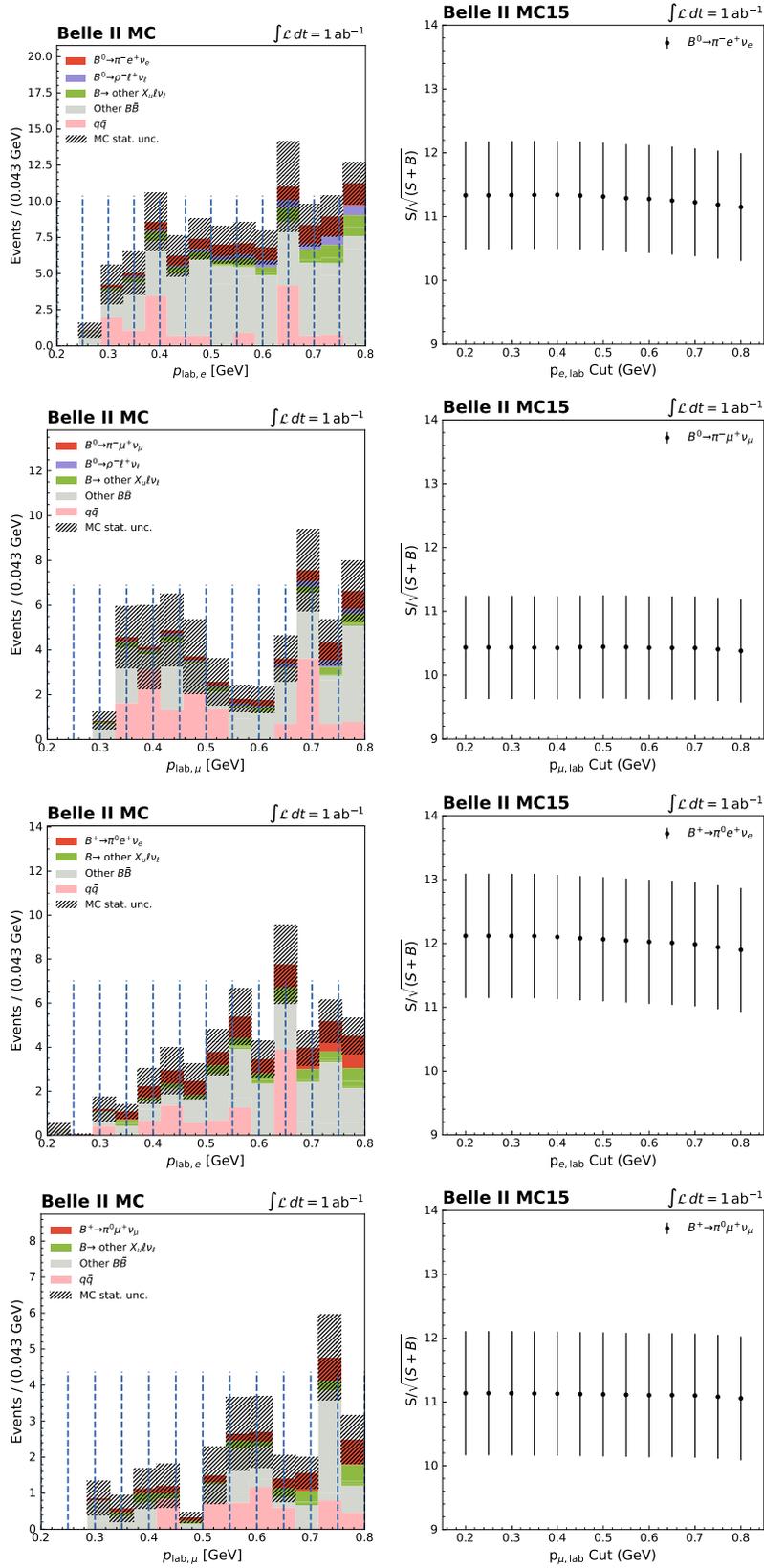Figure 5.9: Distributions of the signal lepton $p_{\ell,\text{lab}}$ (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

$p_{\ell,\mathrm{lab}}$

The 2013 Belle study selected electrons and muons with different minimum thresholds on the magnitudes of their respective lab-frame momenta, $p_{e,\mathrm{lab}} > 0.3$ GeV and $p_{\mu,\mathrm{lab}} > 0.6$ GeV. To establish the optimal selection for this analysis, the $S/\sqrt{S+B}$ significance estimator was similarly evaluated for cut values defined within a window around the nominal selections. Only $p_{\ell,\mathrm{lab}}$ values greater than 0.2 GeV were considered, as no lepton identification corrections were available for reconstructed leptons with momenta lower than this threshold. In any case, the proportion of leptons in this analysis possessing these lower momentum values was negligible after all other analysis selections.

Figure 5.9 displays the $p_{\ell,\mathrm{lab}}$ and $S/\sqrt{S+B}$ distributions for the range of cut values studied, with only a small number of leptons possessing a momentum within this window. Given the relatively flat nature of the significance estimator across this range, it was decided to cut loosely on both the electron and muon momenta to maximise the signal efficiency:

$$p_{\ell,\mathrm{lab}} > 0.2 \text{ GeV}$$

$\psi_{\gamma\gamma}$

To reject backgrounds from fake π^0 candidates in reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, a selection on the photon opening angle introduced in Section 5.5.2 was also implemented. Various selections were investigated around the nominal Belle cut of $\psi_{\gamma\gamma} < 1.318$ rad., equivalent to a selection on the cosine of the angle of $\cos\psi_{\gamma\gamma} > 0.25$. The resultant $S/\sqrt{S+B}$ distributions are shown in Figure 5.10 together with the equivalent $\psi_{\gamma\gamma}$ distributions, for $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays.

The estimated significance was similarly found to not depend heavily on the choice of the $\psi_{\gamma\gamma}$ selection. To maximise the signal efficiency, the Belle selection was therefore loosened for this analysis:

$$\psi_{\gamma\gamma} < 1.5 \text{ rad.}$$

$|z_\ell - z_\pi|$

In the reconstruction of $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, the final-state lepton and charged pion on the signal side originate from the same decay vertex, and fake combinations can similarly be rejected by requiring that the tracks are reconstructed in close proximity. Defining a cylindrical coordinate system with an origin defined at the $e^+ e^-$ interaction point, the Belle analysis imposed a selection on the distance between the points on the pion and lepton tracks that corresponded to the shortest distance from the z-axis, termed the points of closest approach (PCA) [32], $|z_\ell - z_\pi| < 0.10$ cm.

The $z_\ell - z_\pi$ distributions for the reconstructed $B^0 \to \pi^- e^+ \nu_e$ and $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in MC are illustrated in Figure 5.11, together with the $S/\sqrt{S+B}$ distributions resulting from
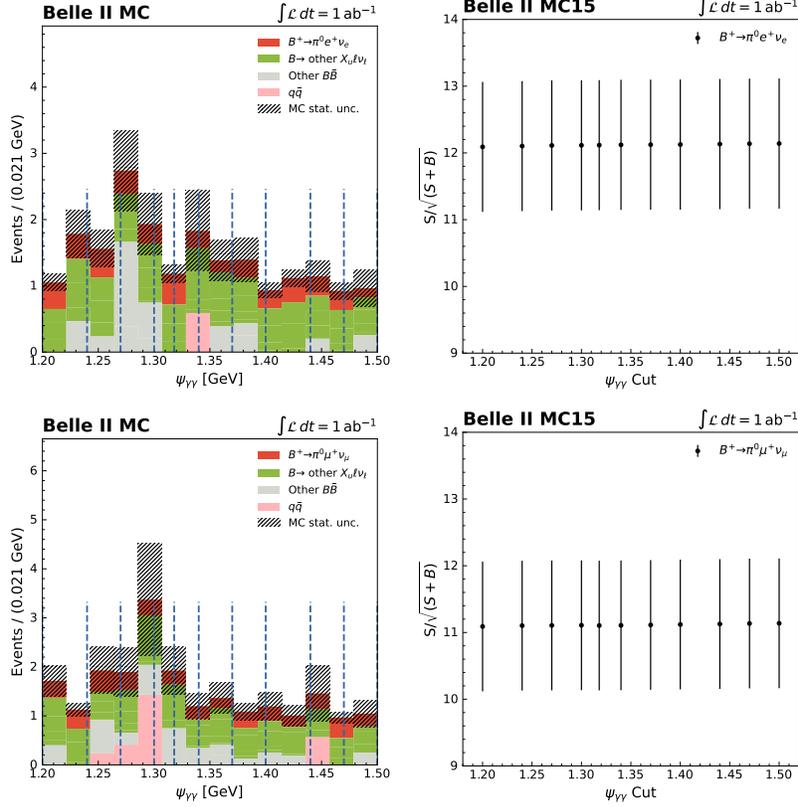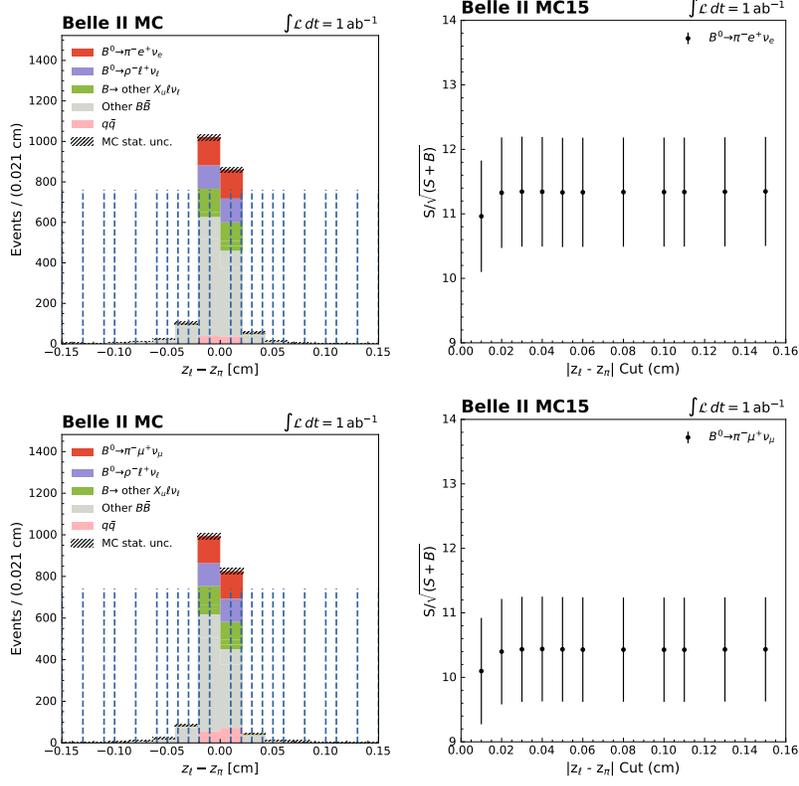
Figure 5.10: Distributions of $\psi_{\gamma\gamma}$ (left) and the significance estimator $S/\sqrt{S + B}$ as a function of the selection (right), for $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

the scan of cut values around the nominal Belle selection. The observed peak is particularly narrow, with few events observed outside of the $|z_\ell - z_\pi| < 0.05$ cm window. As a result, the signal significance varies only slightly for selections outside of this range. On the other hand, cutting too tightly into the signal peak at $|z_\ell - z_\pi| < 0.01$ cm corresponds to a drop in the significance. It was decided to tighten the selection from the Belle case given the absence of events with $|z_\ell - z_\pi| > 0.05$ cm, but to keep the selection relatively loose to retain as many signal events as possible:
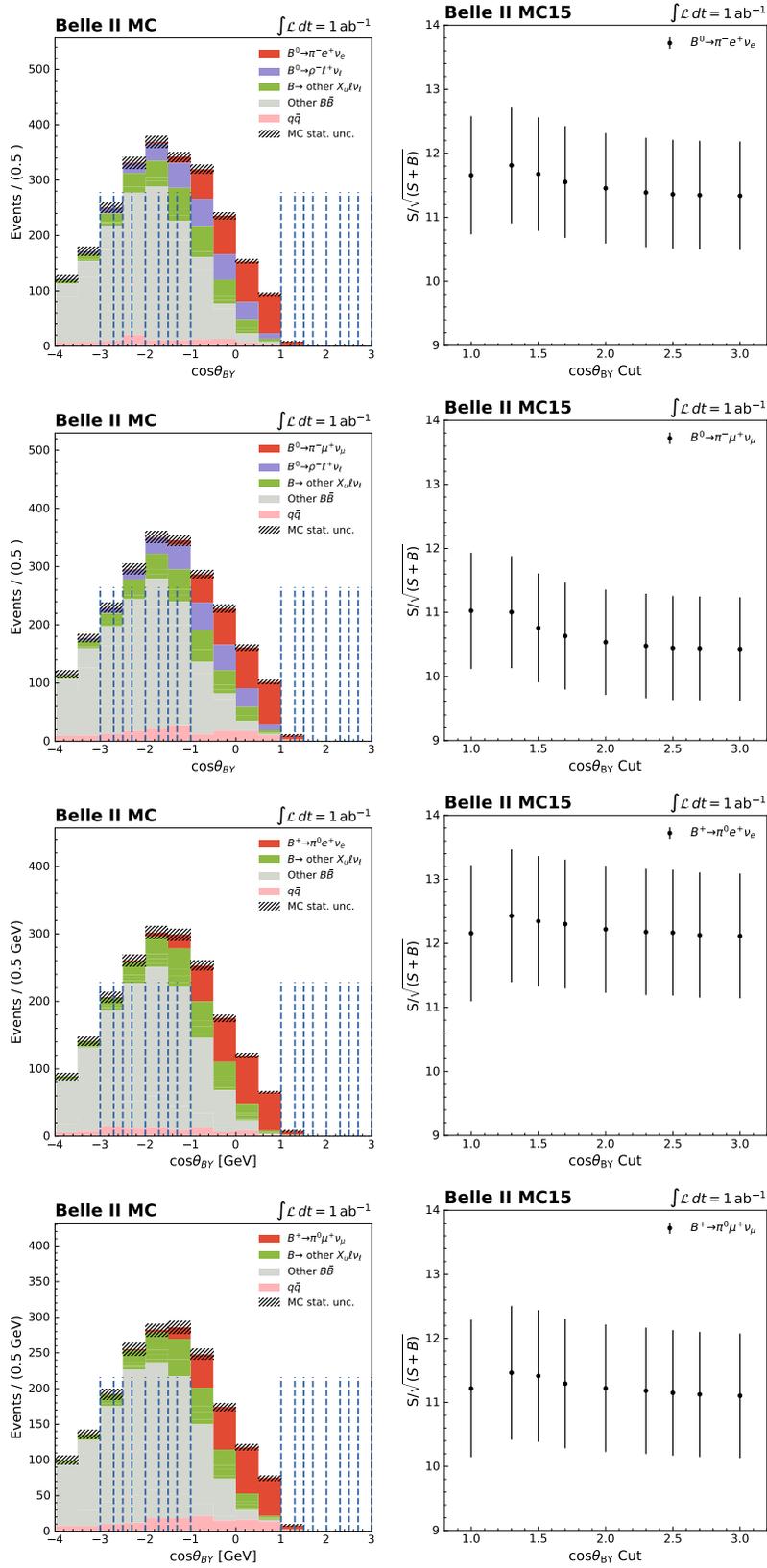
$$|z_\ell - z_\pi| < 0.05 \text{ cm}$$

Figure 5.11: Distributions of $z_\ell - z_\pi$ (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$ and $B^0 \to \pi^- \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

B_{sig} $\cos\theta_{BY}$

Another powerful variable for distinguishing between signal and background events in a semi-leptonic B-meson decay is the cosine of the angle between the flight directions of the B-meson and the pseudo-particle Y, $\cos\theta_{BY}$. Here, the Y system does not represent a real particle, but rather corresponds to the combination of the four-momenta of the leptonic and hadronic B-meson daughters; in this case, the combined pion-lepton system. The value of $\cos\theta_{BY}$ for the $B \to \pi\ell\nu_\ell$ decays relevant to this study can be determined using the following formula [32]:

$$\cos\theta_{BY} = \frac{2E_{B_{\text{sig}}}E_Y - m_B^2 - M_Y^2}{2|\vec{p}_{B_{\text{sig}}}||\vec{p}_Y|},$$

where $E_{B_{\text{sig}}}$ and E_Y are the energies of the signal B-meson and pion-lepton system Y, respectively, m_B and m_Y are the respective invariant masses, and $|\vec{p}_{B_{\text{sig}}}|$ and $|\vec{p}_Y|$ are the magnitudes of the respective 3-momenta. Correctly reconstructed signal events must correspond to physical values of this cosine, with $|\cos\theta_{BY}| < 1$. However, to account for detector resolution effects, the 2013 Belle study loosened this selection to $|\cos\theta_{BY}| \leqslant 3$ [32].

Figure 5.12: Distributions of the signal $\cos\theta_{BY}$ (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

Several cut values around the Belle selection were tested, with the resultant $S/\sqrt{S+B}$ distributions displayed in Figure 5.12 for each of the four reconstructed $B \to \pi\ell\nu_\ell$ modes, together with the equivalent $\cos\theta_{BY}$ distributions. The vast majority of signal events fell within the $|\cos\theta_{BY}| < 1$ range. No large variation was observed in the estimated significance, though slight improvements were found to correspond to tighter selections. In light of this, it was initially decided to tighten the Belle selection further to $|\cos\theta_{BY}| < 1.5$. However, whilst the gain to the significance in the M_{miss}^2 signal region was marginal as demonstrated by the given $S/\sqrt{S+B}$ distributions, this selection also greatly reduced the background levels in the M_{miss}^2 sideband, where $1.0 < M_{\text{miss}}^2 \leqslant 3.0$ GeV^2. This can be clearly observed in Figures 5.13 and 5.14, which depict the M_{miss}^2 distributions in MC for $B^0 \to \pi^-\ell^+\nu_\ell$ and $B^+ \to \pi^0\ell^+\nu_\ell$ decays, respectively, after all default selections, best candidate selections, and nominal 2013 Belle selections as listed in Table 5.2, but with the $\cos\theta_{BY}$ selection tightened to $|\cos\theta_{BY}| < 1.5$. It is necessary to retain a certain background level in the M_{miss}^2 sideband such that a robust template pdf can be formed through which the background shape can be adequately constrained during the fitting process. To accommodate for this in the given study, the $\cos\theta_{BY}$ selection was ultimately loosened to be consistent with the Belle selection:
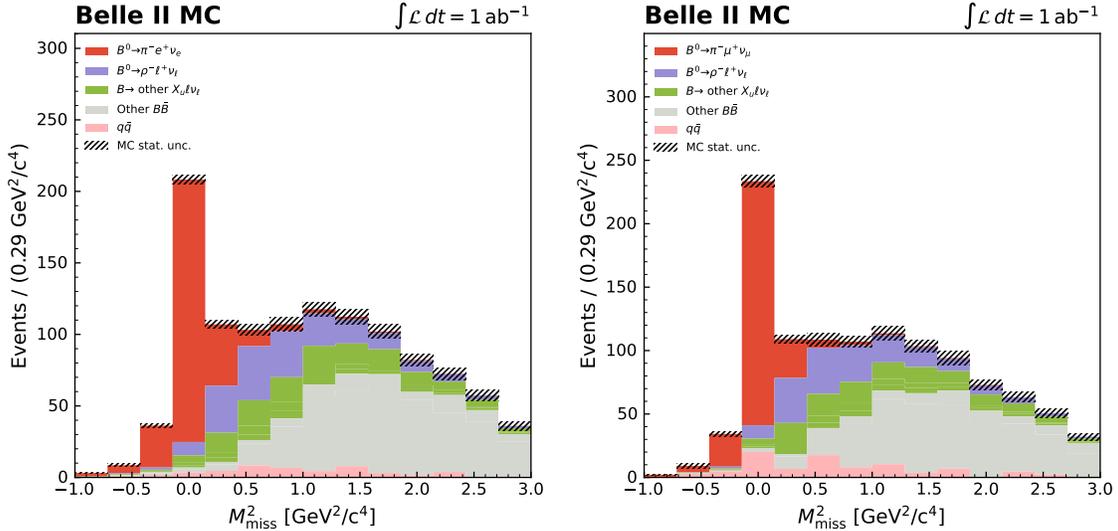
$$|\cos\theta_{BY}| < 3.0$$

Figure 5.13: M_{miss}^2 distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and the nominal selections described in Table 5.2 [32], with the $\cos\theta_{BY}$ selection tightened to $|\cos\theta_{BY}| < 1.5$, for reconstructed $B^0 \to \pi^-e^+\nu_e$ (left) and $B^0 \to \pi^-\mu^+\nu_\mu$ (right) decays.

E_{residual}

The E_{residual} variable, referred to in other literature as E_{ECL} [32], is defined as the sum of the residual energy left in the electromagnetic calorimeter by neutral particles (primarily photons) that are not reconstructed on either the tag or signal sides of the event. The 2013 Belle
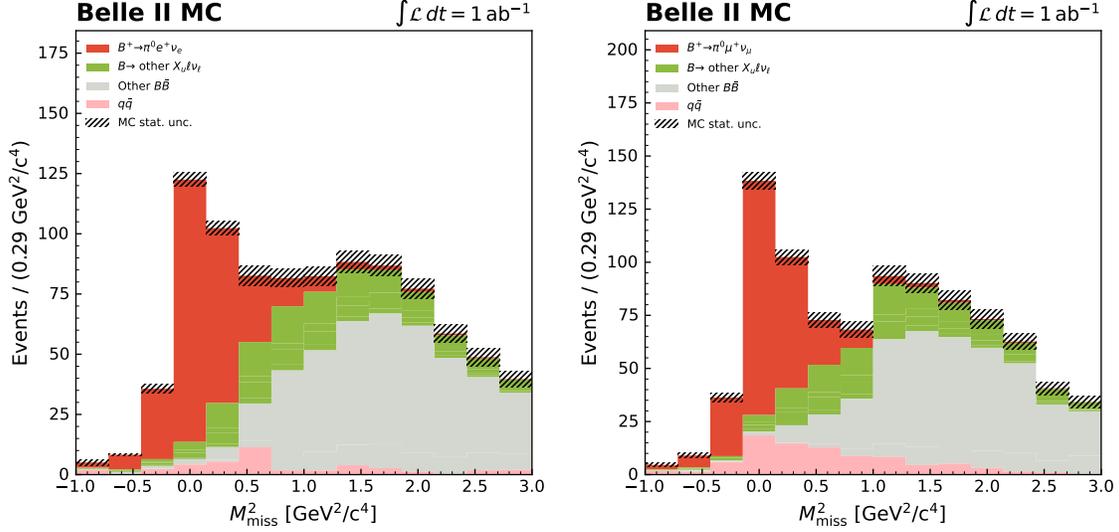
Figure 5.14: M^2_{miss} distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and the nominal selections described in Table 5.2 [32], with the $\cos\theta_{BY}$ selection tightened to $|\cos\theta_{BY}| < 1.5$, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right) decays.

study imposed selections on this residual energy calculated in the CMS frame, optimised for each signal decay mode studied, with $E_{\text{residual}} < 1.0$ GeV for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$, and $E_{\text{residual}} < 0.6$ GeV for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays. For this analysis, the CMS energies from all neutral ECL clusters satisfying the $\Upsilon(4S)$ ROE conditions listed in Section 5.5.1 were summed, and the respective selections were optimised using the $S/\sqrt{S + B}$ distributions given in Figure 5.15.

For the neutral B-meson decays, the resultant distributions were largely flat across the E_{residual} cut values scanned, and it was decided to maintain the Belle 2013 selection:

$$E_{\text{residual}} < 1.0 \text{ GeV } (B^0 \to \pi^- \ell^+ \nu_\ell)$$

More variation in the significance estimator was observed for the $B^+ \to \pi^0 \ell^+ \nu_\ell$ modes, with signal events somewhat skewed towards lower E_{residual} values as seen in the MC distributions, though the results for each cut value were largely consistent within uncertainties. As a result, selections tighter than those used for the Belle study generally corresponded to a lower signal significance. The decision was made to loosen the selection slightly to retain more signal events at no cost to the statistical significance:

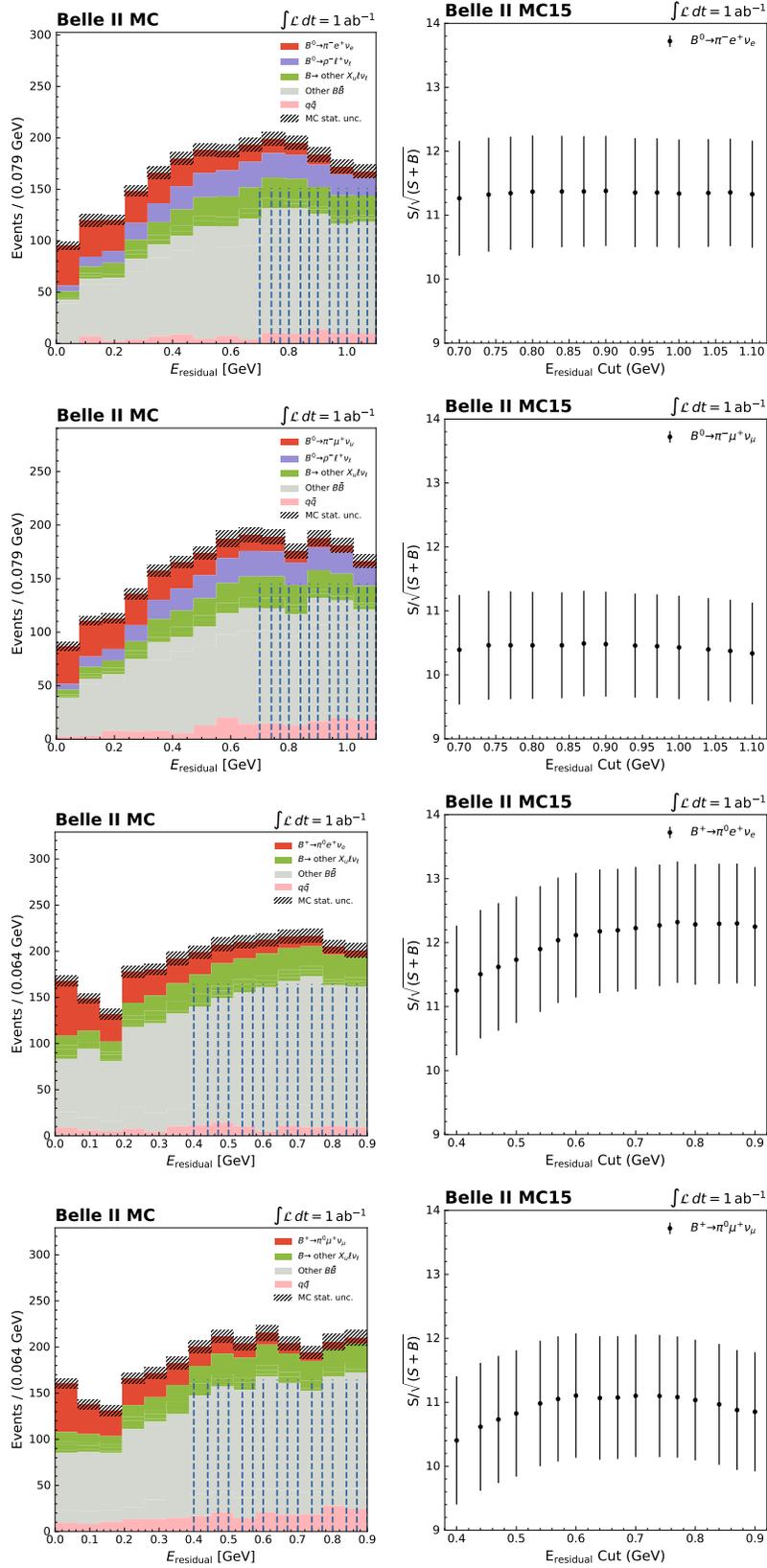$$E_{\text{residual}} < 0.75 \text{ GeV } (B^+ \to \pi^0 \ell^+ \nu_\ell)$$

108

Figure 5.15: Distributions of E_{residual} (left) and the significance estimator $S/\sqrt{S+B}$ as a function of the selection (right), for $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. Selection values are shown as dashed vertical lines on the leftmost plots.

The E_{residual} selections decided upon as described above were initially applied to the MC and data used throughout the subsequent analysis. However, during a later process of internal peer review within the Belle II collaboration for the purpose of publishing the $B \to \pi \ell \nu_\ell$ study as a journal paper, the E_{residual} selections were ultimately removed. Several concerns were raised due to the current mismodelling of this variable with regards to the observed data-MC agreement in other, independent Belle II analyses, and it was suggested to remove any dependency on this variable for the given analysis to avoid potential bias.

The resultant M^2_{miss} distributions after the removal of the E_{residual} selections are shown in Figures 5.16 and 5.17 for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively. All other default selections, best candidate selections, and those selections chosen after the optimisation process above have been applied.
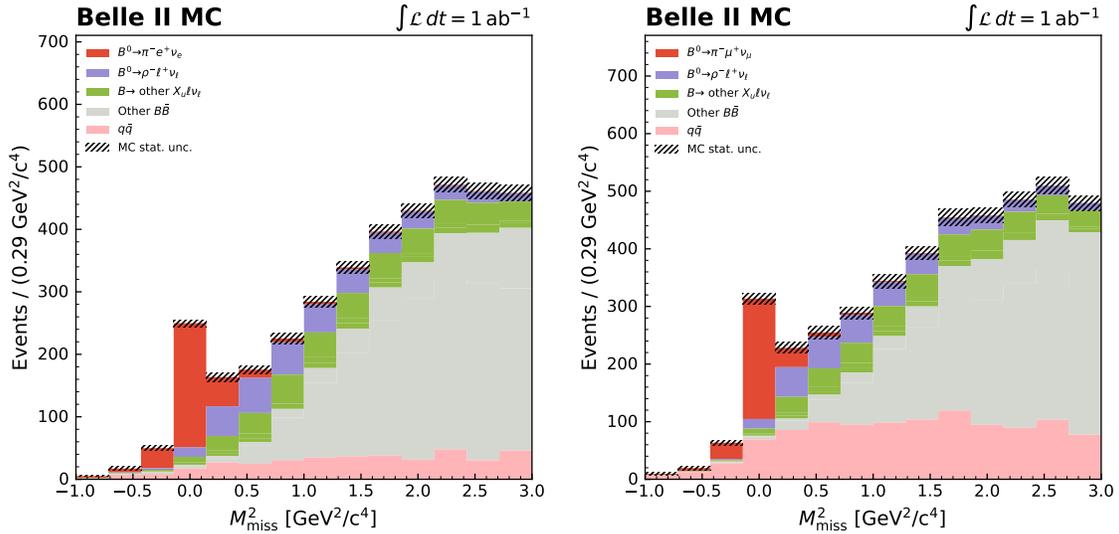
Figure 5.16: M^2_{miss} distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and all optimised selections described in Section 5.5.3 with the exception of the E_{residual} selection, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (right).

As observed, removing the E_{residual} selections resulted in a large increase in the number of background events, both in the signal and sideband regions of M^2_{miss}. In particular, the continuum background within the signal region was observed to rise substantially for both muon channels, $B^0 \to \pi^- \mu^+ \nu_\mu$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$. Despite these enhanced continuum levels, as will be demonstrated in Section 5.6, the application of a continuum suppression Boosted Decision Tree (BDT) post-selections was effective at eliminating the vast majority of the remaining continuum background, and the removal of the E_{residual} selection proved to be a viable decision for the analysis moving forward.
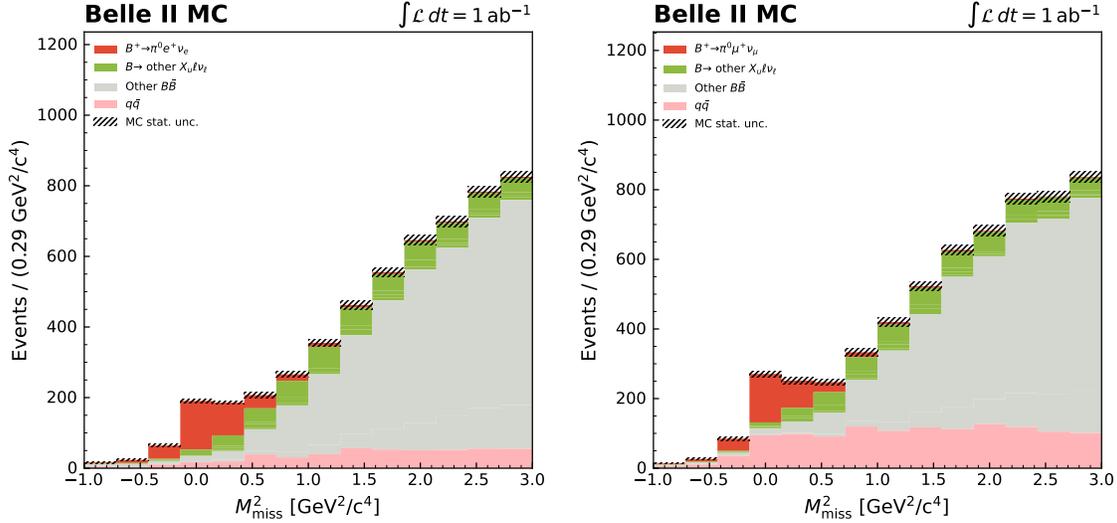
Figure 5.17: M_{miss}^2 distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and all optimised selections described in Section 5.5.3 with the exception of the $E_{residual}$ selection, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right).

M_Y^2

A final selection was implemented after the removal of the $E_{residual}$ selection described above, created specifically for this analysis to target the rejection of continuum events in the M_{miss}^2 signal region. This selection was applied exclusively to the $B^+ \to \pi^0 \mu^+ \nu_\mu$ decay channel, for reasons that will be outlined below.

The M_{miss}^2 distributions shown thus far throughout this chapter have been evaluated over the full range of the square of the four-momentum transfer to the leptonic system, q^2, first introduced in Section 1.2.3. Of particular importance to this analysis, however, is the division of these distributions into bins of q^2, from which the differential decay rates as a function of q^2 and the magnitude of $|V_{ub}|$ will be extracted in the following chapters. With the integrated luminosity of the current data-set, sufficient statistics were available to divide the given M_{miss}^2 distributions into six q^2 bins for the subsequent analysis, namely $q_1 : q^2 \in [0, 4], q_2 : [4, 8], q_3 : [8, 12], q_4 : [12, 16], q_5 : [16, 20]$ and $q_6 : [20, 26.4]$ GeV^2.

Upon investigating the relative background levels in the M_{miss}^2 distributions within each q^2 bin for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, after the removal of the $E_{residual}$ selection, the continuum background was found to peak at 0 GeV^2 in the first q^2 bin alongside the signal MC. Figure 5.18 illustrates the result of splitting the M_{miss}^2 distribution for the $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays given in Figure 5.17 into six q^2 bins, with the peaking nature of the continuum clearly seen in bin q_1. Peaking backgrounds are particularly problematic when considering fits to real data, as the similar background and signal MC templates can potentially result in continuum data events being wrongly assigned as signal events by the fitting algorithm, or vice versa. For this reason, the continuum background in this first q^2 bin was investigated further
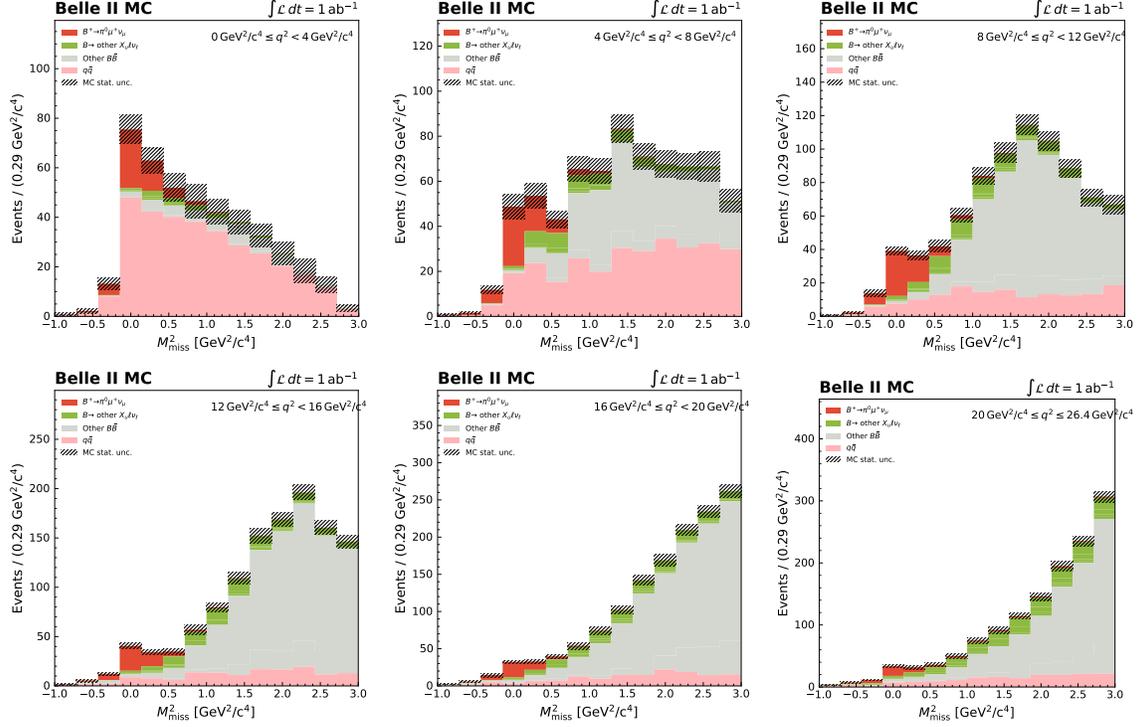
Figure 5.18: M^2_{miss} distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and all optimised selections described in Section 5.5.3 with the exception of the E_{residual} selection, for reconstructed $B^+ \to \pi^0\mu^+\nu_\mu$ decays in six bins of q^2.

in an attempt to remove the peaking nature of this background before the application of the continuum suppression BDT for the overall large-scale reduction of the continuum across all values of q^2.

Figure 5.19 depicts the $B^+ \to \pi^0\mu^+\nu_\mu$ events reconstructed from continuum MC distributed on a 2-dimensional plot of M^2_{miss} versus q^2. The events are split into two categories: those in which the reconstructed muons were real muons in the underlying MC truth, and those in which the muons were fake, misreconstructed from other final-state particles, largely charged pions and kaons. As can be observed, the overwhelming majority (over 85%) of muon candidates reconstructed from the continuum background were fake, with a substantial amount of these fake muons concentrated in the first q^2 bin, $0 \leqslant q^2 < 4$ GeV^2.

For a correctly reconstructed hadronic tagged $B^+ \to \pi^0\mu^+\nu_\mu$ event, the peak at $M^2_{\mathrm{miss}} = 0$ GeV^2 originates from the (approximately) massless neutrino being the only particle in the event unable to be reconstructed. For a continuum background event, however, this premise is entirely invalid, with the production of events in which $M^2_{\mathrm{miss}} = 0$ GeV^2 being due to other underlying causes. As discussed, the pion and muon can be combined into the pseudo-particle Y, with four momentum $p_Y = p_{\pi^0} + p_\mu$. Under the assumption that the four-momentum of the signal B-meson, B_{sig}, is a combination of the Y- and missing four-vectors, $p_{B_{\mathrm{sig}}} \equiv p_Y + p_{\mathrm{miss}}$,
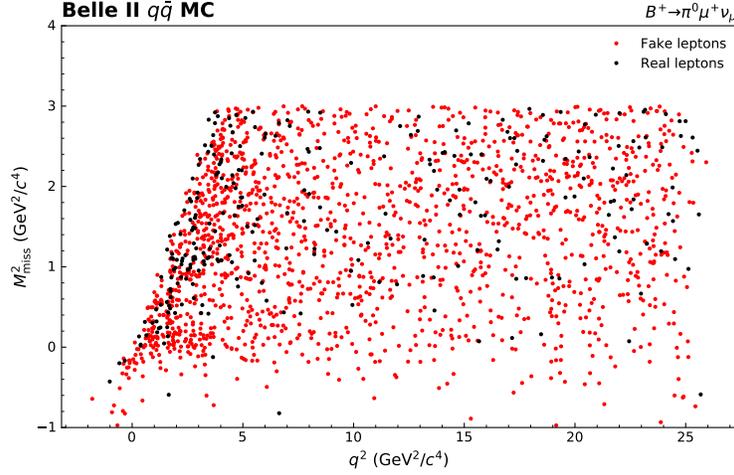
Figure 5.19: Distribution of M^2_{miss} vs. q^2 for $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays reconstructed from $q\bar{q}$ MC, after all FEI skim selections, default analysis selections, best candidate selections and all optimised selections described in Section 5.5.3 with the exception of the E_{residual} selection.
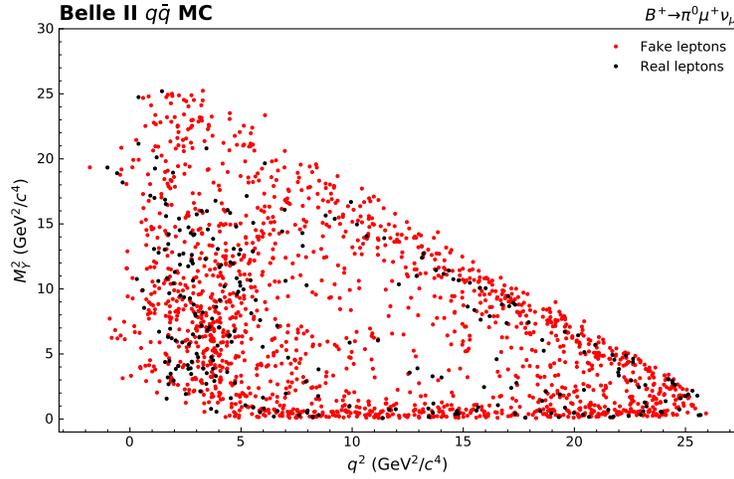

Figure 5.20: Distribution of M^2_Y vs. q^2 for $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays reconstructed from $q\bar{q}$ MC, after all FEI skim selections, default analysis selections, best candidate selections and all optimised selections described in Section 5.5.3 with the exception of the E_{residual} selection.

continuum events in which $M^2_{\mathrm{miss}} \sim 0\ \mathrm{GeV}^2$, in turn implying a null four-vector for the missing momentum, $p_{\mathrm{miss}} \sim 0\ \mathrm{GeV}$, must correspond to events in which $p_Y \sim p_{B_{\mathrm{sig}}}$. It directly follows that the square of the four-momentum of the Y, or the square of the Y invariant mass, $M^2_Y \equiv p^2_Y$, must be consistent with the mass of the signal B-meson, $M^2_Y \sim M^2_B$. In other words, continuum events with low M^2_{miss} values must also possess high M^2_Y values close to the square of the B-meson mass. In this way, the continuum events peaking within the signal region in the first q^2 bin can be directly targeted by restricting M^2_Y to lower values.

Figure 5.20 displays a 2-dimensional plot of the distribution of M^2_Y against q^2 for $B^+ \to$

$\pi^0\mu^+\nu_\mu$ decays reconstructed from continuum MC. In the q^2 region corresponding to bin q_1, the vast majority of events with higher M_Y^2 values above 20 GeV^2 are those in which a fake muon was misreconstructed from a pion or kaon. To reject these problematic events, thereby excluding those events responsible for the peaking nature of the continuum background at $M_{\text{miss}}^2 = 0$ GeV^2, it was decided to make the following selection on M_Y^2:

$$M_Y^2 < 20 \text{ GeV}^2 \ (B^+ \to \pi^0\mu^+\nu_\mu)$$

The result of implementing this selection is demonstrated through the M_{miss}^2 distributions shown in Figure 5.21 for the six q^2 bins studied. It can be clearly observed that the peaking behaviour of the continuum background in the first q^2 bin has been mitigated. The given selection also had little effect on the signal MC in this bin, with a drop in the number of remaining signal events of only $\sim 3\%$. The distributions in the five, higher q^2 bins were larger unaffected by the selection, due to the kinematically motivated absence of events with M_Y^2 values above this range as inferred from Figure 5.20.
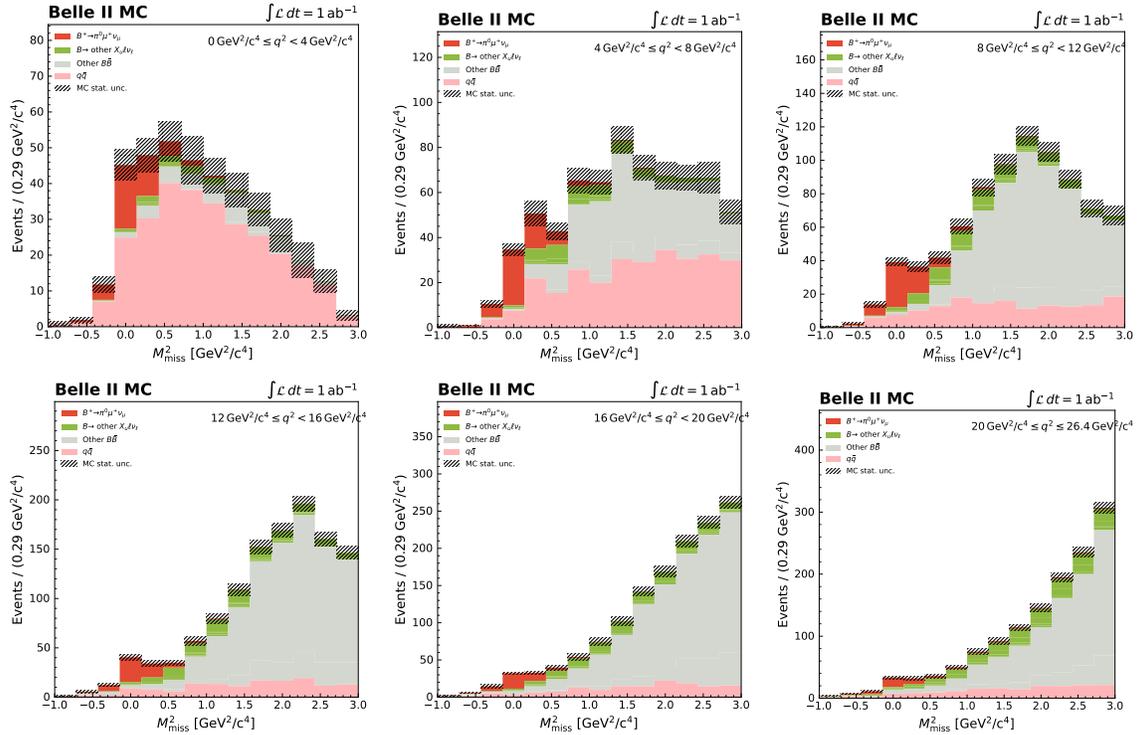

Figure 5.21: M_{miss}^2 distributions in MC after the application of the FEI skim selections, the default analysis selections, the best candidate selections and all optimised selections described in Section 5.5.3 with the exception of the E_{residual} selection and the addition of the M_Y^2 selection, for reconstructed $B^+ \to \pi^0\mu^+\nu_\mu$ decays in six bins of q^2.

5.6   Continuum Suppression

Due to the fact that the event shape differs significantly between continuum and $\Upsilon(4S)$ $\rightarrow B\bar{B}$ events, a common practice at B-factory experiments such as Belle II involves training a classifier to distinguish between continuum and signal events using a number of variables that characterise the event shape and kinematics. This is typically performed after all other analysis selections have been applied. Selections made on the resultant classifier output can then be used to suppress backgrounds from continuum events. As mentioned in the previous sections, dedicated BDT algorithms were trained and implemented after the application of all analysis selections described in Sections 5.5.1 – 5.5.3, for the purpose of rejecting the majority of the remaining continuum background.

Prior to the training process, the continuum and signal MC samples listed in Table 5.1 were split to avoid potential bias, with separate sub-samples used for the BDT training and the subsequent analysis. Of the MC samples used, 50% of the given continuum MC was set aside for the training, with 50% retained for use in the analysis. The $B \rightarrow \pi\ell\nu_\ell$ events were isolated from the $B \rightarrow X_u\ell\nu_\ell$ hybrid samples, and 10% of these events were set aside for the training, leaving 90% of the signal MC events for the analysis. The scaling factors used to normalise the integrated luminosities of the signal and continuum components of the MC in the analysis were updated to reflect these relative percentages.

The `FastBDT` software introduced in Section 2.4.3 was utilised for the training of the continuum suppression BDTs. A number of variables were used in the training, based on Belle II recommendations for well-modelled variables that were found to exhibit reasonable data-MC agreement in an independent study. These included several Cleo Cones [90], measures of the energy flux in various polar angular cones built around the B-meson thrust axes, and Kakuno-Super-Fox-Wolfram (KSFW) moments [46], further analytical functions describing the geometric distribution of particle energies in a given event. The `cosTBTO` variable first introduced in Section 4.1.2 was also used for the training, along with the closely related variable `cosTBz`, in which the cosine of the angle between the B-meson thrust axis and the z-axis as defined by the detector geometry was determined. The full list of variables is given in Table 5.3, and included both those calculated with respect to the reconstructed tag and signal B-mesons. Figures 5.22, 5.23, 5.24 and 5.25 depict the distributions of each training variable in MC, for reconstructed $B^0 \rightarrow \pi^- e^+\nu_e$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B^+ \rightarrow \pi^0 e^+\nu_e$ and $B^+ \rightarrow \pi^0\mu^+\nu_\mu$ decays, respectively.

It was decided to perform the training in each q^2 bin as opposed to doing so over the full q^2 range. As the analysis would ultimately be conducted within the six q^2 bins chosen, this approach meant that the presence of any sharp features in the M^2_{miss} distributions after applying a selection on the BDT classifier output could be avoided. A total of 24 BDTs were trained, one in each q^2 bin for each reconstructed $B \rightarrow \pi\ell\nu_\ell$ mode.

Variable	B_{tag}	B_{sig}
CleoCone(3)	✓	✓
CleoCone(4)	✓	✓
CleoCone(5)	✓	✓
CleoCone(6)	✗	✓
CleoCone(7)	✗	✓
CleoCone(8)	✗	✓
CleoCone(9)	✓	✗
KSFWVariables(hoo4)	✓	✓
KSFWVariables(hso02)	✓	✗
KSFWVariables(hso04)	✓	✓
KSFWVariables(hso12)	✗	✓
KSFWVariables(hso14)	✓	✓
KSFWVariables(hso22)	✗	✓
KSFWVariables(hso24)	✗	✓
cosTBTO	✓	✓
cosTBz	✓	✓

Table 5.3: Variables used in the training of the continuum suppression BDTs for the $B \to \pi\ell\nu_\ell$ analysis. Check marks indicate the inclusion of the given variable calculated with respect to the particular B-meson reconstructed, that is, the B_{tag} or the B_{sig}.

Figures 5.26, 5.27, 5.28 and 5.29 show the MC distributions of the resultant BDT classifier output, $\mathcal{P}_{\text{CS}}$, in each q^2 bin for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, respectively. Clear discrimination between continuum and signal events is observed in all cases. Additionally, the majority of events are skewed towards values of 0 or 1, with few events populating the middle range of $0.2 \leqslant \mathcal{P}_{\text{CS}} \leqslant 0.8$. Given this behaviour, a single selection was chosen for each mode in each q^2 bin for simplicity:

$$\mathcal{P}_{\text{CS}} > 0.6$$

Figure 5.22: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays.
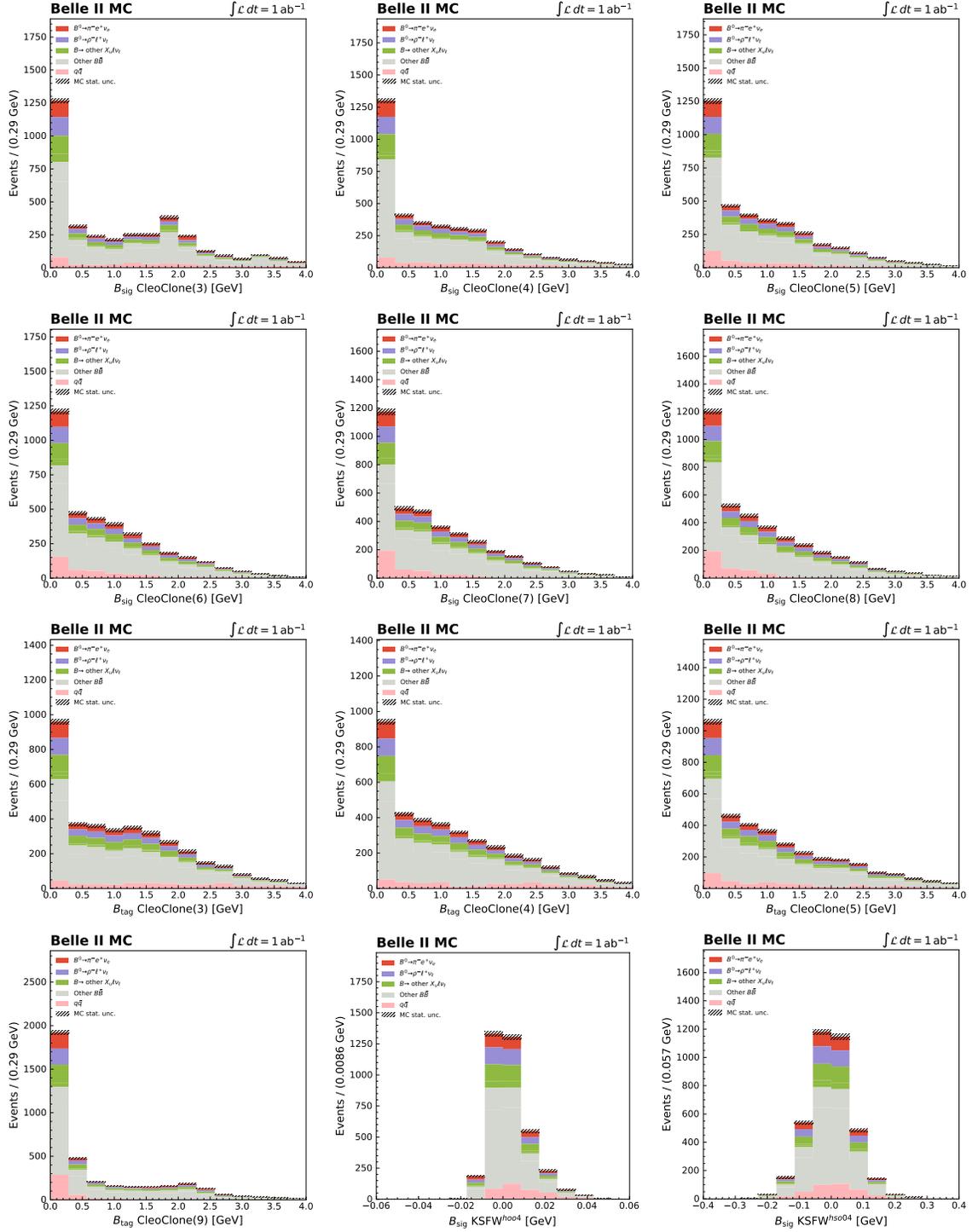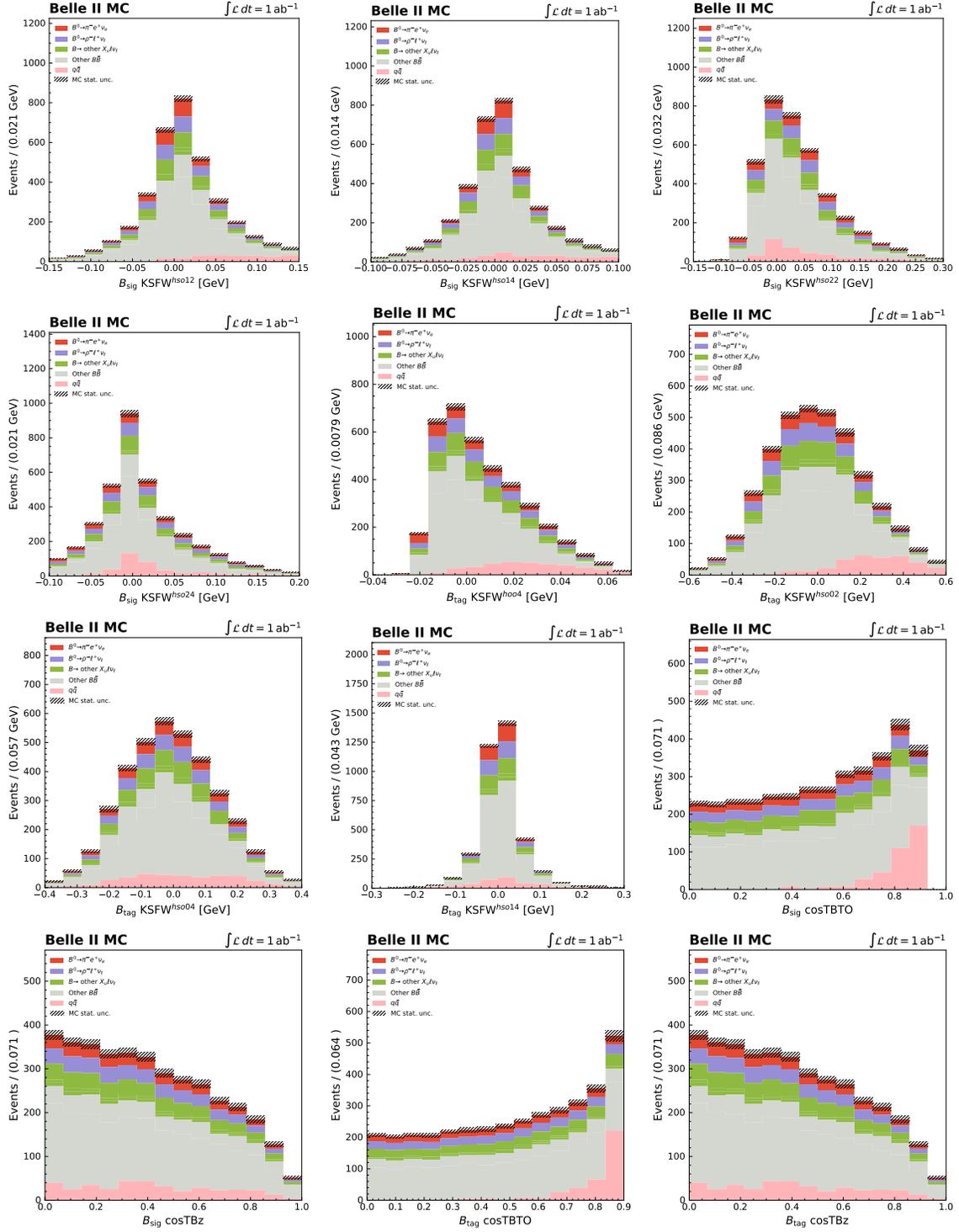
Figure 5.22: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays.
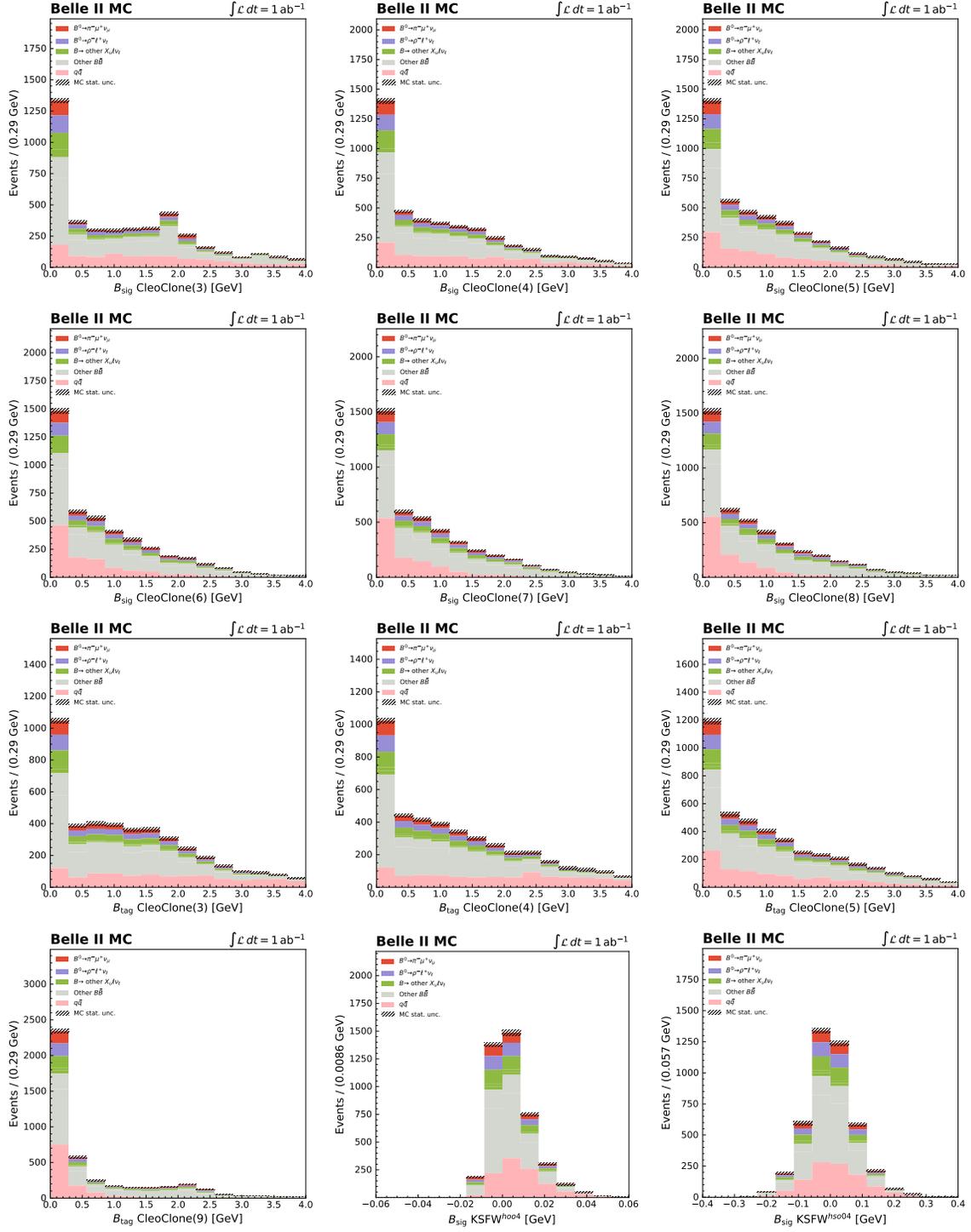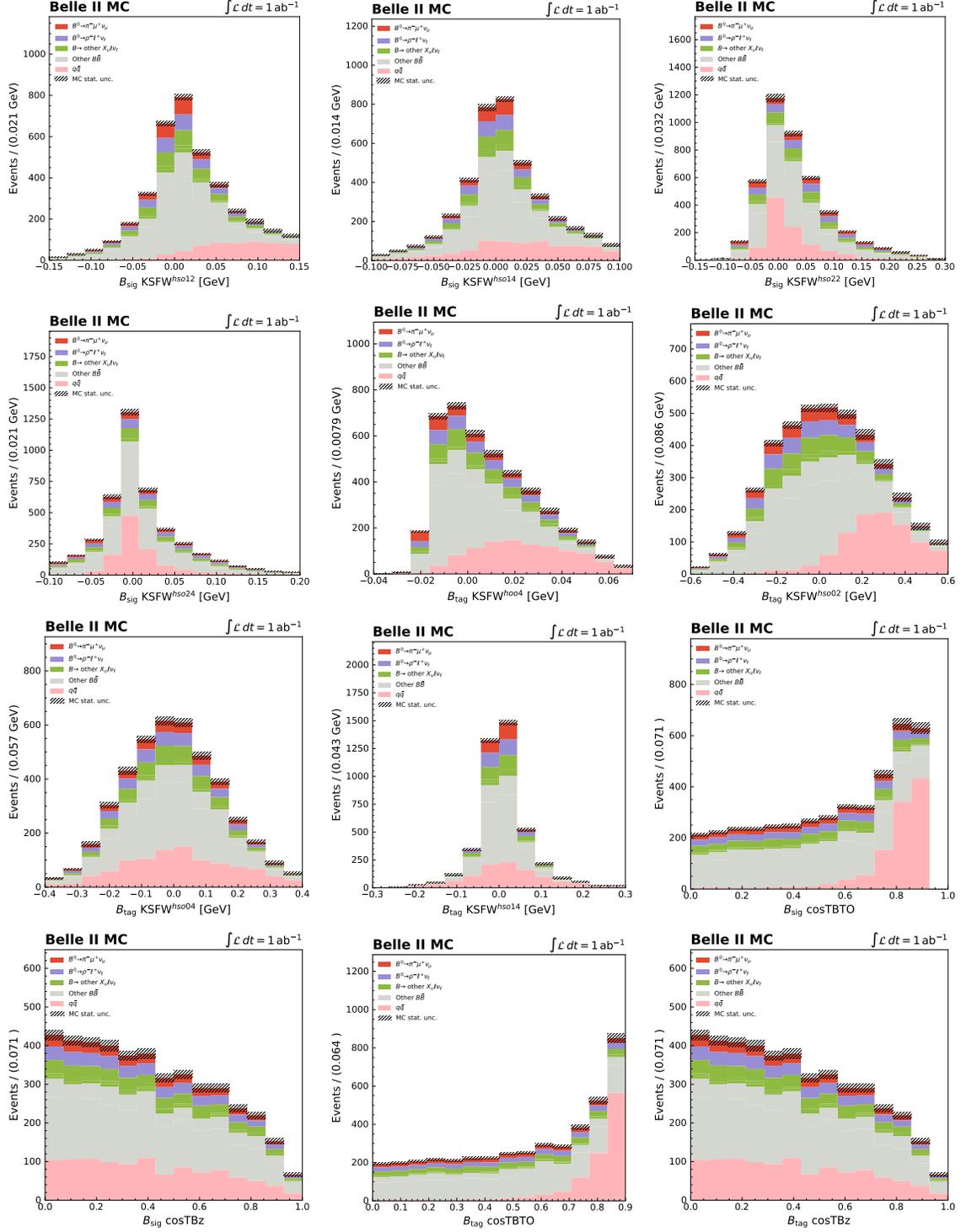
Figure 5.23: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays.
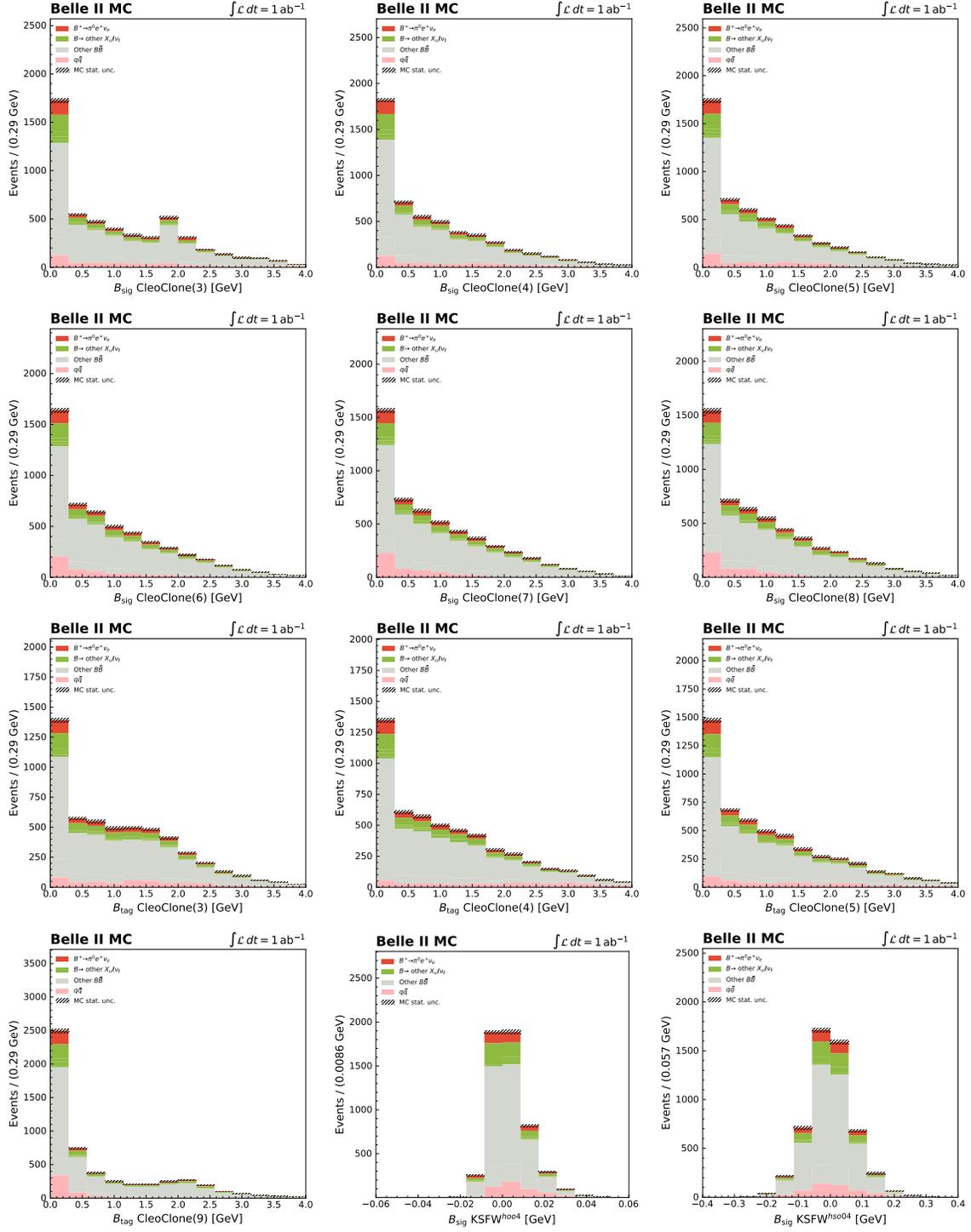
Figure 5.23: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays.

Figure 5.24: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays.
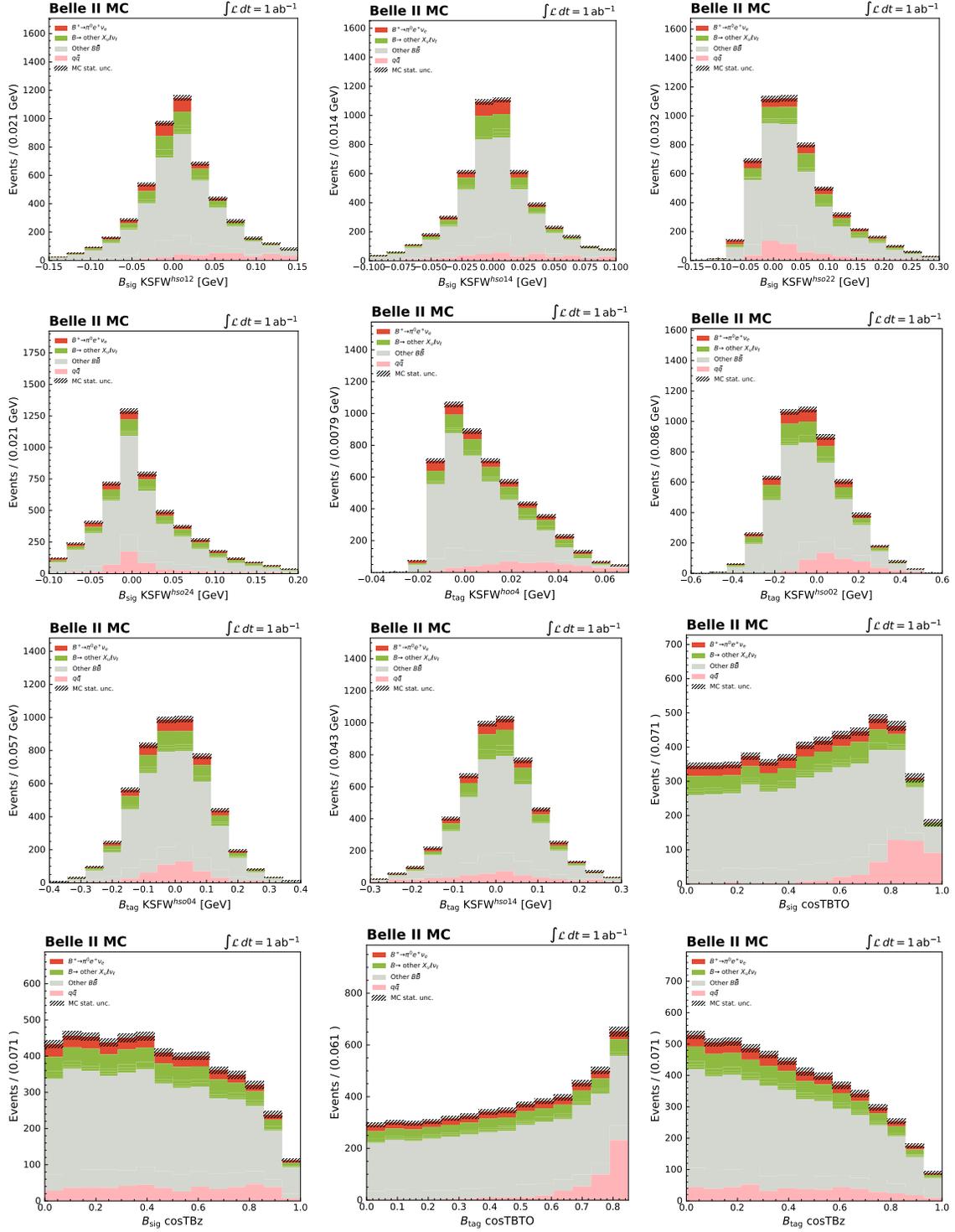
Figure 5.24: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays.
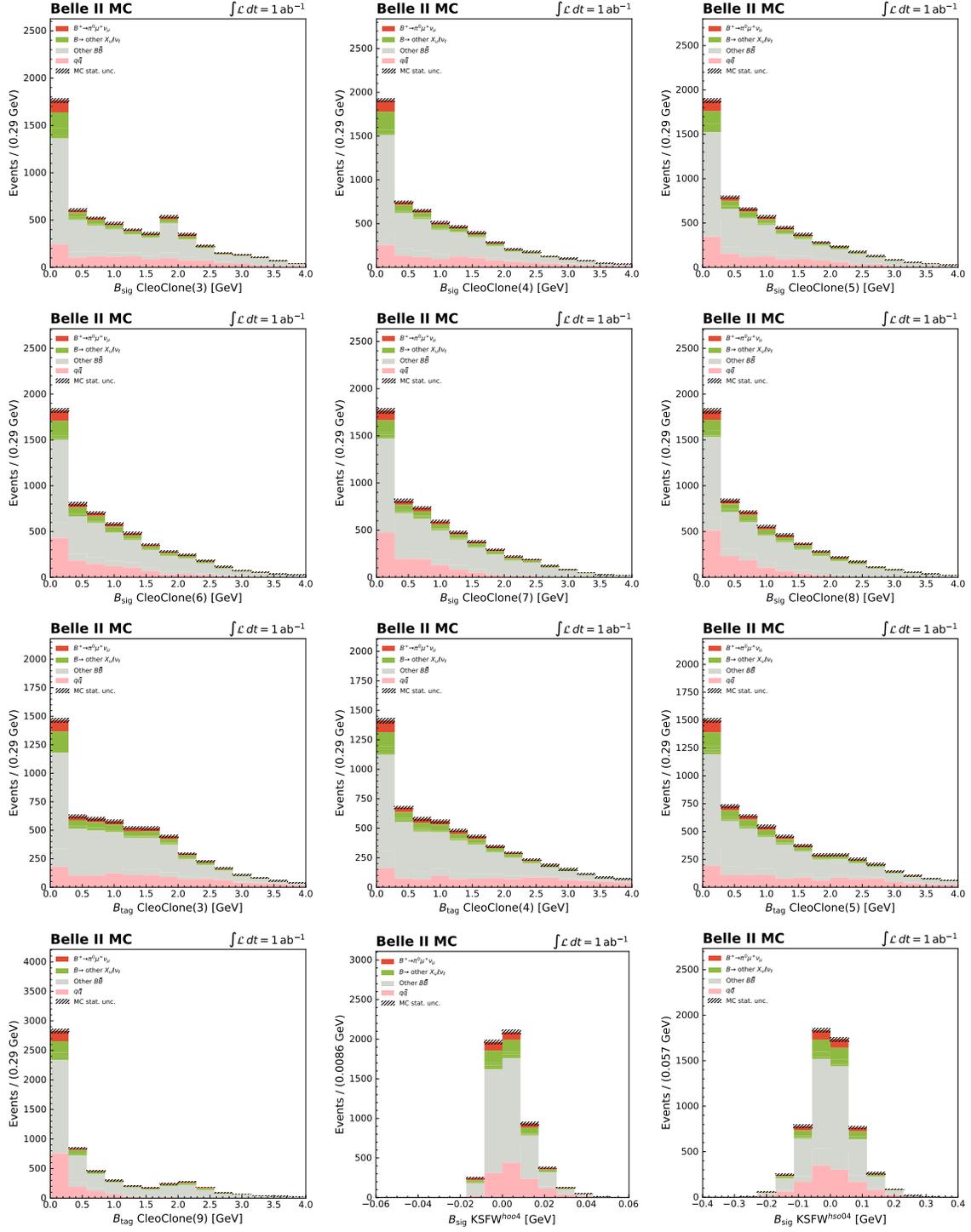
Figure 5.25: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays.
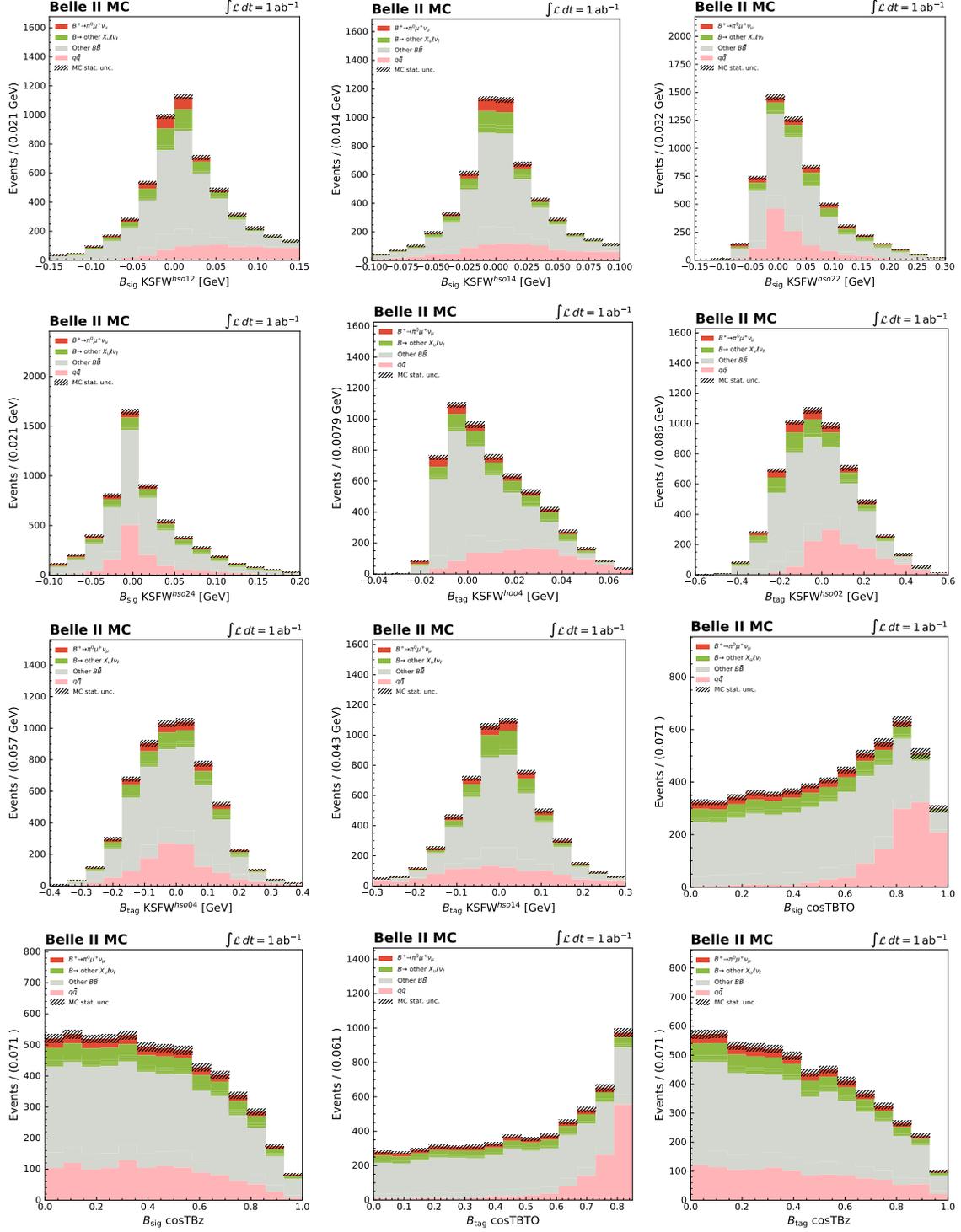
Figure 5.25: Distributions in MC of the variables used to train the continuum suppression BDTs, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays.

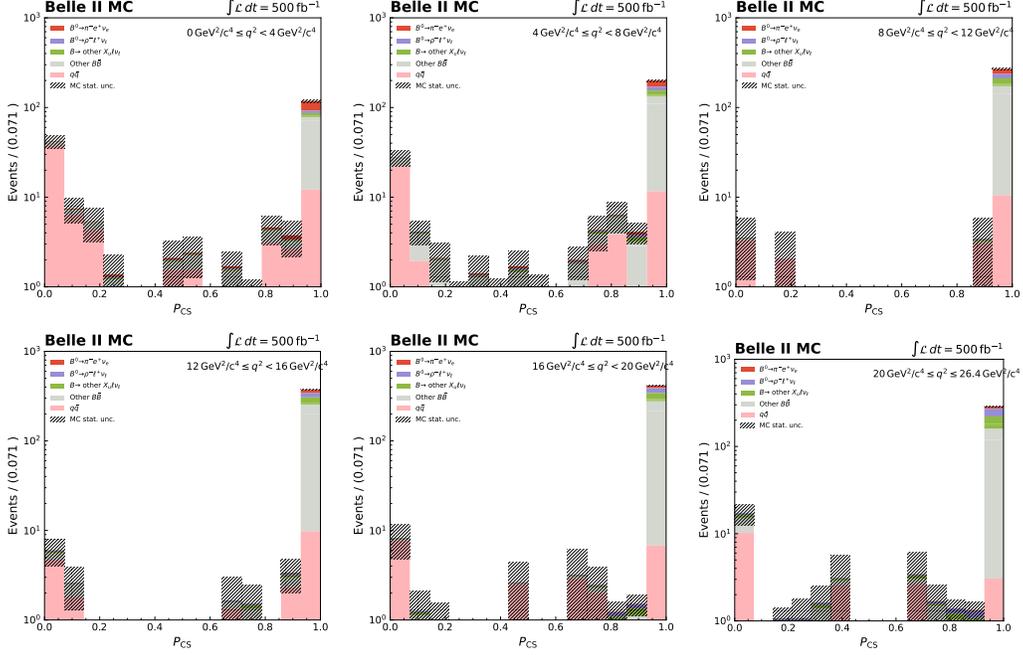Figure 5.26: Distribution of the BDT classifier output P_{CS} in MC for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays, in six bins of q^2. A log scale is used on the y-axis.

Figure 5.27: Distribution of the BDT classifier output P_{CS} in MC for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays, in six bins of q^2. A log scale is used on the y-axis.

125

Figure 5.28: Distribution of the BDT classifier output P_{CS} in MC for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays, in six bins of q^2. A log scale is used on the y-axis.
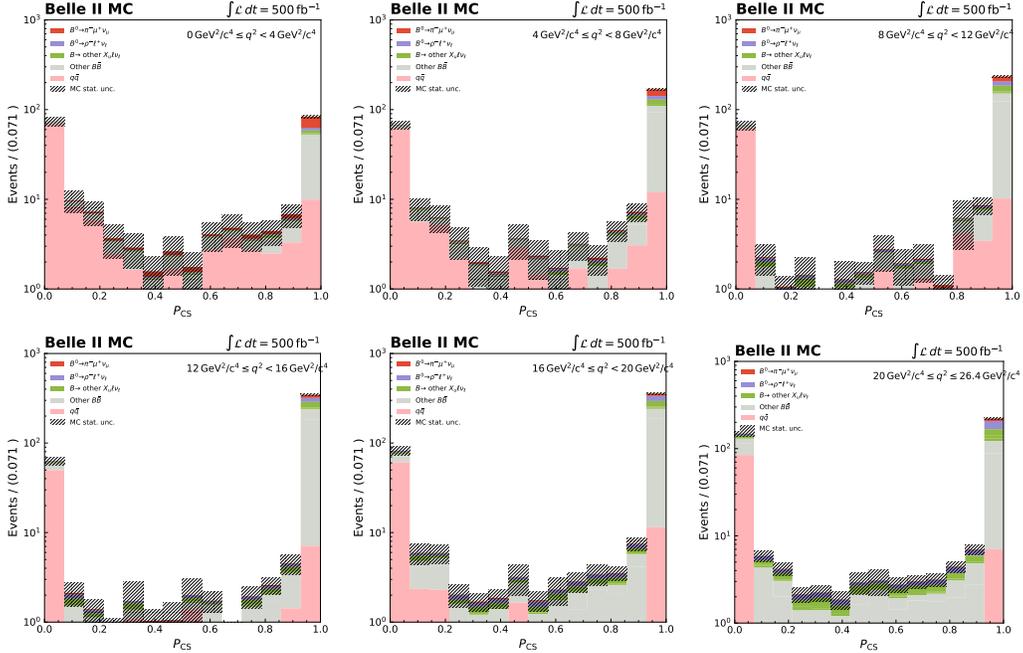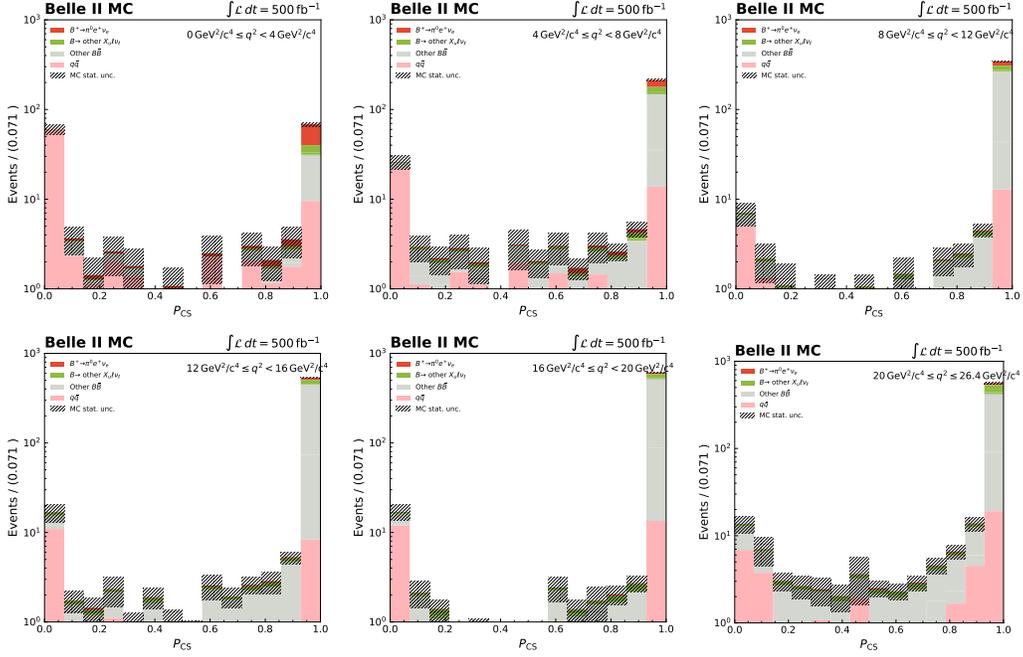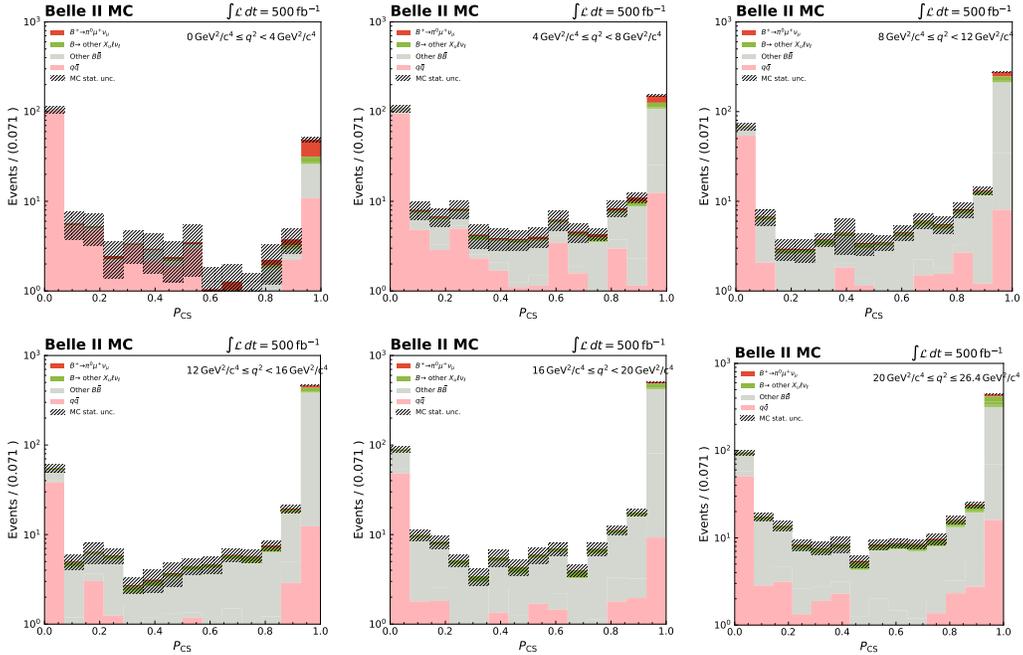

Figure 5.29: Distribution of the BDT classifier output P_{CS} in MC for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, in six bins of q^2. A log scale is used on the y-axis.

5.7 Final Analysis Selections

To summarise the various selections introduced and discussed in the previous sections, the final list of all analysis selections chosen is given in Table 5.4. The final M_{miss}^2 distributions in MC after the application of these selections are displayed in Figures 5.30 and 5.31 over the full q^2 range, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively. The equivalent distributions separated into six q^2 bins are shown in Figures 5.32, 5.33, 5.34, 5.35, 5.36 and 5.37. In each plot, the MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}, equivalent to half of the total continuum sample prior to splitting it into the continuum suppression BDT training sample and the analysis sample. It can be observed that the continuum background has been significantly reduced after the application of the continuum suppression BDT classifier selection. With the analysis selections and the resulting M_{miss}^2 distributions built from MC finalised, the analysis on real data will be explored in the following chapter.

Figure 5.30: M_{miss}^2 distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (top left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (top right). The combined distribution for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, is also shown (bottom). All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.
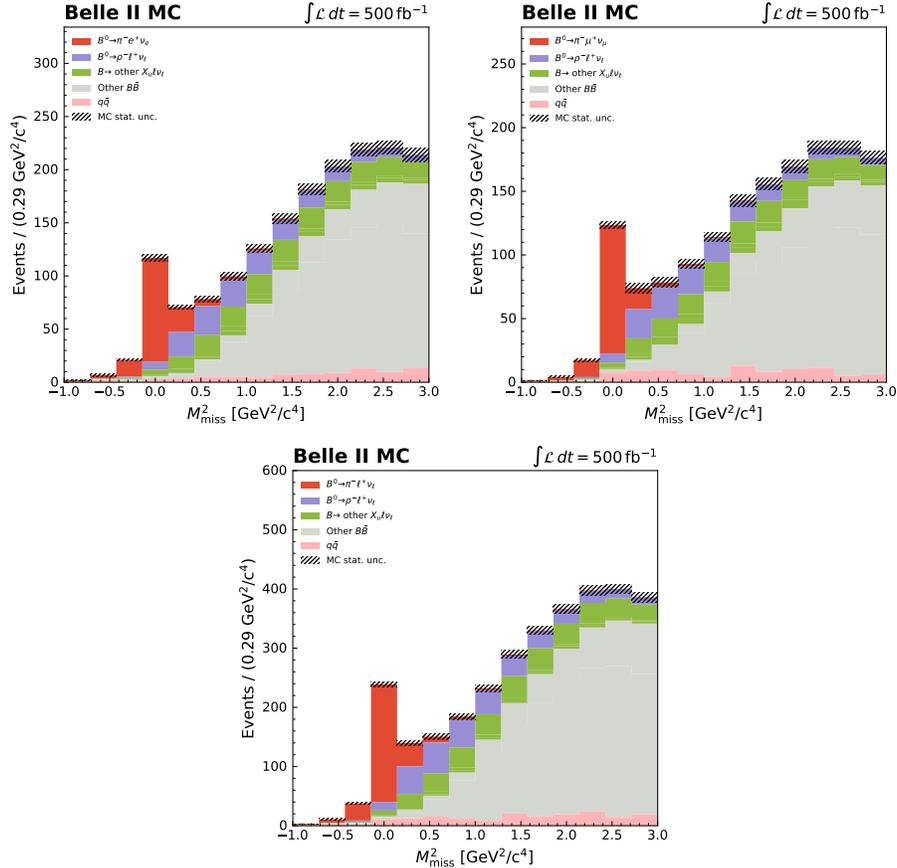
Figure 5.31: M^2_{miss} distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (top left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (top right). The combined distribution for all reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, is also shown (bottom). All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.
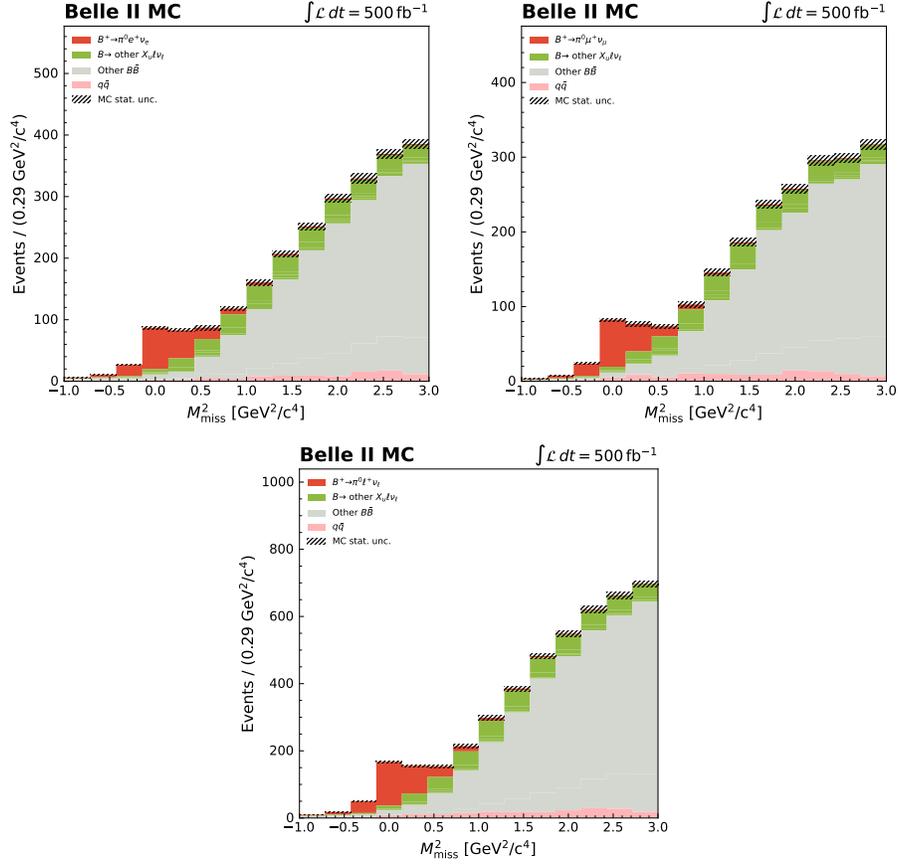
Event	$E_{\text{miss}} > 0.25$ GeV
	No additional tracks in event satisfying: $dr < 2$ cm, $\|dz\| < 5$ cm, $0.297 < \theta < 2.618$ rad., $p_t > 0.2$ GeV, $E > 0.10$ GeV (forward end-cap), $E > 0.09$ GeV (barrel), $E > 0.16$ GeV (backward end-cap)
	$\mathcal{P}_{\text{CS}} > 0.6$
B_{tag}	FEI `SignalProbability` > 0.001
	$M_{\text{bc}} > 5.27$ GeV
	$-0.15 < \Delta E < 0.1$ GeV
	cosTBTO $< 0.90(0.85)$, $B^0_{\text{tag}}(B^+_{\text{tag}})$
	One B_{tag} per event with highest `SignalProbability`
e/μ	$dr < 2$ cm, $\|dz\| < 5$ cm
	$0.297 < \theta < 2.618$ rad.
	$p_{\ell,\text{lab}} > 0.2$ GeV
	`pidChargedBDTScore` > 0.9
	One lepton per event with highest `pidChargedBDTScore`
$\pi^\pm$	$dr < 2$ cm, $\|dz\| < 5$ cm
	$0.297 < \theta < 2.618$ rad.
	`pionID` > 0.6
	`nCDCHits` > 20
γ ($\pi^0 \to \gamma\gamma$)	$0.297 < \theta < 2.618$ rad.
	Cluster from hits to more than one ECL crystal
	$E > 0.080$ GeV (forward end-cap), $E > 0.030$ GeV (barrel), $E > 0.060$ GeV (backward end-cap)
π^0	$0.120 < M < 0.145$ GeV
	$\psi_{\gamma\gamma} < 1.5$ rad.
B^0_{sig}	$-3 < cos\theta_{BY} < 3$
	$\|z_\ell - z_\pi\| < 0.5$ cm
B^+_{sig}	$-3 < cos\theta_{BY} < 3$
	$M^2_Y < 20$ GeV^2 ($B^+ \to \pi^0\mu^+\nu_\mu$ only)
$\Upsilon(4S)$	BCS with highest `pionID` (B^0_{sig}), lowest $\psi_{\gamma\gamma}$ (B^+_{sig})

Table 5.4: Final set of analysis selections chosen for the reconstruction of $B \to \pi\ell\nu_\ell$ decays.

Figure 5.32: M_{miss}^2 distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins. All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.

Figure 5.33: M_{miss}^2 distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins. All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.

Figure 5.34: M^2_{miss} distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, in six q^2 bins. All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.
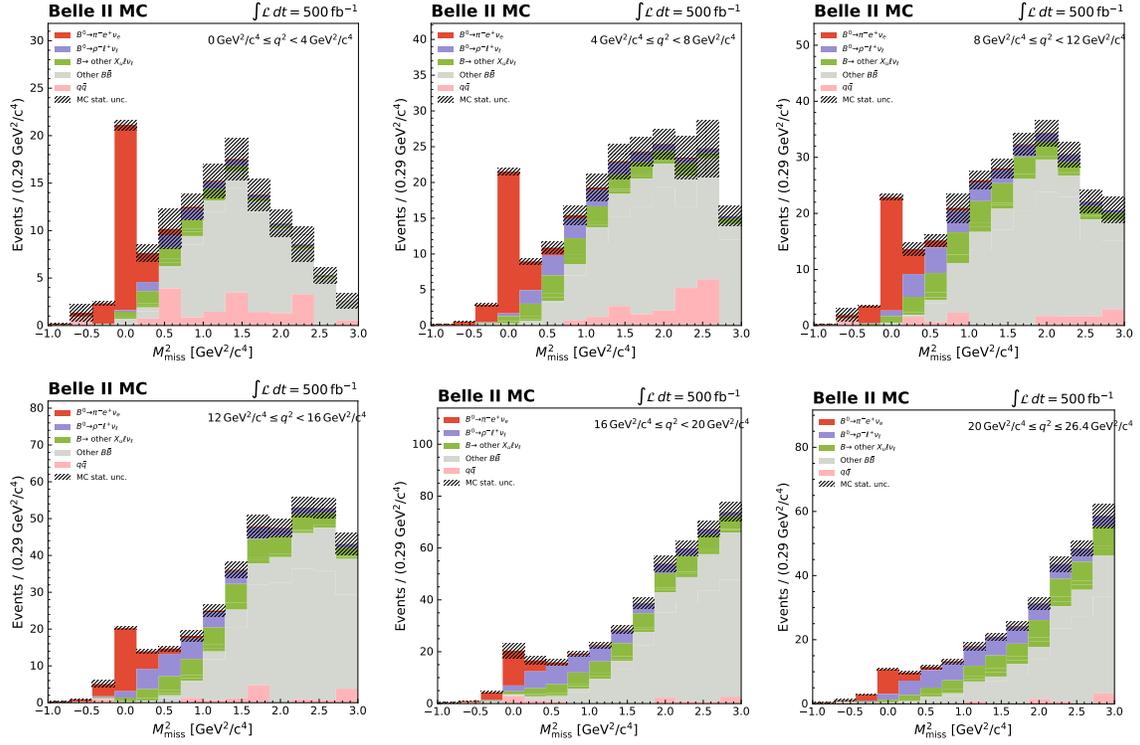
Figure 5.35: M_{miss}^2 distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins. All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.
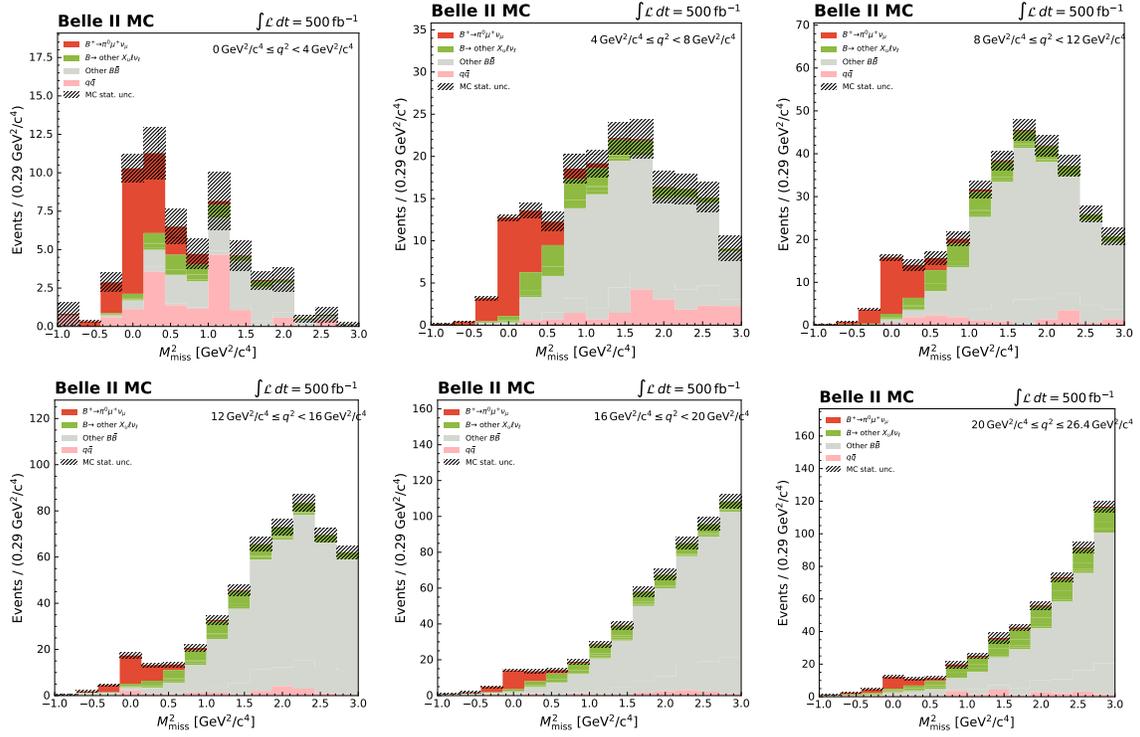
Figure 5.36: M^2_{miss} distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins. All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.
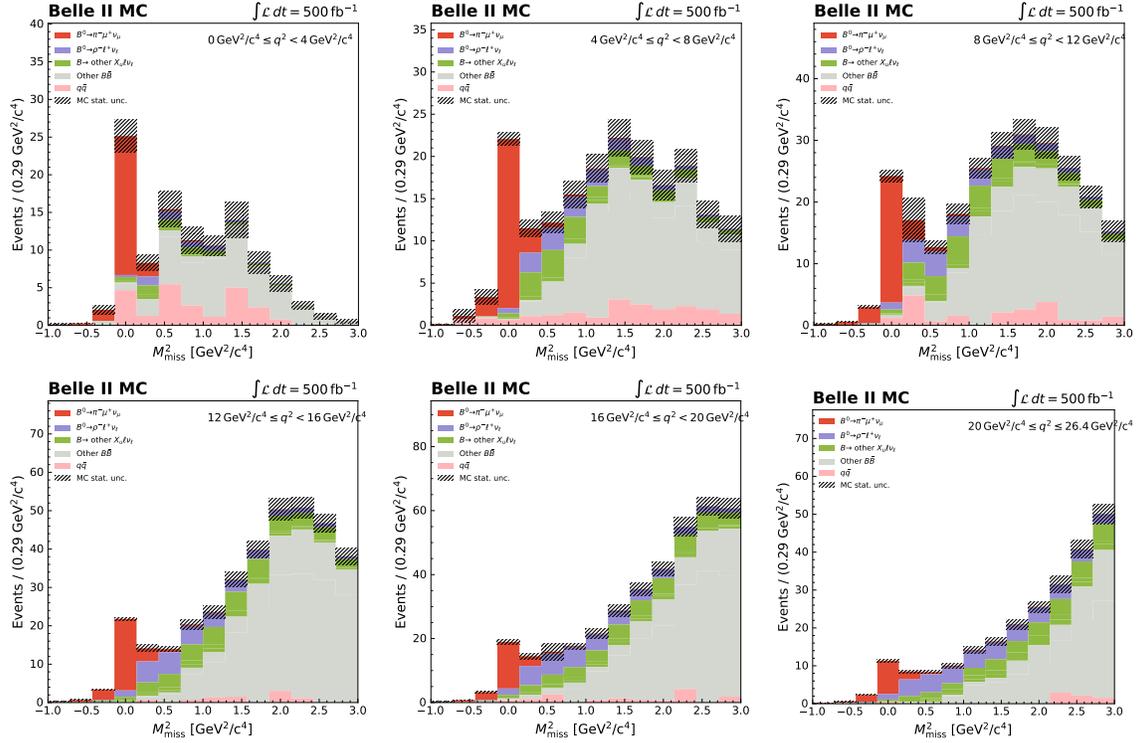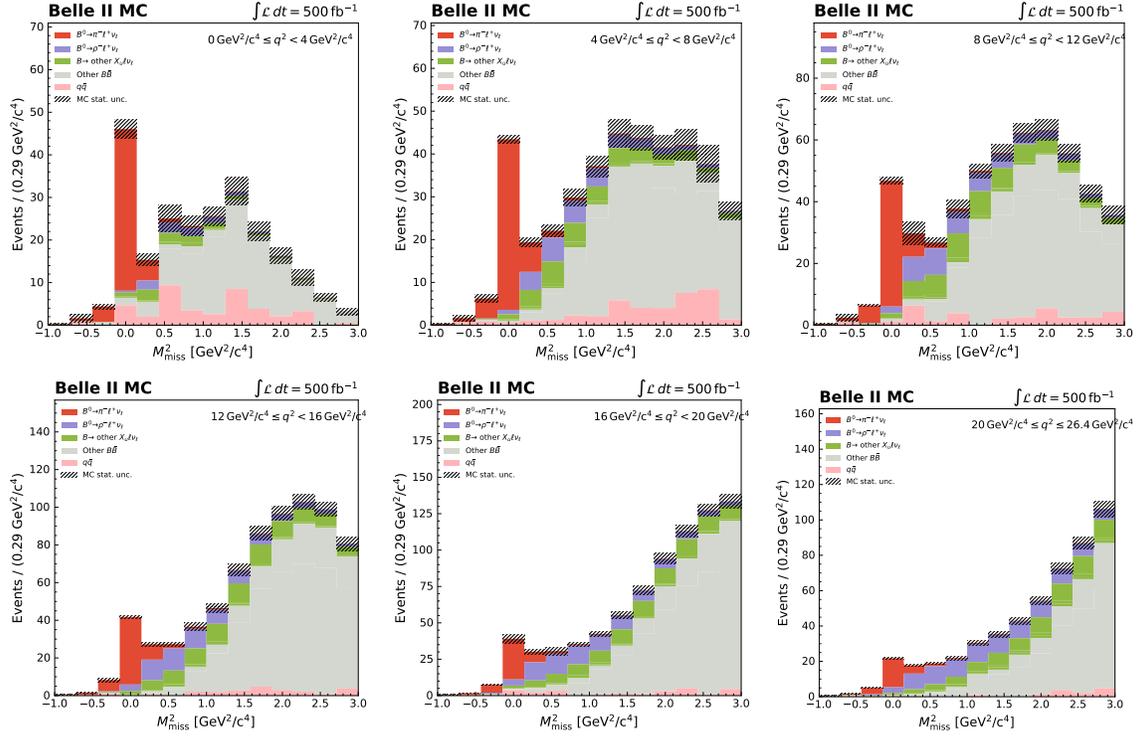
Figure 5.37: M_{miss}^2 distributions in MC after the application of all analysis selections and data-MC corrections, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, in six q^2 bins. All MC has been normalised to the integrated luminosity of the continuum sample used for analysis, 500 fb^{-1}.
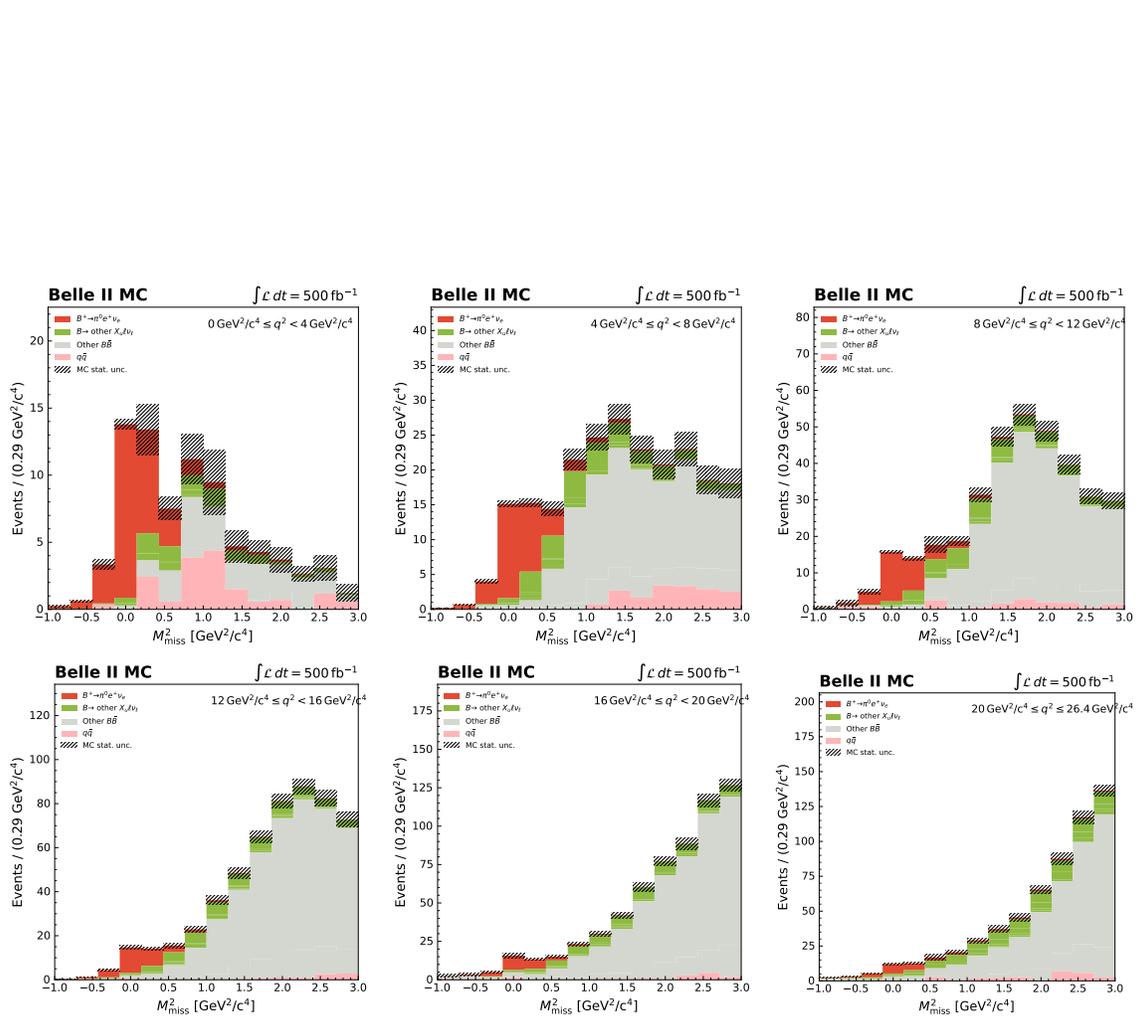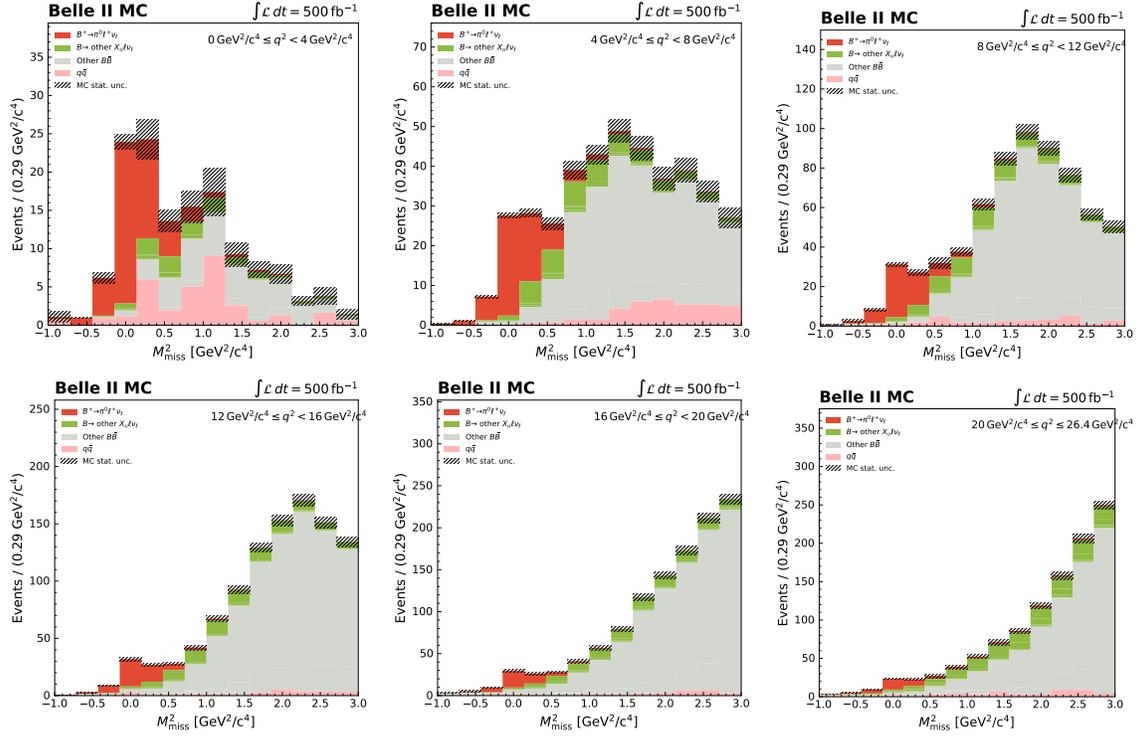
135

Chapter 6

Fitting and Results

The previous chapter established the foundations of the hadronically tagged $B \to \pi \ell \nu_\ell$ analysis, with the description of the Monte Carlo samples used, the relevant data-MC corrections, and ultimately, the development and optimisation of the selections implemented during the reconstruction. The signal extraction variable chosen for the study, M^2_{miss}, was also introduced, and the distributions of this variable in simulated data were shown for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays. M^2_{miss} distributions both over the full q^2 kinematic range and divided into six q^2 bins were presented, in preparation of the measurement of the differential decay rates as a function of q^2 and the extraction of $|V_{\mathrm{ub}}|$.

In this chapter, the fitting procedure used for the extraction of the signal yields from the M^2_{miss} distributions will first be described, including the validation of this procedure using simulated data. The distributions in real data will then be investigated for the first time, and the results of fitting these distributions will be presented and discussed.

6.1   Fitting Strategy

In a similar procedure to that employed for the FEI calibration study detailed in Chapter 4, the binned maximum likelihood method [84] was used to fit the M^2_{miss} distributions for this analysis, implemented via the RooFit software package [85]. The shapes of the signal and background components of the M^2_{miss} distributions in MC after the application of all analysis selections and corrections were used to construct template pdfs for use in the fits. Due to limited statistics, particularly when dividing the M^2_{miss} distributions into bins of q^2, all background components including the continuum, cross-feed from other $B \to X_u \ell \nu_\ell$ decays and other $B\bar{B}$ events were combined into a single background template under the assumption that the relative proportions of each background type were correct. This assumption was reasonable given the statistical limitations, as any potential systematic uncertainty related to the normalisation of the various backgrounds would be insignificant when compared to the total statistical uncertainty. At higher integrated luminosities as more data is collected, this assumption would likely need to be revisited and the fitting procedure revised to incorporate multiple templates for the different background types. Total fit pdfs were subsequently con-

structed from the combination of the signal and background templates, and two-component maximum likelihood fits were performed, returning the fitted signal and background yields with uncertainties taking the size of the fitted samples into account.

The M^2_{miss} distributions of each individual q^2 bin were fitted in this way, extracting the per-bin signal yields. As a secondary measurement and a cross-check, additional fits were also performed to the M^2_{miss} distributions evaluated over the full q^2 kinematic range, extracting the total number of signal events for a given reconstructed $B \to \pi \ell \nu_\ell$ mode. Before applying the fitting procedure to real data, an in-depth study was performed in order to validate the method using MC.

6.2 Fits to Asimov Data

As a first step towards validating the fitting framework, maximum likelihood fits were performed to Asimov data generated from the MC samples at the given data integrated luminosity of 362 fb^{-1}. For each M^2_{miss} distribution, this corresponded to fitting a data-set of equivalent size to the real data sample, but formed from the same combined signal and background MC templates used for the fit. The uncertainty on the number of Asimov data events in each M^2_{miss} bin was taken as the Poisson error on the number of events ($\sqrt{N}$) assuming the given integrated luminosity, resulting in fitted yields with uncertainties of similar magnitude to those expected when fitting real data.

Figures 6.1, 6.2, 6.4 and 6.5 depict the fitted Asimov data distributions in M^2_{miss} for each q^2 bin, for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, respectively. The equivalent distributions fitted over the full q^2 kinematic range are provided in Figures 6.3 and 6.6. In each distribution, the two fit components are delineated by the red and grey histograms representing the signal and background MC templates, with the Asimov data clearly observed to match the total number of signal and background MC events in each M^2_{miss} bin, as expected by construction.

The fitted signal and background yields returned from the maximum likelihood fits to these Asimov data distributions are listed in Table 6.1, alongside the predicted yields determined by counting the raw number of signal and background events in the MC truth, applying all relevant correction and scaling factors, and normalising each MC component to an integrated luminosity of 362 fb^{-1}. All fitted signal and background yields agreed closely with the MC predictions as expected when fitting Asimov data, demonstrating that, when presented with a data-set generated from the exact signal and background distributions given by simulation, the fit was clearly able to distinguish the two components to a high degree of accuracy. In this way, a first step towards the validation of the fitting procedure was achieved, with fits to data-sets under ideal circumstances yielding expected results.

The Asimov yields measured during this study will subsequently be used to validate the methods of extracting the $B \to \pi \ell \nu_\ell$ branching fractions and the magnitude of the matrix element $|V_{\mathrm{ub}}|$ in the following chapter.

Figure 6.1: M_{miss}^2 distributions of Asimov data with MC fit projections overlaid, for recon-
structed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins.

Figure 6.3: M_{miss}^2 distributions of Asimov data with MC fit projections overlaid, for recon-
structed $B^0 \to \pi^- e^+ \nu_e$ (left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (right) decays over the full q^2 kinematic
range.
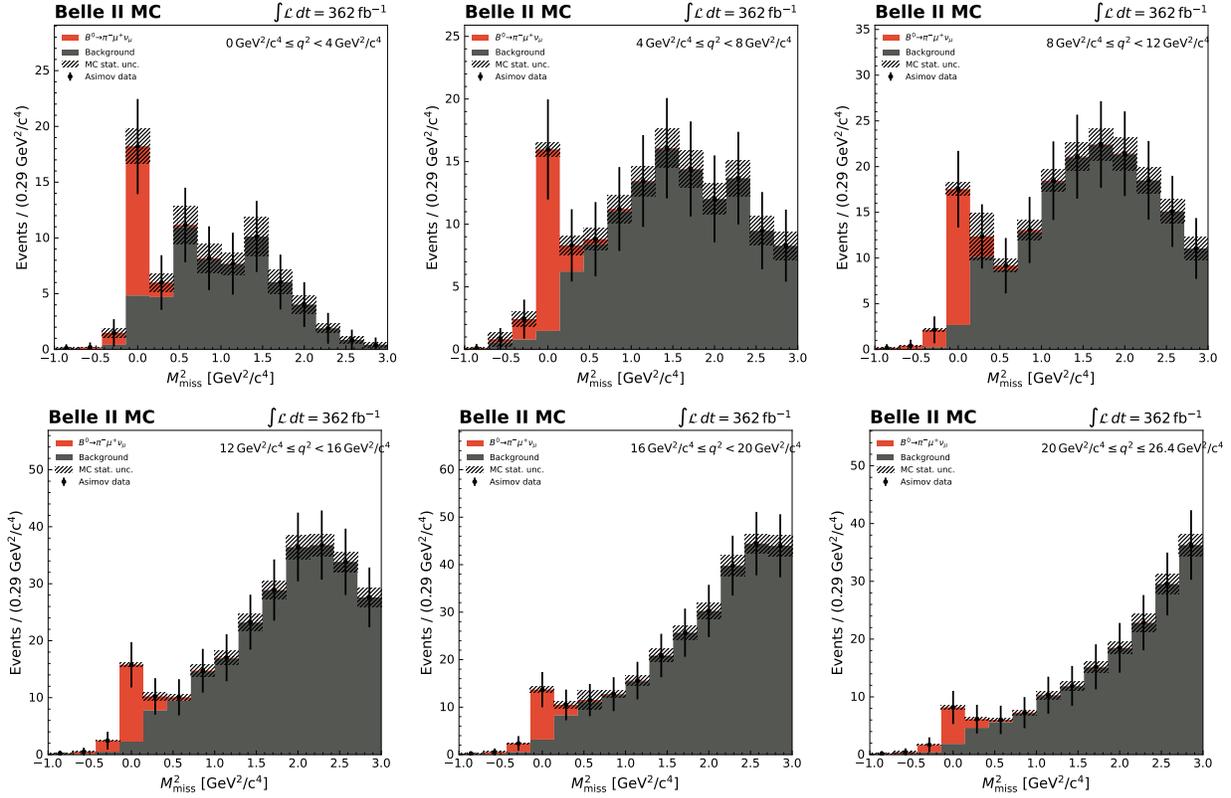
Figure 6.2: M_{miss}^2 distributions of Asimov data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins.
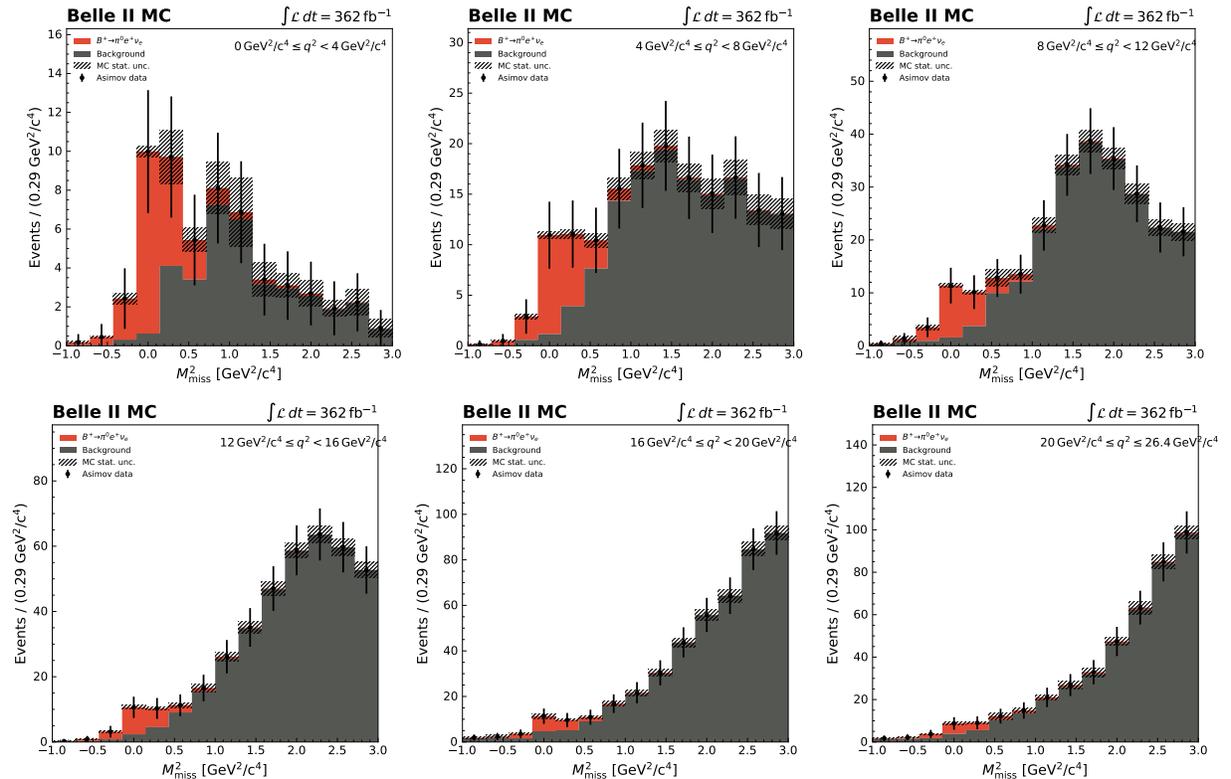
Figure 6.4: M_{miss}^2 distributions of Asimov data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins.

Figure 6.5: M_{miss}^2 distributions of Asimov data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins.
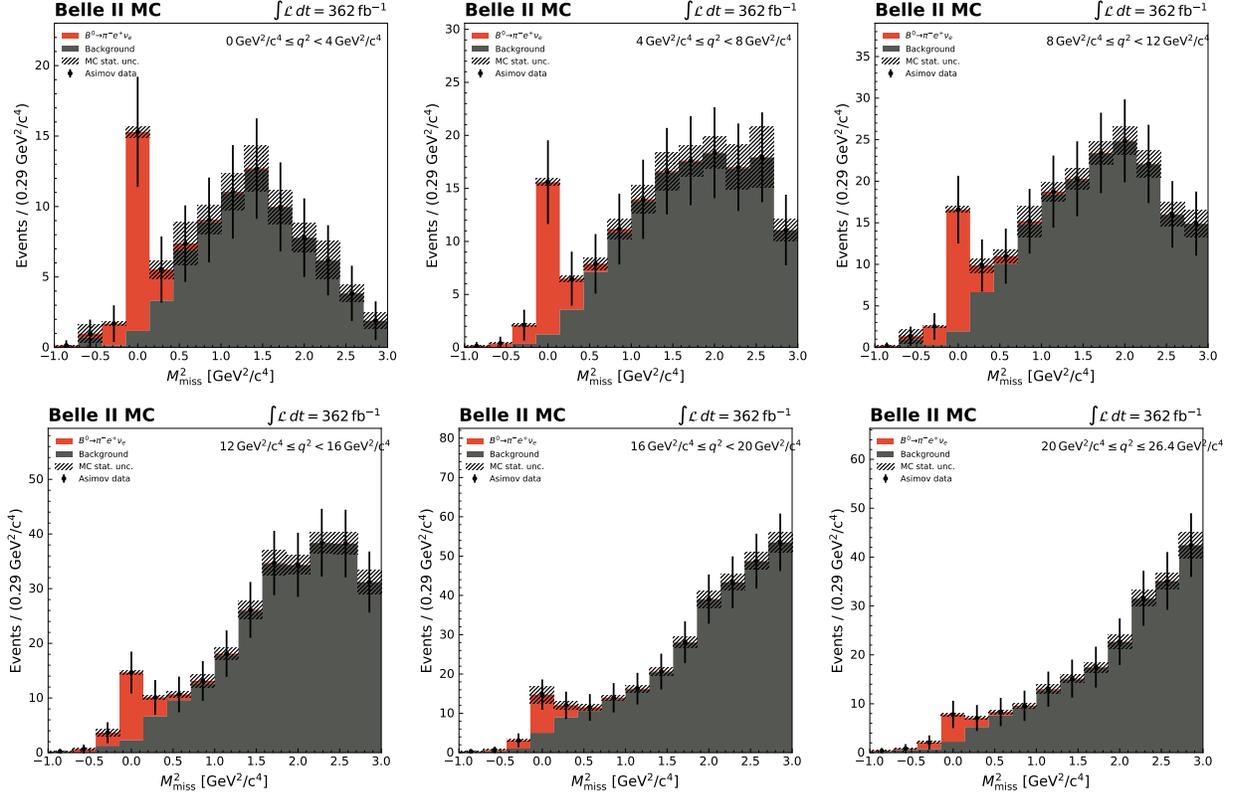

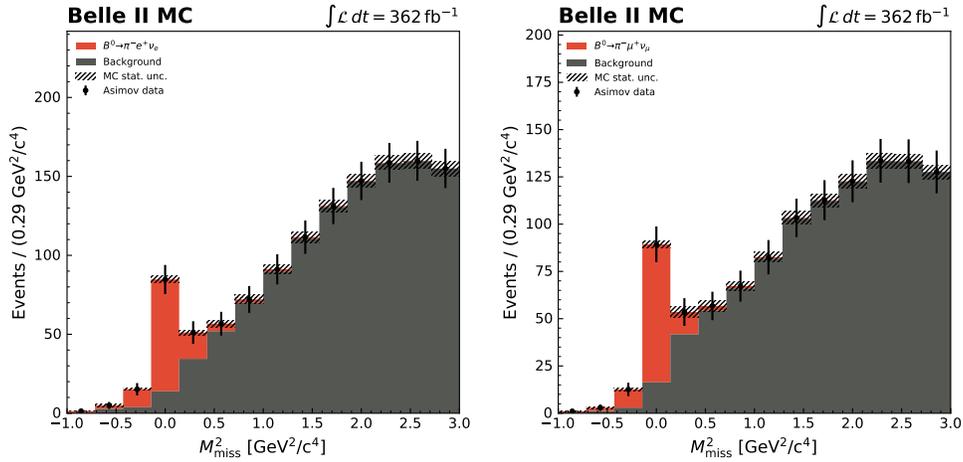
Figure 6.6: M_{miss}^2 distributions of Asimov data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right) decays over the full q^2 kinematic range.
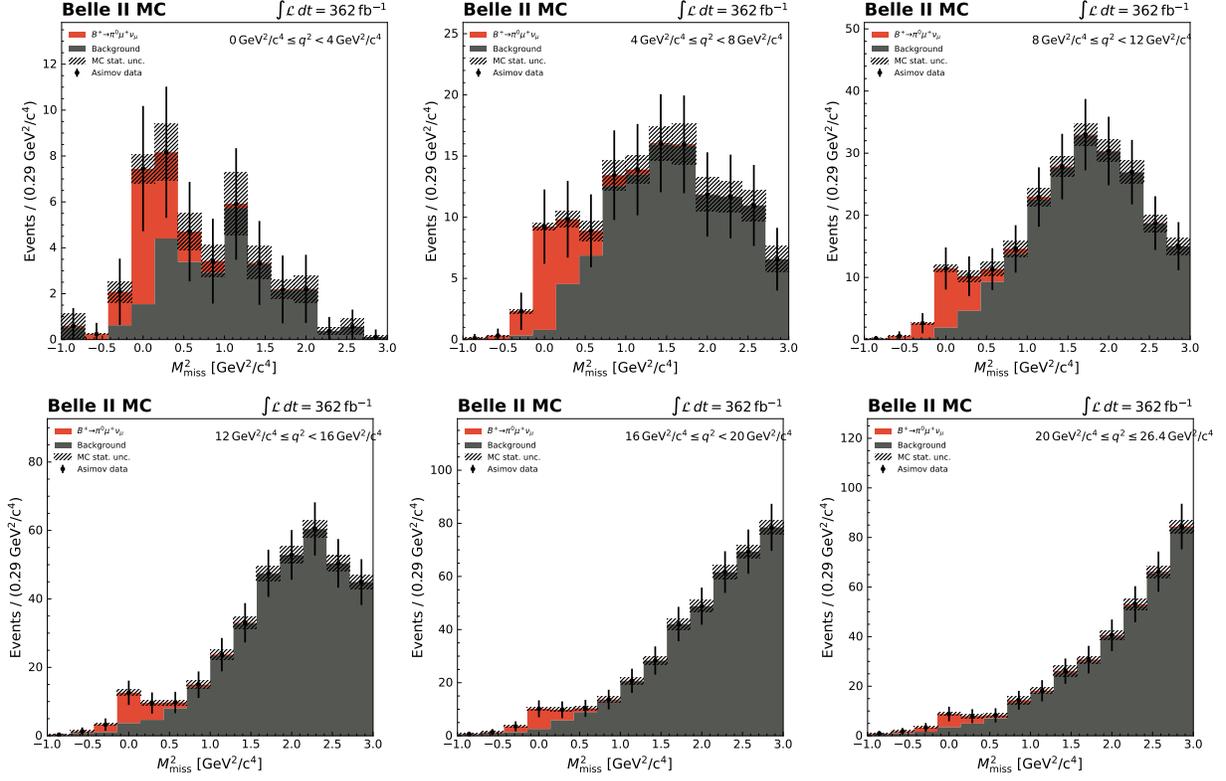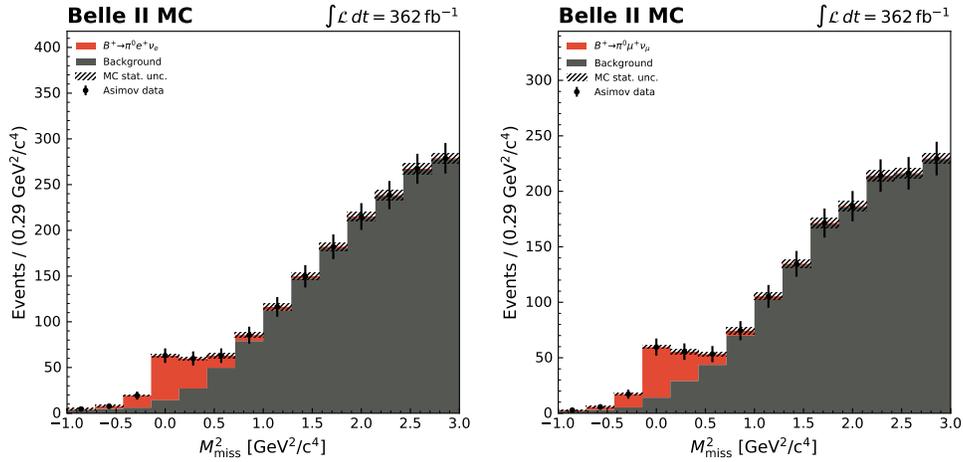
q^2 bin (GeV^2)	Pred. sig. yield	Fit. sig. yield	Pred. bkg. yield	Fit. bkg. yield
		$B^0 \to \pi^- e^+ \nu_e$		
$0 \leqslant q^2 < 4$	19.22	19.19 ± 4.90	74.26	74.35 ± 8.90
$4 \leqslant q^2 < 8$	21.26	21.26 ± 5.26	135.11	135.11 ± 11.89
$8 \leqslant q^2 < 12$	22.67	22.68 ± 5.66	174.34	174.34 ± 13.55
$12 \leqslant q^2 < 16$	21.42	21.42 ± 5.80	252.82	252.82 ± 16.28
$16 \leqslant q^2 < 20$	17.69	17.69 ± 6.01	288.40	288.40 ± 17.52
$20 \leqslant q^2 \leqslant 26.4$	11.09	11.08 ± 4.77	202.43	202.43 ± 14.63
Sum	113.35	113.32 ± 13.28	1127.35	1127.45 ± 34.50
Full range	113.35	113.31 ± 13.19	1127.35	1127.44 ± 34.47
		$B^0 \to \pi^- \mu^+ \nu_\mu$		
$0 \leqslant q^2 < 4$	16.26	16.29 ± 5.03	60.16	60.13 ± 8.31
$4 \leqslant q^2 < 8$	19.44	19.44 ± 5.05	115.65	115.65 ± 11.03
$8 \leqslant q^2 < 12$	20.74	20.74 ± 5.37	162.18	162.18 ± 13.05
$12 \leqslant q^2 < 16$	19.65	19.65 ± 5.27	238.19	238.19 ± 15.69
$16 \leqslant q^2 < 20$	16.13	16.13 ± 5.07	256.04	256.04 ± 16.30
$20 \leqslant q^2 \leqslant 26.4$	10.62	10.62 ± 4.22	163.88	163.87 ± 13.08
Sum	102.84	102.87 ± 12.29	996.09	996.06 ± 32.31
Full range	102.84	102.77 ± 12.27	996.09	996.18 ± 32.31
		$B^+ \to \pi^0 e^+ \nu_e$		
$0 \leqslant q^2 < 4$	21.50	21.51 ± 5.77	35.89	35.89 ± 6.90
$4 \leqslant q^2 < 8$	25.42	25.42 ± 6.35	138.54	138.53 ± 12.39
$8 \leqslant q^2 < 12$	25.42	25.42 ± 6.57	231.22	231.22 ± 15.78
$12 \leqslant q^2 < 16$	21.94	21.94 ± 6.28	374.09	374.09 ± 19.79
$16 \leqslant q^2 < 20$	17.94	17.94 ± 6.68	431.02	431.03 ± 21.39
$20 \leqslant q^2 \leqslant 26.4$	19.96	19.96 ± 8.66	407.72	407.72 ± 21.51
Sum	132.18	132.19 ± 16.61	1618.48	1618.48 ± 41.98
Full range	132.18	132.08 ± 16.11	1618.48	1618.48 ± 41.78
		$B^+ \to \pi^0 \mu^+ \nu_\mu$		
$0 \leqslant q^2 < 4$	13.58	13.55 ± 5.19	27.76	27.80 ± 6.42
$4 \leqslant q^2 < 8$	19.83	19.83 ± 5.54	111.32	111.32 ± 11.05
$8 \leqslant q^2 < 12$	21.91	21.91 ± 5.93	203.86	203.86 ± 14.74
$12 \leqslant q^2 < 16$	20.12	20.12 ± 6.17	344.46	344.46 ± 19.04
$16 \leqslant q^2 < 20$	17.32	17.31 ± 5.80	381.67	381.68 ± 19.95
$20 \leqslant q^2 \leqslant 26.4$	18.08	18.16 ± 7.66	346.58	346.45 ± 19.67
Sum	110.84	110.83 ± 14.94	1415.66	1415.57 ± 39.09
Full range	110.84	110.66 ± 14.57	1415.66	1415.66 ± 38.95

Table 6.1: Fitted signal and background yields obtained from the maximum likelihood fits to Asimov data at 362 fb^{-1}. The predicted yields are also listed.

6.3 Toy MC Studies

Whilst Asimov data was useful for ensuring the fitting procedure performed as expected given a single best case scenario, further validation was required to investigate the fit performance for a wide variety of potential outcomes given the current data integrated luminosity. In order to conduct such an investigation, mock data-sets known as toys were generated from the simulated signal and background M^2_{miss} distributions. Unlike Asimov data, the toy data-sets were not generated to match the MC expectation exactly, but rather, to be consistent with the MC expectation within Poisson uncertainties.

Figure 6.7: Gaussian pdfs fitted to the distribution of signal yields from 1000 MC toys, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins.

For each q^2 bin of each reconstructed $B \to \pi \ell \nu_\ell$ mode, one thousand toy data-sets were generated, each with a number of events randomly pulled from a Poisson distribution centered around the total number of expected signal and background events at an integrated luminosity of 362 fb^{-1}. The event yields in each M^2_{miss} bin were similarly sampled using Poisson statistics, allowing for a variation in the number of events per-bin with respect to the

Figure 6.8: Gaussian pdfs fitted to the distribution of signal yields from 1000 MC toys, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins.

Figure 6.9: Gaussian pdfs fitted to the distribution of signal yields from 1000 MC toys, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (right) decays.

MC expectation. Similar toys were also generated for the M^2_{miss} distributions evaluated over the full q^2 kinematic range. The RooFit software package [85] was utilised for the generation

of all toy data-sets.

Maximum likelihood fits were subsequently performed to the M_{miss}^2 distributions of each toy in the same manner as described in Section 6.1, resulting in a set of 1000 fitted signal and background yields for each toy study. At such a large number of samples, a Poisson distribution can be well approximated by a normal, or Gaussian, distribution [83]. Given this approximation, a Gaussian pdf was then fit to the distribution of these signal yields, with the mean yield and standard deviation returned from the fit. Figures 6.7, 6.8 and 6.9 depict the distributions of the signal yields for each toy overlaid with the fitted Gaussian pdf, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ and $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six bins of q^2 and over the full q^2 kinematic range, respectively, with the equivalent distributions for $B^+ \to \pi^0 \ell^+ \nu_\ell$ shown in Figures 6.10, 6.11 and 6.12. The mean yield and standard deviation returned from each fitted Gaussian pdf are listed in Table 6.2 alongside the predicted yields as per the MC expectation.
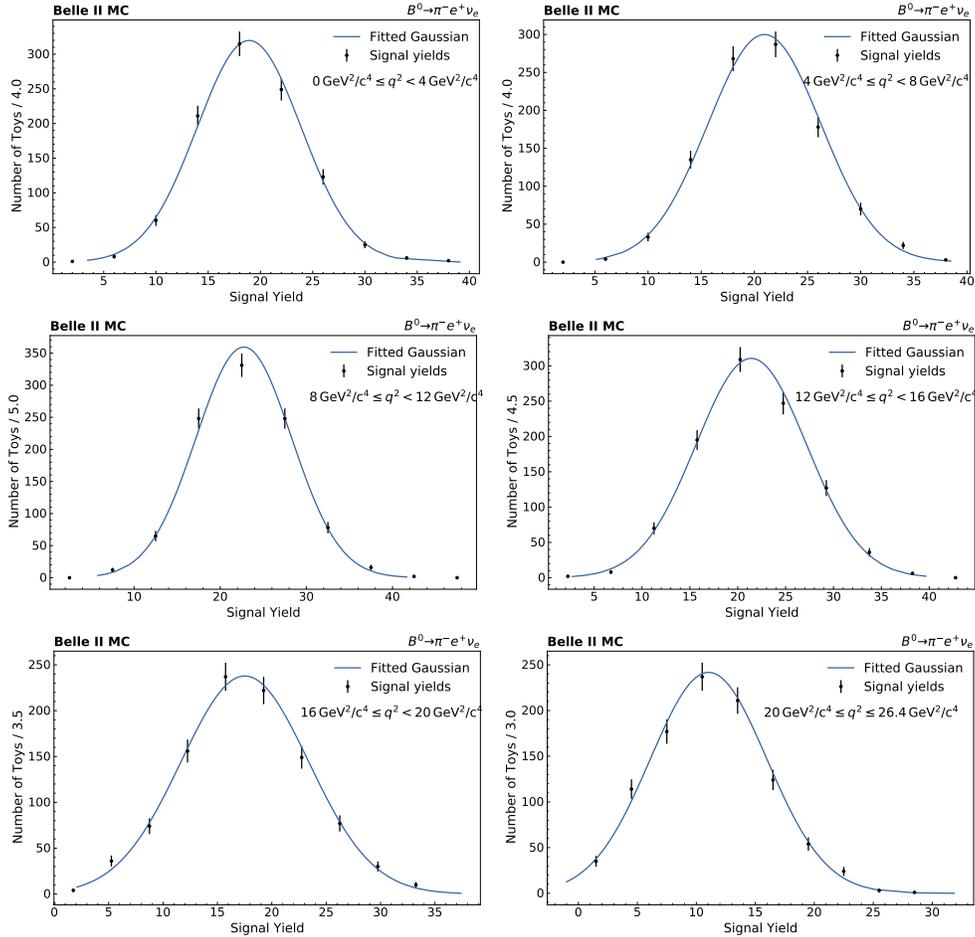

Figure 6.10: Gaussian pdfs fitted to the distribution of signal yields from 1000 MC toys, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins.
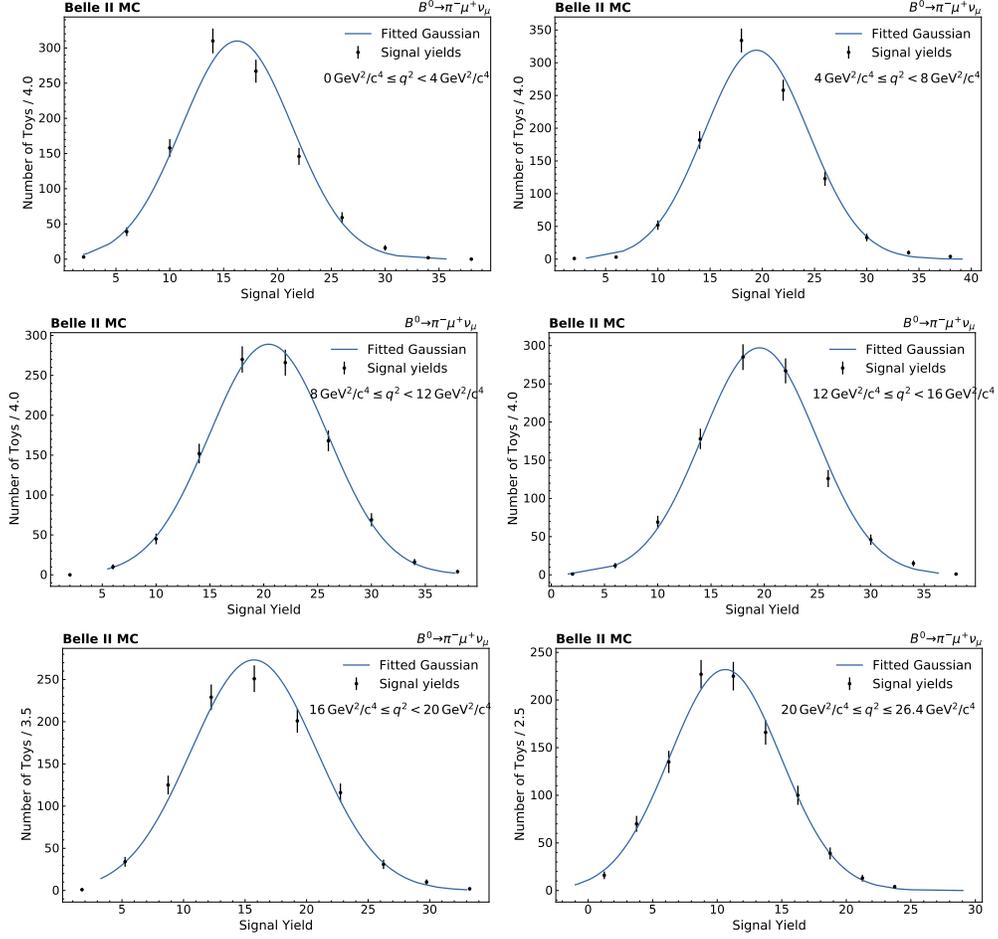
Figure 6.11: Gaussian pdfs fitted to the distribution of signal yields from 1000 MC toys, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins.

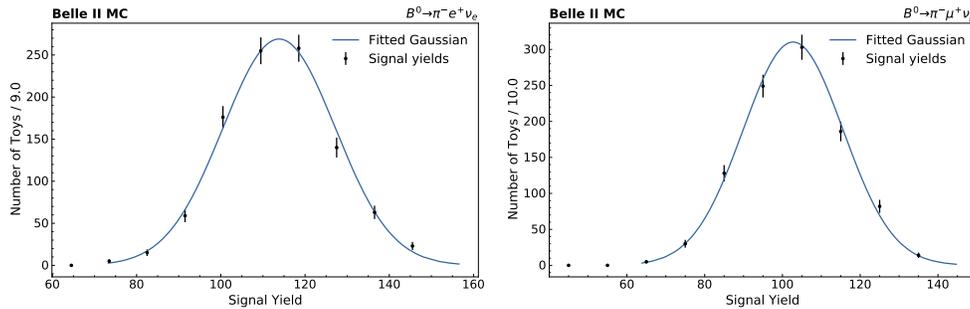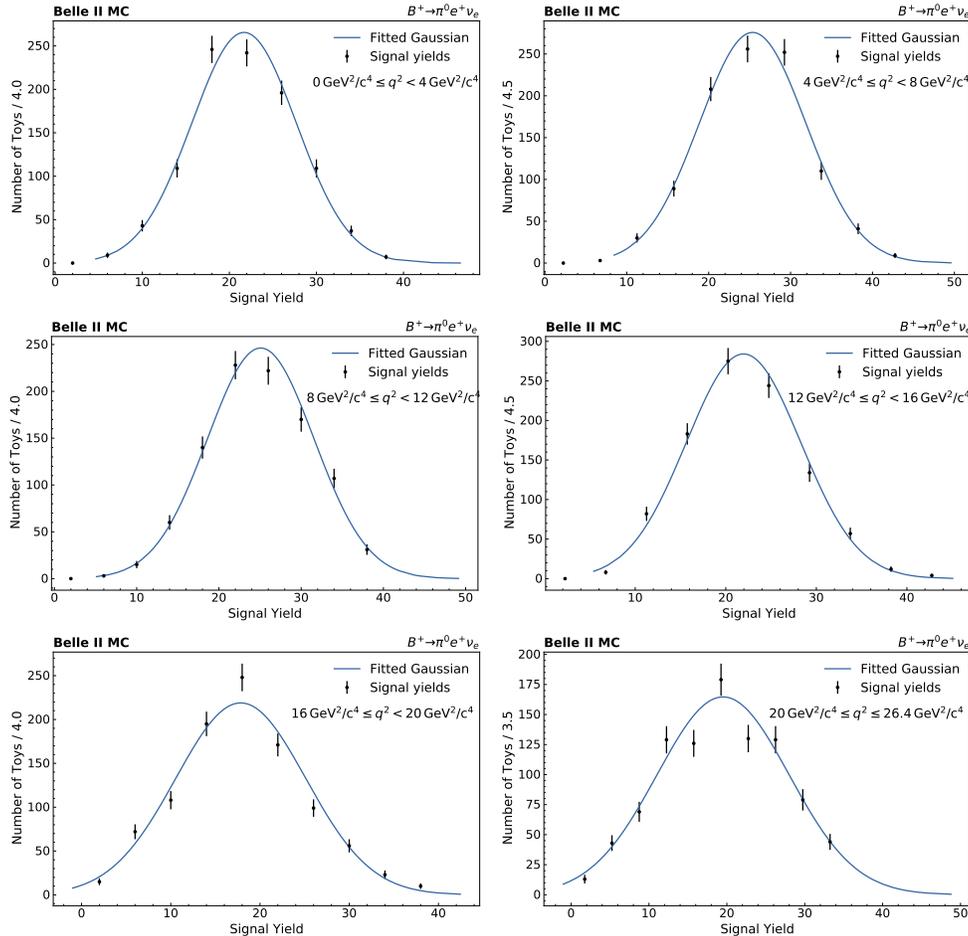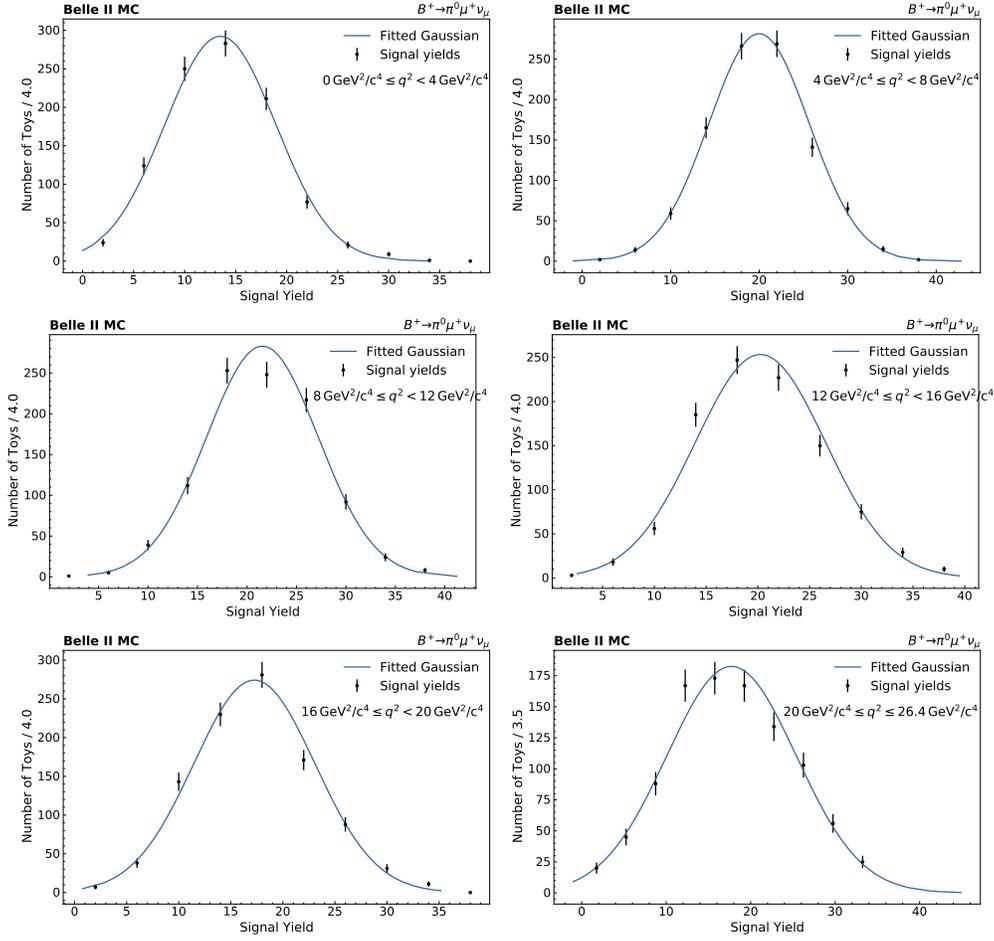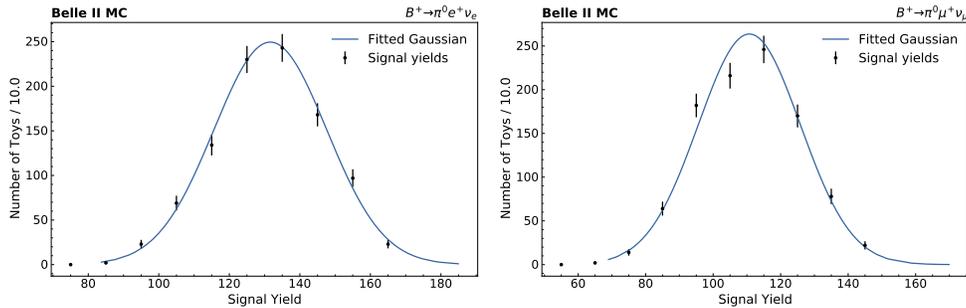Figure 6.12: Gaussian pdfs fitted to the distribution of signal yields from 1000 MC toys, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right) decays.

q^2 bin (GeV^2)	Predicted yield	Fitted μ	Fitted σ
$B^0 \to \pi^- e^+ \nu_e$			
$0 \leqslant q^2 < 4$	19.22	18.92 ± 0.16	4.99 ± 0.11
$4 \leqslant q^2 < 8$	21.26	20.95 ± 0.17	5.32 ± 0.12
$8 \leqslant q^2 < 12$	22.67	22.74 ± 0.18	5.55 ± 0.12
$12 \leqslant q^2 < 16$	21.42	21.41 ± 0.18	5.78 ± 0.13
$16 \leqslant q^2 < 20$	17.69	17.52 ± 0.19	5.87 ± 0.14
$20 \leqslant q^2 \leqslant 26.4$	11.09	11.04 ± 0.17	4.95 ± 0.13
Full q^2 range	113.35	113.84 ± 0.43	13.35 ± 0.32
$B^0 \to \pi^- \mu^+ \nu_\mu$			
$0 \leqslant q^2 < 4$	16.26	16.25 ± 0.16	5.15 ± 0.12
$4 \leqslant q^2 < 8$	19.44	19.44 ± 0.16	5.00 ± 0.11
$8 \leqslant q^2 < 12$	20.74	20.45 ± 0.18	5.52 ± 0.13
$12 \leqslant q^2 < 16$	19.65	19.54 ± 0.17	5.37 ± 0.12
$16 \leqslant q^2 < 20$	16.13	15.72 ± 0.16	5.11 ± 0.12
$20 \leqslant q^2 \leqslant 26.4$	10.62	10.62 ± 0.14	4.30 ± 0.11
Full q^2 range	102.84	102.74 ± 0.41	12.86 ± 0.30
$B^+ \to \pi^0 e^+ \nu_e$			
$0 \leqslant q^2 < 4$	21.50	21.69 ± 0.19	6.01 ± 0.14
$4 \leqslant q^2 < 8$	25.42	25.36 ± 0.21	6.51 ± 0.15
$8 \leqslant q^2 < 12$	25.42	25.08 ± 0.22	6.48 ± 0.17
$12 \leqslant q^2 < 16$	21.94	21.96 ± 0.20	6.32 ± 0.14
$16 \leqslant q^2 < 20$	17.94	17.84 ± 0.24	7.29 ± 0.19
$20 \leqslant q^2 \leqslant 26.4$	19.96	19.51 ± 0.32	8.48 ± 0.30
Full q^2 range	132.18	131.61 ± 0.53	15.99 ± 0.40
$B^+ \to \pi^0 \mu^+ \nu_\mu$			
$0 \leqslant q^2 < 4$	13.58	13.50 ± 0.18	5.46 ± 0.13
$4 \leqslant q^2 < 8$	19.83	19.98 ± 0.18	5.67 ± 0.13
$8 \leqslant q^2 < 12$	21.91	21.56 ± 0.18	5.64 ± 0.13
$12 \leqslant q^2 < 16$	20.12	20.25 ± 0.20	6.30 ± 0.15
$16 \leqslant q^2 < 20$	17.32	17.26 ± 0.19	5.82 ± 0.14
$20 \leqslant q^2 \leqslant 26.4$	18.08	17.73 ± 0.26	7.65 ± 0.23
Full q^2 range	110.84	110.72 ± 0.49	15.13 ± 0.37

Table 6.2: Mean μ and standard deviation σ of the Gaussian pdfs fit to the distributions of signal yields from 1000 toy MC samples. The number of expected signal events is also listed.

It can be clearly observed that the signal yields were well described by the fitted Gaussian pdfs, and were found to be normally distributed around mean yields that were consistent with the MC predictions. Thus, the fitting procedure was able to recover the underlying nature of the variation in the generated sample sizes of the toy distributions, demonstrating its ability to effectively distinguish between signal and background events in a variety of scenarios designed to mimic potential real data.

6.3.1 Bias Testing

Another crucial component of validating a given fitting strategy involves investigating the existence of any potential bias resulting from the fit. Using the same toy data-sets described above, a bias test was performed by determining the pull associated with each toy, defined as the ratio of the difference between the predicted and fitted signal yields and the fitted yield uncertainty σ_{yield}, (pred. – yield / σ_{yield}). In a valid toy study, the pulls should similarly be described by a Gaussian distribution centered at 0, the value for which the predicted and fitted signal yields are identical. If the mean of the distribution of pulls deviates from 0, this indicates the tendency of the fit to more frequently return signal yields below (if the mean pull is positive) or above (if the mean pull is negative) the MC prediction, implying the existence of a fit bias.

For this analysis, the pulls of each toy were determined and, for each individual toy study, a Gaussian pdf was fitted to the distribution of 1000 pulls. The resultant fitted Gaussians plotted against the measured pulls are illustrated in Figures 6.13, 6.14, 6.16 and 6.17 for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins, respectively, and Figures 6.15 and 6.18 for the equivalent distributions evaluated over the full q^2 kinematic range. The mean pull and standard deviation returned from each fit are provided in Table 6.3.

A slight asymmetry can be observed in the pull distributions across most of the q^2 bins studied for all four reconstructed $B \to \pi \ell \nu_\ell$ modes. In these cases, the pull distributions appear to be somewhat skewed towards positive values, a trend reflected in the values of the fitted mean pulls in Table 6.3, which tend towards values closer to ~ 0.1. Thus, these fits were determined to give rise to a slight positive bias, resulting in signal yields below the MC prediction being returned at a rate higher than expected. This behaviour was not observed, however, for all fits performed to those M_{miss}^2 distributions evaluated over the full q^2 kinematic range, which exhibited no signs of bias.

Fit bias can arise due to a number of reasons, and is not necessarily indicative of an underlying problem in the fitting algorithm. One common example of this is in the case of limited statistics, where fitting to data-sets of sufficiently small sample sizes can introduce a level of bias due to the difficulty in distinguishing signal and background events in cases where the content in each bin of the fitted distribution is limited, possessing large uncertainties. Given that only the fits performed to the M_{miss}^2 distributions restricted to certain q^2 ranges were found to exhibit a slight bias, unlike the fits evaluated over the full q^2 range which were far more statistically powerful, it was hypothesised that limited statistics were respon-

sible for the observed bias, and this was subsequently investigated through a dedicated study.

Figure 6.13: Gaussian pdfs fitted to the distribution of pulls from 1000 MC toys, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins.

The toy data-sets used in the above studies were generated from MC to correspond to an integrated luminosity of 362 fb^{-1}, in order to represent the potential fit performance at the size of the current real data-set. To investigate whether the observed fit bias was due to limited statistics, new toy data-sets were generated in an identical manner, but at sample sizes ranging from two to fifty times the size of the original toys. For each reconstructed $B \to \pi \ell \nu_\ell$ mode, the q^2 bin that corresponded with the largest observed bias was chosen for this test, and 1000 toy data-sets were generated each for effective sample sizes of $2\times$, $5\times$, $10\times$, $20\times$ and $50\times$ the sample sizes of the original toys. The maximum likelihood fits to the M^2_{miss} distributions were subsequently performed for each new toy, the pulls were calculated, and a Gaussian pdf fit to the pull distributions for each case. Table 6.4 summarises the results of these tests, with the fitted mean pull and standard deviation given for each effective sample size.
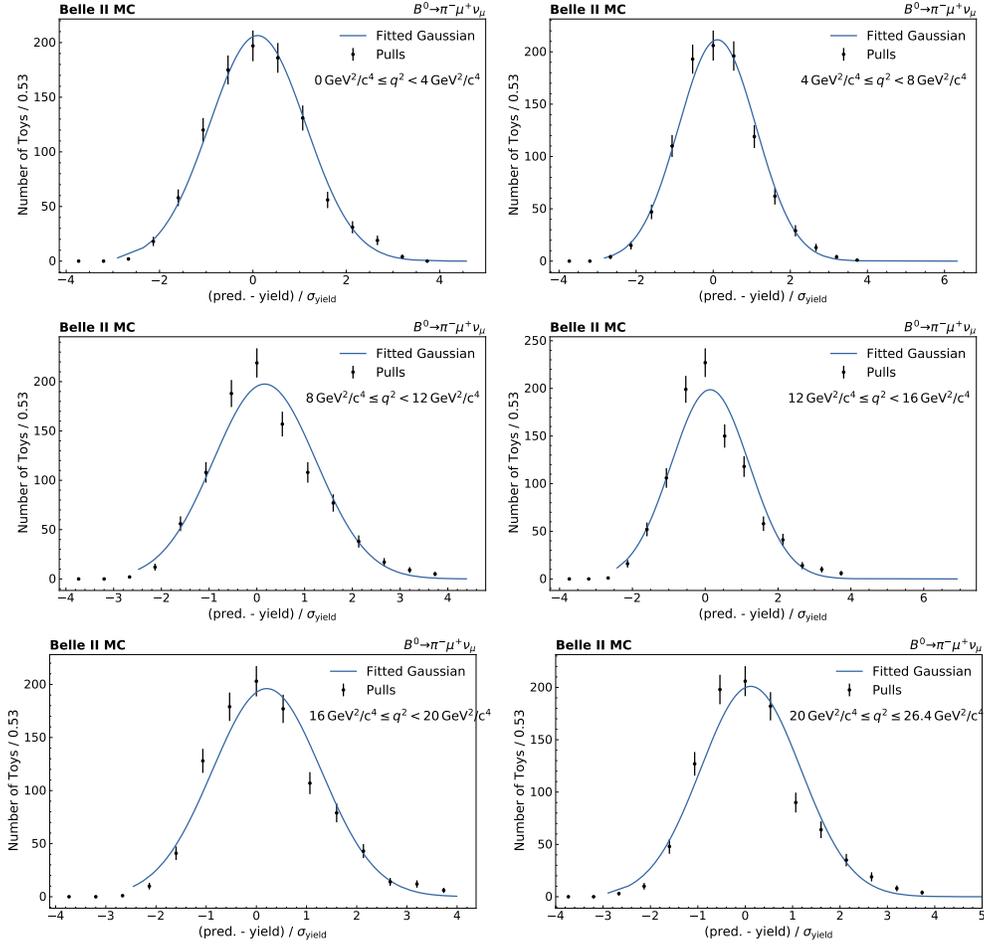
Figure 6.14: Gaussian pdfs fitted to the distribution of pulls from 1000 MC toys, for reconstructed $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins.

Figure 6.15: Gaussian pdfs fitted to the distribution of pulls from 1000 MC toys, for reconstructed $B^0 \rightarrow \pi^- e^+ \nu_e$ (left) and $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (right) decays.

It can be clearly observed that the observed bias disappeared as the effective sample size was increased in all cases, with a large reduction obtained even at a simple doubling of the
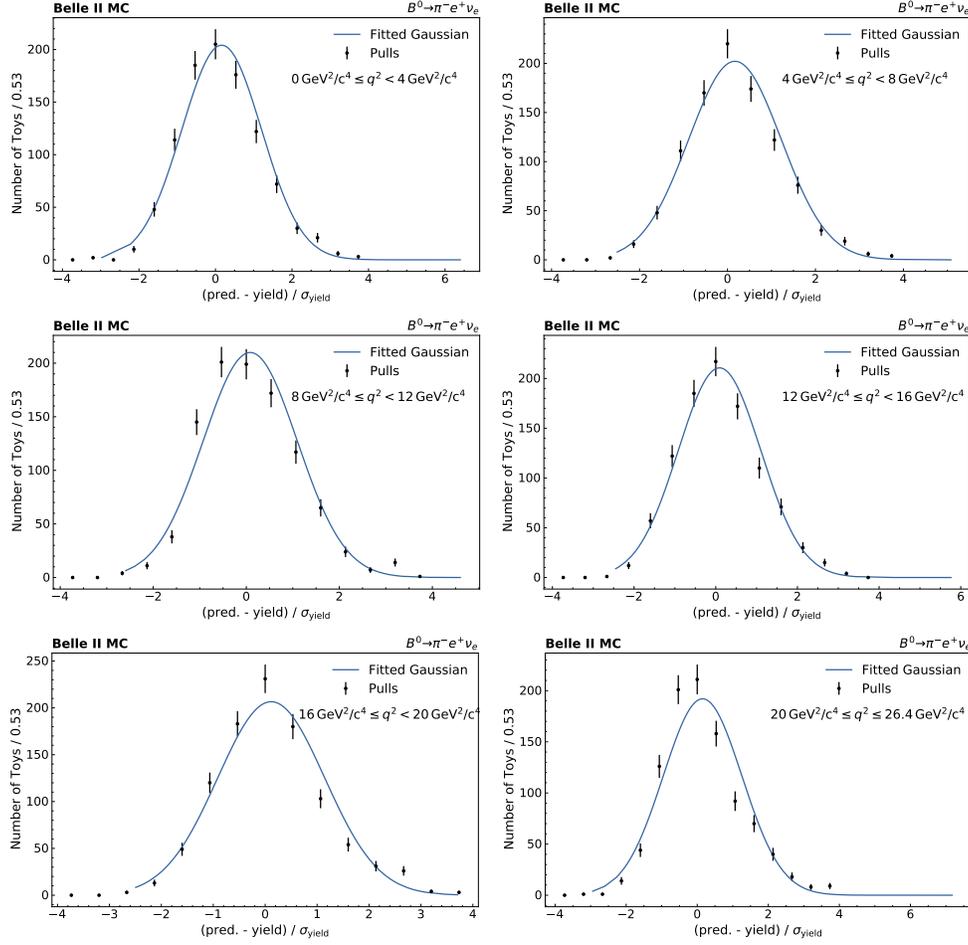
Figure 6.16: Gaussian pdfs fitted to the distribution of pulls from 1000 MC toys, for reconstructed $B^+ \rightarrow \pi^0 e^+ \nu_e$ decays in six q^2 bins.

original toy data-set sample size. It was thus concluded that the fit bias originated solely from the limited statistics available when dividing the M^2_{miss} distributions into several q^2 bins, with no evidence for any underlying issues with the fitting procedure. Given this conclusion, together with the the fact that the magnitude of the observed fit bias was relatively small, it was decided that the current fitting strategy was appropriate for use on real data.

Figure 6.17: Gaussian pdfs fitted to the distribution of pulls from 1000 MC toys, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins.

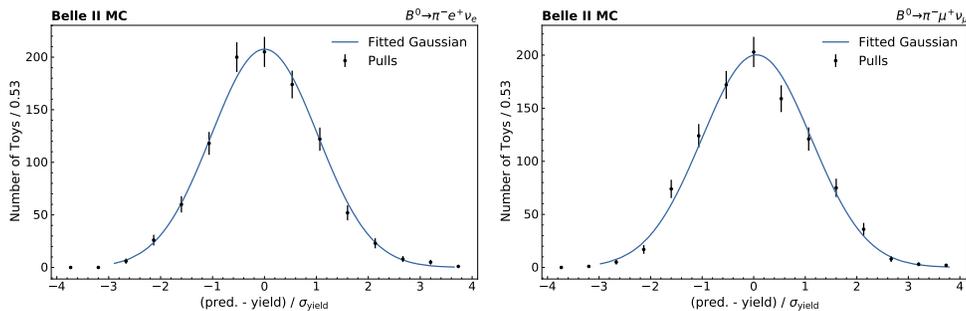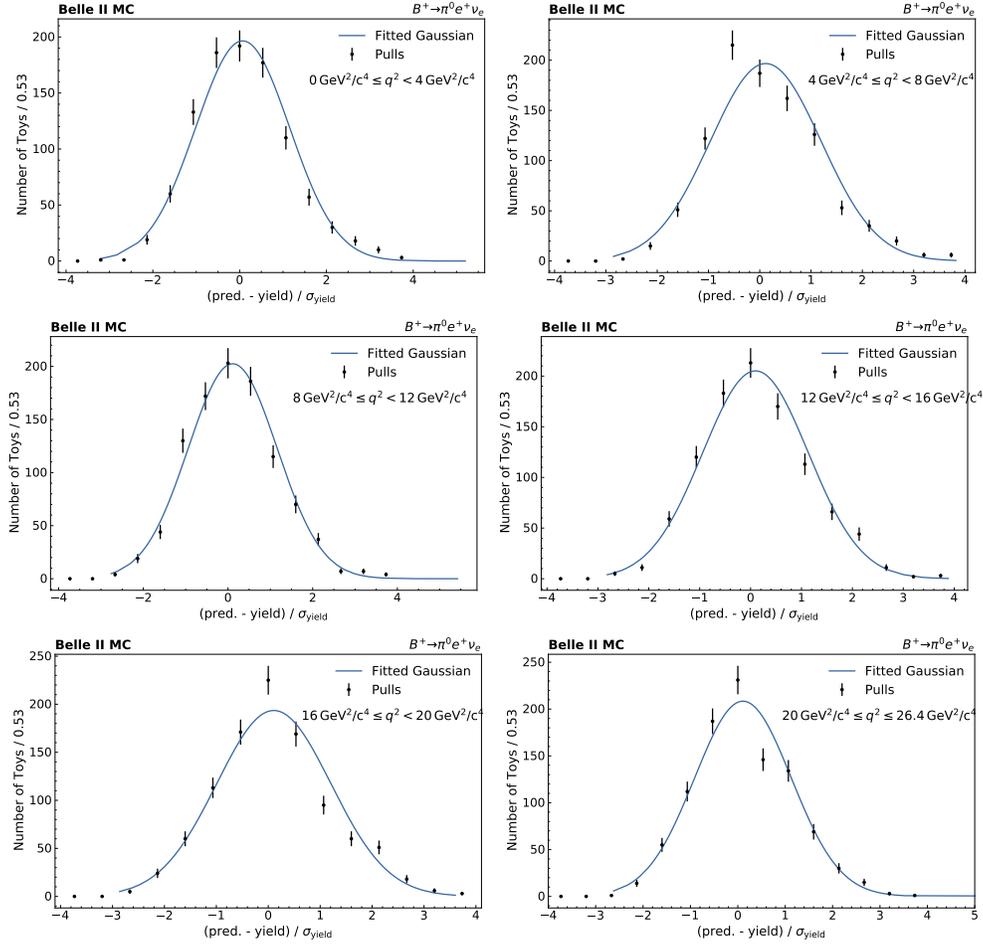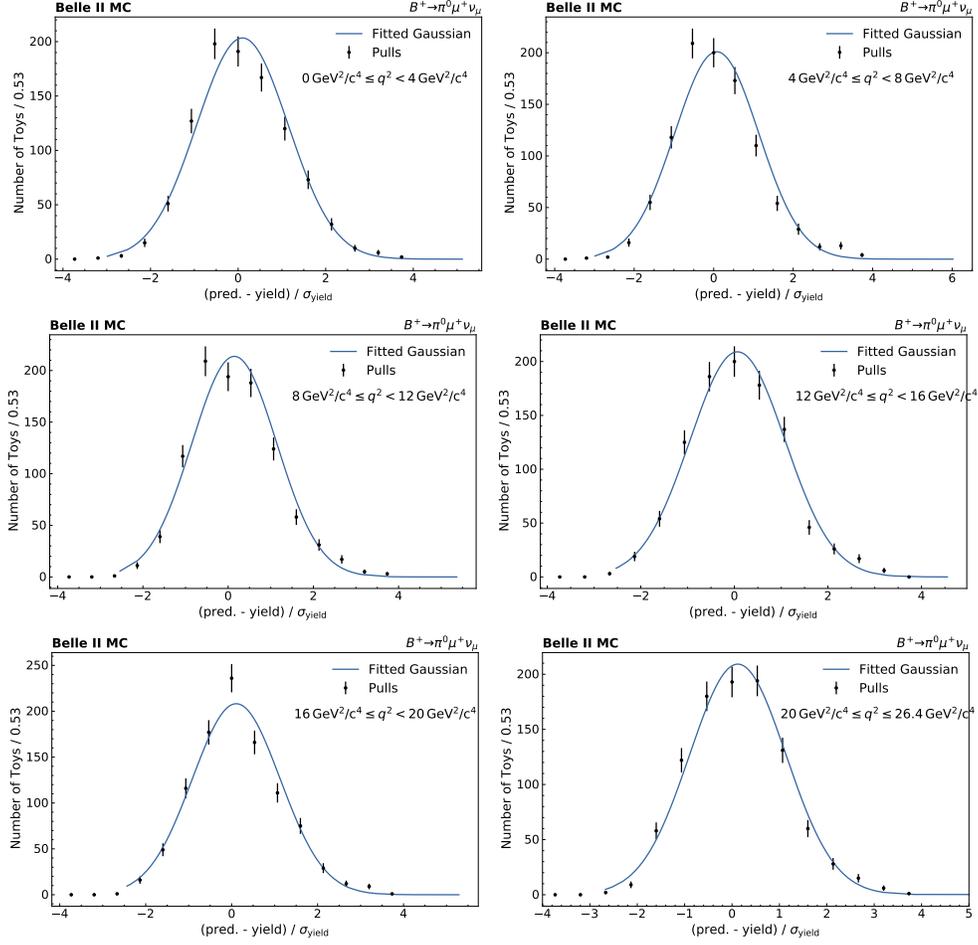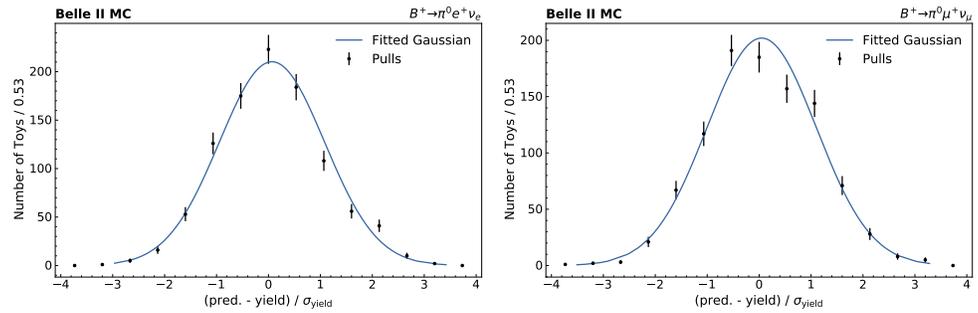Figure 6.18: Gaussian pdfs fitted to the distribution of pulls from 1000 MC toys, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right) decays.

6.3.2   Significance Estimation

Finally, in addition to using the toy MC studies for the purpose of validating the fitting strategy, the generated toy data-sets were also utilised in order to evaluate the expected sig-

q^2 bin (GeV^2)	Fitted μ	Fitted σ
$B^0 \to \pi^- e^+ \nu_e$		
$0 \leqslant q^2 < 4$	0.161 ± 0.033	1.042 ± 0.024
$4 \leqslant q^2 < 8$	0.169 ± 0.033	1.052 ± 0.024
$8 \leqslant q^2 < 12$	0.084 ± 0.032	1.013 ± 0.023
$12 \leqslant q^2 < 16$	0.092 ± 0.032	1.010 ± 0.023
$16 \leqslant q^2 < 20$	0.116 ± 0.033	1.030 ± 0.023
$20 \leqslant q^2 \leqslant 26.4$	0.150 ± 0.035	1.108 ± 0.025
Full q^2 range	0.002 ± 0.032	1.025 ± 0.023
$B^0 \to \pi^- \mu^+ \nu_\mu$		
$0 \leqslant q^2 < 4$	0.100 ± 0.033	1.031 ± 0.023
$4 \leqslant q^2 < 8$	0.108 ± 0.032	1.006 ± 0.023
$8 \leqslant q^2 < 12$	0.163 ± 0.034	1.077 ± 0.024
$12 \leqslant q^2 < 16$	0.136 ± 0.034	1.071 ± 0.024
$16 \leqslant q^2 < 20$	0.207 ± 0.034	1.085 ± 0.025
$20 \leqslant q^2 \leqslant 26.4$	0.117 ± 0.034	1.059 ± 0.024
Full q^2 range	0.056 ± 0.034	1.063 ± 0.024
$B^+ \to \pi^0 e^+ \nu_e$		
$0 \leqslant q^2 < 4$	0.069 ± 0.034	1.083 ± 0.025
$4 \leqslant q^2 < 8$	0.115 ± 0.034	1.082 ± 0.024
$8 \leqslant q^2 < 12$	0.111 ± 0.033	1.051 ± 0.024
$12 \leqslant q^2 < 16$	0.096 ± 0.033	1.037 ± 0.023
$16 \leqslant q^2 < 20$	0.107 ± 0.035	1.100 ± 0.025
$20 \leqslant q^2 \leqslant 26.4$	0.108 ± 0.032	1.021 ± 0.023
Full q^2 range	0.067 ± 0.032	1.012 ± 0.023
$B^+ \to \pi^0 \mu^+ \nu_\mu$		
$0 \leqslant q^2 < 4$	0.099 ± 0.033	1.046 ± 0.024
$4 \leqslant q^2 < 8$	0.077 ± 0.034	1.059 ± 0.024
$8 \leqslant q^2 < 12$	0.145 ± 0.032	0.996 ± 0.022
$12 \leqslant q^2 < 16$	0.067 ± 0.032	1.018 ± 0.023
$16 \leqslant q^2 < 20$	0.107 ± 0.032	1.022 ± 0.023
$20 \leqslant q^2 \leqslant 26.4$	0.121 ± 0.032	1.017 ± 0.023
Full q^2 range	0.048 ± 0.033	1.053 ± 0.024

Table 6.3: Mean and standard deviation of the Gaussian pdfs fit to the distributions of pulls from 1000 toy MC samples.

Effective sample size	Fitted μ	Fitted σ
$B^0 \to \pi^- e^+ \nu_e$, $4 \leqslant q^2 < 8$ GeV^2		
1×	0.169 ± 0.033	1.052 ± 0.024
2×	0.102 ± 0.034	1.079 ± 0.024
5×	0.015 ± 0.033	1.036 ± 0.023
10×	0.032 ± 0.031	0.979 ± 0.022
20×	0.066 ± 0.031	0.987 ± 0.022
50×	0.002 ± 0.033	1.044 ± 0.024
$B^0 \to \pi^- \mu^+ \nu_\mu$, $16 \leqslant q^2 < 20$ GeV^2		
1×	0.207 ± 0.034	1.085 ± 0.025
2×	0.072 ± 0.032	1.013 ± 0.023
5×	0.061 ± 0.033	1.050 ± 0.024
10×	0.027 ± 0.032	0.996 ± 0.022
20×	0.037 ± 0.033	1.047 ± 0.024
50×	0.042 ± 0.031	0.979 ± 0.022
$B^+ \to \pi^0 e^+ \nu_e$, $4 \leqslant q^2 < 8$ GeV^2		
1×	0.115 ± 0.034	1.082 ± 0.024
2×	0.079 ± 0.033	1.046 ± 0.024
5×	0.006 ± 0.032	1.012 ± 0.023
10×	0.082 ± 0.031	0.991 ± 0.022
20×	0.076 ± 0.032	0.999 ± 0.022
50×	0.010 ± 0.031	0.992 ± 0.022
$B^+ \to \pi^0 \mu^+ \nu_\mu$, $8 \leqslant q^2 < 12$ GeV^2		
1×	0.145 ± 0.032	0.996 ± 0.022
2×	0.071 ± 0.033	1.031 ± 0.023
5×	0.044 ± 0.031	0.985 ± 0.022
10×	0.059 ± 0.032	1.004 ± 0.023
20×	0.012 ± 0.032	1.020 ± 0.023
50×	0.062 ± 0.032	1.000 ± 0.022

Table 6.4: Mean and standard deviation of the Gaussian pdfs fit to the distributions of pulls from 1000 toy MC samples, for various effective sample sizes.

nal statistical significance of real data given the current integrated luminosity. The previous chapter introduced an estimator for the signal significance, the quantity $S/\sqrt{S+B}$, which was used for the optimisation of the event selections for the analysis. In reality, several significance estimators exist for a variety of applications. One such estimator can be derived from what is known as the likelihood ratio λ:

$$\lambda = \frac{\mathcal{L}_B}{\mathcal{L}_{S+B}},$$

where $\mathcal{L}_B$ and $\mathcal{L}_{S+B}$ are the maximised likelihoods corresponding to fits performed under the background-only and signal+background hypotheses, respectively. The fits performed thus far refer to the latter, where the total pdfs used to fit the generated toy data-sets were formed from the combination of the signal and background MC templates. To determine the likelihoods under the background-only hypothesis, additional maximum likelihood fits were performed to the same toys using only the background MC templates as the fit pdfs. The estimator of significance Σ was then determined for each toy from the likelihood ratio as per the following relation [91]:

$$\Sigma = \sqrt{-2\ln\lambda}.$$

Figures 6.19, 6.20, 6.22 and 6.23 illustrate the distributions of the significance estimator Σ obtained from the toy studies performed in six bins of q^2 for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, respectively, with the equivalent distributions evaluated over the full q^2 kinematic range given in Figures 6.21 and 6.24. In each distribution, the median significance across all toys is represented by a vertical line, with these values summarised in Table 6.5.

At the current data integrated luminosity of 362 fb^{-1}, a signal statistical significance of $> 10\sigma$ can be expected for each reconstructed $B \to \pi\ell\nu_\ell$ mode, a value substantially greater than the 5σ threshold typically used for claiming discovery. This result is unsurprising given that the $B \to \pi\ell\nu_\ell$ decays are well established, with branching fractions that have been measured multiple times by prior experiments. Furthermore, a signal significance of at least $\sim 3\sigma$ can be expected for each individual q^2 bin when performing the analysis on real data.

q^2 bin (GeV^2)	Median significance			
	$B^0 \to \pi^- e^+ \nu_e$	$B^0 \to \pi^- \mu^+ \nu_\mu$	$B^+ \to \pi^0 e^+ \nu_e$	$B^+ \to \pi^0 \mu^+ \nu_\mu$
$0 \leqslant q^2 < 4$	7.1σ	4.3σ	6.3σ	3.3σ
$4 \leqslant q^2 < 8$	7.3σ	6.8σ	6.3σ	6.2σ
$8 \leqslant q^2 < 12$	6.9σ	6.3σ	5.9σ	5.7σ
$12 \leqslant q^2 < 16$	5.8σ	6.1σ	5.0σ	4.4σ
$16 \leqslant q^2 < 20$	3.8σ	4.6σ	3.3σ	4.1σ
$20 \leqslant q^2 \leqslant 26.4$	3.2σ	3.7σ	2.8σ	2.9σ
Full q^2 range	13.6σ	13.0σ	11.4σ	10.7σ

Table 6.5: Median significance estimated from 1000 toy MC samples.

Figure 6.19: Distribution of the estimator of significance for 1000 MC toys, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins. The median significance is shown in each plot by the vertical blue line.

6.4 Data-MC Comparison

Whilst conducting the $B \to \pi \ell \nu_\ell$ analysis presented throughout this study, a process of internal peer review within the Belle II Collaboration was also proceeding for the purpose of evaluating the analysis for publication as a journal paper at a later date. Under collaboration guidelines, this was implemented through a strategy known as 'blind analysis' [28], where several tests, validation measures and cross-checks were required to be performed before permission to study real Belle II data would be given. In this way, the introduction of any potential bias by designing or altering the analysis to better produce a set of desired results could be avoided. In the context of this study, this included performing the selection optimisation using only simulated data and the toy MC studies used for the fit validation and significance estimation detailed in the previous sections. In addition to this, the data-MC agreement in the M_{miss}^2 sideband, $1.0 < M_{\mathrm{miss}}^2 \leqslant 3.0$ GeV^2, where few signal events were expected in real data, was required to be checked before 'unblinding' the full M_{miss}^2 distributions and observing the data in the signal region, $-1.0 \leqslant M_{\mathrm{miss}}^2 \leqslant 1.0$ GeV^2. After an extensive review in which the data-MC agreement was found to be reasonable in the M_{miss}^2 sideband, permission to unblind the full distributions was received, and all plots presented throughout the remainder of this thesis will portray final, unblinded results.

156

Figure 6.20: Distribution of the estimator of significance for 1000 MC toys, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins. The median significance is shown in each plot by the vertical blue line.

Figure 6.21: Distribution of the estimator of significance for 1000 MC toys, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (right) decays. The median significance is shown in each plot by the vertical blue line.
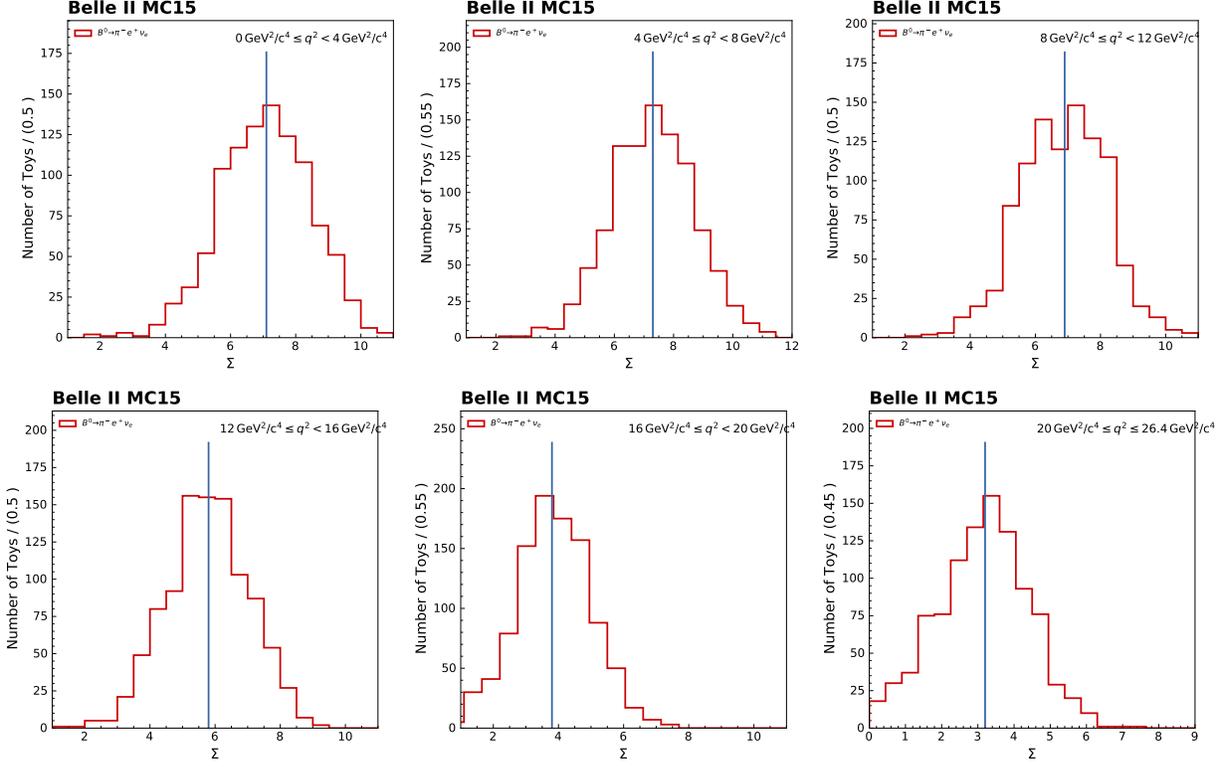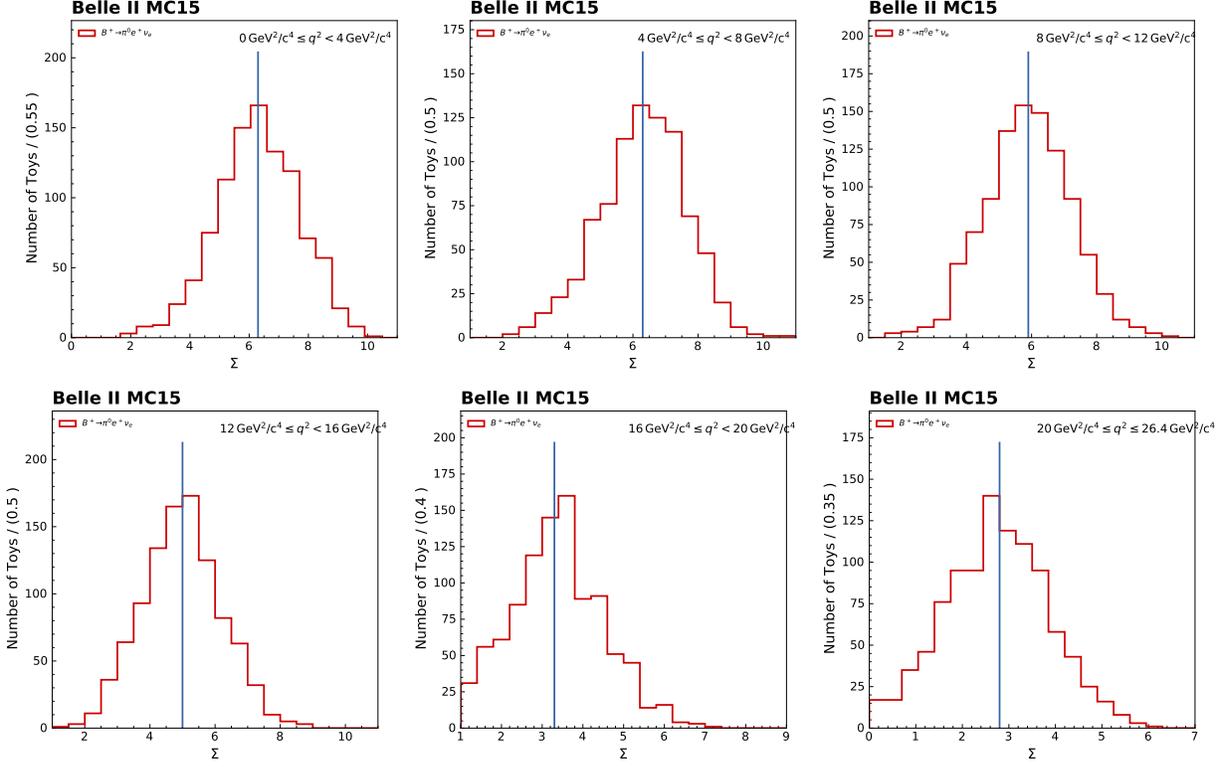
Figure 6.22: Distribution of the estimator of significance for 1000 MC toys, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins. The median significance is shown in each plot by the vertical blue line.

Figures 6.25, 6.26, 6.29 and 6.30 depict the pre-fit M^2_{miss} distributions in six q^2 bins of 362 fb^{-1} of Belle II data against the MC prediction, for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, respectively. In each plot, all relevant data and MC corrections have been applied, with the MC normalised to the given data integrated luminosity. To illustrate the data-MC agreement prior to fitting, the bottom panel of each plot portrays the residuals as defined in Section 4.1.3, calculated for each M^2_{miss} bin. To summarise the overall agreement, the chi-squared (χ^2) between the shapes of the data and the MC prediction was computed from the sum of the square of the residuals in each bin i:

$$\chi^2 = \sum_{i=1}^{\mathrm{ndf}} \left(\frac{N^i_{\mathrm{Data}} - N^i_{\mathrm{MC}}}{\sqrt{\sigma^i_{\mathrm{Data}}{}^2 + \sigma^i_{\mathrm{MC}}{}^2}} \right)^2,$$

where the number of degrees of freedom (ndf) corresponded to the number of bins in the M^2_{miss} distribution. The resultant χ^2 values were subsequently used to determine the p-values for each distribution, demonstrating the probability of the given MC modeling producing the observed results. These p-values were computed from the cumulative distribution function $D_{\mathrm{ndf}}(x)$ for the given value of $x = \chi^2$ [59]:
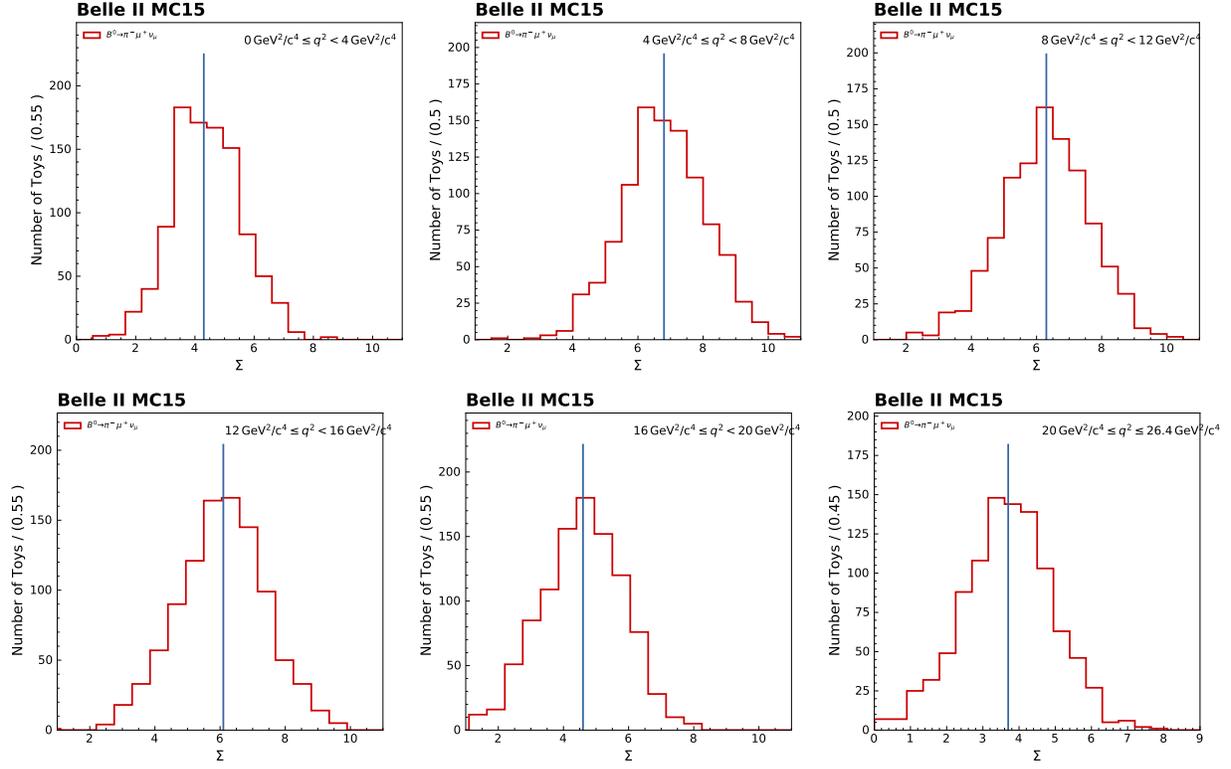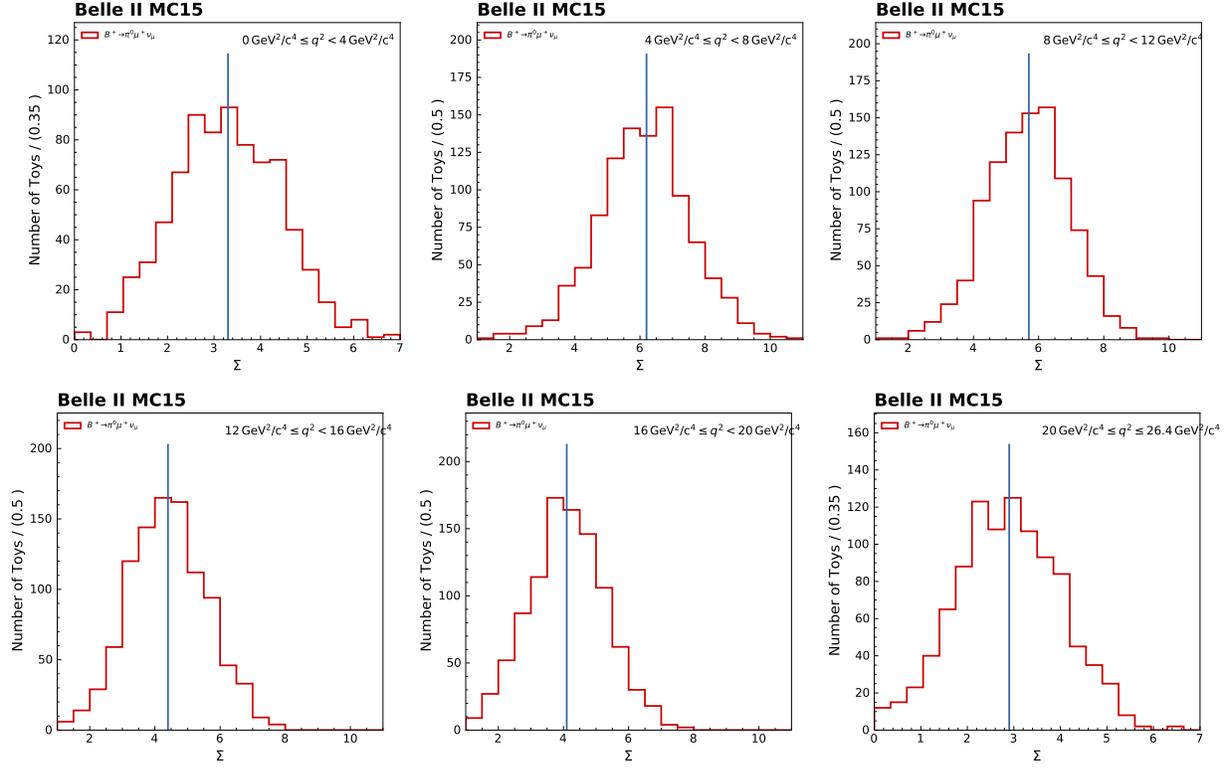
Figure 6.23: Distribution of the estimator of significance for 1000 MC toys, for reconstructed $B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins. The median significance is shown in each plot by the vertical blue line.
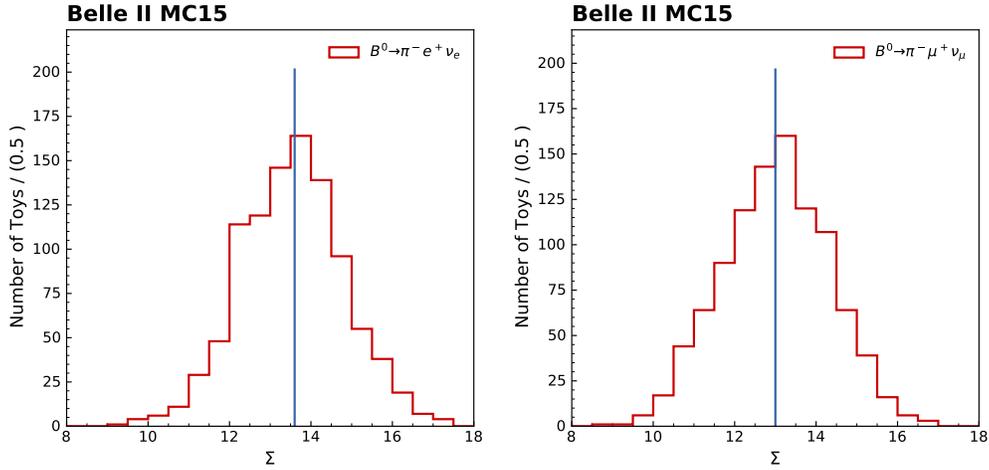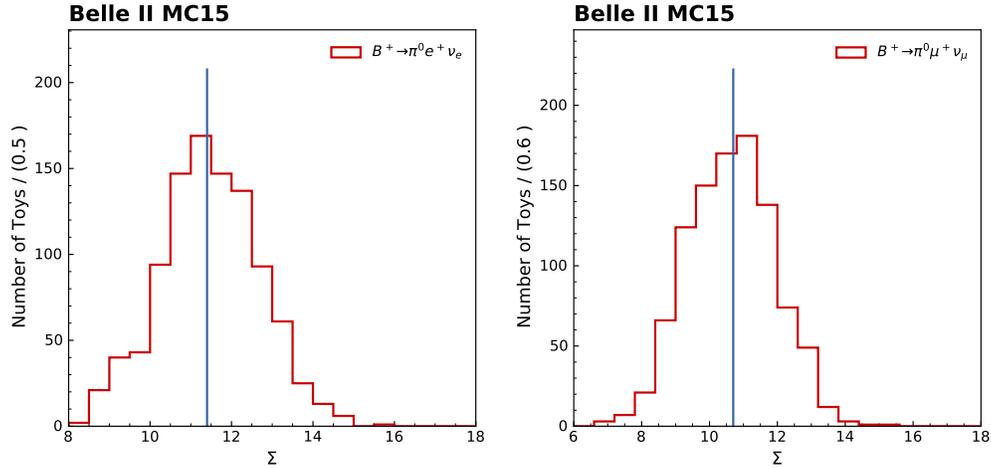

Figure 6.24: Distribution of the estimator of significance for 1000 MC toys, for reconstructed $B^+ \rightarrow \pi^0 e^+ \nu_e$ (left) and $B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$ (right) decays. The median significance is shown in each plot by the vertical blue line.

$$D_{\mathrm{ndf}}(x) = \int_{x}^{+\infty} \frac{x'^{(\mathrm{ndf}/2)-1}e^{-x'/2}}{\Gamma(\mathrm{ndf}/2)2^{\mathrm{ndf}/2}}dx,$$

where $\Gamma(\mathrm{ndf}/2)$ is the gamma function $\Gamma(\mathrm{ndf}/2) = ((\mathrm{ndf}/2)-1)!$. Both the p-values and the χ^2 values divided by the number of degrees of freedom are included in the plotted figures.

Figure 6.25: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins.

The pre-fit data-MC agreement is reasonable overall, with p-values $> 1\%$ observed across the majority of the given distributions. It should be noted that whilst the residuals in some cases seem to imply a deviation of the order of $\sim 5\sigma$ from the MC prediction, these values are somewhat overestimated as the calculation of the residuals relies on a Gaussian-like assumption of the underlying Poisson-like data, an assumption which does not hold as well for limited statistics.

Equivalent M_{miss}^2 distributions were also produced for the combined lepton case in which either an electron or muon on the signal side was accepted, effectively doubling the statistics by combining the given electron-only and muon-only plots shown in Figures 6.25, 6.26, 6.29 and 6.30. These distributions are displayed in Figures 6.27 and 6.31 for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively, where $\ell = e$ or μ. As can be observed, the data-MC agreement prior to fitting in the combination of the two lepton modes reflects

Figure 6.26: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins.

the situation observed in the individual lepton modes.

Finally, the pre-fit M_{miss}^2 distributions evaluated over the full q^2 kinematic range were also considered, as demonstrated in Figures 6.28 and 6.32. Whilst the shapes of the data and MC distributions were largely similar, the MC prediction appeared to overestimate the data in almost all cases, resulting in relatively low observed p-values. This is typical for pre-fit distributions, however, with the underlying signal-background normalisation ultimately determined via the fitting process detailed in the following section. After fitting the distributions, the resultant p-values will be subsequently reevaluated.

Figure 6.27: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, in six q^2 bins.

Figure 6.28: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (top left), $B^0 \to \pi^- \mu^+ \nu_\mu$ (top right) and all $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, over the full q^2 kinematic range.
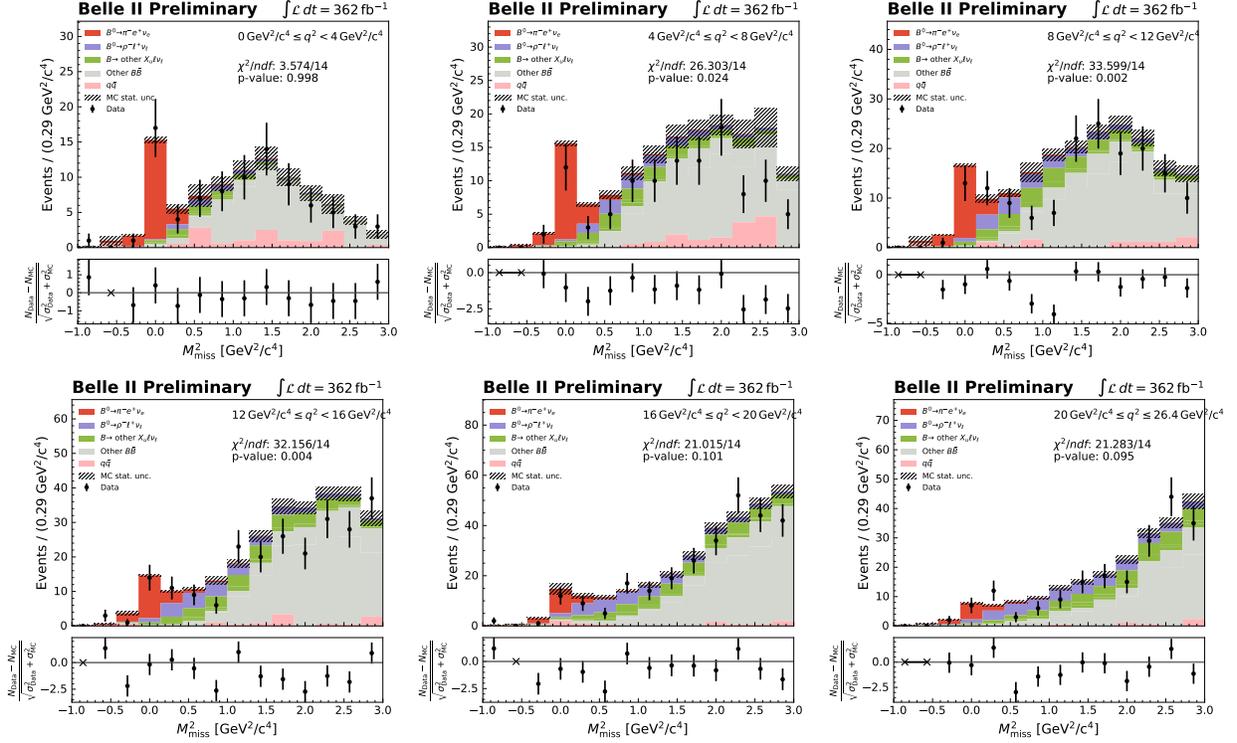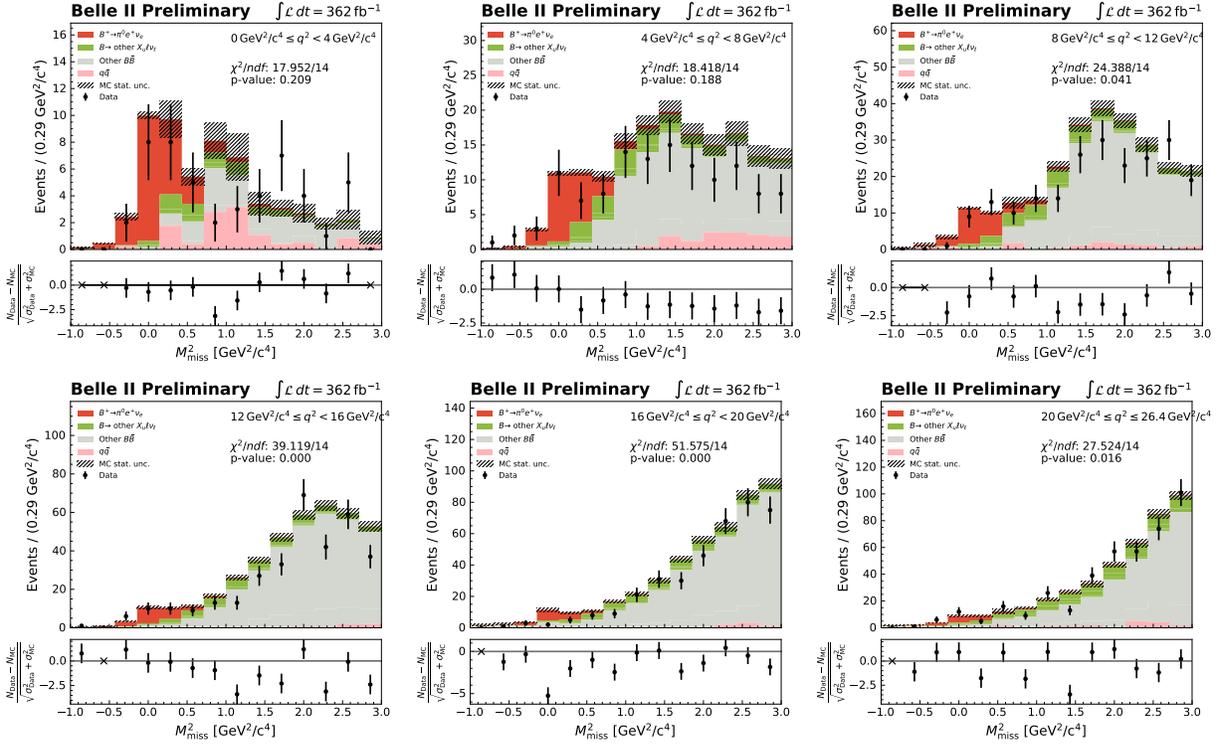
Figure 6.29: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins.
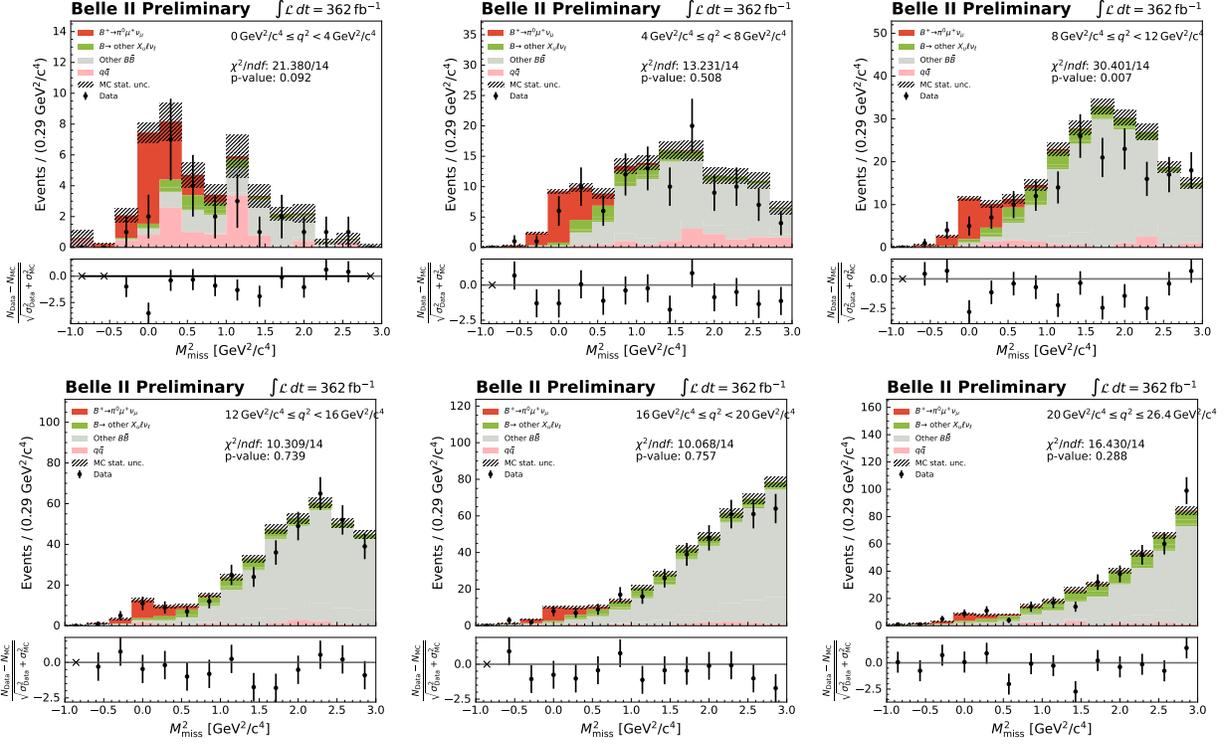
Figure 6.30: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins.
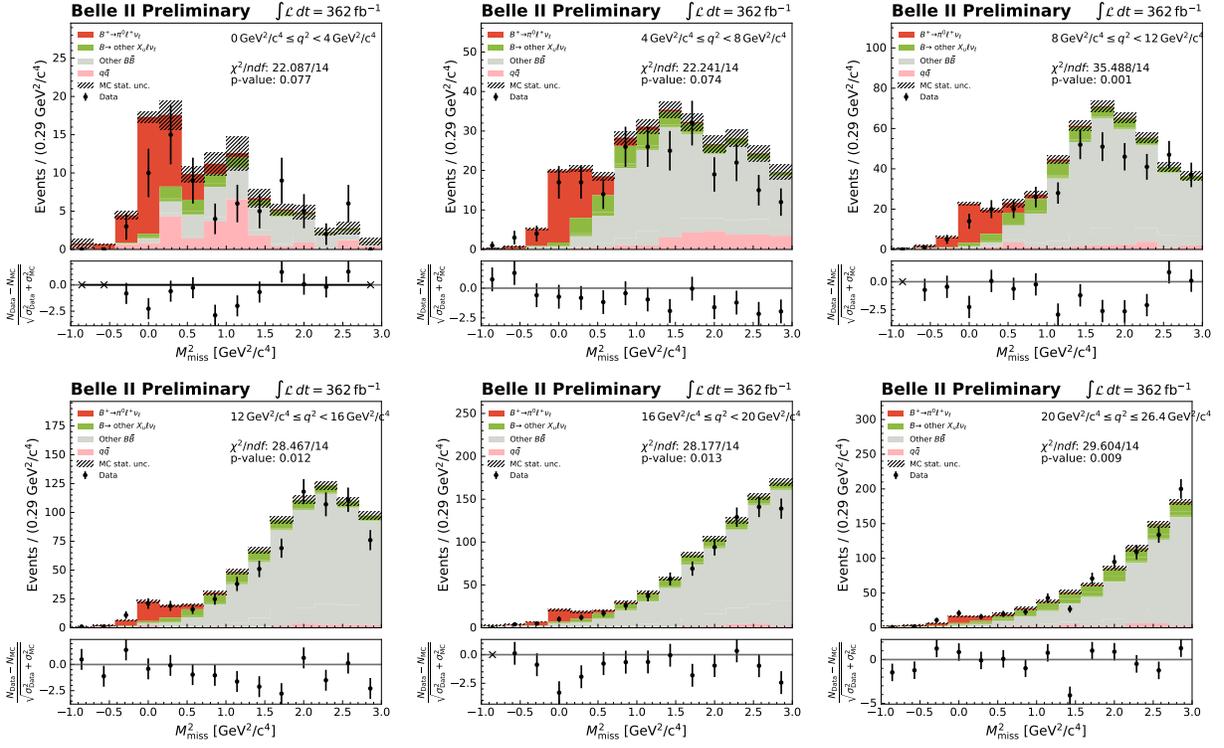
Figure 6.31: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for all reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, in six q^2 bins.
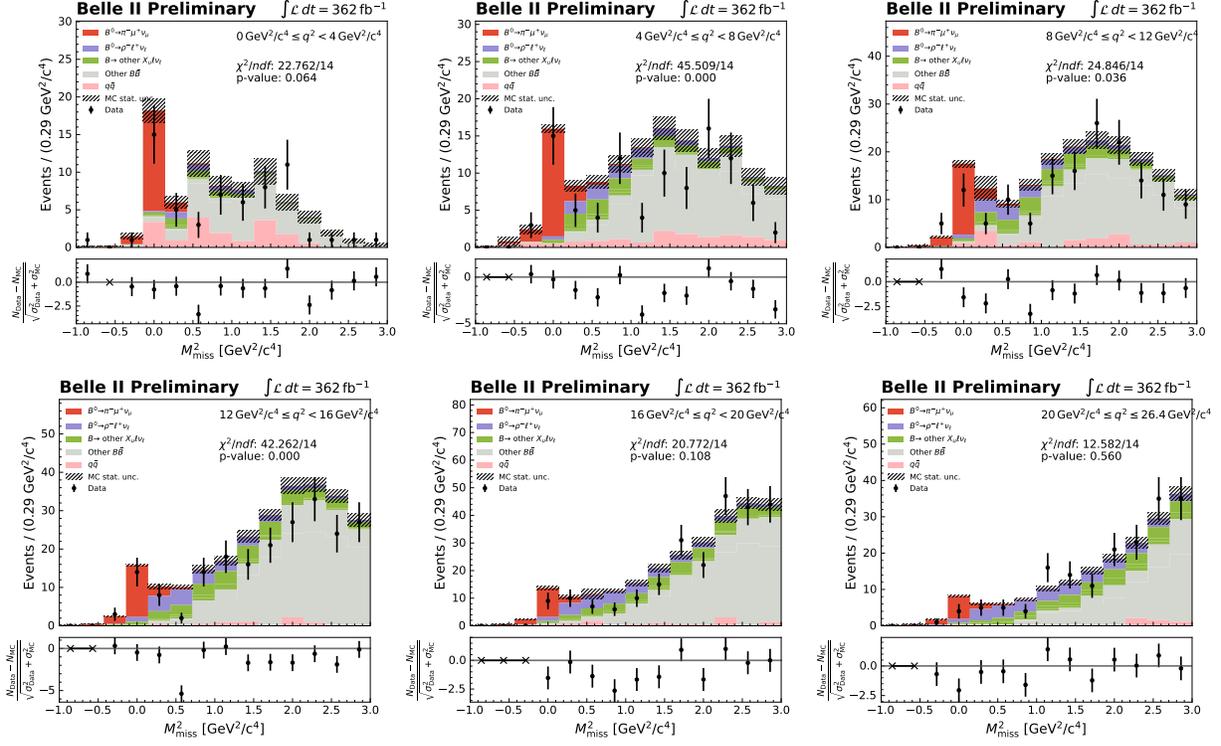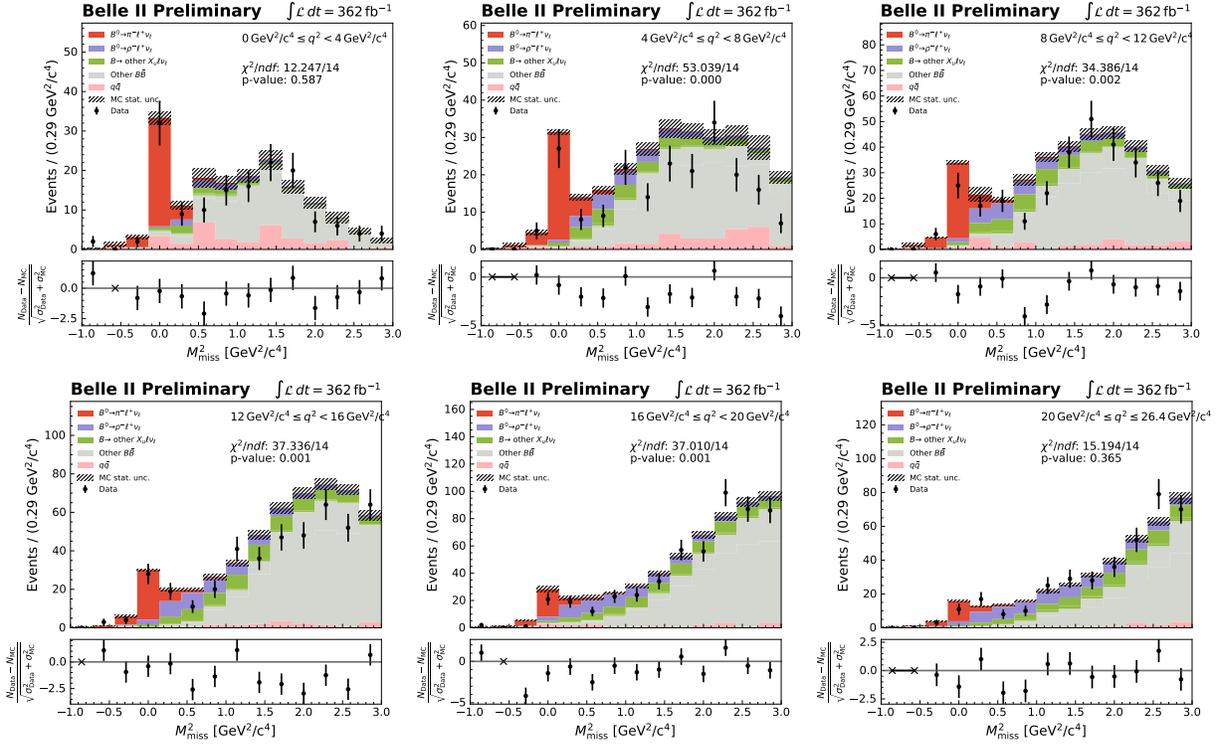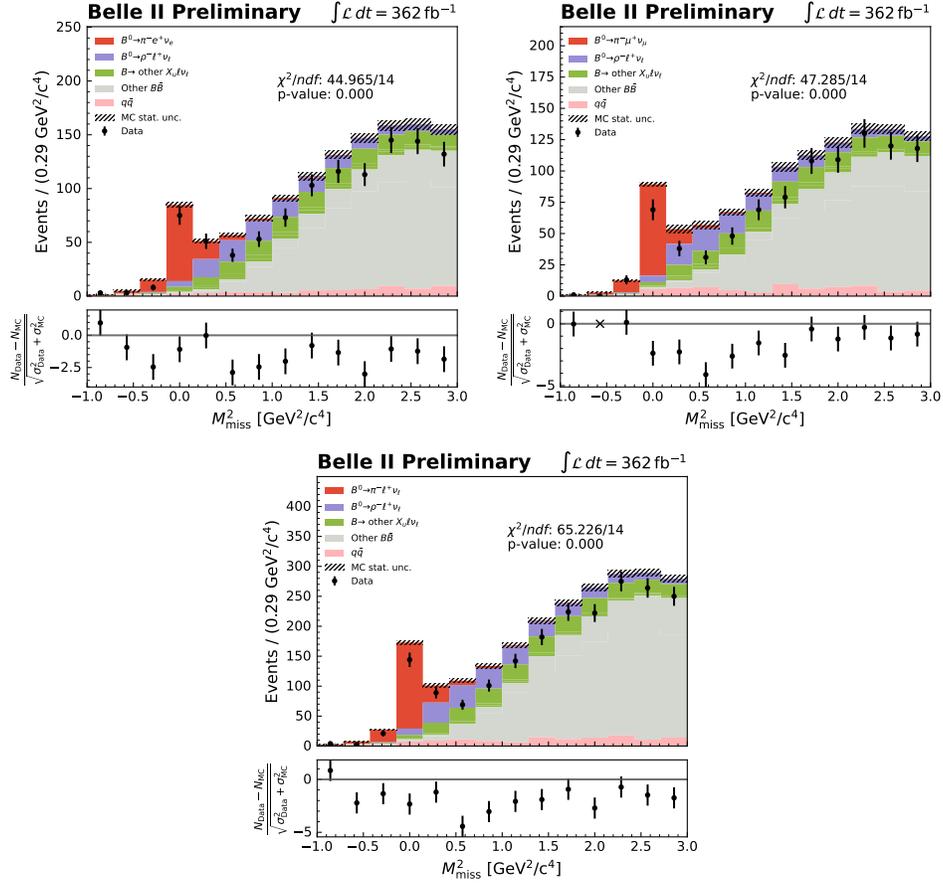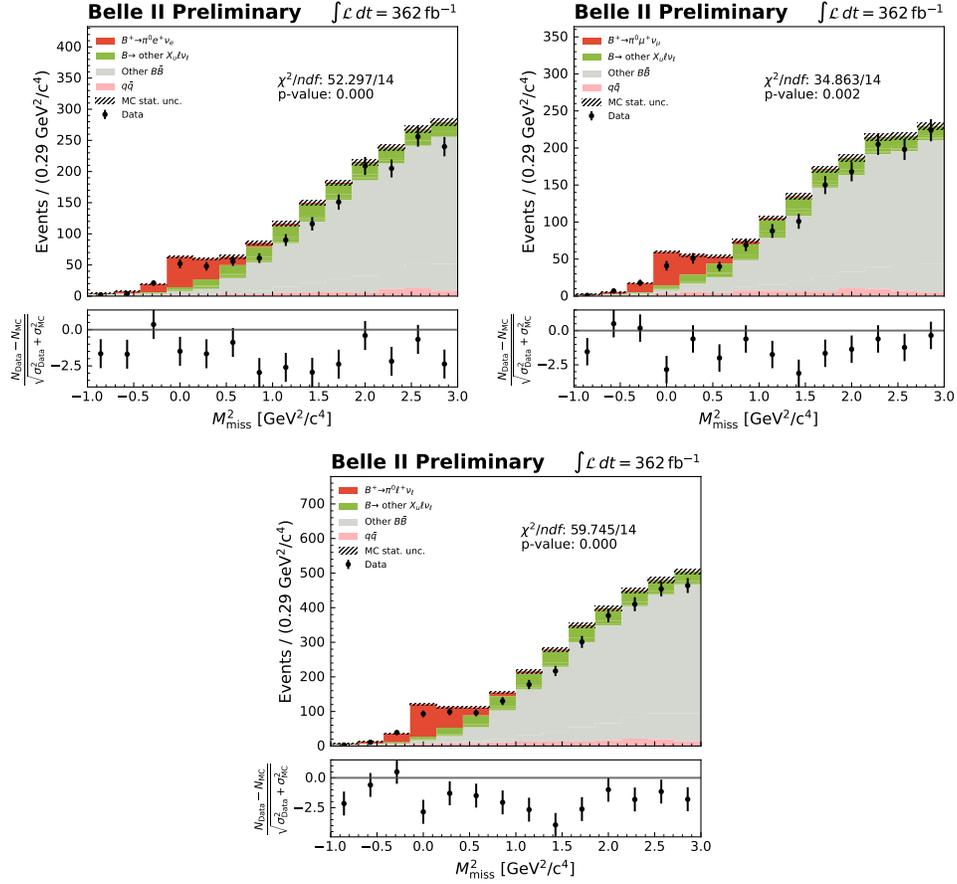
Figure 6.32: Pre-fit M_{miss}^2 distributions of 362 fb^{-1} of Belle II data against the MC expectation, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (top left), $B^+ \to \pi^0 \mu^+ \nu_\mu$ (top right) and all $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, over the full q^2 kinematic range.

6.5 Results and Fitted Yields

Using the fitting strategy described in Section 6.1, maximum likelihood fits were performed to each of the M_{miss}^2 distributions built from real Belle II data presented in the previous section. The post-fit distributions in six q^2 bins illustrating the fitted signal and background components are depicted in Figures 6.33, 6.34, 6.35, 6.37, 6.38 and 6.39, for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^0 \to \pi^- \ell^+ \nu_\ell$, $B^+ \to \pi^0 e^+ \nu_e$, $B^+ \to \pi^0 \mu^+ \nu_\mu$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively, where $\ell = e$ or μ. The equivalent distributions evaluated over the full q^2 kinematic range are provided in Figures 6.36 and 6.40. The post-fit distributions of other key kinematic variables can be found in Appendix B.

In comparing the pre- and post-fit distributions, an improvement in the data-MC agreement after performing the fits can be observed in all cases, demonstrated by the larger p-values obtained when comparing the shapes of the data and post-fit MC templates. This effect is particularly pronounced for those cases in which the pre-fit p-values were very small, such as the M_{miss}^2 distributions over the full q^2 range, now with post-fit values ranging from $17 - 64$ %.

The signal and background yields returned from the fits are summarised in Tables 6.6 and 6.7 for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively. The individual yields corresponding to each q^2 bin are listed, together with the sum of these yields across all q^2 bins and the results returned from the fits evaluated over the full q^2 kinematic range. For each reconstructed $B \to \pi \ell \nu_\ell$ mode, the total yields from the fits over the full q^2 spectrum were consistent with the sum of the individual per-bin yields within uncertainty.

These results illustrate an overall trend with the data yields typically falling below the MC expectation as derived from the latest world average measurements in the majority of cases, particularly when considering the more statistically powerful combined lepton cases. In considering all q^2 bins with the data and MC yields evaluated over the full q^2 range, these overall discrepancies were found to be of the order of 2.6σ for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and 2.4σ for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays.

Figure 6.33: M_{miss}^2 distributions of 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins.

Figure 6.34: M_{miss}^2 distributions of 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins.

Figure 6.35: M_{miss}^2 distributions of $362\,\mathrm{fb}^{-1}$ of Belle II data with MC fit projections overlaid, for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, in six q^2 bins.

Figure 6.36: M^2_{miss} distributions of 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (top left), $B^0 \to \pi^- \mu^+ \nu_\mu$ (top right) and all $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, over the full q^2 kinematic range.

q^2 bin (GeV^2)	Pred. sig. yield	Fitted sig. yield	Fitted bkg. yield
$B^0 \to \pi^- e^+ \nu_e$			
$0 \leqslant q^2 < 4$	19.22	20.29 ± 4.95	67.71 ± 8.48
$4 \leqslant q^2 < 8$	21.26	15.44 ± 4.41	93.56 ± 9.88
$8 \leqslant q^2 < 12$	22.67	17.44 ± 5.08	141.56 ± 12.24
$12 \leqslant q^2 < 16$	21.42	22.06 ± 5.78	207.95 ± 14.81
$16 \leqslant q^2 < 20$	17.69	11.35 ± 5.14	265.69 ± 16.76
$20 \leqslant q^2 \leqslant 26.4$	11.09	11.30 ± 4.80	182.71 ± 13.94
Sum	113.35	97.88 ± 12.35	959.18 ± 31.84
Full q^2 range	113.35	97.66 ± 12.24	959.28 ± 31.80
$B^0 \to \pi^- \mu^+ \nu_\mu$			
$0 \leqslant q^2 < 4$	16.26	13.56 ± 4.57	47.45 ± 7.41
$4 \leqslant q^2 < 8$	19.44	18.57 ± 4.78	78.43 ± 9.09
$8 \leqslant q^2 < 12$	20.74	16.39 ± 4.64	133.61 ± 11.78
$12 \leqslant q^2 < 16$	19.65	17.49 ± 4.86	189.51 ± 13.99
$16 \leqslant q^2 < 20$	16.13	6.84 ± 3.75	237.02 ± 15.63
$20 \leqslant q^2 \leqslant 26.4$	10.62	3.26 ± 2.91	170.74 ± 13.26
Sum	102.84	76.11 ± 10.56	856.76 ± 29.87
Full q^2 range	102.84	75.96 ± 10.46	857.03 ± 29.88
$B^0 \to \pi^- \ell^+ \nu_\ell$			
$0 \leqslant q^2 < 4$	35.48	33.57 ± 6.75	115.47 ± 11.29
$4 \leqslant q^2 < 8$	40.70	34.42 ± 6.55	171.58 ± 13.42
$8 \leqslant q^2 < 12$	43.41	33.32 ± 6.80	275.79 ± 16.99
$12 \leqslant q^2 < 16$	41.07	39.43 ± 7.51	397.52 ± 20.36
$16 \leqslant q^2 < 20$	33.82	18.33 ± 6.33	502.81 ± 22.91
$20 \leqslant q^2 \leqslant 26.4$	21.71	13.89 ± 5.47	354.17 ± 19.24
Sum	216.19	172.96 ± 16.16	1817.34 ± 43.66
Full q^2 range	216.19	173.51 ± 16.13	1816.61 ± 43.63

Table 6.6: Yields of reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays obtained from maximum likelihood fits to 362 fb^{-1} of Belle II data.

q^2 bin (GeV^2)	Pred. sig. yield	Fitted sig. yield	Fitted bkg. yield
$B^+ \to \pi^0 e^+ \nu_e$			
$0 \leqslant q^2 < 4$	21.50	15.88 ± 4.99	33.00 ± 6.48
$4 \leqslant q^2 < 8$	25.42	26.94 ± 6.21	97.06 ± 10.42
$8 \leqslant q^2 < 12$	25.42	22.11 ± 6.12	192.02 ± 14.40
$12 \leqslant q^2 < 16$	21.94	25.46 ± 6.35	303.54 ± 17.84
$16 \leqslant q^2 < 20$	17.94	0.00 ± 6.60	379.04 ± 19.47
$20 \leqslant q^2 \leqslant 26.4$	19.96	23.76 ± 8.63	392.24 ± 21.05
Sum	132.18	114.15 ± 16.10	1396.90 ± 38.71
Full q^2 range	132.18	107.34 ± 14.51	1403.66 ± 38.82
$B^+ \to \pi^0 \mu^+ \nu_\mu$			
$0 \leqslant q^2 < 4$	13.58	4.67 ± 4.08	20.33 ± 24.66
$4 \leqslant q^2 < 8$	19.83	14.90 ± 4.94	94.08 ± 10.18
$8 \leqslant q^2 < 12$	21.91	14.75 ± 4.94	159.25 ± 13.00
$12 \leqslant q^2 < 16$	20.12	19.22 ± 5.91	315.77 ± 18.21
$16 \leqslant q^2 < 20$	17.32	12.02 ± 5.12	349.00 ± 19.06
$20 \leqslant q^2 \leqslant 26.4$	18.08	21.38 ± 7.77	335.62 ± 19.36
Sum	110.84	86.94 ± 13.67	1274.05 ± 44.16
Full q^2 range	110.84	83.27 ± 13.11	1277.74 ± 36.96
$B^+ \to \pi^0 \ell^+ \nu_\ell$			
$0 \leqslant q^2 < 4$	35.08	20.54 ± 6.34	53.46 ± 8.55
$4 \leqslant q^2 < 8$	45.25	43.06 ± 7.97	189.94 ± 14.51
$8 \leqslant q^2 < 12$	47.33	36.01 ± 7.90	351.96 ± 19.45
$12 \leqslant q^2 < 16$	42.06	45.87 ± 8.71	618.09 ± 25.46
$16 \leqslant q^2 < 20$	35.26	11.26 ± 6.59	728.76 ± 27.58
$20 \leqslant q^2 \leqslant 26.4$	38.04	47.69 ± 11.46	725.31 ± 28.51
Sum	243.02	204.43 ± 20.41	2667.52 ± 53.70
Full q^2 range	243.02	196.42 ± 19.64	2675.39 ± 53.52

Table 6.7: Yields of reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays obtained from maximum likelihood fits to 362 fb^{-1} of Belle II data.
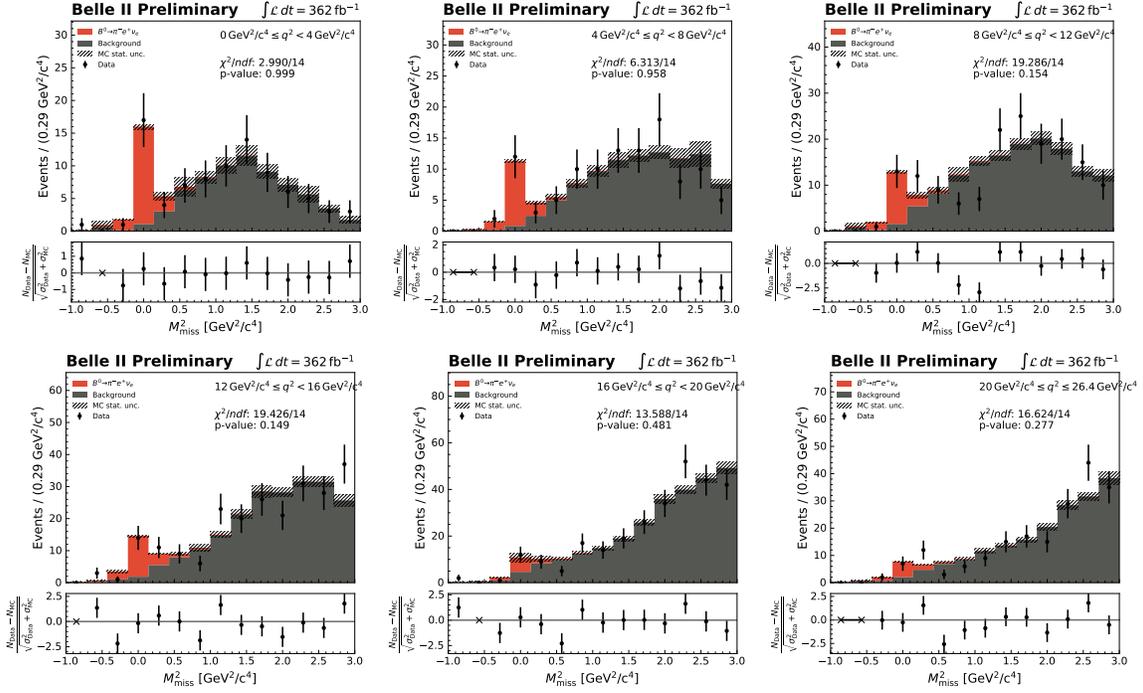
Figure 6.37: M^2_{miss} distributions of $362\,\mathrm{fb}^{-1}$ of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins.
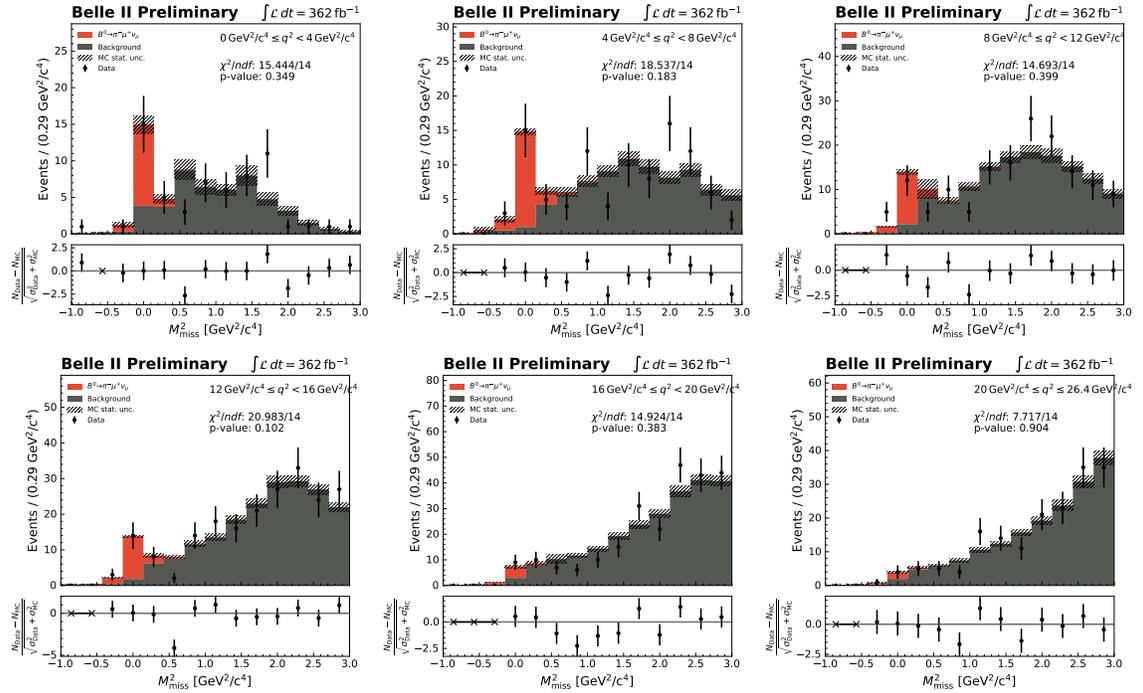

Figure 6.38: M^2_{miss} distributions of $362\,\mathrm{fb}^{-1}$ of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins.
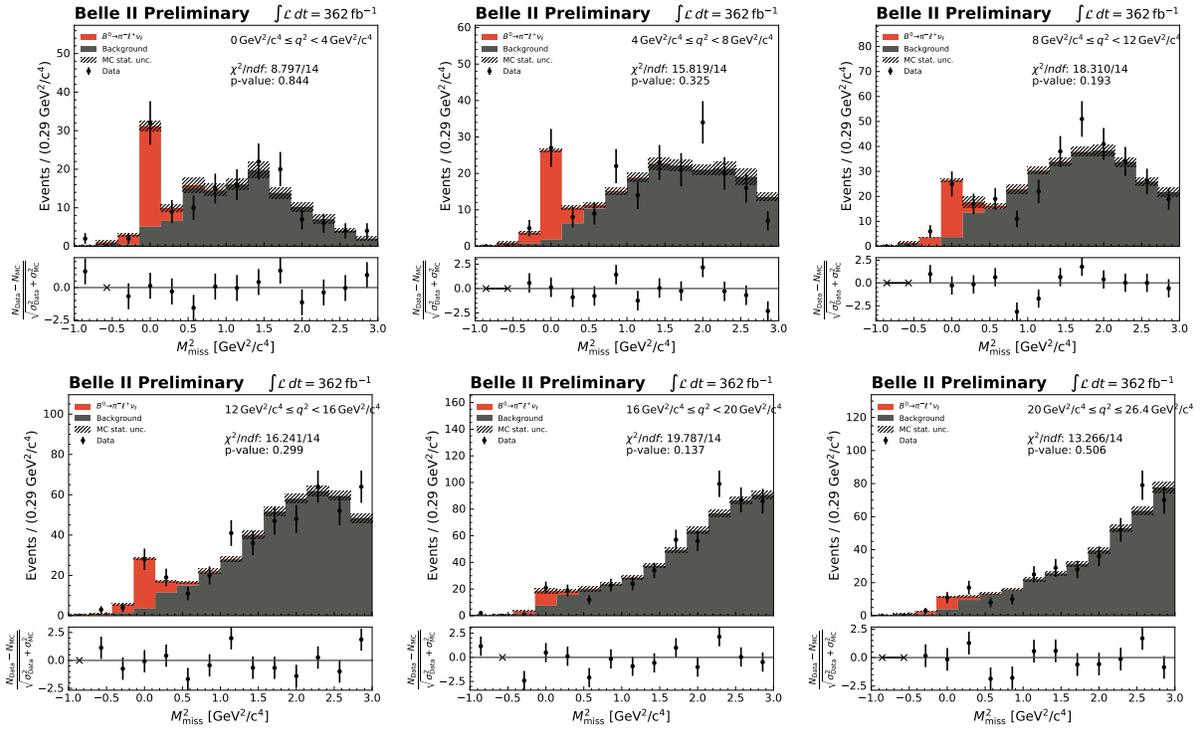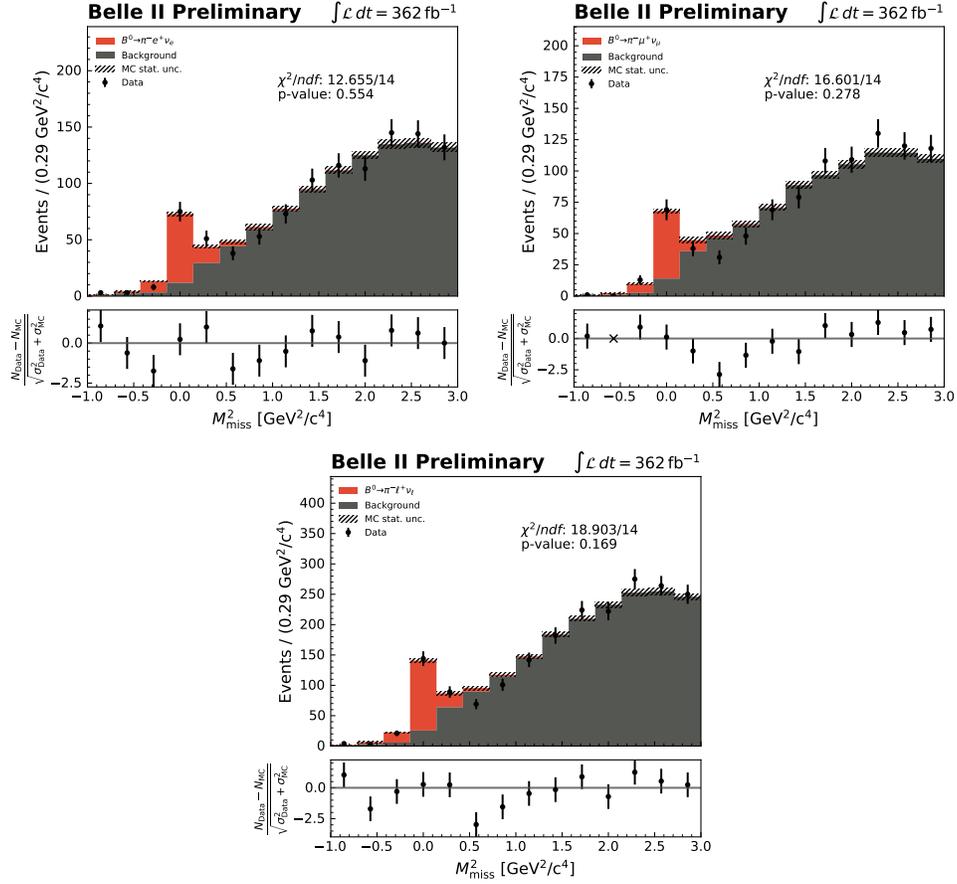
Figure 6.39: M^2_{miss} distributions of 362 fb^{-1} of Belle II data with MC fit projections overlaid, for all reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ, in six q^2 bins.

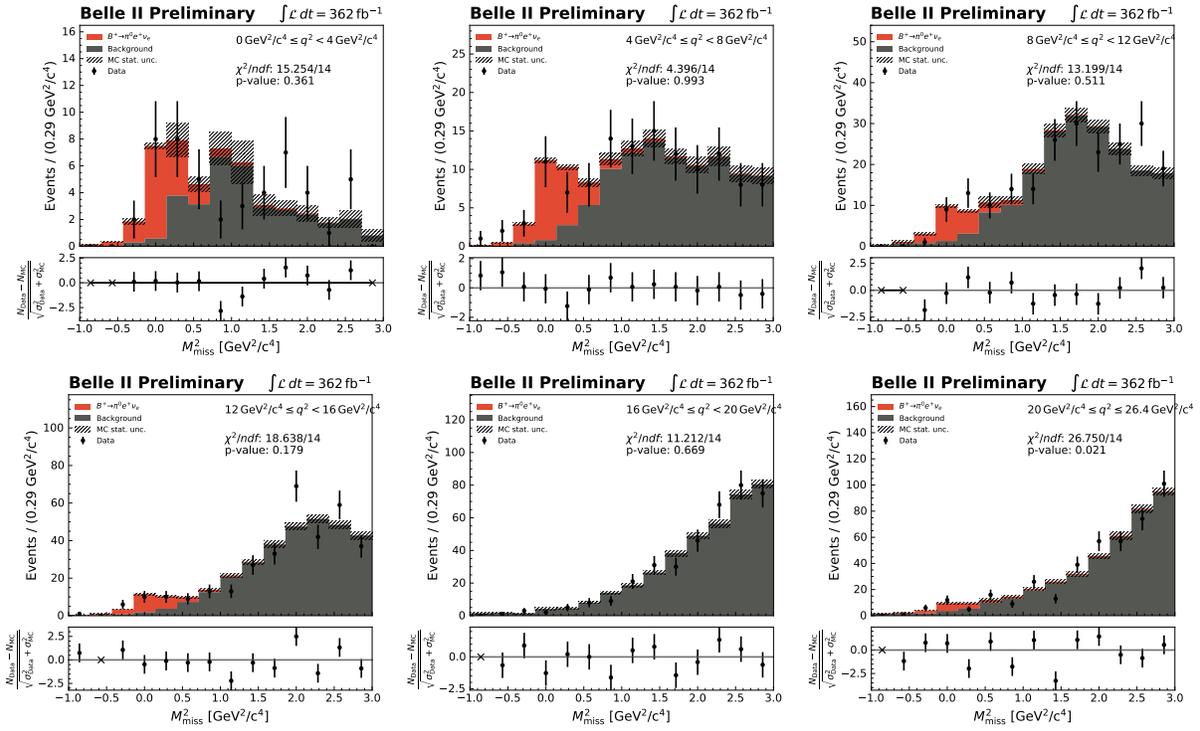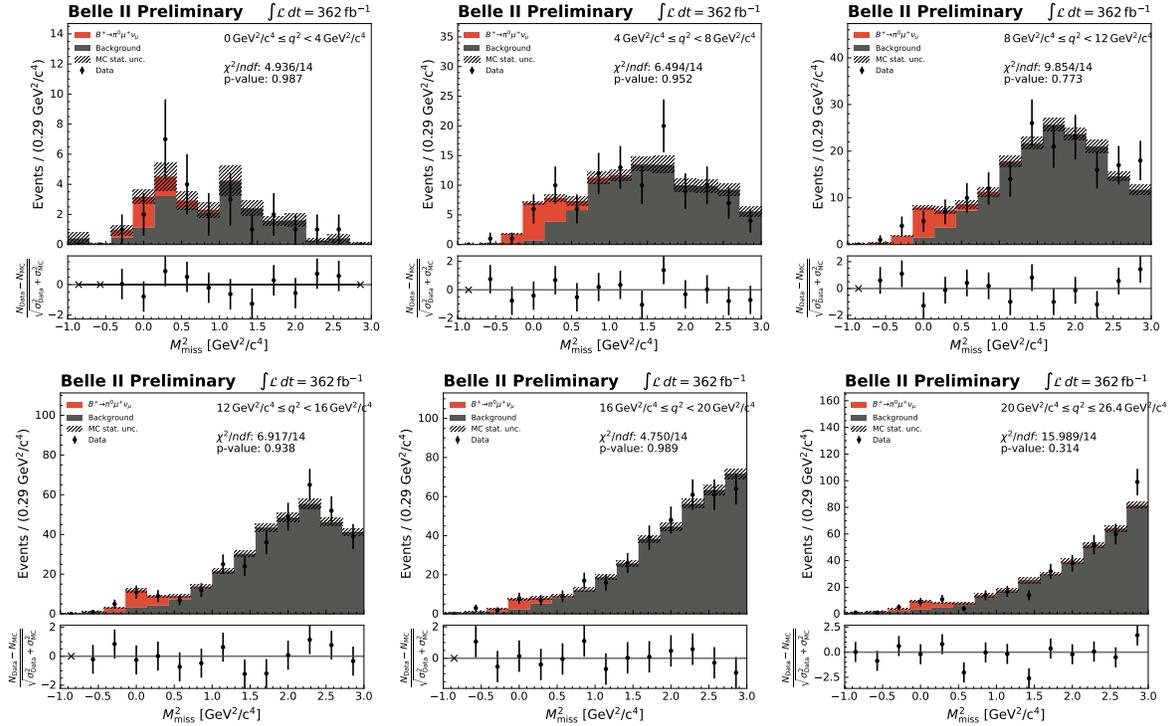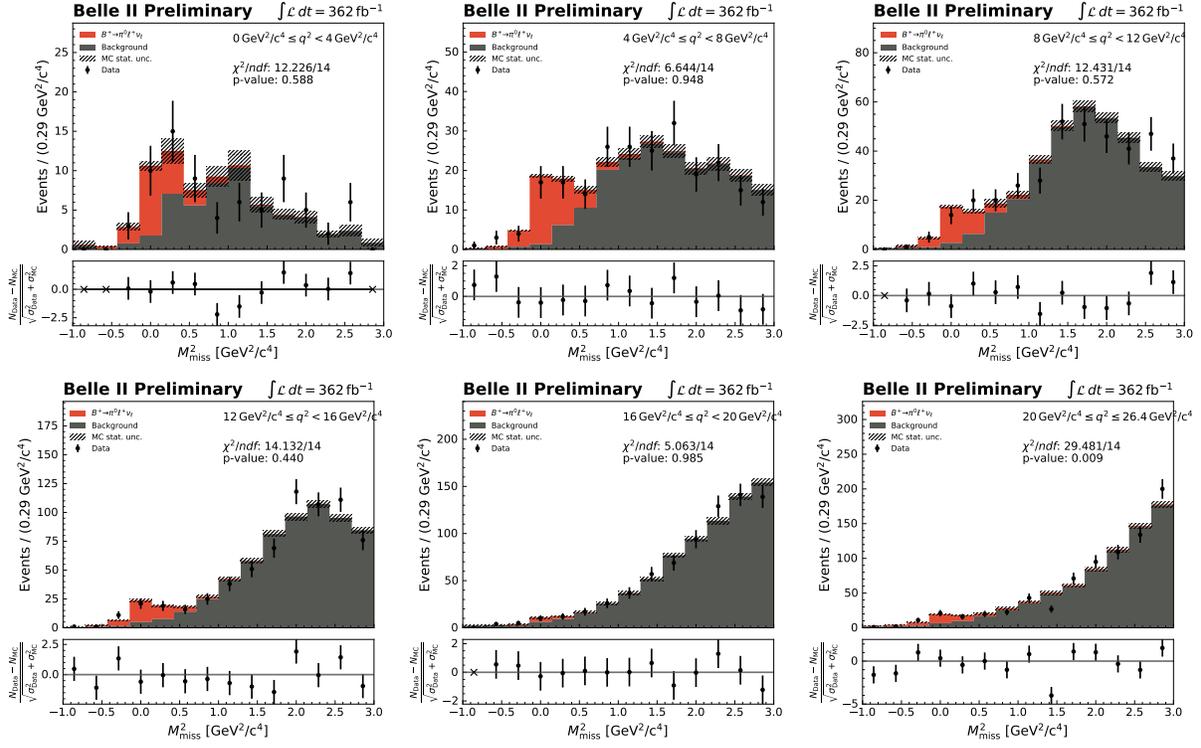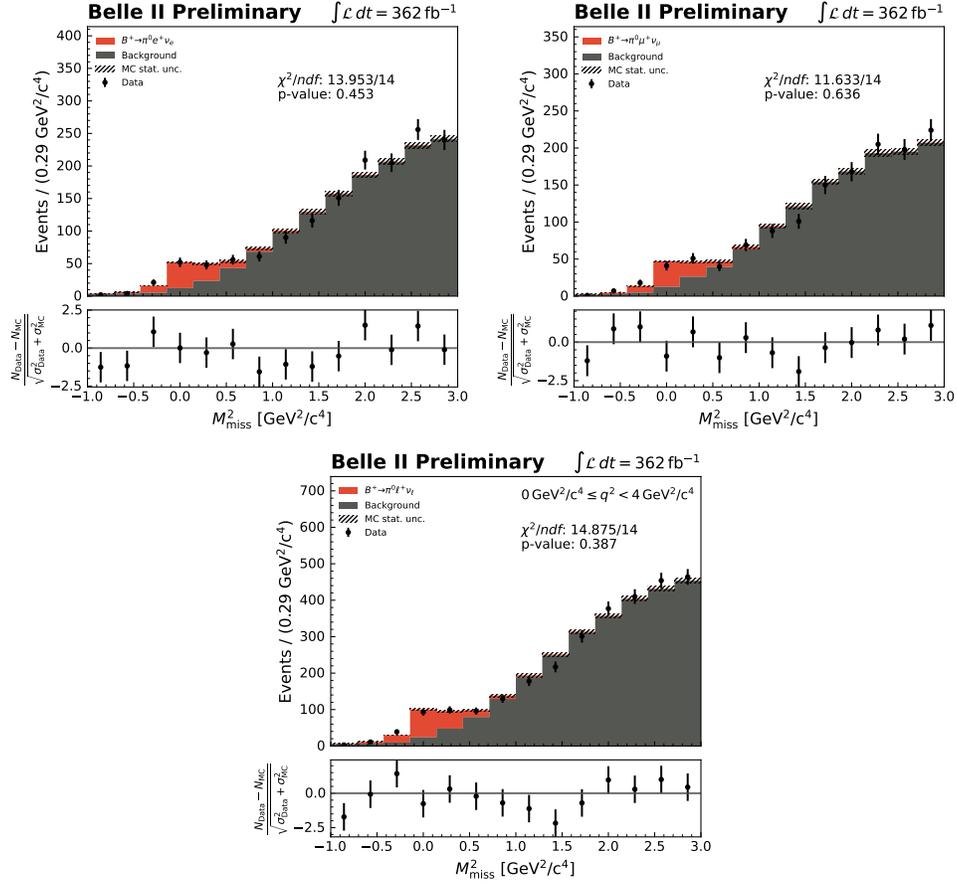Figure 6.40: M^2_{miss} distributions of 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (top left), $B^+ \to \pi^0 \mu^+ \nu_\mu$ (top right) and all $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, over the full q^2 kinematic range.

6.5.1 Further Tests

Further investigation into the fitted data yields was subsequently conducted in order to verify that no other underlying issues were contributing to the observed results. One potential hypothesis for the given behaviour was related to the appropriateness of the current fitting strategy in terms of the number of fit components used. As described in Section 6.1, the total fit pdf was constructed from only two components, with all of the various background contributions combined into a single template due to limited statistics at the current data integrated luminosity. In the prior analysis completed on the full Belle data-set (711 fb^{-1}) [32] from which the current study was heavily inspired, multiple fit components were used to represent the total background contribution. Of particular interest in the reconstruction of $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, the yield of the $B^0 \to \rho^- \ell^+ \nu_\ell$ background component was fixed in the fit to the value corresponding to the number of expected $B^0 \to \rho^- \ell^+ \nu_\ell$ events assuming the $B^0 \to \rho^- \ell^+ \nu_\ell$ branching fraction measured within the same study [32]. It was hypothesised that in the current study, as illustrated in Figures 6.25 and 6.26, given the shape and size of the $B^0 \to \rho^- \ell^+ \nu_\ell$ background components in the M^2_{miss} signal region, it was possible that the fits were struggling to distinguish between signal and $B^0 \to \rho^- \ell^+ \nu_\ell$ background events in particular, in turn inflating the total background yields at the expense of the signal yields.

To determine whether or not this hypothesis was correct, additional sets of maximum likelihood fits were performed to the M^2_{miss} distributions in six q^2 bins built from reconstructed $B^0 \to \pi^- e^+ \nu_e$ and $B^0 \to \pi^- \mu^+ \nu_\mu$ decays, this time with a total of three fit components. These corresponded to the signal, the $B^0 \to \rho^- \ell^+ \nu_\ell$ background, and the sum of the remaining background contributions, with the $B^0 \to \rho^- \ell^+ \nu_\ell$ background events isolated from the original background template. To fix the $B^0 \to \rho^- \ell^+ \nu_\ell$ yields in the fits, a recent Belle II measurement[1] of the $B^0 \to \rho^- \ell^+ \nu_\ell$ branching fraction was used. For each M^2_{miss} distribution, in an identical manner to the branching fraction reweighting described in Section 4.1.3, the branching fraction of the $B^0 \to \rho^- \ell^+ \nu_\ell$ background was reweighted to this recent measurement by applying a scale factor to the $B^0 \to \rho^- \ell^+ \nu_\ell$ events equivalent to $\mathcal{B}(B^0 \to \rho^- \ell^+ \nu_\ell)_{\mathrm{Belle\ II}}/\mathcal{B}(B^0 \to \rho^- \ell^+ \nu_\ell)_{\mathrm{MC}} = (2.53 \times 10^{-4})/(2.94 \times 10^{-4})$. The number of $B^0 \to \rho^- \ell^+ \nu_\ell$ events in each q^2 bin was fixed in the fits to the expected number of these events in MC after all analysis corrections including this additional reweighting.

The yields obtained from these three-component maximum likelihood fits are listed in Table 6.8, with the equivalent fitted distributions given in Figures 6.41 and 6.42. The post-fit data-MC agreement was observed to worsen slightly with the increase in the number of fit components. The signal yields under this new fit configuration were found to be consistent with those obtained from the two-component fits within uncertainties, varying only slightly from these values. The ratios between the signal yields returned from both fits are also included in Table 6.8 as a rough indication of the observed variation, though the uncertainties on these values are substantial. A similar test was performed for $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays,

---

[1]This measurement is yet to be published publicly at the time of writing this thesis. The $B \to \pi \ell \nu_\ell$ analysis presented in this study was performed in collaboration with a $B \to \rho \ell \nu_\ell$ analysis headed by Moritz Bauer of the Karlsruhe Institute of Technology. A joint journal paper containing both analyses is currently proceeding through a process of internal peer review by the Belle II Collaboration.

but the observed variations were negligible due to the relatively small size of the $B \rightarrow \rho \ell \nu_\ell$ background components.

With the results of these tests, it can be concluded that the fit configuration, namely in terms of the number of fit components used, was not responsible for the observed deficiency in the data signal yields with respect to the MC prediction. Additionally, the consistency with which the results of the maximum likelihood fits evaluated over the full q^2 range agreed with the sum of the yields obtained from the fits to the individual q^2 bins lends further confidence to the ability of the current fitting strategy to discriminate between signal and background events. Further discussion on the low observed yields will be presented in the coming chapters, both in the context of the extraction of the $B \rightarrow \pi \ell \nu_\ell$ branching fractions and $|V_{ub}|$, and in the concluding remarks of this thesis. Thus far, no evidence has been established to indicate that the observed discrepancies are due to anything other than statistical fluctuations.

q^2 bin (GeV^2)	Fitted sig. yield	Fixed $B^0 \rightarrow \rho^- \ell^+ \nu_\ell$ yield	Fitted other bkg. yield	Sig. yield (2 comp.) /Sig. yield (3 comp.)
$B^0 \rightarrow \pi^- e^+ \nu_e$				
$0 \leqslant q^2 < 4$	20.18 ± 4.93	4.94	63.10 ± 8.42	1.01 ± 0.49
$4 \leqslant q^2 < 8$	14.98 ± 4.35	13.66	80.75 ± 9.72	1.03 ± 0.59
$8 \leqslant q^2 < 12$	16.38 ± 4.93	21.82	121.82 ± 11.91	1.06 ± 0.63
$12 \leqslant q^2 < 16$	20.27 ± 5.60	31.03	180.31 ± 14.32	1.09 ± 0.59
$16 \leqslant q^2 < 20$	10.13 ± 5.00	40.82	229.31 ± 16.30	1.12 ± 1.06
$20 \leqslant q^2 \leqslant 26.4$	9.81 ± 4.59	44.03	143.64 ± 13.37	1.15 ± 1.03
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$				
$0 \leqslant q^2 < 4$	13.45 ± 4.54	4.12	43.85 ± 7.36	1.01 ± 0.68
$4 \leqslant q^2 < 8$	18.09 ± 4.72	11.92	68.24 ± 8.95	1.03 ± 0.53
$8 \leqslant q^2 < 12$	15.89 ± 4.59	18.31	118.06 ± 11.60	1.03 ± 0.59
$12 \leqslant q^2 < 16$	16.39 ± 4.75	27.29	167.25 ± 13.68	1.07 ± 0.61
$16 \leqslant q^2 < 20$	6.01 ± 3.65	36.83	207.52 ± 15.28	1.14 ± 1.32
$20 \leqslant q^2 \leqslant 26.4$	2.91 ± 2.84	36.62	136.19 ± 12.85	1.12 ± 2.09

Table 6.8: Yields of reconstructed $B^0 \rightarrow \pi^- e^+ \nu_e$ and $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays obtained from 3-component maximum likelihood fits to 362 fb^{-1} of Belle II data. The $B^0 \rightarrow \rho^- \ell^+ \nu_\ell$ yields have been fixed to the data-driven predictions taken from a recent Belle II measurement of the $B^0 \rightarrow \rho^- \ell^+ \nu_\ell$ branching fraction.
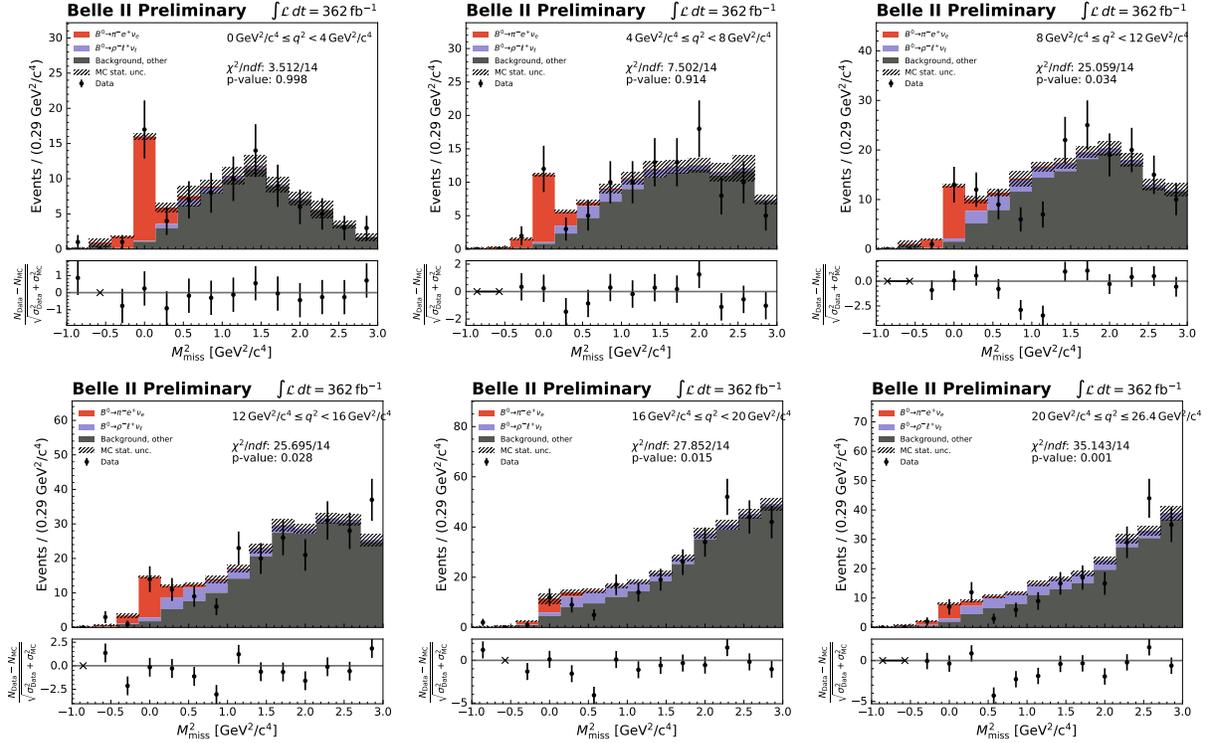
Figure 6.41: M_{miss}^2 distributions of 362 fb^{-1} of Belle II data for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins, with 3-component MC fit projections overlaid. The $B^0 \to \rho^- \ell^+ \nu_\ell$ components have been fixed to the data-driven predictions taken from a recent Belle II measurement of the $B^0 \to \rho^- \ell^+ \nu_\ell$ branching fraction.
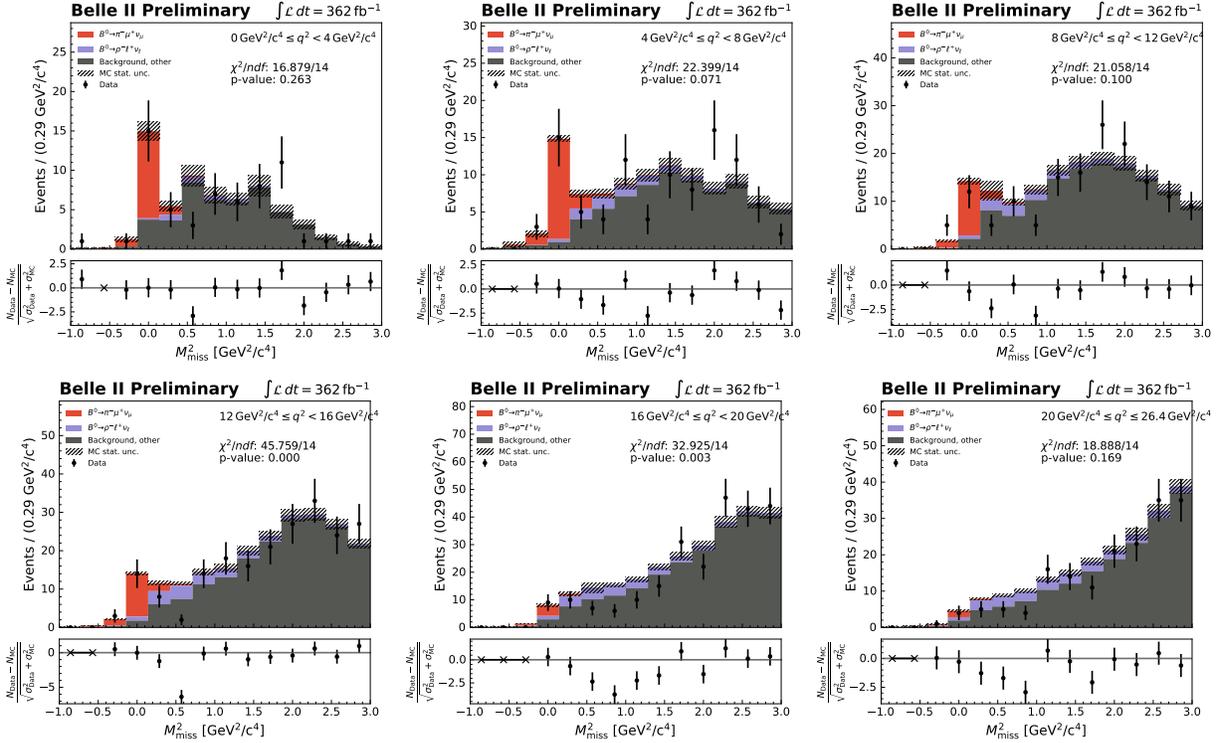
Figure 6.42: M_{miss}^2 distributions of 362 fb^{-1} of Belle II data for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins, with 3-component MC fit projections overlaid. The $B^0 \to \rho^- \ell^+ \nu_\ell$ components have been fixed to the data-driven predictions taken from a recent Belle II measurement of the $B^0 \to \rho^- \ell^+ \nu_\ell$ branching fraction.

Chapter 7

Branching Fraction and $|V_{\mathrm{ub}}|$ Determination

Using a robust fitting framework built and validated upon simulated data, the number of signal $B \to \pi \ell \nu_\ell$ events in real Belle II data was able to be estimated, with the previous chapter detailing the various results obtained. In addition to determining the total number of signal events present in the given data-set, the proportion of events falling within each of the six distinct q^2 regions studied was also evaluated, in preparation for the extraction of the $B \to \pi \ell \nu_\ell$ branching fractions, and subsequently, $|V_{\mathrm{ub}}|$.

In this chapter, measurements of the branching fractions for the semi-leptonic $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays will first be presented, including a detailed description of the various quantities necessary for the calculations. The partial branching fractions evaluated for each q^2 bin will be extracted, with these summed to evaluate the total branching fraction for each decay. The latter sections of this chapter will focus on the extraction of the magnitude of the CKM-matrix element $|V_{\mathrm{ub}}|$ using the distributions of the measured partial branching fractions. With each evaluation performed first using simulated Asimov data for the purpose of validating the given procedures, a summary and discussion of the results subsequently obtained from real Belle II data will then be presented.

7.1 Measuring the Partial Branching Fractions

The branching fraction of a particular decay, first introduced in Section 1.2, represents the relative proportion of particles decaying via that specified decay mode. In the context of $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, where ℓ refers to either an electron or muon, these refer to the fraction of B^0 mesons decaying via the mode $B^0 \to \pi^- \ell^+ \nu_\ell$ and the fraction of B^+ mesons decaying via $B^+ \to \pi^0 \ell^+ \nu_\ell$, respectively. The total branching fractions for these decays can be measured in two distinct ways. The first method involves using the number of estimated signal events within each q^2 bin to extract what are known as the partial branching fractions for each bin. These partial branching fractions can then be summed to evaluate the total branching fraction for the given decay, ensuring all correlations in the uncertainties between q^2 bins are taken into account. Alternatively, the total branching fractions can be

measured directly from the total number of estimated signal events evaluated over the full q^2 kinematic range. The former method is most desirable, due to the fact that the individual fits performed within each q^2 bin can better capture the variation in the signal shape and yield at low versus high q^2. Additionally, it is the distribution of the partial branching fractions as a function of q^2 that is required for the computation of $|V_{ub}|$, as will be described in detail in Section 7.2. For this study, the partial branching fractions in the six q^2 bins studied will be evaluated and used to determine the total branching fractions, with the branching fractions evaluated over the full q^2 range provided as a secondary measurement and as a cross-check.

The partial branching fractions for $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, $\Delta\mathcal{B}_i(B^0 \to \pi^- \ell^+ \nu_\ell)$ and $\Delta\mathcal{B}_i(B^+ \to \pi^0 \ell^+ \nu_\ell)$, respectively, can be determined for each q^2 bin i via the following formulae [92]:

$$\Delta\mathcal{B}_i(B^0 \to \pi^- \ell^+ \nu_\ell) = \frac{N_{\text{sig},i}(1 + f_{+0})}{2 \times N_{e/\mu} \times N_{B\bar{B}} \times \epsilon_i} \ ,$$

$$\Delta\mathcal{B}_i(B^+ \to \pi^0 \ell^+ \nu_\ell) = \frac{N_{\text{sig},i}(1 + f_{+0})}{2 \times N_{e/\mu} \times N_{B\bar{B}} \times f_{+0} \times \epsilon_i} \ ,$$

where $N_{\text{sig},i}$ is the estimated signal yield in each q^2 bin, f_{+0} is the ratio between the branching fractions of the decays of the $\Upsilon(4S)$ meson to pairs of charged and neutral B-mesons [93], $N_{B\bar{B}}$ is the number of B-meson pairs counted in the data-set at the current integrated luminosity, $N_{e/\mu}$ is the number of lepton flavours used for the calculation, and ϵ is the weighted signal reconstruction efficiency for each bin. The factor of two present in the denominator accounts for the two B-mesons in the $\Upsilon(4S)$ decay. In a similar manner, the $N_{e/\mu}$ parameter supports the determination of the partial branching fractions using a single lepton flavour, $N_{e/\mu} = 1$, such as in the reconstruction of $B^0 \to \pi^- e^+ \nu_e$ or $B^0 \to \pi^- \mu^+ \nu_\mu$, or in the reconstruction of either lepton flavour, $N_{e/\mu} = 2$, where all $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays are reconstructed. In this study, the branching fractions will be evaluated using both the individual lepton channels and the combined lepton channels, with the latter forming the typical measurement presented in literature. By also determining $\mathcal{B}(B^0 \to \pi^- e^+ \nu_e)$, $\mathcal{B}(B^0 \to \pi^- \mu^+ \nu_\mu)$, $\mathcal{B}(B^+ \to \pi^0 e^+ \nu_e)$ and $\mathcal{B}(B^+ \to \pi^0 \mu^+ \nu_\mu)$, some insight can be garnered into the relative capability of the current Belle II detector and analysis software in the reconstruction of electron versus muon modes.

The formulae for the total branching fractions evaluated over the full q^2 kinematic range, $\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell)$ and $\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell)$, are largely similar to those used for the derivation of the partial branching fractions:

$$\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell) = \frac{N_{\text{sig}}(1 + f_{+0})}{2 \times N_{e/\mu} \times N_{B\bar{B}} \times \epsilon} \ ,$$

$$\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell) = \frac{N_{\text{sig}}(1 + f_{+0})}{2 \times N_{e/\mu} \times N_{B\bar{B}} \times f_{+0} \times \epsilon} \ ,$$

where N_{sig} and ϵ now refer to the estimated signal yield and the weighted signal reconstruction efficiency evaluated over the full q^2 range, respectively.

q^2 bin (GeV^2)	MC signal efficiency $\epsilon_{(i)}$
$B^0 \to \pi^- e^+ \nu_e$	
$0 \leqslant q^2 < 4$	$(0.194 \pm 0.004)\%$
$4 \leqslant q^2 < 8$	$(0.209 \pm 0.004)\%$
$8 \leqslant q^2 < 12$	$(0.226 \pm 0.004)\%$
$12 \leqslant q^2 < 16$	$(0.231 \pm 0.004)\%$
$16 \leqslant q^2 < 20$	$(0.225 \pm 0.005)\%$
$20 \leqslant q^2 \leqslant 26.4$	$(0.176 \pm 0.004)\%$
Full q^2 range	$(0.212 \pm 0.002)\%$
$B^0 \to \pi^- \mu^+ \nu_\mu$	
$0 \leqslant q^2 < 4$	$(0.164 \pm 0.003)\%$
$4 \leqslant q^2 < 8$	$(0.192 \pm 0.004)\%$
$8 \leqslant q^2 < 12$	$(0.207 \pm 0.004)\%$
$12 \leqslant q^2 < 16$	$(0.212 \pm 0.004)\%$
$16 \leqslant q^2 < 20$	$(0.205 \pm 0.004)\%$
$20 \leqslant q^2 \leqslant 26.4$	$(0.169 \pm 0.004)\%$
Full q^2 range	$(0.192 \pm 0.002)\%$
$B^0 \to \pi^- \ell^+ \nu_\ell$	
$0 \leqslant q^2 < 4$	$(0.179 \pm 0.003)\%$
$4 \leqslant q^2 < 8$	$(0.201 \pm 0.003)\%$
$8 \leqslant q^2 < 12$	$(0.216 \pm 0.003)\%$
$12 \leqslant q^2 < 16$	$(0.222 \pm 0.003)\%$
$16 \leqslant q^2 < 20$	$(0.215 \pm 0.003)\%$
$20 \leqslant q^2 \leqslant 26.4$	$(0.172 \pm 0.003)\%$
Full q^2 range	$(0.202 \pm 0.001)\%$

Table 7.1: Weighted signal reconstruction efficiencies derived from MC for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays.

The f_{+0} parameter used in each calculation was taken from a recent Belle study representing the most precise measurement made to date, with $f_{+0} = \mathcal{B}(\Upsilon(4S) \to B^+ B^-)/\mathcal{B}(\Upsilon(4S) \to B^0 \bar{B}^0) = 1.065 \pm 0.052$ [93]. The number of $\Upsilon(4S) \to B\bar{B}$ pairs in the current 362 fb^{-1} Belle II data-set, $N_{B\bar{B}} = (387 \pm 6) \times 10^6$, was determined from an independent Belle II study in which B-meson events were reconstructed from both on- and off-resonance data, with the number of B-meson pairs in the on-resonance data-set estimated using a data subtraction procedure.

The signal reconstruction efficiencies used for the branching fraction determinations were

evaluated from MC, both for each q^2 bin and over the full q^2 kinematic range. In each case, the efficiency was defined as the ratio between the number of signal MC events after the application of all analysis selections and data-MC corrections, and the total number of signal MC events generated in the original samples. As the numerator is weighted by the various correction factors including the FEI calibration, lepton and pion identification corrections and so on, the resultant values are referred to as the weighted signal reconstruction efficiencies. Furthermore, each efficiency technically represents the combination of the tagging efficiency ϵ_{tag}, the proportion of events for which a hadronic B_{tag} was able to be reconstructed, and the selection efficiency ϵ_{sel}, the proportion of events remaining after all analysis selections, with $\epsilon = \epsilon_{\text{tag}} \times \epsilon_{\text{sel}}$. A binomial uncertainty $\sigma_{\epsilon_{(i)}}$ was assigned to each efficiency measurement in order to represent the sampling error when taking a proportion of a statistical population [94], $\sigma_{\epsilon_{(i)}} = \sqrt{\epsilon_{(i)}(1 - \epsilon_{(i)})/N_{\text{gen}_{(i)}}}$, where $N_{\text{gen}_{(i)}}$ is the number of generated events relevant to the calculation.

Tables 7.1 and 7.2 list the signal reconstruction efficiencies used in the branching fraction calculations, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively. The efficiencies for both the individual lepton and combined lepton reconstructions are included in each table. It can be observed that for both B-meson flavours, the muon channels were reconstructed with a lower efficiency than the electron channels. In addition, the reconstruction efficiencies for the charged B-meson channels were significantly higher than those for the neutral B-meson channels, a feature driven largely by the fact that the tagging efficiencies for B^+ tags have been found to exceed those for B^0 tags in independent studies of the application of the FEI [13].

The final elements required for the branching fraction calculations were the values of the estimated signal yields. Whilst the determination of these yields was described in the previous chapter, one further correction was required before using the yields measured in each q^2 bin to extract the partial branching fractions, related to the resolution of the measured q^2 spectrum. This correction is known as the unfolding of the q^2 distribution, and will be explained in detail in the following section.

7.1.1 Unfolding the q^2 Distribution

As explained in Section 2.4.2, when generating particle decays in MC (via the EvtGen software [60]), the kinematic quantities corresponding to the various particles involved in the decays are precisely known, including those of the energy and momentum. However, in simulating the detection and reconstruction of these generated particles by Belle II (via the Geant4 software [61]), the finite resolution of the various detector components is taken into account, resulting in reconstructed energy and momentum values that are slightly shifted with respect to the underlying MC truth. It directly follows that these resolution effects would subsequently impact the determination of q^2, which is derived using the four-momenta of the leptonic daughters in the semi-leptonic $B \to \pi \ell \nu_\ell$ decay. In this way, a discrepancy between the true and reconstructed values of q^2 for a particular decay could result in it being ascribed to a different q^2 bin in the reconstruction than it belongs in truth, a feature referred

q^2 bin (GeV^2)	MC signal efficiency $\epsilon_{(i)}$
$B^+ \to \pi^0 e^+ \nu_e$	
$0 \leqslant q^2 < 4$	$(0.377 \pm 0.007)\%$
$4 \leqslant q^2 < 8$	$(0.428 \pm 0.007)\%$
$8 \leqslant q^2 < 12$	$(0.455 \pm 0.008)\%$
$12 \leqslant q^2 < 16$	$(0.424 \pm 0.008)\%$
$16 \leqslant q^2 < 20$	$(0.396 \pm 0.008)\%$
$20 \leqslant q^2 \leqslant 26.4$	$(0.562 \pm 0.011)\%$
Full q^2 range	$(0.433 \pm 0.003)\%$
$B^+ \to \pi^0 \mu^+ \nu_\mu$	
$0 \leqslant q^2 < 4$	$(0.238 \pm 0.005)\%$
$4 \leqslant q^2 < 8$	$(0.334 \pm 0.006)\%$
$8 \leqslant q^2 < 12$	$(0.392 \pm 0.007)\%$
$12 \leqslant q^2 < 16$	$(0.388 \pm 0.007)\%$
$16 \leqslant q^2 < 20$	$(0.383 \pm 0.008)\%$
$20 \leqslant q^2 \leqslant 26.4$	$(0.509 \pm 0.010)\%$
Full q^2 range	$(0.363 \pm 0.003)\%$
$B^+ \to \pi^0 \ell^+ \nu_\ell$	
$0 \leqslant q^2 < 4$	$(0.307 \pm 0.004)\%$
$4 \leqslant q^2 < 8$	$(0.381 \pm 0.005)\%$
$8 \leqslant q^2 < 12$	$(0.423 \pm 0.005)\%$
$12 \leqslant q^2 < 16$	$(0.406 \pm 0.005)\%$
$16 \leqslant q^2 < 20$	$(0.389 \pm 0.006)\%$
$20 \leqslant q^2 \leqslant 26.4$	$(0.535 \pm 0.007)\%$
Full q^2 range	$(0.398 \pm 0.002)\%$

Table 7.2: Weighted signal reconstruction efficiencies derived from MC for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays.

to as q^2 bin migration. Such a scenario would thus result in a potential distortion of the relative signal yields distributed across the q^2 bins studied, and in turn, the resultant partial branching fractions. Whilst it is impossible to access any underlying truth information for real data, it can be safely assumed that the same resolution effects would result in a similar distortion in the measured partial branching fractions.

In simulated data, the magnitude of the bin migration can be evaluated explicitly due to the fact that both the true and reconstructed kinematic information can be accessed. A q^2 bin migration matrix can be defined for each reconstructed $B \to \pi \ell \nu_\ell$ decay, M_{ij}, which lists the proportion of events that fall in each reconstructed bin i for each true bin j. Figures 7.1 and 7.2 display the migration matrices built from MC for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively. A Python implementation of the `RooUnfold` software

package [95] was used to construct each matrix. It can be clearly observed that in all cases, the vast majority of MC events were reconstructed within the correct q^2 bin, with minimal bin migration found. Since only six q^2 bins were used for the analysis at present due to limited statistics, these results were largely unsurprising given the relatively broad width of $\geqslant 4$ GeV^2 for each bin. In potential future iterations of the analysis where a finer granularity in q^2 could be investigated, bin migration effects would be expected to be more significant.

Figure 7.1: The q^2 bin migration matrices evaluated from MC, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (top left), $B^0 \to \pi^- \mu^+ \nu_\mu$ (top right), and $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ.

For reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, where both the individual and combined lepton cases were considered, less than 4% of events in each q^2 bin were found to correspond to a different bin in the MC truth. Furthermore, almost all of the migrated events were reconstructed within the q^2 bin directly above or below the bin in the underlying MC truth.

186

Slightly higher rates of bin migration were observed for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, due to the poorer resolution pertaining to reconstructing the $\pi^0 \to \gamma\gamma$ decay in contrast with the relatively clean charged pion track, with the highest observed migration in a single q^2 bin of $\sim 18\%$. These migrated events were similarly concentrated within the reconstructed bins directly above or below the true bins, although the spread of events into reconstructed bins further away was likewise more substantial than the $B^0 \to \pi^- \ell^+ \nu_\ell$ case. In addition to this, a clear asymmetric behaviour in the bin migration was observed in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ case, with migrated events tending to fall more often within higher q^2 bins than lower q^2 bins with respect to the MC truth, implying a tendency for signal events to be reconstructed with q^2 values higher than the underlying reality.
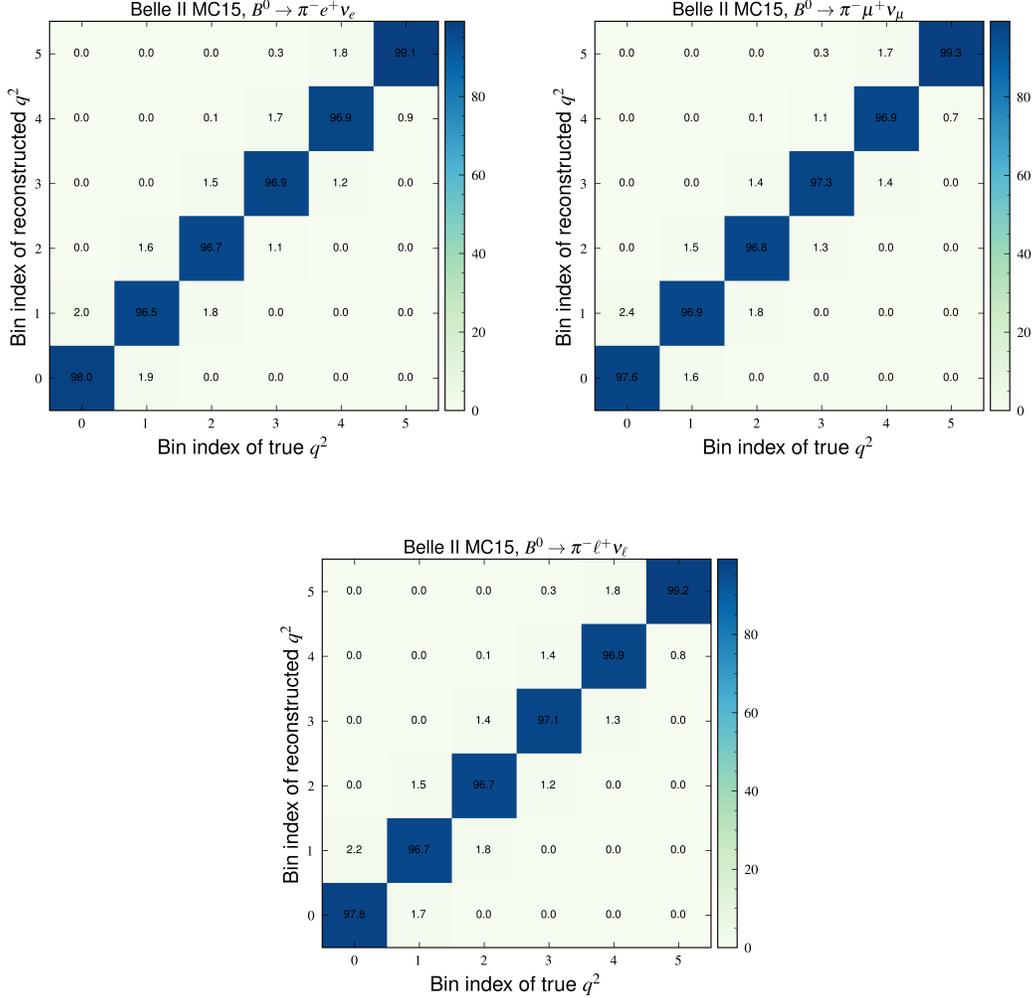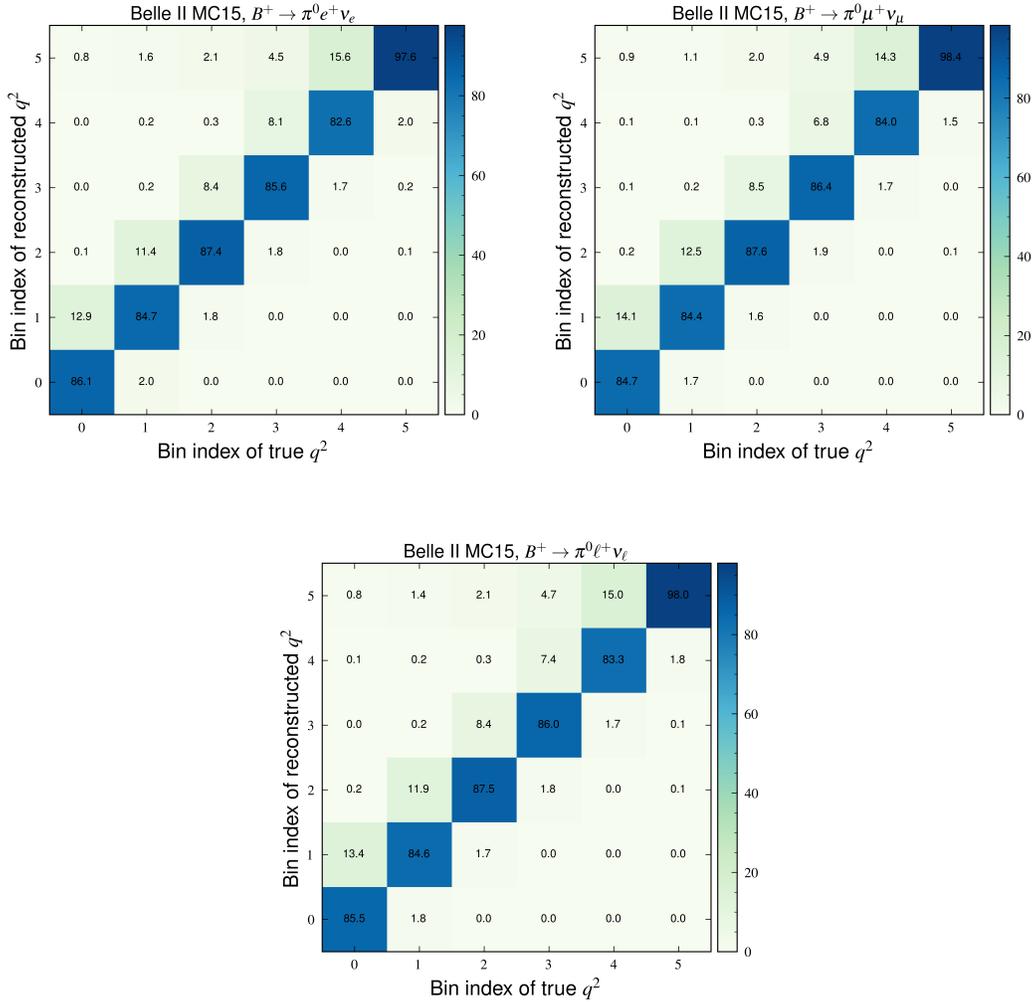

Figure 7.2: The q^2 bin migration matrices evaluated from MC, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (top left), $B^+ \to \pi^0 \mu^+ \nu_\mu$ (top right), and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ.

q^2 bin (GeV^2)	Asimov data		Belle II data	
	Fitted sig. yield	Unfolded sig. yield	Fitted sig. yield	Unfolded sig. yield
$B^0 \to \pi^- e^+ \nu_e$				
$0 \leqslant q^2 < 4$	19.19 ± 4.90	19.18 ± 4.90	20.29 ± 4.95	20.28 ± 4.95
$4 \leqslant q^2 < 8$	21.26 ± 5.26	21.20 ± 5.24	15.44 ± 4.41	15.40 ± 4.40
$8 \leqslant q^2 < 12$	22.68 ± 5.66	22.85 ± 5.70	17.44 ± 5.08	17.57 ± 5.12
$12 \leqslant q^2 < 16$	21.42 ± 5.80	21.54 ± 5.83	22.06 ± 5.78	22.17 ± 5.81
$16 \leqslant q^2 < 20$	17.69 ± 6.01	17.77 ± 6.04	11.35 ± 5.14	11.40 ± 5.16
$20 \leqslant q^2 \leqslant 26.4$	11.08 ± 4.77	10.78 ± 4.64	11.30 ± 4.80	10.99 ± 4.66
Sum	113.32 ± 13.28	113.32 ± 13.26	97.88 ± 12.35	97.81 ± 12.34
$B^0 \to \pi^- \mu^+ \nu_\mu$				
$0 \leqslant q^2 < 4$	16.29 ± 5.03	16.36 ± 5.05	13.56 ± 4.57	13.62 ± 4.59
$4 \leqslant q^2 < 8$	19.44 ± 5.05	19.28 ± 5.01	18.57 ± 4.78	18.42 ± 4.74
$8 \leqslant q^2 < 12$	20.74 ± 5.37	20.86 ± 5.40	16.39 ± 4.64	16.48 ± 4.67
$12 \leqslant q^2 < 16$	19.65 ± 5.27	19.68 ± 5.28	17.49 ± 4.86	17.52 ± 4.86
$16 \leqslant q^2 < 20$	16.13 ± 5.07	16.32 ± 5.13	6.84 ± 3.75	6.92 ± 3.80
$20 \leqslant q^2 \leqslant 26.4$	10.62 ± 4.22	10.36 ± 4.12	3.26 ± 2.91	3.19 ± 2.84
Sum	102.87 ± 12.29	102.86 ± 12.29	76.11 ± 10.56	76.15 ± 10.56
$B^0 \to \pi^- \ell^+ \nu_\ell$				
$0 \leqslant q^2 < 4$	35.49 ± 7.03	35.56 ± 7.04	33.57 ± 6.75	33.63 ± 6.76
$4 \leqslant q^2 < 8$	40.70 ± 7.30	40.49 ± 7.26	34.42 ± 6.55	34.24 ± 6.52
$8 \leqslant q^2 < 12$	43.41 ± 7.79	43.71 ± 7.85	33.32 ± 6.80	33.55 ± 6.85
$12 \leqslant q^2 < 16$	41.07 ± 7.79	41.21 ± 7.82	39.43 ± 7.51	39.56 ± 7.54
$16 \leqslant q^2 < 20$	33.81 ± 7.79	34.09 ± 7.85	18.33 ± 6.33	18.47 ± 6.39
$20 \leqslant q^2 \leqslant 26.4$	21.71 ± 6.28	21.15 ± 6.12	13.89 ± 5.47	13.53 ± 5.33
Sum	216.19 ± 18.01	216.21 ± 18.00	172.96 ± 16.16	172.98 ± 16.16

Table 7.3: The unfolded signal yields compared with the fitted signal yields, for all $B^0 \to \pi^- \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data.

Whilst the magnitude of the q^2 bin migration was observed to be small for the given analysis, a correction procedure was nevertheless implemented to adjust the relative proportions of signal events in each q^2 bin to better reflect the true underlying yield distribution. In order to implement these corrections, the q^2 bin migration matrices were inverted in a process referred to as q^2 unfolding. Multiple methods exist for the unfolding process, with the simplest implementation resulting in the derivation of a set of bin-by-bin correction factors. Given the minimal bin migration, this simple approach was adopted for the given analysis. In this method, the true (N_{true}) and reconstructed ($N_{\text{reconstructed}}$) signal yields are used to define a weight for each q^2 bin k, where $w_k = N_{\text{true},k}/N_{\text{reconstructed},k}$. The fitted signal yields in each bin can then simply be corrected by multiplying them by the given weight. For this analysis, the `RooUnfold` package [95] was used to perform a bin-by-bin unfolding of the q^2 spectrum using the true and reconstructed information given in the migration matrices built from MC.

Before applying the unfolding procedure to real data for which the underlying truth is

q^2 bin (GeV^2)	Asimov data		Belle II data	
	Fitted sig. yield	Unfolded sig. yield	Fitted sig. yield	Unfolded sig. yield
$B^+ \to \pi^0 e^+ \nu_e$				
$0 \leqslant q^2 < 4$	21.51 ± 5.77	24.39 ± 6.54	15.88 ± 4.99	18.01 ± 5.66
$4 \leqslant q^2 < 8$	25.42 ± 6.35	25.76 ± 6.44	26.94 ± 6.21	27.31 ± 6.29
$8 \leqslant q^2 < 12$	25.42 ± 6.57	25.18 ± 6.51	22.11 ± 6.12	21.91 ± 6.07
$12 \leqslant q^2 < 16$	21.94 ± 6.28	22.67 ± 6.49	25.46 ± 6.35	26.31 ± 6.56
$16 \leqslant q^2 < 20$	17.94 ± 6.68	18.97 ± 7.06	0.00 ± 6.60	0.00 ± 6.98
$20 \leqslant q^2 \leqslant 26.4$	19.96 ± 8.66	15.21 ± 6.60	23.76 ± 8.63	18.10 ± 6.58
Sum	132.19 ± 16.61	132.18 ± 16.19	114.15 ± 16.10	111.64 ± 15.60
$B^+ \to \pi^0 \mu^+ \nu_\mu$				
$0 \leqslant q^2 < 4$	13.55 ± 5.19	15.61 ± 5.98	4.67 ± 4.08	5.38 ± 4.70
$4 \leqslant q^2 < 8$	19.83 ± 5.54	20.45 ± 5.72	14.90 ± 4.94	15.37 ± 5.09
$8 \leqslant q^2 < 12$	21.91 ± 5.93	21.61 ± 5.85	14.75 ± 4.94	14.55 ± 4.87
$12 \leqslant q^2 < 16$	20.12 ± 6.17	20.74 ± 6.36	19.22 ± 5.91	19.81 ± 6.09
$16 \leqslant q^2 < 20$	17.31 ± 5.80	18.55 ± 6.21	12.02 ± 5.12	12.88 ± 5.48
$20 \leqslant q^2 \leqslant 26.4$	18.16 ± 7.66	13.89 ± 5.86	21.38 ± 7.77	16.35 ± 5.95
Sum	110.83 ± 14.94	110.85 ± 14.70	86.94 ± 13.67	84.34 ± 13.20
$B^+ \to \pi^0 \ell^+ \nu_\ell$				
$0 \leqslant q^2 < 4$	35.01 ± 7.73	39.94 ± 8.81	20.54 ± 6.34	23.43 ± 7.23
$4 \leqslant q^2 < 8$	45.24 ± 8.38	46.21 ± 8.56	43.06 ± 7.97	43.99 ± 8.14
$8 \leqslant q^2 < 12$	47.32 ± 8.80	46.79 ± 8.70	36.01 ± 7.90	35.60 ± 7.81
$12 \leqslant q^2 < 16$	42.03 ± 8.75	43.38 ± 9.03	45.87 ± 8.71	47.35 ± 9.00
$16 \leqslant q^2 < 20$	35.26 ± 8.76	37.53 ± 9.32	11.26 ± 6.59	11.99 ± 7.01
$20 \leqslant q^2 \leqslant 26.4$	38.05 ± 11.32	29.05 ± 8.64	47.69 ± 11.46	36.40 ± 8.75
Sum	242.91 ± 22.11	242.90 ± 21.67	204.43 ± 20.41	198.76 ± 19.65

Table 7.4: The unfolded signal yields compared with the fitted signal yields, for all $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data.

unknown, the process was validated using Asimov data built from MC at the integrated luminosity of the current data-set, 362 fb^{-1}. The signal yields in each q^2 bin returned from the fits to Asimov data described in the previous chapter (Table 6.1, Section 6.2) were corrected using the bin-by-bin method, with the resultant unfolded yields given in Tables 7.3 and 7.4 for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively, presented alongside the raw fitted yields for comparison. As expected due to the minimal bin migration observed, the yields were found to shift only slightly after unfolding for the reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays. Somewhat larger shifts were similarly observed as expected in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ case, though these differences largely fell within the range of the fitted yield uncertainties, further reinforcing the minimal impact of the q^2 unfolding at the given integrated luminosity and number of q^2 bins. To illustrate the effectiveness of the unfolding procedure, the respective fitted and unfolded Asimov yields listed in Tables 7.3 and 7.4 are depicted graphically in Figures 7.3 and 7.5 together with the underlying true yields. In every case, correcting the reconstructed signal yields using the unfolded q^2 spectrum recovered

the true yields, with this behaviour demonstrated particularly well in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ case where the distinction between the true and reconstructed yields was visibly more pronounced.

Figure 7.3: The unfolded signal yields derived from 362 fb^{-1} of Asimov data, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (top left), $B^0 \to \pi^- \mu^+ \nu_\mu$ (top right), and $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ.

Given the clear effectiveness of the unfolding procedure on Asimov data, the correction factors were subsequently applied to the real data yields detailed in Tables 6.6 and 6.7 (Section 6.5) in the previous chapter. The resultant unfolded yields compared with the fitted yields are also listed in Tables 7.3 and 7.4 for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively, with the respective yield plots depicted in Figures 7.4 and 7.6. The magnitude of the given shifts in the signal yields post-unfolding can be clearly seen to be of the same order as those observed for Asimov data. With the determination of the unfolded signal yields representing the final remaining component required for the calculations, the following section will detail the final results obtained for the $B \to \pi \ell \nu_\ell$ partial branching fractions.

7.1.2 Final Results

Finally, using the formulae given in Section 7.1, the partial and total branching fractions for each $B \to \pi \ell \nu_\ell$ decay mode studied were determined from both Asimov and real Belle II data at an integrated luminosity of 362 fb^{-1}. To extract the partial branching fractions, the

Figure 7.4: The unfolded signal yields derived from $362\,\mathrm{fb}^{-1}$ of Belle II data, for reconstructed $B^0 \rightarrow \pi^- e^+ \nu_e$ (top left), $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (top right), and $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ.

unfolded signal yields in each q^2 bin were used as input into the formulae, together with the signal reconstruction efficiencies derived from MC and the values of f_{+0} and $N_{B\bar{B}}$. For each decay mode, the partial branching fractions were summed to determine the total branching fractions. To extract the total branching fractions evaluated over the full q^2 range as a secondary measurement, the total fitted signal yields were used in lieu of the unfolded per-bin yields in the relevant equations. Both the branching fractions evaluated using a single lepton flavour ($\mathcal{B}(B^0 \rightarrow \pi^- e^+ \nu_e)$, $\mathcal{B}(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)$, $\mathcal{B}(B^+ \rightarrow \pi^0 e^+ \nu_e)$ and $\mathcal{B}(B^+ \rightarrow \pi^0 \mu^+ \nu_\mu)$) and those evaluated for the combined lepton case ($\mathcal{B}(B^0 \rightarrow \pi^- \ell^+ \nu_\ell)$ and $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$, where $\ell = e$ or μ) were determined. The final results are summarised in Tables 7.5 and 7.6 for all reconstructed $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays, respectively. In each table, the world average measurement listed by the Particle Data Group [20] is included for comparison with the measured branching fractions.

Each branching fraction presented in the given tables is associated with two distinct uncertainties, the former being statistical and the latter systematic. The statistical uncertainties for each partial branching fraction were determined by replacing the unfolded signal yields in the relevant formulae with the uncertainties on these unfolded yields. A similar method was adopted for the statistical uncertainties on the total branching fractions evaluated over the full q^2 range, with the uncertainties on the signal yields returned from the M^2_{miss} fits over the full q^2 spectrum used in place of the signal yields in the calculation.
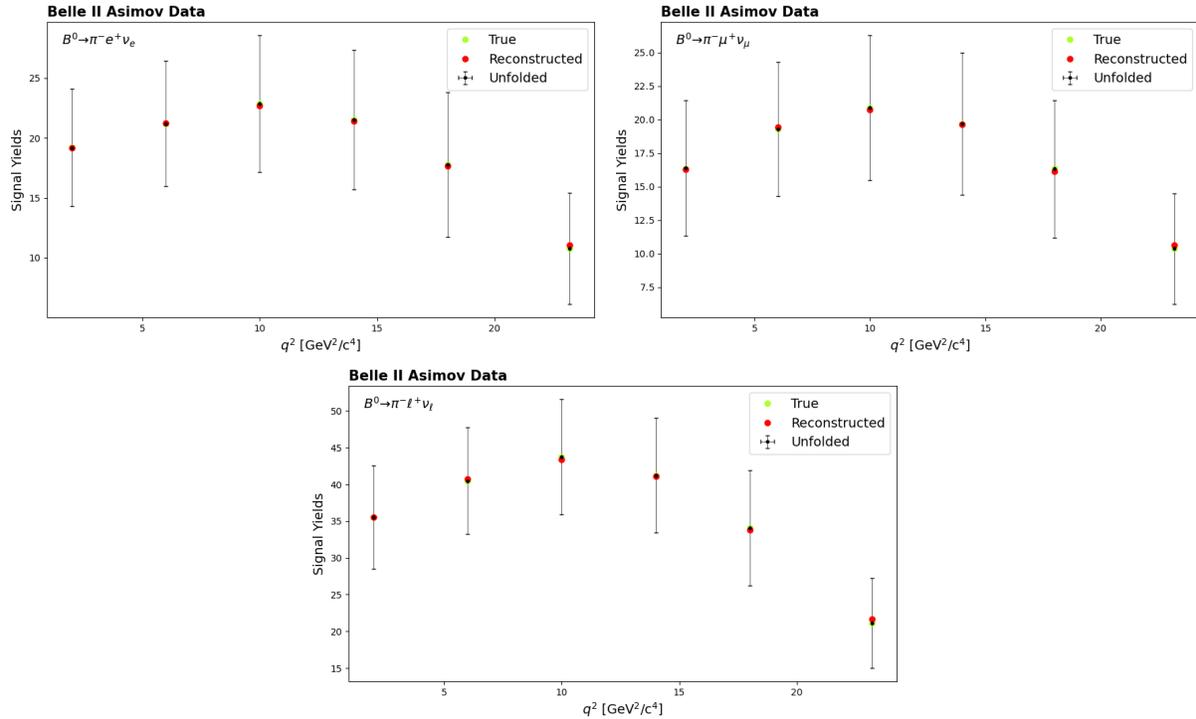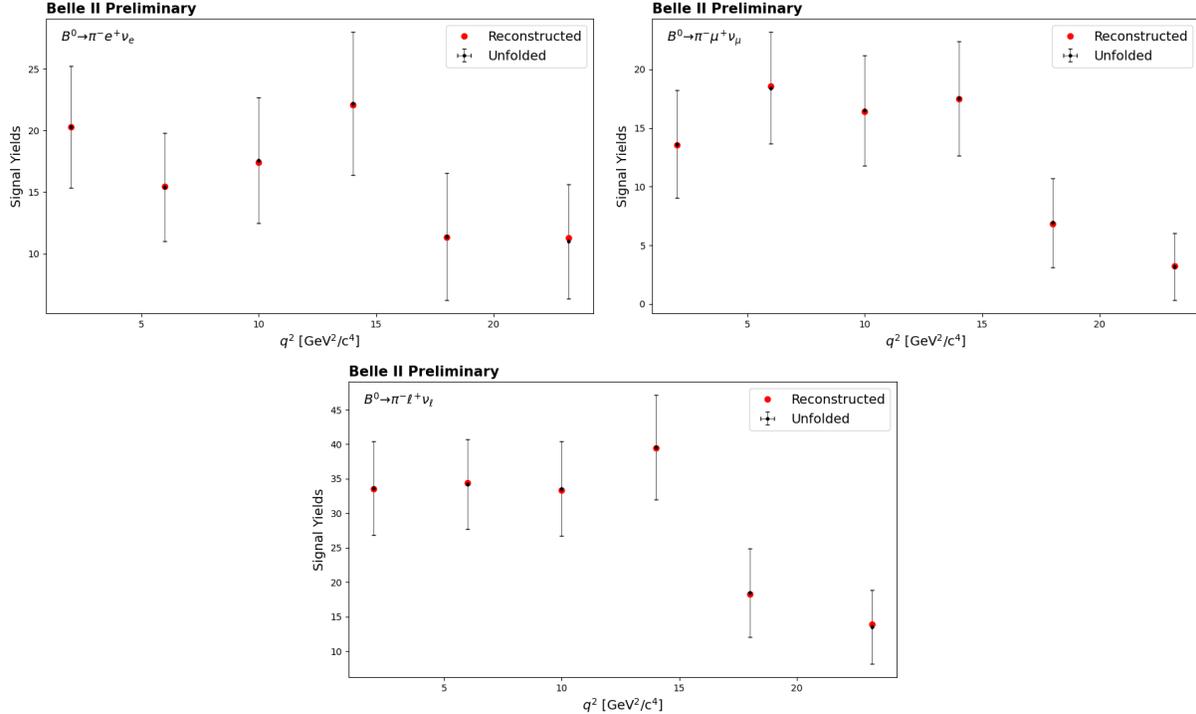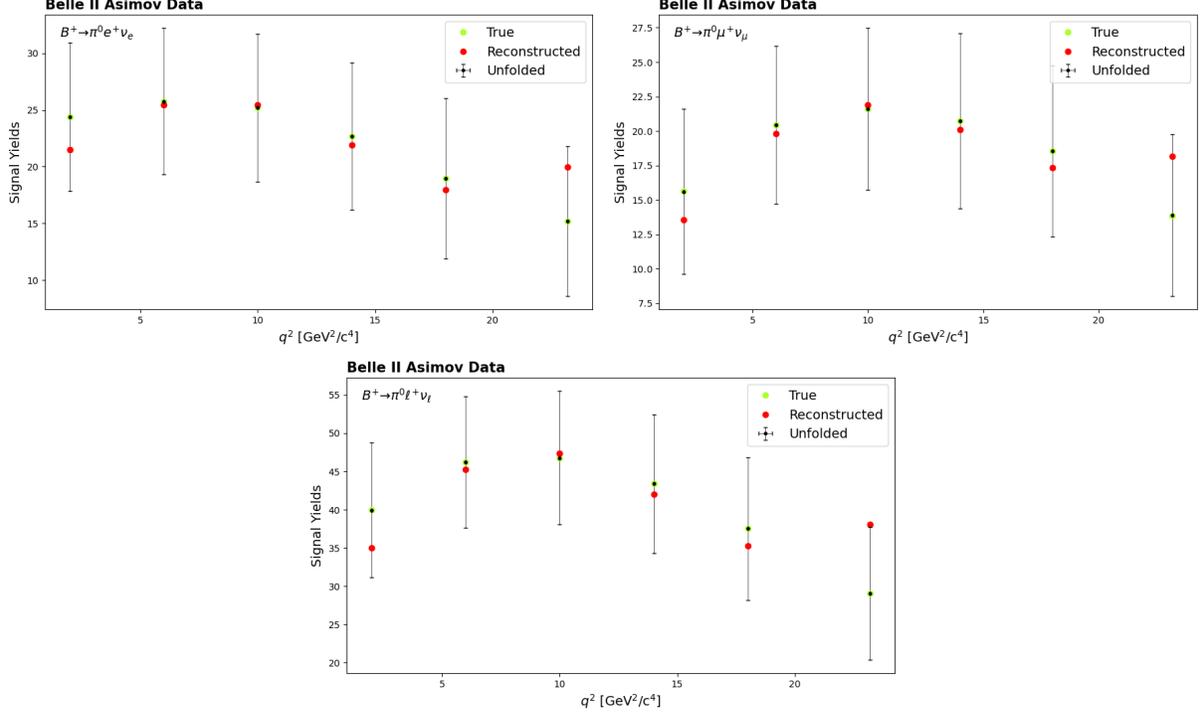
Figure 7.5: The unfolded signal yields derived from 362 fb^{-1} of Asimov data, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (top left), $B^+ \to \pi^0 \mu^+ \nu_\mu$ (top right), and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ.

To determine the statistical uncertainty on the sum of the partial branching fractions for each reconstructed mode, the statistical uncertainties of each partial branching fraction were combined in quadrature under the assumption of statistical independence between the subsets of data in each q^2 bin, with $\sigma_{\text{stat, sum}} = \sqrt{\sum_{i=1}^{6} \sigma_{\text{stat},i}^2}$, where $\sigma_{\text{stat},i}$ refers to the statistical uncertainty of each partial branching fraction. A discussion of the systematic uncertainties assigned to each branching fraction is left to the following section.

When considering the results obtained from Asimov data, several notable observations can be made. Firstly, in the reconstruction of all $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, the total branching fractions obtained through the sum of the partial branching fractions and the evaluation over the full q^2 range are largely indistinguishable, with these consistent results obtained through two independent measurements lending further confidence to the fitting and unfolding procedures utilised within the analysis. Both sets of results are likewise consistent with the world average value [20], exhibiting the expected behaviour given that the Asimov data was produced from MC that was generated and corrected for according to this same world average branching fraction. Similar trends are observed for $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from Asimov data, with both methods of determining the total branching fractions producing results consistent with one another as well as the world average value [20]. It should be noted
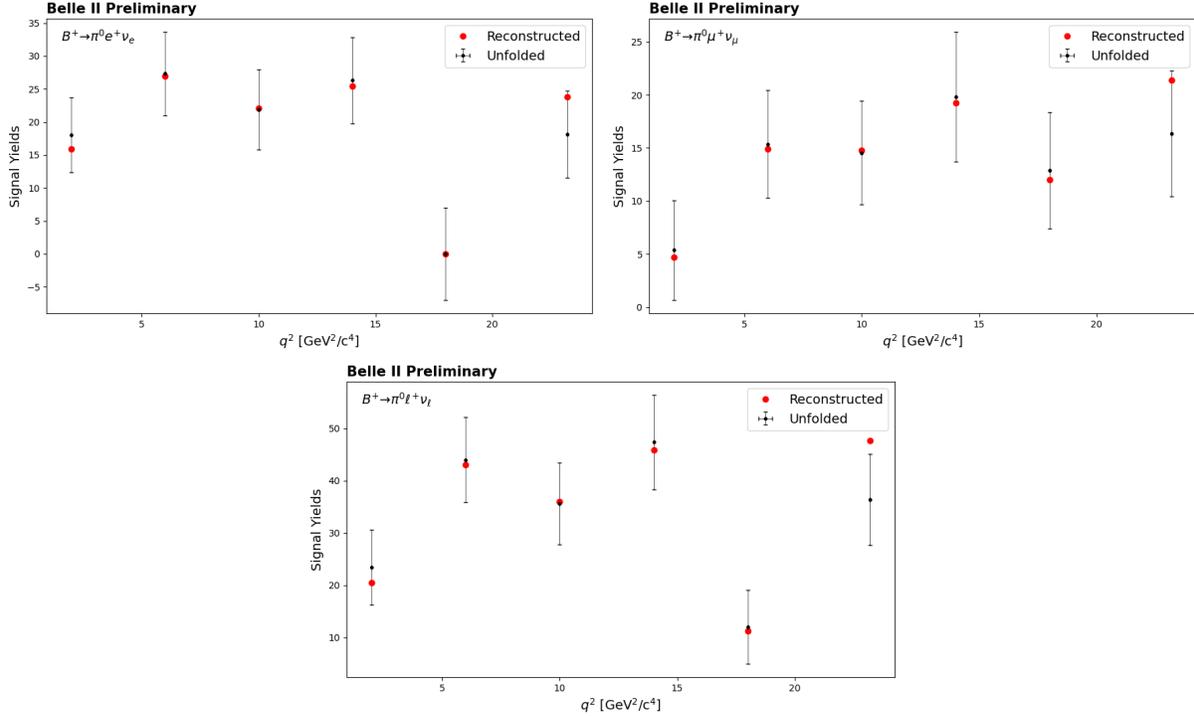
Figure 7.6: The unfolded signal yields derived from $362\,\mathrm{fb}^{-1}$ of Belle II data, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (top left), $B^+ \to \pi^0 \mu^+ \nu_\mu$ (top right), and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ.

that slightly more variation is observed between the sum of the partial branching fractions and the total branching fractions evaluated over the full q^2 range than in the $B^0 \to \pi^- \ell^+ \nu_\ell$ case, although both are still consistent well within the stated uncertainties. No significant discrepancies are observed between the partial branching fractions derived from the electron and muon channels in the reconstruction of both $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays.

Reflecting on the results obtained from Belle II data, it can firstly be observed that the consistency achieved between the two independent measurements of the total branching fractions in Asimov data is mirrored in the real data results, both for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays. However, when considering the world average values, each branching fraction measurement was found to be lower than expected, with a total deviation for the combined lepton measurements of $\sim 2.7\sigma$ for $B^0 \to \pi^- \ell^+ \nu_\ell$ and $\sim 2.0\sigma$ for $B^+ \to \pi^0 \ell^+ \nu_\ell$ taking into account both statistical and systematic uncertainties. These results are altogether unsurprising as per the discussion in Section 6.5, where discrepancies of this order were first observed in the fitted signal yields obtained from data when compared with the expected values derived from MC.

Furthermore, a higher degree of fluctuation between the branching fractions measured from the electron versus the muon channels can be observed in real data with respect to Asimov data, with this behaviour present for both reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$

	Asimov data	Belle II data
q^2 bin (GeV^2)	$\Delta\mathcal{B}_{(i)}$	
$B^0 \to \pi^- e^+ \nu_e$		
$0 \leqslant q^2 < 4$	$(0.26 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.28 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$
$4 \leqslant q^2 < 8$	$(0.27 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.20 \pm 0.06_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$8 \leqslant q^2 < 12$	$(0.27 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.21 \pm 0.06_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$12 \leqslant q^2 < 16$	$(0.25 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.26 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$
$16 \leqslant q^2 < 20$	$(0.21 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.14 \pm 0.06_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$
$20 \leqslant q^2 \leqslant 26.4$	$(0.16 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.17 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$
Sum	$(1.43 \pm 0.17_{\text{stat}} \pm 0.09_{\text{sys}}) \times 10^{-4}$	$(1.24 \pm 0.16_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$
Over full q^2 range	$(1.43 \pm 0.17_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$	$(1.23 \pm 0.15_{\text{stat}} \pm 0.07_{\text{sys}}) \times 10^{-4}$
$B^0 \to \pi^- \mu^+ \nu_\mu$		
$0 \leqslant q^2 < 4$	$(0.27 \pm 0.08_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.22 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$4 \leqslant q^2 < 8$	$(0.27 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.26 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$8 \leqslant q^2 < 12$	$(0.27 \pm 0.07_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.21 \pm 0.06_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$12 \leqslant q^2 < 16$	$(0.25 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.22 \pm 0.06_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$16 \leqslant q^2 < 20$	$(0.21 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.09 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$20 \leqslant q^2 \leqslant 26.4$	$(0.16 \pm 0.07_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.05 \pm 0.04_{\text{stat}} \pm 0.00_{\text{sys}}) \times 10^{-4}$
Sum	$(1.43 \pm 0.17_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$	$(1.05 \pm 0.15_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$
Over full q^2 range	$(1.43 \pm 0.17_{\text{stat}} \pm 0.09_{\text{sys}}) \times 10^{-4}$	$(1.06 \pm 0.15_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$
$B^0 \to \pi^- \ell^+ \nu_\ell$		
$0 \leqslant q^2 < 4$	$(0.27 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.25 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$4 \leqslant q^2 < 8$	$(0.27 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.23 \pm 0.04_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$8 \leqslant q^2 < 12$	$(0.27 \pm 0.05_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.21 \pm 0.04_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$12 \leqslant q^2 < 16$	$(0.25 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.24 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$16 \leqslant q^2 < 20$	$(0.21 \pm 0.05_{\text{stat}} \pm 0.02_{\text{sys}}) \times 10^{-4}$	$(0.11 \pm 0.04_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
$20 \leqslant q^2 \leqslant 26.4$	$(0.16 \pm 0.05_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$	$(0.10 \pm 0.04_{\text{stat}} \pm 0.01_{\text{sys}}) \times 10^{-4}$
Sum	$(1.43 \pm 0.12_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$	$(1.14 \pm 0.11_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$
Over full q^2 range	$(1.43 \pm 0.12_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$	$(1.15 \pm 0.11_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$
World average [20]	$(1.50 \pm 0.06) \times 10^{-4}$	

Table 7.5: The partial branching fractions in six q^2 bins, extracted from $B^0 \to \pi^- \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data. The total branching fractions derived from the sum of the partial branching fractions and the evaluation over the full q^2 kinematic range are also listed.

decays to different extents. In the $B^0 \to \pi^- \ell^+ \nu_\ell$ case, the branching fraction measured for the muon mode, $\mathcal{B}(B^0 \to \pi^- \mu^+ \nu_\mu)$, is smaller than and slightly inconsistent within uncertainties with that for the electron mode, $\mathcal{B}(B^0 \to \pi^- e^+ \nu_e)$. A similar trend is observed for the central values of $\mathcal{B}(B^+ \to \pi^0 e^+ \nu_e)$ and $\mathcal{B}(B^+ \to \pi^0 \mu^+ \nu_\mu)$, though the branching fractions in this

	Asimov data	Belle II data
q^2 bin (GeV^2)	$\Delta\mathcal{B}_{(i)}$	
$B^+ \to \pi^0 e^+ \nu_e$		
$0 \leqslant q^2 < 4$	$(1.62 \pm 0.43_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(1.20 \pm 0.38_{\text{stat}} \pm 0.09_{\text{sys}}) \times 10^{-5}$
$4 \leqslant q^2 < 8$	$(1.51 \pm 0.38_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.60 \pm 0.37_{\text{stat}} \pm 0.14_{\text{sys}}) \times 10^{-5}$
$8 \leqslant q^2 < 12$	$(1.39 \pm 0.36_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.21 \pm 0.33_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$
$12 \leqslant q^2 < 16$	$(1.34 \pm 0.38_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.56 \pm 0.39_{\text{stat}} \pm 0.15_{\text{sys}}) \times 10^{-5}$
$16 \leqslant q^2 < 20$	$(1.20 \pm 0.45_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(0.00 \pm 0.44_{\text{stat}} \pm 0.00_{\text{sys}}) \times 10^{-5}$
$20 \leqslant q^2 \leqslant 26.4$	$(0.68 \pm 0.29_{\text{stat}} \pm 0.07_{\text{sys}}) \times 10^{-5}$	$(0.81 \pm 0.29_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-5}$
Sum	$(7.73 \pm 0.94_{\text{stat}} \pm 0.68_{\text{sys}}) \times 10^{-5}$	$(6.37 \pm 0.91_{\text{stat}} \pm 0.56_{\text{sys}}) \times 10^{-5}$
Over full q^2 range	$(7.63 \pm 0.93_{\text{stat}} \pm 0.65_{\text{sys}}) \times 10^{-5}$	$(6.20 \pm 0.84_{\text{stat}} \pm 0.53_{\text{sys}}) \times 10^{-5}$
$B^+ \to \pi^0 \mu^+ \nu_\mu$		
$0 \leqslant q^2 < 4$	$(1.64 \pm 0.63_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(0.57 \pm 0.50_{\text{stat}} \pm 0.04_{\text{sys}}) \times 10^{-5}$
$4 \leqslant q^2 < 8$	$(1.53 \pm 0.43_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.15 \pm 0.38_{\text{stat}} \pm 0.10_{\text{sys}}) \times 10^{-5}$
$8 \leqslant q^2 < 12$	$(1.38 \pm 0.37_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(0.93 \pm 0.31_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-5}$
$12 \leqslant q^2 < 16$	$(1.34 \pm 0.41_{\text{stat}} \pm 0.14_{\text{sys}}) \times 10^{-5}$	$(1.28 \pm 0.39_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$
$16 \leqslant q^2 < 20$	$(1.21 \pm 0.41_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(0.84 \pm 0.36_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-5}$
$20 \leqslant q^2 \leqslant 26.4$	$(0.68 \pm 0.29_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-5}$	$(0.81 \pm 0.29_{\text{stat}} \pm 0.10_{\text{sys}}) \times 10^{-5}$
Sum	$(7.79 \pm 1.07_{\text{stat}} \pm 0.70_{\text{sys}}) \times 10^{-5}$	$(5.58 \pm 0.93_{\text{stat}} \pm 0.52_{\text{sys}}) \times 10^{-5}$
Over full q^2 range	$(7.63 \pm 1.00_{\text{stat}} \pm 0.67_{\text{sys}}) \times 10^{-5}$	$(5.74 \pm 0.90_{\text{stat}} \pm 0.50_{\text{sys}}) \times 10^{-5}$
$B^+ \to \pi^0 \ell^+ \nu_\ell$		
$0 \leqslant q^2 < 4$	$(1.63 \pm 0.36_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(0.96 \pm 0.29_{\text{stat}} \pm 0.07_{\text{sys}}) \times 10^{-5}$
$4 \leqslant q^2 < 8$	$(1.52 \pm 0.28_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.45 \pm 0.27_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$
$8 \leqslant q^2 < 12$	$(1.38 \pm 0.26_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.05 \pm 0.23_{\text{stat}} \pm 0.10_{\text{sys}}) \times 10^{-5}$
$12 \leqslant q^2 < 16$	$(1.34 \pm 0.28_{\text{stat}} \pm 0.13_{\text{sys}}) \times 10^{-5}$	$(1.46 \pm 0.28_{\text{stat}} \pm 0.14_{\text{sys}}) \times 10^{-5}$
$16 \leqslant q^2 < 20$	$(1.21 \pm 0.30_{\text{stat}} \pm 0.12_{\text{sys}}) \times 10^{-5}$	$(0.39 \pm 0.23_{\text{stat}} \pm 0.04_{\text{sys}}) \times 10^{-5}$
$20 \leqslant q^2 \leqslant 26.4$	$(0.68 \pm 0.20_{\text{stat}} \pm 0.07_{\text{sys}}) \times 10^{-5}$	$(0.85 \pm 0.20_{\text{stat}} \pm 0.09_{\text{sys}}) \times 10^{-5}$
Sum	$(7.76 \pm 0.69_{\text{stat}} \pm 0.68_{\text{sys}}) \times 10^{-5}$	$(6.15 \pm 0.62_{\text{stat}} \pm 0.55_{\text{sys}}) \times 10^{-5}$
Over full q^2 range	$(7.64 \pm 0.68_{\text{stat}} \pm 0.65_{\text{sys}}) \times 10^{-5}$	$(6.17 \pm 0.62_{\text{stat}} \pm 0.53_{\text{sys}}) \times 10^{-5}$
World average [20]	$(7.80 \pm 0.27) \times 10^{-5}$	

Table 7.6: The partial branching fractions in six q^2 bins, extracted from $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data. The total branching fractions derived from the sum of the partial branching fractions and the evaluation over the full q^2 kinematic range are also listed.

case remain consistent with one another within uncertainties. Whilst it is not inconceivable that an unconsidered source of underlying discrepancy may exist between the reconstruction of electrons and muons in real data, given the limited sample size, these observations can likely just be the result of regular statistical fluctuations. As can be seen in the relevant

tables, the branching fractions determined for the combined lepton cases fall between the individual electron and muon results as a natural consequence of these differences. Further discussion on the implications of the low measured $B \to \pi\ell\nu_\ell$ branching fractions and the potential for extended studies will be explored in detail in the concluding chapter.

7.1.3   Systematic Uncertainties

In a similar manner to that first reported in Section 4.1.6 for the $B \to X\ell\nu_\ell$ FEI calibration study, a number of sources of systematic uncertainty were likewise identified for the current $B \to \pi\ell\nu_\ell$ analysis. Each uncertainty was evaluated as a percentage of the given partial or total branching fractions, with each source described in detail below.

f_{+0}

The value of the f_{+0} parameter used as input in the branching fraction calculations was taken from a recent Belle measurement quoting a central value and associated uncertainty of $f_{+0} = 1.065 \pm 0.052$ [93]. To evaluate the effect of this uncertainty on the branching fraction measurements, the relative uncertainty was computed for $B^0 \to \pi^-\ell^+\nu_\ell$ and $B^+ \to \pi^0\ell^+\nu_\ell$ decays according to the contribution of f_{+0} to the respective formulae. For $B^0 \to \pi^-\ell^+\nu_\ell$ decays, this corresponded to a systematic error $\sigma_{\text{sys},f_{+0}}$ of $\sigma_{\text{sys},f_{+0}} = \sigma_{f_{+0}}/(1 + f_{+0}) = 2.52\%$, where $\sigma_{f_{+0}}$ is the given uncertainty on the Belle measurement. For $B^+ \to \pi^0\ell^+\nu_\ell$ decays, due to the presence of both a $(1 + f_{+0})$ term in the numerator and an f_{+0} term in the denominator of the branching fraction formulae, the relative uncertainty was computed as $\sigma_{\text{sys},f_{+0}} = \sigma_{f_{+0}}/(1 + f_{+0})(f_{+0})^{-1} = 2.68\%$.

$N_{B\bar{B}}$

The number of $\Upsilon(4S) \to B\bar{B}$ events in the current data-set, $N_{B\bar{B}} = (387 \pm 6) \times 10^6$, also constituted a source of systematic uncertainty due to the contribution of this parameter to the branching fraction determinations. The magnitude of this uncertainty was similarly evaluated through computing the relative uncertainty on the $N_{B\bar{B}}$ parameter, $\sigma_{\text{sys},N_{B\bar{B}}} = \sigma_{N_{B\bar{B}}}/N_{B\bar{B}} = 1.55\%$.

Tracking

Per official Belle II recommendations as described in Section 4.1.6, an additional systematic uncertainty of 0.24% per reconstructed charged particle track was required to be included. The total systematic uncertainty on the branching fractions due to the track reconstruction efficiency thus corresponded to 0.48% for reconstructed $B^0 \to \pi^-\ell^+\nu_\ell$, and 0.24% for reconstructed $B^+ \to \pi^0\ell^+\nu_\ell$ decays.

Signal Reconstruction Efficiency $\epsilon_{(i)}$

Comprising another element of the branching fraction calculations, the uncertainties on the weighted signal reconstruction efficiencies were also taken into account. The systematic

uncertainties on the partial and total branching fractions were determined by evaluating the relative uncertainties on each respective efficiency measured in q^2 bins or across the full q^2 spectrum, with $\sigma_{\text{sys},\epsilon_{(i)}} = \sigma_{\epsilon_{(i)}}/\epsilon_{(i)}$.

FEI Calibration

The official FEI calibration factors used for this study, as described in Section 5.2, were applied by weighting each individual event according to the decay mode in which the hadronic B_{tag} was reconstructed. Each calibration factor was associated with both a statistical and systematic uncertainty, evaluated in a similar manner to that described for the calibration study presented in Chapter 4. To account for the correlation between the uncertainties of the individual tag decay modes, statistical and systematic covariance matrices were constructed using these uncertainties, and these were summed to build a total covariance matrix. The statistical uncertainties were assumed to be completely independent between tag decay modes, and thus the statistical covariance matrix was constructed as a diagonal matrix with diagonal entries of $\sigma_{\text{stat},j}^2$, where j is the index of the tag decay mode, and off-diagonal entries set to zero. To construct the systematic covariance matrix, all systematic uncertainties on the calibration factors were assumed to be 100% correlated between tag decay modes, resulting in a non-diagonal matrix with entries of $\sigma_{\text{sys},j} \times \sigma_{\text{sys},k}$, where j and k represent the respective decay mode indices. The statistical and systematic covariance matrices were then summed, under the assumption of complete independence between the two, to produce the total uncertainty covariance matrix between the tag modes.

To study the effect of these calibration factor uncertainties on the $B \to \pi \ell \nu_\ell$ branching fractions, a set of alternate event weights was produced using these uncertainties. A multivariate Gaussian distribution was constructed using the total uncertainty matrix, and 200 variations on the nominal calibration factors were sampled from this distribution. Additional toy MC studies were then performed to investigate the magnitude of the effect of these varied weights on the fitted signal yields, where the M_{miss}^2 fits (both in each q^2 bin and over the full q^2 range) were redone using each of these alternate weights in place of the nominal FEI calibration weight. The standard deviation of the distribution of the 200 signal yields returned from these toy fits was then used to determine the size of the total systematic uncertainty $\sigma_{\text{sys, cal.}}$, with $\sigma_{\text{sys, cal.}} = \text{SD}_{\text{sig, cal.}}/N_{\text{sig}}$, where SD_{sig} is the standard deviation of the varied signal yields returned from the fits, and N_{sig} is the nominal signal yield using the standard event weights.

Lepton and Charged Pion Identification

The lepton and pion identification correction factors, the latter applying only to the $B^0 \to \pi^- \ell^+ \nu_\ell$ case, were similarly applied per-event according to the momentum p and polar angle θ of the reconstructed lepton or charged pion. These factors included both the efficiency and fake-rate corrections, and each individual factor likewise possessed both a statistical and systematic uncertainty. To estimate the effect of this on the measured $B \to \pi \ell \nu_\ell$ branching fractions, the same method as described above for the systematic uncertainty due to the FEI calibration was employed. Using the software package in Ref. [80], a total covariance matrix

was built from the statistical and systematic uncertainties between the individual bins, a set of 200 alternate lepton or pion identification weights were sampled from the respective multivariate Gaussian distributions, and toy studies were used to evaluate the effect on the fitted signal yields. The total systematic uncertainties $\sigma_{\text{sys, LID}}$ and $\sigma_{\text{sys, PID}}$ were then computed as $\sigma_{\text{sys, LID}} = \text{SD}_{\text{sig, LID}}/N_{\text{sig}}$ and $\sigma_{\text{sys, PID}} = \text{SD}_{\text{sig, PID}}/N_{\text{sig}}$.

π^0 Reconstruction Efficiency

In the $B^+ \to \pi^0\ell^+\nu_\ell$ case, the systematic uncertainties on the $B \to \pi\ell\nu_\ell$ branching fractions due to the π^0 reconstruction efficiency were likewise evaluated through toy MC studies as above. As mentioned in Section 5.2, the momentum-dependent π^0 efficiency correction factors were associated with a statistical uncertainty as well as two distinct systematic uncertainties, one arising from the factors influencing the D^0 study from which the factors were derived, and one due to the difference observed between the factors obtained from the D^0 study and an alternate 3-prong τ study. Both of these sources of systematic uncertainty were added in quadrature together with the statistical uncertainties to form the total covariance matrix. The final systematic uncertainty $\sigma_{\text{sys},\pi^0}$ was evaluated from the set of 200 varied signal yields as $\sigma_{\text{sys, }\pi^0} = \text{SD}_{\text{sig, }\pi^0}/N_{\text{sig}}$.

Background $B \to X_u\ell\nu_\ell$ and $B \to X_c\ell\nu_\ell$ Branching Fractions

As first discussed in detail in Section 4.1.3, the branching fractions of the various $B \to X_u\ell\nu_\ell$ and $B \to X_c\ell\nu_\ell$ decay modes in the MC were reweighted to updated values, each with a corresponding uncertainty. In considering the individual $B \to X_u\ell\nu_\ell$ and $B \to X_c\ell\nu_\ell$ contributions to the background template in the M_{miss}^2 fits, varying their branching fractions within uncertainties, and thus, varying the size of these contributions, likewise altered the values of the signal yields returned from the fits. The magnitude of these effects was taken into account in a similar manner as for the sources above, where alternate branching fraction correction weights were sampled from Gaussian distributions centered around the nominal branching fractions with the given uncertainties as the widths of the distributions. For each individual background $B \to X_u\ell\nu_\ell$ and $B \to X_c\ell\nu_\ell$ decay mode, 200 alternate weights were sampled in this way and toy fits were performed, with the magnitude of the total systematic uncertainty for each mode, $\sigma_{\text{sys, BF}}$, given as $\sigma_{\text{sys, BF}} = \text{SD}_{\text{sig, BF}}/N_{\text{sig}}$. Systematic uncertainties associated with varying the branching fractions of the signal components ($\mathcal{B}(B^0 \to \pi^-\ell^+\nu_\ell)$ for reconstructed $B^0 \to \pi^-\ell^+\nu_\ell$ and $\mathcal{B}(B^+ \to \pi^0\ell^+\nu_\ell)$ for reconstructed $B^+ \to \pi^0\ell^+\nu_\ell$) were not determined as these were the quantities measured in the analysis.

$B \to X_u\ell\nu_\ell$ and $B \to X_c\ell\nu_\ell$ Form Factors

Another source of systematic uncertainty for the analysis arose from the uncertainties on the form factors of the various $B \to X_u\ell\nu_\ell$ and $B \to X_c\ell\nu_\ell$ modes responsible for the relative shapes of these contributions. As described in Section 4.1.3, the form factors for the $B \to \rho\ell\nu_\ell$, $B^+ \to \omega\ell^+\nu_\ell$, $B^+ \to \eta\ell^+\nu_\ell$, $B^+ \to \eta'\ell^+\nu_\ell$, $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ decays were reweighted using the eFFORT software package [72] by updating the values of the parameters used to parameterise the form factors in the MC generation of each mode. Each updated parameter was associated with a particular uncertainty. The $B \to \pi\ell\nu_\ell$ form factors were

not reweighted due to already possessing up-to-date parameter values, but these parameters likewise possessed associated uncertainties.

To estimate the systematic uncertainties due to the $B \to X_u \ell \nu_\ell$ and $B \to X_c \ell \nu_\ell$ form factors, the effect on the signal yields returned from the M_{miss}^2 fits was similarly evaluated by modifying the shapes of the various $B \to X_u \ell \nu_\ell$ and $B \to X_c \ell \nu_\ell$ decays through varying the form factor parameters within uncertainties. For each decay mode, alternative form factor weights were determined by recalculating the form factors with each parameter modified one-by-one by adding the given uncertainty, termed the up-variation, and subtracting the uncertainty, termed the down-variation, resulting in a number of modified form factor weights for each $B \to X_u \ell \nu_\ell$ and $B \to X_c \ell \nu_\ell$ mode equal to twice the number of parameters used in the MC decay model. In practice, this corresponded to a total of 8 varied weights for $B \to D \ell \nu_\ell$ (from 4 BGL parameters [77]), 12 varied weights for $B \to D^* \ell \nu_\ell$ (from 6 BGL parameters [78]), 38 varied weights for $B \to \rho \ell \nu_\ell$ and $B^+ \to \omega \ell^+ \nu_\ell$ (from 19 BCL parameters [73]), 4 varied weights for $B^+ \to \eta \ell^+ \nu_\ell$ and $B^+ \to \eta' \ell^+ \nu_\ell$ (from 2 DM parameters [75]), and 16 varied weights for $B \to \pi \ell \nu_\ell$ (from 8 BCL parameters [73]). Unlike the systematic uncertainties due to the $B \to \pi \ell \nu_\ell$ branching fractions, which were the target measurement for the analysis, the shape uncertainties due to the $B \to \pi \ell \nu_\ell$ form factors were considered as a source of systematic uncertainty.

For each varied form factor weight, MC toys were likewise generated and new M_{miss}^2 fits were performed. In this case, rather than computing the standard deviation of the resultant signal yields, given the lower number of toys, the maximum variation in the signal yields observed with respect to the nominal yield was taken as a conservative estimate of the uncertainty. The total systematic uncertainties for each decay mode, $\sigma_{\mathrm{sys,\ FF}}$, were thus evaluated as $\sigma_{\mathrm{sys,\ FF}} = \mathrm{Max}(|N_{\mathrm{sig,\ FF},i} - N_{\mathrm{sig}}|)/N_{\mathrm{sig}}$, where $N_{\mathrm{sig,\ FF},i}$ refers to the signal yields returned from the M_{miss}^2 fits for each varied form factor weight i.

Continuum Normalisation

As mentioned in Section 5.2, unlike the FEI calibration study, the continuum MC was unable to be rescaled to the prediction from off-resonance data for the $B \to \pi \ell \nu_\ell$ analysis due to limited statistics. With only a handful of off-resonance events remaining in the relatively small 42.3 fb^{-1} sample after the application of all analysis selections, no valid prediction was able to be made. In order to accommodate for this lack of rescaling, a systematic uncertainty related to the normalisation of the continuum MC background was introduced instead. It was initially decided to assign a flat, conservative uncertainty of 50% on the size of the continuum background as was done in the 2013 Belle analysis [32] that inspired the current study. The systematic uncertainties on the resultant branching fractions would then have been evaluated in a similar manner through toys, with each toy corresponding to a certain normalisation of the continuum background contribution sampled from a Gaussian distribution centered around the nominal normalisation with a standard deviation of 50%.

However, during the peer review process, it was ultimately decided to utilise the available off-resonance data sample to estimate the uncertainty on the continuum normalisation in-
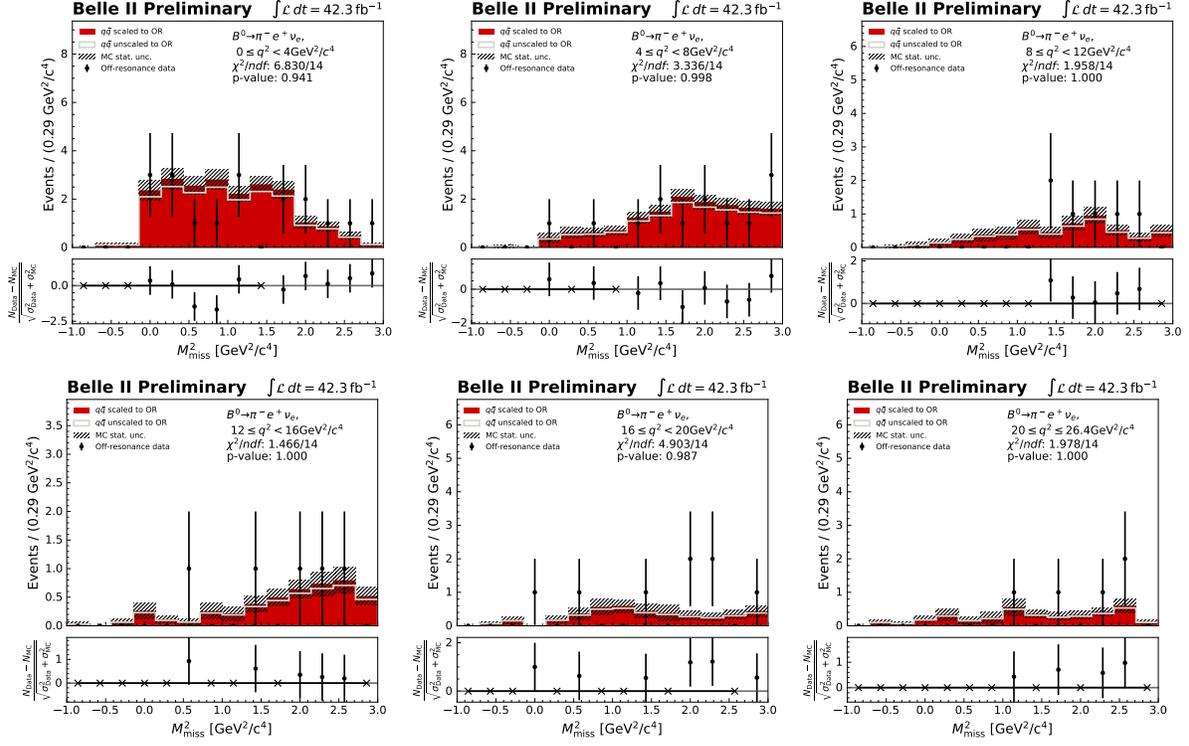
Figure 7.7: M_{miss}^2 distributions of 42.3 fb^{-1} of off-resonance Belle II data against continuum MC in a continuum-enhanced sample, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays in six q^2 bins. Histograms for both continuum MC that has and has not been scaled to the off-resonance prediction are plotted.

stead. Since too few events were present in this off-resonance sample post-selections to even produce such a rough estimate, separate samples of the available off-resonance data and continuum MC were produced with several analysis selections loosened in order to retain a higher proportion of $e^+ e^- \to q\bar{q}$ events. To produce these continuum-enhanced samples, the B_{tag} M_{bc}, ΔE and `cosTBTO` selections listed in Section 5.7 were relaxed to $M_{\mathrm{bc}} > 5.23$ GeV, $|\Delta E| < 0.3$ GeV and cosTBTO < 1.0 for both $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays. In addition, no continuum suppression BDT selection was applied to these samples.

Figures 7.7, 7.8, 7.10 and 7.11 display the M_{miss}^2 distributions of the off-resonance data against continuum MC in the continuum-enhanced samples, with the 1 ab^{-1} continuum MC sample normalised to the data integrated luminosity, for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins, respectively. The equivalent distributions over the full q^2 spectrum are provided in Figures 7.9 and 7.12. In an identical manner to that described in Section 4.1.3, both the continuum MC distributions that have and have not been re-scaled to the off-resonance prediction are included in the plots for illustrative purposes, though the re-scaling was not applied as discussed for the aforementioned reasons.
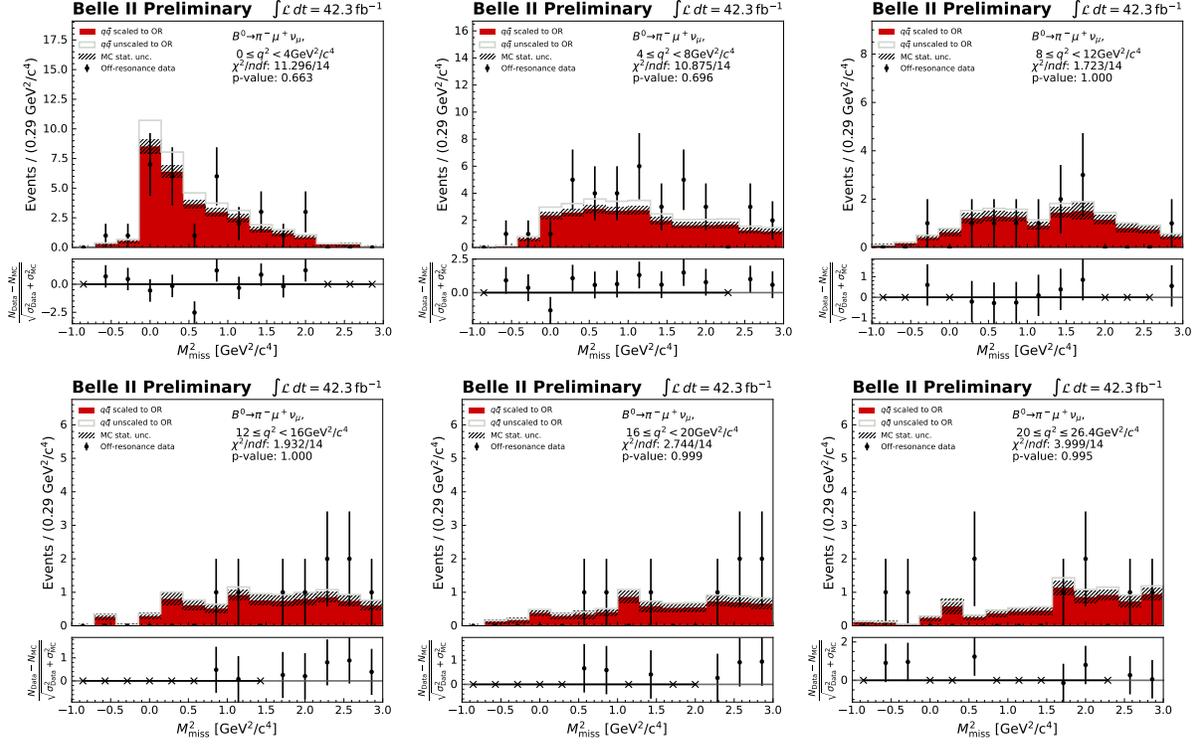
Figure 7.8: M_{miss}^2 distributions of 42.3 fb^{-1} of off-resonance Belle II data against continuum MC in a continuum-enhanced sample, for reconstructed $B^0 \to \pi^- \mu^+ \nu_\mu$ decays in six q^2 bins. Histograms for both continuum MC that has and has not been scaled to the off-resonance prediction are plotted.

As can be observed, the number of off-resonance events surviving these loosened selections was still fairly minimal, particularly when splitting the M_{miss}^2 distributions into q^2 bins, though this number was substantial enough to at least roughly compare the event yields in the off-resonance data and continuum MC samples. As a result, it was decided to use the fractional difference between the number of off-resonance and continuum events in each distribution to represent the total uncertainty on the continuum normalisation in lieu of the conservative 50% estimate. This decision did rely, however, on the assumption that the magnitude of the differences between the continuum MC and off-resonance data made using the continuum-enhanced samples would also reflect the reality under the nominal analysis selections. Given that the data-MC agreement in Figures 7.7 – 7.12 was observed to be quite reasonable, this assumption was also taken to be relatively sound.

For each reconstructed $B \to \pi \ell \nu_\ell$ mode, evaluated both for each individual q^2 bin and over the full q^2 range, the fractional difference between the number of events in the off-resonance data $N_{\text{off-res.}}$ and continuum MC $N_{q\bar{q}}$ in the continuum-enhanced samples was computed as $|N_{q\bar{q}} - N_{\text{off-res.}}|/N_{\text{off-res.}}$. The resultant values are summarised in Table 7.7. For those entries in which the observed difference was < 1%, the values were rounded up to a 1% minimum threshold. Gaussian distributions were subsequently defined for each
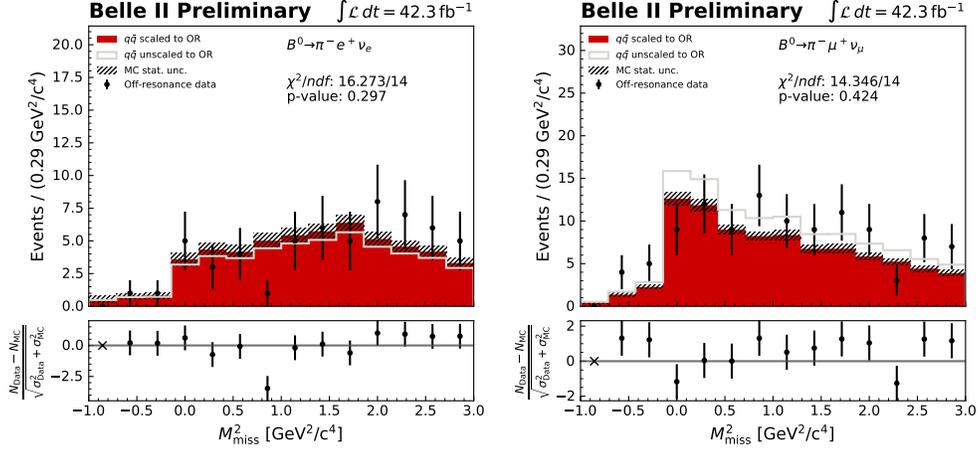
Figure 7.9: M_{miss}^2 distributions of 42.3 fb^{-1} of off-resonance Belle II data against continuum MC in a continuum-enhanced sample, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ (left) and $B^0 \to \pi^- \mu^+ \nu_\mu$ (right) decays, over the full q^2 kinematic range. Histograms for both continuum MC that has and has not been scaled to the off-resonance prediction are plotted.

case, centered around the nominal continuum normalisation with widths corresponding to the given uncertainties, and 200 variations on the normalisation were sampled from these Gaussian distributions for each case. In a similar manner as for the other sources, the systematic uncertainties due to the continuum normalisation, $\sigma_{\mathrm{sys},\, q\bar{q}}$, were determined from the standard deviation of the signal yields resulting from M_{miss}^2 fits to these new toys, with $\sigma_{\mathrm{sys},\, q\bar{q}} = \mathrm{SD}_{\mathrm{sig},\, q\bar{q}}/N_{\mathrm{sig}}$. It should be noted that this systematic uncertainty only covered effects due to the relative normalisation of the continuum MC background, and does not provide any insight into potential additional sources of uncertainty due to the shape of this component with respect to the off-resonance data. Due to limited statistics and the inability to confidently make a similar assumption of equivalence between the shapes of the continuum MC in the continuum-enhanced and nominal analysis samples, no such uncertainty was included in the analysis. However, as stated prior, no significant discrepancies were observed in the data-MC agreement in the continuum-enhanced samples to create any particular cause of concern regarding these potential shape uncertainties.

Total Systematic Uncertainties

Tables 7.8, 7.9, 7.11 and 7.12 list the systematic uncertainties due to each of the above sources as percentages of the measured partial and total branching fractions, for reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays, respectively. For ease of communication, several branching fraction and form factor uncertainties have been combined into distinct categories by adding the uncertainties in quadrature, assuming all sources of systematic uncertainty to be completely independent from one another. All $B \to \rho \ell \nu_\ell$, non-resonant $B \to X_u \ell \nu_\ell$, $B \to D \ell \nu_\ell$ and $B \to D^* \ell \nu_\ell$ entries contain both charged and neutral B-meson decay contributions, such as the $B \to \rho \ell \nu_\ell$ branching fraction category which was formed from the combination of the $B^+ \to \rho^0 \ell^+ \nu_\ell$ and $B^0 \to \rho^- \ell^+ \nu_\ell$

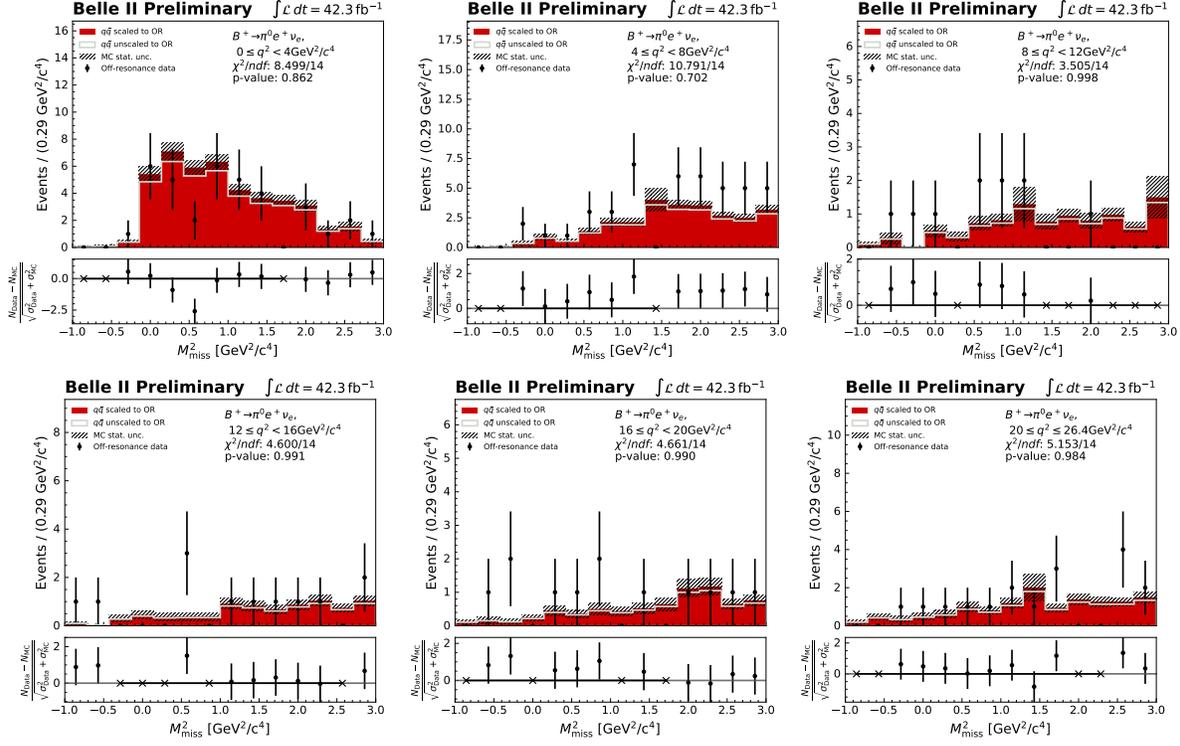Figure 7.10: M_{miss}^2 distributions of 42.3 fb^{-1} of off-resonance Belle II data against continuum MC in a continuum-enhanced sample, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays in six q^2 bins. Histograms for both continuum MC that has and has not been scaled to the off-resonance prediction are plotted.

branching fraction uncertainties. The $B \to \pi \ell \nu_\ell$ form factor category likewise contains contributions from both $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^0 \to \pi^- \ell^+ \nu_\ell$, though the corresponding branching fraction uncertainties are limited to the single background $B^+ \to \pi^0 \ell^+ \nu_\ell$ contribution for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$, and vice versa. The $D^{(*)}\pi\pi\ell\nu_\ell$ branching fraction entries are comprised of all resonant $B \to D^{**}\ell\nu_\ell$ contributions as described in Section 4.1.3, namely those from $B^0 \to D_1^- \ell^+ \nu_\ell$, $B^+ \to D_1^0 \ell^+ \nu_\ell$, $B^0 \to D_0^{*-} \ell^+ \nu_\ell$, $B^+ \to D_0^{*0} \ell^+ \nu_\ell$, $B^0 \to D_1'^- \ell^+ \nu_\ell$, $B^+ \to D_1'^0 \ell^+ \nu_\ell$, $B^0 \to D_2^{*-} \ell^+ \nu_\ell$, $B^+ \to D_2^{*0} \ell^+ \nu_\ell$, $B^0 \to D\pi\pi\ell^+\nu_\ell$, $B^+ \to D\pi\pi\ell^+\nu_\ell$, $B^0 \to D^*\pi\pi\ell^+\nu_\ell$ and $B^+ \to D^*\pi\pi\ell^+\nu_\ell$ decays. All uncertainties due to the branching fractions of the $B \to X_c \ell \nu_\ell$ gap modes are likewise formed from the combination of the individual uncertainties of the $B^0 \to D^- \eta \ell^+ \nu_\ell$, $B^+ \to \bar{D}^0 \eta \ell^+ \nu_\ell$, $B^0 \to D^{*-} \eta \ell^+ \nu_\ell$, $B^+ \to \bar{D}^{*0} \eta \ell^+ \nu_\ell$, $B^+ \to D_s^- K^+ \ell^+ \nu_\ell$ and $B^+ \to D_s^{*-} K^+ \ell^+ \nu_\ell$ decay contributions.

Tables 7.10 and 7.13 summarise the equivalent uncertainties obtained for the combined lepton cases, $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$, respectively. To obtain the values listed in these tables, the individual uncertainties from the electron and muon channels were combined with all correlations taken into account. The lepton identification systematic uncertainties were considered to be completely independent between the two channels given the distinct correction factors determined for the two lepton flavours, and these were there-
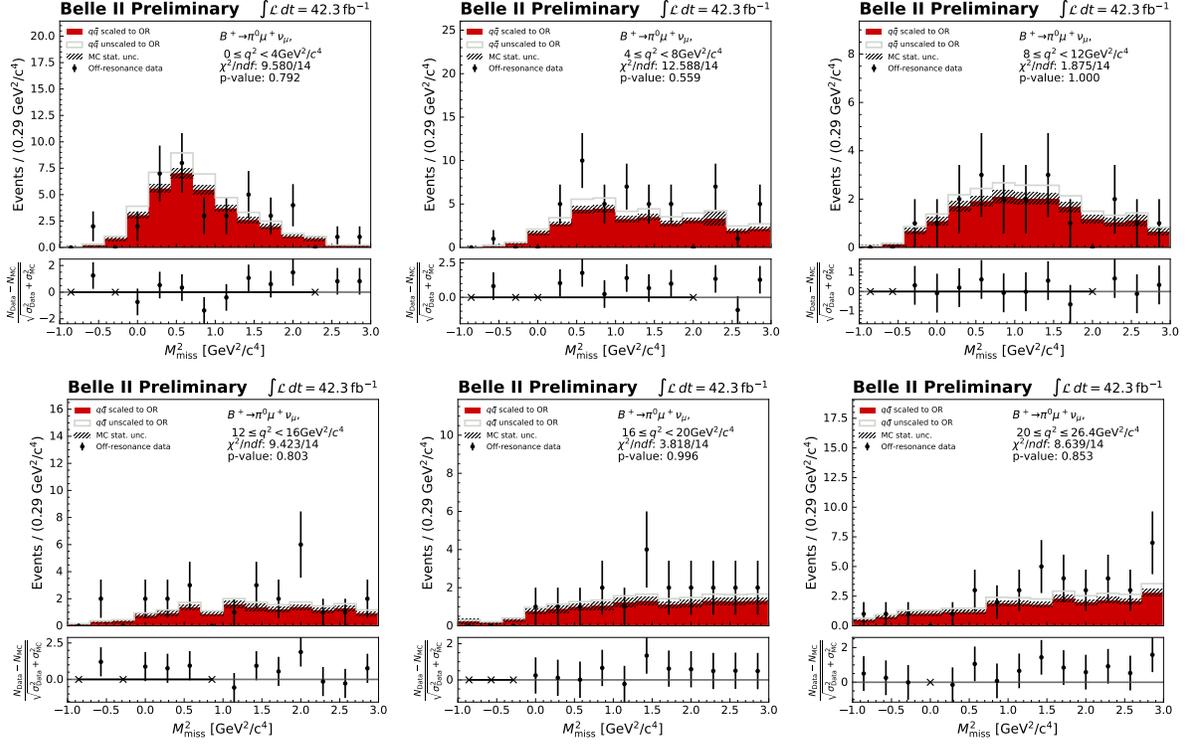
Figure 7.11: M_{miss}^2 distributions of 42.3 fb^{-1} of off-resonance Belle II data against continuum MC in a continuum-enhanced sample, for reconstructed $B^+ \to \pi^0 \mu^+ \nu_\mu$ decays in six q^2 bins. Histograms for both continuum MC that has and has not been scaled to the off-resonance prediction are plotted.

fore combined in quadrature and halved to produce the resultant systematic uncertainty for the combined lepton case. All other systematic uncertainties were taken to be completely correlated between the two lepton channels, with the systematic uncertainties on the branching fractions averaged over the two lepton flavours, $\sigma_{\mathrm{sys},\,e/\mu}$, computed as $\sigma_{\mathrm{sys},\,e/\mu} = \sqrt{\sigma_{\mathrm{sys},\,e}^2 + \sigma_{\mathrm{sys},\,\mu}^2 + 2 \times \sigma_{\mathrm{sys},\,e} \times \sigma_{\mathrm{sys},\,\mu}}/2$, where $\sigma_{\mathrm{sys},\,e}$ and $\sigma_{\mathrm{sys},\,\mu}$ are the systematic uncertainties corresponding to the electron and muon channels, respectively.

For each reconstructed $B \to \pi \ell \nu_\ell$ mode, the total systematic uncertainties were subsequently computed by combining all of the individual sources in quadrature, once again assuming complete independence between each source. The respective total systematic uncertainties are also listed in Tables 7.8 – 7.13. The uncertainties due to the FEI calibration and, in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ case, the π^0 reconstruction efficiency, form the dominant contributions to the total systematic uncertainties. To compare the magnitude of the total systematic and statistical uncertainties on the measured branching fractions derived from real data, the statistical uncertainties as measured in the previous section are also included in the given tables, together with the total uncertainties formed from the combination of the systematic and statistical uncertainties in quadrature. It is clear that at the current data integrated luminosity, the statistical uncertainties dominate over the systematic uncertainties in all cases.
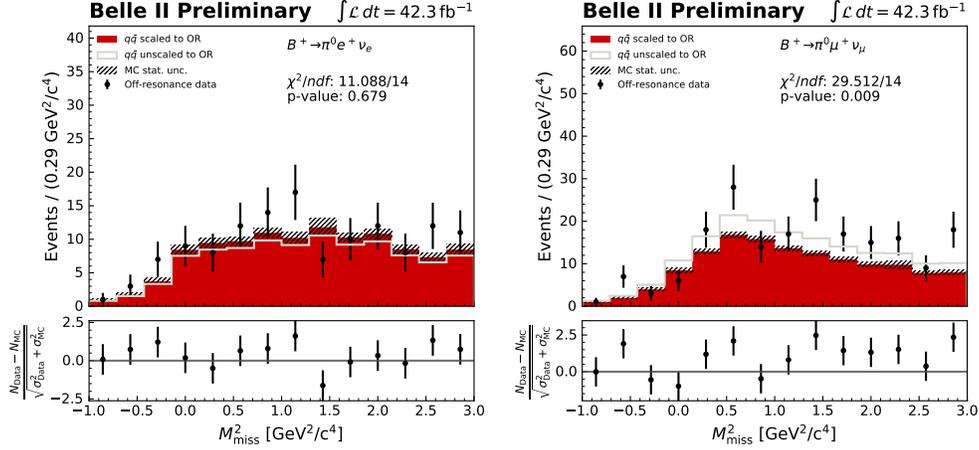
Figure 7.12: M_{miss}^2 distributions of $42.3\,\mathrm{fb}^{-1}$ of off-resonance Belle II data against continuum MC in a continuum-enhanced sample, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ (left) and $B^+ \to \pi^0 \mu^+ \nu_\mu$ (right) decays, over the full q^2 kinematic range. Histograms for both continuum MC that has and has not been scaled to the off-resonance prediction are plotted.

Decay mode	$|N_{q\bar{q}} - N_{\mathrm{off\text{-}res.}}|/N_{\mathrm{off\text{-}res.}}$ (%)						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
$B^0 \to \pi^- e^+ \nu_e$	1	4	20	19	54	33	16
$B^0 \to \pi^- \mu^+ \nu_\mu$	18	23	37	11	1	1	1
$B^+ \to \pi^0 e^+ \nu_e$	7	45	14	40	46	29	23
$B^+ \to \pi^0 \mu^+ \nu_\mu$	6	16	16	31	17	25	12

Table 7.7: Percentage difference between the total number of off-resonance data and normalised continuum MC events in continuum-enhanced samples.

This behaviour is particularly enhanced when considering the branching fractions derived from the individual lepton channels as opposed to the combined lepton cases, with the former evaluated using roughly half the statistics of the latter.

Following the detailed discussion on the measured branching fractions and their uncertainties above, the remainder of the current chapter will demonstrate the extraction of the final and most significant result of the given analysis, the magnitude of the CKM-matrix element $|V_{\mathrm{ub}}|$, using the distributions of the partial branching fractions obtained for the various $B \to \pi \ell \nu_\ell$ modes.

Source	% of $\Delta\mathcal{B}_{(i)}(B^0 \to \pi^- e^+ \nu_e)$						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
f_{+0}	2.52						
$N_{B\bar{B}}$	1.55						
Tracking	0.48						
Sig. recon. eff. ϵ_i	1.93	1.84	1.78	1.83	2.02	2.55	0.80
FEI calibration	3.98	4.12	4.45	4.92	5.24	5.26	4.55
Lepton ID	0.45	0.54	1.62	0.39	2.35	5.88	1.05
Pion ID	0.55	0.46	0.42	0.39	0.37	0.21	0.40
$\rho\ell\nu_\ell$ BF	0.17	0.47	0.73	1.47	2.01	2.46	0.87
$\omega\ell\nu_\ell$ BF	0.00	0.00	0.01	0.00	0.03	0.05	0.00
$\eta\ell\nu_\ell$ BF	0.00	0.00	0.00	0.02	0.01	0.06	0.00
$\eta'\ell\nu_\ell$ BF	0.00	0.00	0.02	0.00	0.02	0.05	0.00
$\pi^0\ell\nu_\ell$ BF	0.00	0.00	0.01	0.00	0.03	0.06	0.01
Non-res. $X_u\ell\nu_\ell$ BF	0.49	0.76	0.70	0.79	0.55	0.43	0.64
$D\ell\nu_\ell$ BF	0.05	0.05	0.25	0.25	0.34	0.20	0.22
$D^*\ell\nu_\ell$ BF	0.00	0.03	0.15	0.32	0.62	0.51	0.22
$D^{(*)}\pi\pi\ell\nu_\ell$ BF	0.01	0.08	0.11	0.11	0.18	0.23	0.07
$X_c\ell\nu_\ell$ gap BF	0.00	0.11	0.23	0.21	0.19	0.08	0.11
$\rho\ell\nu_\ell$ FF	0.22	0.20	0.03	0.02	0.14	0.13	0.03
$\omega\ell\nu_\ell$ FF	0.16	0.14	0.01	0.00	0.04	0.07	0.01
$\eta\ell\nu_\ell$ FF	0.16	0.14	0.01	0.00	0.05	0.04	0.01
$\eta'\ell\nu_\ell$ FF	0.16	0.14	0.02	0.00	0.04	0.01	0.01
$\pi\ell\nu_\ell$ FF	0.22	0.20	0.01	0.00	0.05	0.10	0.02
$D\ell\nu_\ell$ FF	0.24	0.22	0.04	0.06	0.05	0.07	0.01
$D^*\ell\nu_\ell$ FF	0.22	0.19	0.09	0.15	0.20	0.14	0.09
$q\bar{q}$ norm.	0.00	0.10	0.44	0.30	8.97	0.67	0.51
Total sys.	5.44	5.56	6.01	6.32	11.47	9.21	5.76
Stat.	25.00	30.00	28.57	26.92	42.86	41.18	12.20
Total	25.6	30.5	29.2	27.7	44.4	42.2	13.5

Table 7.8: Sources of uncertainty and their percentages of the measured partial and total branching fractions, for reconstructed $B^0 \to \pi^- e^+ \nu_e$ decays. The statistical and total uncertainties listed are calculated with respect to the branching fractions derived from 362 fb^{-1} of Belle II data.

Source	% of $\Delta\mathcal{B}_{(i)}(B^0 \to \pi^-\mu^+\nu_\mu)$						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
f_{+0}	2.52						
$N_{B\bar{B}}$	1.55						
Tracking	0.48						
Sig. recon. eff. ϵ_i	2.10	1.92	1.86	1.91	2.11	2.60	0.84
FEI calibration	3.37	3.49	3.74	3.83	3.96	4.00	3.64
Lepton ID	1.78	0.80	1.54	0.73	1.22	3.09	3.65
Pion ID	0.56	0.46	0.33	0.33	0.24	0.16	0.36
$\rho\ell\nu_\ell$ BF	0.21	0.34	0.59	1.11	1.42	1.87	0.64
$\omega\ell\nu_\ell$ BF	0.00	0.00	0.00	0.01	0.01	0.00	0.01
$\eta\ell\nu_\ell$ BF	0.00	0.00	0.00	0.03	0.01	0.02	0.00
$\eta'\ell\nu_\ell$ BF	0.00	0.00	0.00	0.01	0.01	0.01	0.00
$\pi^0\ell\nu_\ell$ BF	0.00	0.00	0.00	0.00	0.01	0.04	0.01
Non-res. $X_u\ell\nu_\ell$ BF	0.44	0.44	0.24	0.75	0.54	0.26	0.38
$D\ell\nu_\ell$ BF	0.13	0.15	0.20	0.20	0.28	0.10	0.23
$D^*\ell\nu_\ell$ BF	0.01	0.05	0.11	0.26	0.48	0.39	0.24
$D^{(*)}\pi\pi\ell\nu_\ell$ BF	0.13	0.03	0.12	0.09	0.10	0.32	0.05
$X_c\ell\nu_\ell$ gap BF	0.00	0.09	0.08	0.13	0.03	0.02	0.06
$\rho\ell\nu_\ell$ FF	0.23	0.00	0.01	0.02	0.02	0.01	0.07
$\omega\ell\nu_\ell$ FF	0.16	0.00	0.01	0.00	0.00	0.01	0.04
$\eta\ell\nu_\ell$ FF	0.16	0.00	0.01	0.00	0.01	0.01	0.05
$\eta'\ell\nu_\ell$ FF	0.16	0.00	0.01	0.00	0.00	0.01	0.04
$\pi\ell\nu_\ell$ FF	0.23	0.01	0.01	0.00	0.00	0.01	0.06
$D\ell\nu_\ell$ FF	0.51	0.02	0.04	0.05	0.06	0.03	0.04
$D^*\ell\nu_\ell$ FF	0.23	0.03	0.09	0.12	0.17	0.13	0.06
$q\bar{q}$ norm.	2.64	1.18	2.32	0.01	0.04	0.01	0.02
Total sys.	6.00	5.24	5.90	5.47	5.77	6.72	6.08
Stat.	31.82	26.92	28.57	27.27	55.56	80.00	14.15
Total	32.3	27.4	29.2	27.8	55.9	80.3	15.4

Table 7.9: Sources of uncertainty and their percentages of the measured partial and total branching fractions, for reconstructed $B^0 \to \pi^-\mu^+\nu_\mu$ decays. The statistical and total uncertainties listed are calculated with respect to the branching fractions derived from 362 fb^{-1} of Belle II data.

Source	% of $\Delta\mathcal{B}_{(i)}(B^0 \to \pi^- \ell^+ \nu_\ell)$						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
f_{+0}	2.52						
$N_{B\bar{B}}$	1.55						
Tracking	0.48						
Sig. recon. eff. ϵ_i	1.42	1.33	1.29	1.32	1.46	1.82	0.58
FEI calibration	3.68	3.81	4.10	4.38	4.60	4.63	4.09
Lepton ID	0.92	0.48	1.12	0.41	1.32	3.32	1.90
Pion ID	0.56	0.46	0.38	0.36	0.31	0.19	0.38
$\rho\ell\nu_\ell$ BF	0.19	0.40	0.66	1.29	1.71	2.17	0.75
$\omega\ell\nu_\ell$ BF	0.00	0.00	0.01	0.01	0.02	0.03	0.00
$\eta\ell\nu_\ell$ BF	0.00	0.00	0.00	0.02	0.01	0.04	0.00
$\eta'\ell\nu_\ell$ BF	0.00	0.00	0.01	0.01	0.02	0.03	0.00
$\pi^0\ell\nu_\ell$ BF	0.00	0.00	0.01	0.00	0.02	0.05	0.01
Non-res. $X_u\ell\nu_\ell$ BF	0.46	0.60	0.47	0.77	0.55	0.35	0.51
$D\ell\nu_\ell$ BF	0.09	0.10	0.23	0.23	0.31	0.15	0.22
$D^*\ell\nu_\ell$ BF	0.01	0.04	0.13	0.29	0.55	0.45	0.23
$D^{(*)}\pi\pi\ell\nu_\ell$ BF	0.07	0.05	0.11	0.10	0.14	0.28	0.06
$X_c\ell\nu_\ell$ gap BF	0.00	0.10	0.16	0.17	0.11	0.05	0.09
$\rho\ell\nu_\ell$ FF	0.22	0.10	0.02	0.02	0.08	0.07	0.05
$\omega\ell\nu_\ell$ FF	0.16	0.07	0.01	0.00	0.02	0.04	0.03
$\eta\ell\nu_\ell$ FF	0.16	0.07	0.01	0.00	0.03	0.03	0.03
$\eta'\ell\nu_\ell$ FF	0.16	0.07	0.01	0.00	0.03	0.03	0.03
$\pi\ell\nu_\ell$ FF	0.22	0.11	0.01	0.00	0.03	0.05	0.04
$D\ell\nu_\ell$ FF	0.38	0.12	0.04	0.05	0.05	0.05	0.02
$D^*\ell\nu_\ell$ FF	0.22	0.11	0.09	0.14	0.19	0.14	0.08
$q\bar{q}$ norm.	1.32	0.64	1.38	0.16	4.51	0.34	0.26
Total sys.	5.29	5.17	5.61	5.71	7.62	7.07	5.55
Stat.	20.00	17.39	19.05	20.83	36.36	40.00	9.57
Total	20.7	18.1	19.9	21.6	37.1	40.6	11.1

Table 7.10: Sources of uncertainty and their percentages of the measured partial and total branching fractions, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ. The statistical and total uncertainties listed are calculated with respect to the branching fractions derived from 362 fb^{-1} of Belle II data.

Source	% of $\Delta \mathcal{B}_{(i)}(B^+ \to \pi^0 e^+ \nu_e)$						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
f_{+0}	2.68						
$N_{B\bar{B}}$	1.55						
Tracking	0.24						
Sig. recon. eff. ϵ_i	1.83	1.68	1.68	1.81	2.00	1.90	0.74
FEI calibration	4.07	4.39	4.81	4.79	5.01	6.04	4.73
Lepton ID	0.95	0.42	2.99	0.57	0.66	1.48	0.75
π^0 efficiency	4.52	5.90	6.67	6.76	6.68	6.58	5.87
$\rho\ell\nu_\ell$ BF	0.10	0.20	0.26	0.26	0.41	0.49	0.29
$\omega\ell\nu_\ell$ BF	0.01	0.05	0.14	0.13	0.25	0.65	0.14
$\eta\ell\nu_\ell$ BF	0.02	0.03	0.08	0.11	0.16	0.52	0.09
$\eta'\ell\nu_\ell$ BF	0.02	0.01	0.05	0.04	0.04	0.05	0.01
$\pi^-\ell\nu_\ell$ BF	0.00	0.00	0.04	0.07	0.25	2.52	0.32
Non-res. $X_u\ell\nu_\ell$ BF	0.67	1.74	1.83	2.19	2.35	2.37	1.92
$D\ell\nu_\ell$ BF	0.13	0.14	0.17	0.11	0.15	0.14	0.09
$D^*\ell\nu_\ell$ BF	0.03	0.05	0.25	0.34	0.56	0.65	0.42
$D^{(*)}\pi\pi\ell\nu_\ell$ BF	0.04	0.47	0.21	0.20	0.29	0.34	0.16
$X_c\ell\nu_\ell$ gap BF	0.00	0.08	0.20	0.15	0.25	0.19	0.12
$\rho\ell\nu_\ell$ FF	0.00	0.01	0.10	0.03	0.02	0.01	0.10
$\omega\ell\nu_\ell$ FF	0.00	0.00	0.13	0.01	0.15	0.05	0.06
$\eta\ell\nu_\ell$ FF	0.01	0.01	0.01	0.06	0.00	0.50	0.03
$\eta'\ell\nu_\ell$ FF	0.00	0.01	0.12	0.02	0.00	0.12	0.07
$\pi\ell\nu_\ell$ FF	0.00	0.01	0.14	0.04	0.06	0.20	0.10
$D\ell\nu_\ell$ FF	0.10	0.07	0.03	0.05	0.05	0.04	0.09
$D^*\ell\nu_\ell$ FF	0.02	0.05	0.20	0.22	0.37	0.34	0.10
$q\bar{q}$ norm.	0.05	2.01	0.54	0.70	3.02	1.05	0.84
Total sys.	7.17	8.61	9.65	9.36	9.97	10.50	8.51
Stat.	31.67	23.13	27.27	25.00	–	35.80	13.55
Total	32.5	24.7	28.9	26.7	–	37.3	16.0

Table 7.11: Sources of uncertainty and their percentages of the measured partial and total branching fractions, for reconstructed $B^+ \to \pi^0 e^+ \nu_e$ decays. The statistical and total uncertainties listed are calculated with respect to the branching fractions derived from 362 fb^{-1} of Belle II data.

Source	% of $\Delta\mathcal{B}_{(i)}(B^+ \to \pi^0\mu^+\nu_\mu)$						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
f_{+0}	2.68						
$N_{B\bar{B}}$	1.55						
Tracking	0.24						
Sig. recon. eff. ϵ_i	2.30	1.91	1.81	1.89	2.04	1.99	0.81
FEI calibration	4.35	4.19	4.33	4.86	4.58	5.39	4.50
Lepton ID	0.89	0.86	0.85	1.81	2.04	5.62	1.46
π^0 efficiency	4.65	5.97	6.60	6.92	7.16	7.21	6.24
$\rho\ell\nu_\ell$ BF	0.07	0.24	0.22	0.24	0.33	0.44	0.28
$\omega\ell\nu_\ell$ BF	0.01	0.03	0.08	0.16	0.20	0.28	0.12
$\eta\ell\nu_\ell$ BF	0.03	0.03	0.06	0.09	0.18	0.45	0.10
$\eta'\ell\nu_\ell$ BF	0.00	0.01	0.01	0.03	0.09	0.09	0.00
$\pi^-\ell\nu_\ell$ BF	0.00	0.00	0.01	0.08	0.33	2.39	0.36
Non-res. $X_u\ell\nu_\ell$ BF	0.34	1.31	1.28	1.75	1.74	1.91	1.53
$D\ell\nu_\ell$ BF	0.35	0.19	0.19	0.11	0.18	0.19	0.12
$D^*\ell\nu_\ell$ BF	0.02	0.12	0.28	0.45	0.40	0.72	0.48
$D^{(*)}\pi\pi\ell\nu_\ell$ BF	0.06	0.06	0.12	0.19	0.35	0.31	0.07
$X_c\ell\nu_\ell$ gap BF	0.00	0.37	0.17	0.25	0.31	0.40	0.26
$\rho\ell\nu_\ell$ FF	0.32	0.05	0.00	0.08	0.06	0.58	0.16
$\omega\ell\nu_\ell$ FF	0.25	0.03	0.00	0.07	0.13	0.36	0.12
$\eta\ell\nu_\ell$ FF	0.25	0.02	0.00	0.03	0.12	1.00	0.04
$\eta'\ell\nu_\ell$ FF	0.23	0.04	0.00	0.06	0.03	0.45	0.11
$\pi\ell\nu_\ell$ FF	0.32	0.05	0.00	0.08	0.08	0.61	0.16
$D\ell\nu_\ell$ FF	0.52	0.11	0.05	0.05	0.07	0.63	0.13
$D^*\ell\nu_\ell$ FF	0.31	0.03	0.16	0.19	0.19	0.89	0.14
$q\bar{q}$ norm.	0.69	0.47	1.34	3.86	0.99	3.34	1.27
Total sys.	7.60	8.33	8.92	10.32	9.75	12.30	8.73
Stat.	87.72	33.04	33.33	30.47	42.86	35.80	15.68
Total	88.1	34.1	34.5	32.2	44.0	37.9	17.9

Table 7.12: Sources of uncertainty and their percentages of the measured partial and total branching fractions, for reconstructed $B^+ \to \pi^0\mu^+\nu_\mu$ decays. The statistical and total uncertainties listed are calculated with respect to the branching fractions derived from 362 fb^{-1} of Belle II data.

Source	% of $\Delta\mathcal{B}_{(i)}(B^+ \to \pi^0\ell^+\nu_\ell)$						
	q_1	q_2	q_3	q_4	q_5	q_6	full q^2 range
f_{+0}	2.68						
$N_{B\bar{B}}$	1.55						
Tracking	0.24						
Sig. recon. eff. ϵ_i	1.43	1.26	1.23	1.31	1.43	1.37	0.54
FEI calibration	4.21	4.29	4.57	4.83	4.80	5.71	4.61
Lepton ID	0.65	0.48	1.55	0.95	1.07	2.91	0.82
π^0 efficiency	4.59	5.94	6.64	6.84	6.92	6.90	6.06
$\rho\ell\nu_\ell$ BF	0.09	0.22	0.24	0.25	0.37	0.46	0.28
$\omega\ell\nu_\ell$ BF	0.01	0.04	0.11	0.15	0.23	0.47	0.13
$\eta\ell\nu_\ell$ BF	0.02	0.03	0.07	0.10	0.17	0.49	0.09
$\eta'\ell\nu_\ell$ BF	0.01	0.01	0.03	0.03	0.07	0.07	0.01
$\pi^-\ell\nu_\ell$ BF	0.00	0.00	0.03	0.08	0.29	2.46	0.34
Non-res. $X_u\ell\nu_\ell$ BF	0.51	1.53	1.56	1.97	2.05	2.14	1.72
$D\ell\nu_\ell$ BF	0.24	0.17	0.18	0.11	0.17	0.17	0.10
$D^*\ell\nu_\ell$ BF	0.02	0.08	0.27	0.40	0.48	0.69	0.45
$D^{(*)}\pi\pi\ell\nu_\ell$ BF	0.05	0.26	0.17	0.19	0.32	0.33	0.11
$X_c\ell\nu_\ell$ gap BF	0.00	0.23	0.19	0.20	0.28	0.30	0.19
$\rho\ell\nu_\ell$ FF	0.16	0.03	0.05	0.05	0.04	0.30	0.13
$\omega\ell\nu_\ell$ FF	0.13	0.02	0.07	0.04	0.14	0.21	0.09
$\eta\ell\nu_\ell$ FF	0.13	0.02	0.01	0.05	0.06	0.75	0.03
$\eta'\ell\nu_\ell$ FF	0.11	0.03	0.06	0.04	0.02	0.29	0.09
$\pi\ell\nu_\ell$ FF	0.16	0.03	0.07	0.06	0.07	0.40	0.13
$D\ell\nu_\ell$ FF	0.31	0.09	0.04	0.05	0.06	0.34	0.11
$D^*\ell\nu_\ell$ FF	0.17	0.04	0.18	0.21	0.28	0.62	0.12
$q\bar{q}$ norm.	0.37	1.24	0.94	2.28	2.01	2.20	1.06
Total sys.	7.18	8.32	9.06	9.58	9.63	10.88	8.56
Stat.	30.21	18.62	21.90	19.18	58.97	23.53	10.05
Total	31.1	20.4	23.7	21.4	59.8	25.9	13.2

Table 7.13: Sources of uncertainty and their percentages of the measured partial and total branching fractions, for reconstructed $B^+ \to \pi^0\ell^+\nu_\ell$ decays, where $\ell = e$ or μ. The statistical and total uncertainties listed are calculated with respect to the branching fractions derived from 362 fb^{-1} of Belle II data.

7.2 $|V_{\mathrm{ub}}|$ Extraction

In the introduction to the semi-leptonic $B \to \pi \ell \nu_\ell$ decay given in Section 1.2.3, the differential decay rate as a function of q^2 was defined as the probability of the decay occurring per unit time. The determination of this quantity is dependent on both $|V_{\mathrm{ub}}|$ and the vector form factor $f_+(q^2)$, with the relevant formula given in Section 1.2.3 reproduced here for convenience:

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} = \frac{G_F^2}{24\pi^3}|V_{\mathrm{ub}}|^2|f_+(q^2)|^2|\vec{p}_\pi|^3.$$

Using this relation, a measurement of the $B \to \pi \ell \nu_\ell$ differential decay rate can in turn be used to extract a measurement of $|V_{\mathrm{ub}}|$, given a suitable parameterisation of the form factor $f_+(q^2)$. The analysis presented in the previous sections detailed the extraction of the partial branching fractions in bins of q^2, which are directly related to but not exactly equivalent to the differential decay rate. As defined in Section 1.2, the branching fraction of a particular decay mode corresponds to the proportion of decaying particles that decay via the stated mode, or, in other words, the ratio between the specified decay rate Γ and the total decay rate of all decaying particles Γ_{tot}. In the context of $B \to \pi \ell \nu_\ell$ decays, the total branching fraction can thus be computed as $\mathcal{B}(B \to \pi \ell \nu_\ell) = \Gamma/\Gamma_{\mathrm{tot}}$, where Γ refers to the $B \to \pi \ell \nu_\ell$ decay rate, and Γ_{tot} refers to the total B-meson decay rate, a quantity equal to the inverse of the B-meson lifetime τ_B, $\Gamma_{\mathrm{tot}} = 1/\tau_B$. It directly follows that a transformation between the $B \to \pi \ell \nu_\ell$ branching fraction and the decay rate can be made through multiplying or dividing by a factor of Γ_{tot}, with this relation also holding true for the partial branching fractions and differential decay rate as a function of q^2:

$$\frac{d\mathcal{B}(B \to \pi \ell \nu_\ell)}{dq^2} = \frac{d\Gamma}{dq^2} \times \tau_B.$$

Using this relation and the formula for the $B \to \pi \ell \nu_\ell$ differential decay rate above, a set of predicted partial branching fractions was simulated for the purpose of comparing these with the measured results. As per the above formulae, both an initial value of $|V_{\mathrm{ub}}|$ and an evaluation of the form factor $f_+(q^2)$ were required for the generation of these partial branching fractions. The world average measurement for exclusive $|V_{\mathrm{ub}}|$ was used for the initial value, $|V_{\mathrm{ub}}| = (3.70 \pm 0.10 \pm 0.12) \times 10^{-3}$, where the first uncertainty is experimental and the second theoretical [20]. As described in Section 1.2.3, using a BCL parameterisation [34], a determination of $f_+(q^2)$ over the full q^2 kinematic range can be computed using lattice quantum chromodynamics (LQCD). In this analysis, two separate predictions for the partial branching fractions were made using two leading LQCD results for the values of the BCL parameters, the first by the Fermilab/MILC collaborations in 2015 [96], and the second from the 2021 FLAG Review [97], which combined the Fermilab/MILC results with additional constraints from the RBC/UKQCD [98] collaborations. In the Fermilab/MILC determination, the z-expansion given in Section 1.2.3 was performed to fourth order, resulting in a total of eight BCL parameters [96]. A third-order z-expansion was performed for the FLAG Review computation, with one parameter fixed via additional constraints, resulting in five BCL parameters in total [97]. The values of each of these parameters are listed in Table

7.14, together with the respective covariance matrices representing the magnitude of the uncertainties on each parameter with all correlations taken into account. Each parameter in the table is assigned the same name as quoted in the respective papers, with all b (Fermilab/MILC) and a (FLAG) parameters of the same subscript and superscript being equivalent (for example, $b_0^+ \equiv a_0^+$). All partial branching fraction predictions were made, using these parameters and the above formulae, via the private Python software package in Ref.[99]. Using the two sets of predictions for the partial branching fractions, the following section will detail the extraction of $|V_{ub}|$ via a χ^2 fitting procedure used to compare the measured partial branching fractions with these theoretical predictions.

7.2.1   Fitting Strategy

A common statistical tool for comparing a set of observed results with a particular theoretical model is the χ^2 goodness-of-fit test. In such a test, given a specified number of categories n for which the data can be sorted, the χ^2 can be computed as [100]:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - e_i)^2}{\sigma_i^2},$$

where x_i and e_i refer to the observed and expected values for each category i, respectively, and σ_i refers to the uncertainty on the observed result for each category. In the context of the given analysis, a χ^2 can be constructed for the measured partial branching fractions against a set of predicted partial branching fractions in each q^2 bin, with the six q^2 bins representing six categories used in the computation. For a certain number of degrees of freedom (ndf) equal to one less than the number of categories, the lower the value of χ^2/ndf, the better the agreement between the observed and predicted results.

Whilst the χ^2 test can simply be used to compare a set of observations with various theoretical models as above, an added benefit of the χ^2 test lies in its ability to estimate a set of optimal parameters for a given theoretical model. In this implementation, using a set of initial parameters to define the theoretical model used to describe the observed data, a χ^2 fit can be performed whereby these parameters can be modified until those values that result in the smallest χ^2 value are found. It is this method that was employed for the purpose of extracting $|V_{ub}|$ in this analysis, where the optimal values of $|V_{ub}|$ and the BCL parameters used to predict the partial branching fractions described in the previous section were found through minimising the χ^2 in the fit.

As was done for validation purposes throughout the analysis, the χ^2 fits were first performed using Asimov data, specifically the Asimov partial branching fractions listed in Tables 7.5 and 7.6, before extracting $|V_{ub}|$ using the partial branching fractions measured from real Belle II data. Individual χ^2 fits were performed for each $B \to \pi \ell \nu_\ell$ decay mode studied including both the single and combined lepton cases, namely using the partial branching fractions measured from $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$, $B^0 \to \pi^- \ell^+ \nu_\ell$, $B^+ \to \pi^0 e^+ \nu_e$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, where $\ell = e$ or μ. Separate χ^2 fits were performed for the two sets of LQCD predictions, with the χ^2 calculated as above between the mea-

sured and predicted partial branching fractions. The uncertainties on the measured partial branching fractions were incorporated into the χ^2 calculation as covariance matrices to ensure the correlations between the partial branching fractions in each q^2 bin were properly taken into account. These covariance matrices were built in a similar manner to those described for the FEI calibration systematic uncertainties in Section 7.1.3, by first defining statistical and systematic covariance matrices which were then combined. The statistical uncertainties were assumed to be completely independent between each q^2 bin, resulting in a diagonal covariance matrix with diagonal entries equal to $\sigma^2_{\text{stat},j}$, where j is the q^2 bin index. All systematic uncertainties were taken to be completely correlated between q^2 bins, resulting in a covariance matrix with entries of $\sigma_{\text{sys},j} \times \sigma_{\text{sys},k}$ for bin indices j and k. The total covariance matrices built from the sum of the individual statistical and systematic contributions are given in Tables 7.15 and 7.16 for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively, both those evaluated from Asimov and real Belle II data.

	Fermilab/MILC 2015 [96]							
p_{BCL}	b_0^+	b_1^+	b_2^+	b_3^+	b_0^0	b_1^0	b_2^0	b_3^0
	0.41 ± 0.02	-0.65 ± 0.16	-0.46 ± 0.88	0.40 ± 1.30	0.51 ± 0.02	-1.77 ± 0.18	1.27 ± 0.81	4.20 ± 1.40
	Covariance matrix							$(\times 10^2)$
	b_0^+	b_1^+	b_2^+	b_3^+	b_0^0	b_1^0	b_2^0	b_3^0
b_0^+	0.02	0.11	0.21	0.2	0.01	0.09	0.35	0.45
b_1^+	0.11	2.57	10.66	13.83	0.15	2.35	11.07	15.66
b_2^+	0.21	10.66	77.83	113.03	0.93	13.42	67.79	97.94
b_3^+	0.2	13.83	113.03	169.84	1.38	19.49	96.14	129.95
b_0^0	0.01	0.15	0.93	1.38	0.05	0.18	0.64	0.58
b_1^0	0.09	2.35	13.42	19.49	0.18	3.26	12.06	12.6
b_2^0	0.35	11.07	67.79	96.14	0.64	12.06	65.94	95.03
b_3^0	0.45	15.66	97.94	129.95	0.58	12.6	95.03	196.98

	FLAG 2021 [97]				
p_{BCL}	a_0^+	a_1^+	a_2^+	a_0^0	a_1^0
	0.40 ± 0.01	-0.68 ± 0.13	-0.86 ± 0.61	0.49 ± 0.02	-1.61 ± 0.16
	Covariance matrix				$(\times 10^2)$
	a_0^+	a_1^+	a_2^+	a_0^0	a_1^0
a_0^+	0.02	0.07	0.09	0.01	0.07
a_1^+	0.07	1.7	5.88	0.08	1.87
a_2^+	0.09	5.88	37.4	0.47	8.65
a_0^0	0.01	0.08	0.47	0.04	0.08
a_1^0	0.07	1.87	8.65	0.08	2.57

Table 7.14: Values of the BCL parameters according to the Fermilab/MILC 2015 [96] and 2021 FLAG Review [97] determinations resulting from fourth and third order z-expansions, respectively. The covariance matrices between the given parameters are also included.

Using a modified version of the same private Python software package mentioned above [99], a Python interface to the Minuit tool [101] was utilised to perform the χ^2 minimisation, where the values of the input $|V_{\text{ub}}|$ and BCL parameters used to simulate the expected partial branching fractions were modified to produce the smallest possible χ^2, thereby fitting the predicted partial branching fractions to the measured values. An additional constraint was enforced during each fit, with the BCL parameters and their uncertainties given in Table 7.14 used to constrain the fitted parameters from exceeding the boundaries of the stated

Asimov data						Belle II data					
$B^0 \to \pi^- e^+ \nu_e$					$(\times 10^{12})$	$B^0 \to \pi^- e^+ \nu_e$					$(\times 10^{12})$
q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6

q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6
q_1	2.97	0.13	0.14	0.14	0.12	0.06	q_1	3.04	0.1	0.11	0.15	0.08	0.07
q_2	0.13	2.93	0.15	0.14	0.14	0.07	q_2	0.1	2.03	0.08	0.11	0.06	0.05
q_3	0.14	0.15	3.	0.15	0.16	0.08	q_3	0.11	0.08	2.38	0.12	0.08	0.07
q_4	0.14	0.14	0.15	2.99	0.15	0.07	q_4	0.15	0.11	0.12	2.98	0.1	0.08
q_5	0.12	0.14	0.16	0.15	3.58	0.09	q_5	0.08	0.06	0.08	0.1	2.5	0.06
q_6	0.06	0.07	0.08	0.07	0.09	1.26	q_6	0.07	0.05	0.07	0.08	0.06	1.28

Asimov data						Belle II data					
$B^0 \to \pi^- \mu^+ \nu_\mu$					$(\times 10^{12})$	$B^0 \to \pi^- \mu^+ \nu_\mu$					$(\times 10^{12})$
q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6
q_1	4.37	0.13	0.16	0.12	0.11	0.06	q_1	3.59	0.11	0.1	0.09	0.04	0.01
q_2	0.13	3.17	0.14	0.11	0.1	0.05	q_2	0.11	2.84	0.1	0.1	0.04	0.02
q_3	0.16	0.14	3.19	0.12	0.11	0.06	q_3	0.1	0.1	2.37	0.09	0.04	0.01
q_4	0.12	0.11	0.12	2.87	0.1	0.05	q_4	0.09	0.1	0.09	2.43	0.04	0.01
q_5	0.11	0.1	0.11	0.1	2.88	0.05	q_5	0.04	0.04	0.04	0.04	1.54	0.01
q_6	0.06	0.05	0.06	0.05	0.05	1.07	q_6	0.01	0.02	0.01	0.01	0.01	0.5

Asimov data						Belle II data					
$B^0 \to \pi^- \ell^+ \nu_\ell$					$(\times 10^{12})$	$B^0 \to \pi^- \ell^+ \nu_\ell$					$(\times 10^{12})$
q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6
q_1	1.85	0.12	0.13	0.12	0.13	0.06	q_1	1.7	0.1	0.1	0.11	0.06	0.03
q_2	0.12	1.58	0.13	0.12	0.12	0.06	q_2	0.1	1.26	0.08	0.1	0.05	0.03
q_3	0.13	0.13	1.61	0.13	0.14	0.06	q_3	0.1	0.08	1.2	0.09	0.06	0.03
q_4	0.12	0.12	0.13	1.51	0.12	0.06	q_4	0.11	0.1	0.09	1.4	0.06	0.04
q_5	0.13	0.12	0.14	0.12	1.65	0.06	q_5	0.06	0.05	0.06	0.06	1.03	0.02
q_6	0.06	0.06	0.06	0.06	0.06	0.58	q_6	0.03	0.03	0.03	0.04	0.02	0.43

Table 7.15: The total covariance matrices for each set of partial branching fractions derived from $B^0 \to \pi^- \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data.

uncertainties. This constraint was implemented by defining an augmented χ^2_{tot} built from the sum of the nominal χ^2 defined above and an additional constraint term χ^2_{LQCD}:

$$\chi^2_{\text{LQCD}} = \sum_{\ell=1}^{L} \frac{(p_{\text{fit}, \ell} - p_{\text{BCL}, \ell})^2}{\sigma^2_{\text{BCL}, \ell}},$$

where $p_{\text{fit}, \ell}$ refers to the values of each BCL parameter ℓ floating within the fit, $p_{\text{BCL}, \ell}$ refers to the nominal parameters as determined in the literature and given in Table 7.14, $\sigma_{\text{BCL}, \ell}$ refers to the respective uncertainties on the nominal parameters, implemented through a covariance matrix taking the correlations between parameters into account, and L is the total number of parameters. Prior to fitting, $\chi^2_{\text{LQCD}} = 0$ as the initial fit parameters were set to the nominal parameter values. However, as the given parameters were modified during the fit, this term deviated from zero and applied a constraint on the χ^2 minimisation. By including this additional χ^2_{LQCD} value in the total χ^2 to be minimised, χ^2_{tot}, the fit was

prohibited from returning a value of χ^2_{tot} below the threshold corresponding to the case in which the measured partial branching fractions exactly matched the theoretical predictions.

	Asimov data						Belle II data				
	$B^+ \to \pi^0 e^+ \nu_e$					$(\times 10^{12})$	$B^+ \to \pi^0 e^+ \nu_e$					$(\times 10^{12})$
q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q_1	q_2	q_3	q_4	q_5	q_6
q_1	1.27	0.09	0.09	0.09	0.08	0.03	0.93	0.07	0.06	0.08	0.	0.03
q_2	0.09	0.99	0.1	0.1	0.1	0.03	0.07	0.96	0.09	0.12	0.	0.04
q_3	0.09	0.1	0.91	0.1	0.09	0.04	0.06	0.09	0.78	0.1	0.	0.04
q_4	0.09	0.1	0.1	1.02	0.09	0.03	0.08	0.12	0.1	1.07	0.	0.05
q_5	0.08	0.1	0.09	0.09	1.34	0.03	0.	0.	0.	0.	1.22	0.
q_6	0.03	0.03	0.04	0.03	0.03	0.22	0.03	0.04	0.04	0.05	0.	0.23
	Asimov data						Belle II data				
	$B^+ \to \pi^0 \mu^+ \nu_\mu$					$(\times 10^{12})$	$B^+ \to \pi^0 \mu^+ \nu_\mu$					$(\times 10^{12})$
q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q_1	q_2	q_3	q_4	q_5	q_6
q_1	2.58	0.1	0.09	0.1	0.09	0.04	1.54	0.03	0.02	0.03	0.02	0.01
q_2	0.1	1.25	0.1	0.1	0.09	0.04	0.03	0.97	0.05	0.07	0.05	0.03
q_3	0.09	0.1	0.97	0.1	0.09	0.04	0.02	0.05	0.65	0.07	0.04	0.03
q_4	0.1	0.1	0.1	1.17	0.1	0.04	0.03	0.07	0.07	1.07	0.06	0.05
q_5	0.09	0.09	0.09	0.1	1.12	0.04	0.02	0.05	0.04	0.06	0.85	0.03
q_6	0.04	0.04	0.04	0.04	0.04	0.22	0.01	0.03	0.03	0.05	0.03	0.23
	Asimov data						Belle II data				
	$B^+ \to \pi^0 \ell^+ \nu_\ell$					$(\times 10^{12})$	$B^+ \to \pi^0 \ell^+ \nu_\ell$					$(\times 10^{12})$
q^2 bin	q_1	q_2	q_3	q_4	q_5	q_6	q_1	q_2	q_3	q_4	q_5	q_6
q_1	0.89	0.09	0.09	0.09	0.08	0.03	0.57	0.05	0.04	0.06	0.02	0.02
q_2	0.09	0.59	0.1	0.1	0.09	0.03	0.05	0.54	0.07	0.1	0.03	0.04
q_3	0.09	0.1	0.51	0.1	0.09	0.03	0.04	0.07	0.39	0.08	0.02	0.03
q_4	0.09	0.1	0.1	0.59	0.09	0.04	0.06	0.1	0.08	0.6	0.03	0.05
q_5	0.08	0.09	0.09	0.09	0.65	0.03	0.02	0.03	0.02	0.03	0.33	0.01
q_6	0.03	0.03	0.03	0.04	0.03	0.11	0.02	0.04	0.03	0.05	0.01	0.12

Table 7.16: The total covariance matrices for each set of partial branching fractions derived from $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data.

Whilst the individual χ^2 fits performed to each $B \to \pi \ell \nu_\ell$ decay mode were informative, the real statistical power lay in performing a simultaneous χ^2 fit to both sets of partial branching fraction distributions, $d\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell)/dq^2$ and $d\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell)/dq^2$. By combining both sets of partial branching fractions in a single measurement of $|V_{\text{ub}}|$, a higher precision could be obtained. To perform the simultaneous fit, the χ^2 values determined for each set of measured partial branching fractions as well as the LQCD constraint were combined together to form the total fit χ^2 to be minimised, χ^2_{tot}:

$$\chi^2_{\text{tot}} = \sum_{i=1}^{n} \frac{(x_{i,B^0} - e_{i,B^0})^2}{\sigma^2_{i,B^0}} + \sum_{i=1}^{n} \frac{(x_{i,B^+} - e_{i,B^+})^2}{\sigma^2_{i,B^+}} + \chi^2_{\text{LQCD}},$$

where x_{i,B^0} and x_{i,B^+} are the measured $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ partial branching fractions, respectively, e_{i,B^0} and e_{i,B^+} are the respective predicted partial branching fractions, and σ_{i,B^0} and σ_{i,B^+} are the respective uncertainties on the measured values.

The output returned from each χ^2 fit consisted of the post-fit values of each BCL parameter and, most importantly, $|V_{\mathrm{ub}}|$, that corresponded to the minimum χ^2 found by the fitting algorithm. Both the experimental uncertainties comprised of the statistical and systematic uncertainties on the measured partial branching fractions as well as the theoretical uncertainties on the LQCD parameters were taken into account by the fitting algorithm and contributed to the total uncertainties assigned to the post-fit parameters. As a goodness-of-fit indication, the post-fit χ^2 was also able to be accessed directly. All fit results together with a detailed discussion of the findings are included in the following section.

7.2.2 Results

All results pertaining to the χ^2 fits performed to the partial branching fraction distributions measured from Asimov data are summarised in Figures 7.13 – 7.18. In each figure, both the pre- and post-fit distributions of the measured partial branching fractions (data points) against the predicted distributions built from the LQCD parameters (red histograms) are included. An additional post-fit plot is also included for each case, likewise illustrating the measured partial branching fractions against the LQCD predictions, but with these represented as a central red curve surrounded by the 1σ, 2σ and 3σ uncertainty bands derived from the uncertainties on the post-fit parameters. Figures 7.13 and 7.14 depict the relevant plots for the individual χ^2 fits performed to the Asimov distributions corresponding to reconstructed $B^0 \to \pi^- e^+ \nu_e$, $B^0 \to \pi^- \mu^+ \nu_\mu$ and all $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, for the LQCD predictions from Fermilab/MILC [96] and the FLAG Review [97], respectively. The equivalent distributions for all reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays are included in Figures 7.15 and 7.16. Finally, the post-fit distributions of the Asimov partial branching fractions of $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays resulting from the simultaneous χ^2 fit to both sets of results are given in Figures 7.17 and 7.18 for the Fermilab/MILC [96] and FLAG Review LQCD constraints, respectively. The values of $|V_{\mathrm{ub}}|$ and the BCL parameters returned from each fit are summarised in Tables 7.17 and 7.18, respectively, together with the post-fit χ^2 divided by the corresponding number of degrees of freedom.

It is evident that in each case, the pre-fit agreement between the Asimov data and the LQCD predicted partial branching fractions was quite poor, with this improving significantly post-fit. As noted in Table 7.14, several BCL parameters possessed relatively large theoretical uncertainties, allowing the predicted partial branching fractions to fluctuate by a fair amount during the fit even under the imposed LQCD constraints. This behaviour was further reflected in the observed variation in the post-fit BCL parameters with respect to the nominal pre-fit values, though the effect of the LQCD constraints in the fits was also clear, with these values remaining consistent within uncertainties. For each case, the post-fit distributions illustrating the 1–3σ uncertainty bands on the post-fit predictions largely depicted the central values of the measured partial branching fractions falling within the 1σ uncertainty band, with respective uncertainties extending to the 2σ and 3σ bands given the

large uncertainties consistent with the limited statistics of the Asimov data sample. The reasonable post-fit agreement observed for Asimov data in each plot was similarly reflected in the magnitude of the minimised χ^2 returned from each fit, with small χ^2/ndf values listed in Table 7.17 consistent with p-values $\geqslant 88\%$ for the Fermilab/MILC determinations, and $\geqslant 67\%$ for the FLAG Review results.
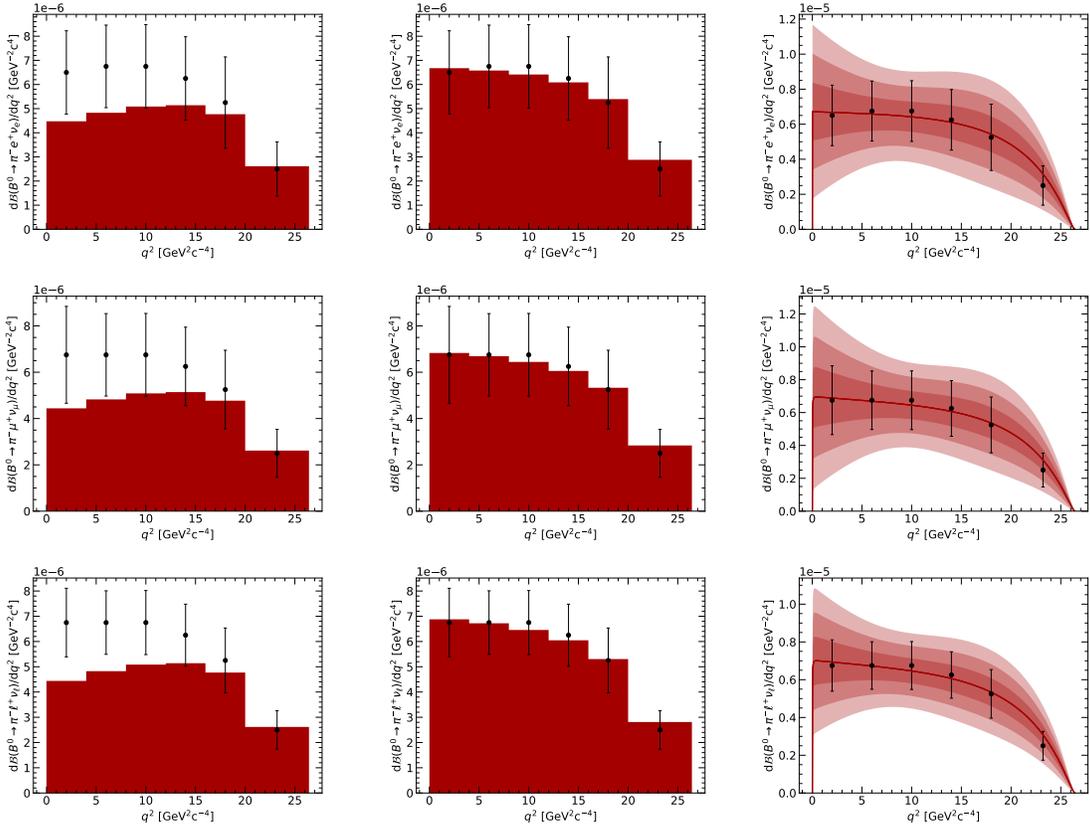

Figure 7.13: Partial branching fractions as a function of q^2 derived from $B^0 \to \pi^- e^+ \nu_e$ (top), $B^0 \to \pi^- \mu^+ \nu_\mu$ (middle) and all $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Asimov data against predictions from lattice QCD using Fermilab/MILC 2015 constraints [96]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.

All resultant $|V_{ub}|$ values extracted from each χ^2 fit to Asimov data shown in Table 7.17 were consistent with the world average value [20]. It can also be observed that in all cases, the central values extracted from the fits using the FLAG LQCD predictions were larger than those extracted using the Fermilab/MILC predictions. In addition, those central values extracted from the individual fits to the Asimov partial branching fractions of $B^+ \to \pi^0 \ell^+ \nu_\ell$ modes were typically found to be lower than those extracted from $B^0 \to \pi^- \ell^+ \nu_\ell$ modes. The most significant results, those $|V_{ub}|$ values extracted from the simultaneous χ^2 fits to
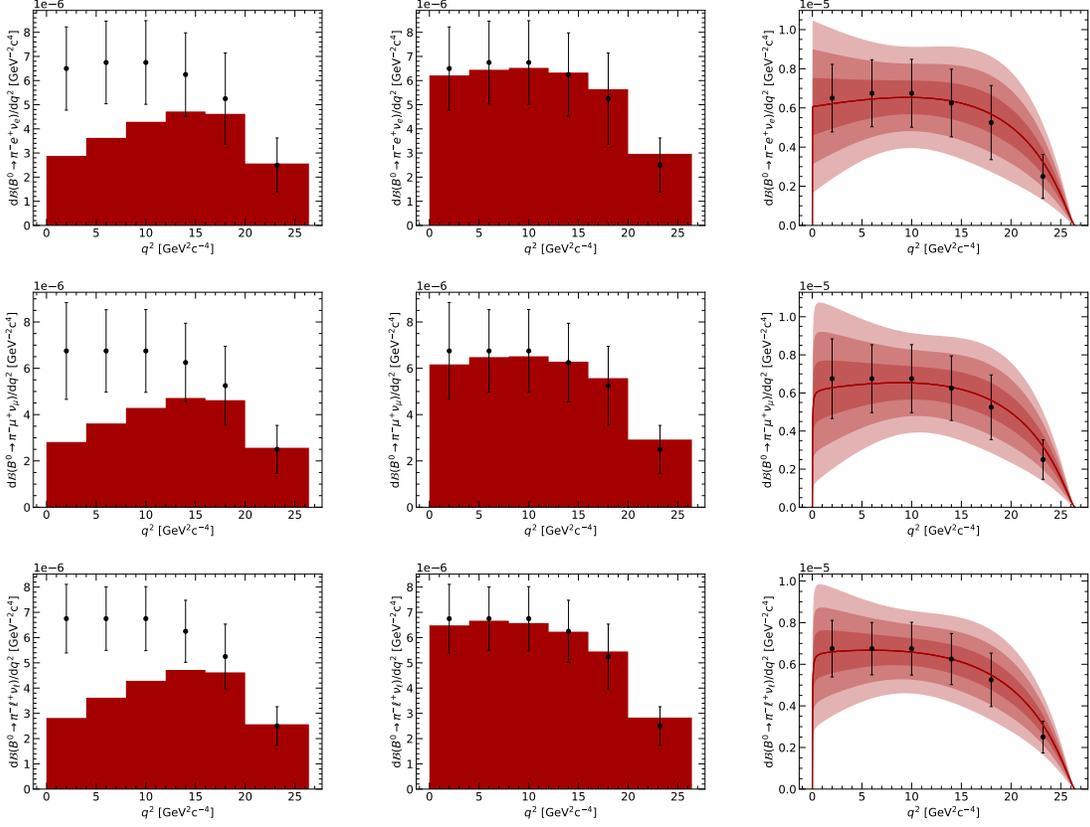
Figure 7.14: Partial branching fractions as a function of q^2 derived from $B^0 \to \pi^- e^+ \nu_e$ (top), $B^0 \to \pi^- \mu^+ \nu_\mu$ (middle) and all $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Asimov data against predictions from lattice QCD using FLAG 2021 constraints [97]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.

both sets of Asimov partial branching fractions, were clearly superior in their precision when compared with the fits to the individual modes as expected. Utilising the combined statistics of both the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ samples, the total uncertainty on $|V_{ub}|$ measured from Asimov data was $\sim 7\%$. With the $|V_{ub}|$ values extracted from the fits to Asimov data agreeing with the world average and demonstrating a reasonable goodness-of-fit as per the post-fit χ^2/ndf results for each set of LQCD constraints imposed during the fits, the χ^2 fitting procedure was thus validated for use on real Belle II data.

The partial branching fraction distributions corresponding to Belle II data are included in Figures 7.19 – 7.24, mirroring the structure of the respective Asimov plots, with Figures 7.19(7.21) and 7.20(7.22) representing the individual χ^2 fits for all $B^0 \to \pi^- \ell^+ \nu_\ell(B^+ \to \pi^0 \ell^+ \nu_\ell)$ decays to the Fermilab/MILC and FLAG LQCD predictions, respectively. The equivalent plots for the simultaneous χ^2 fits to both the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$
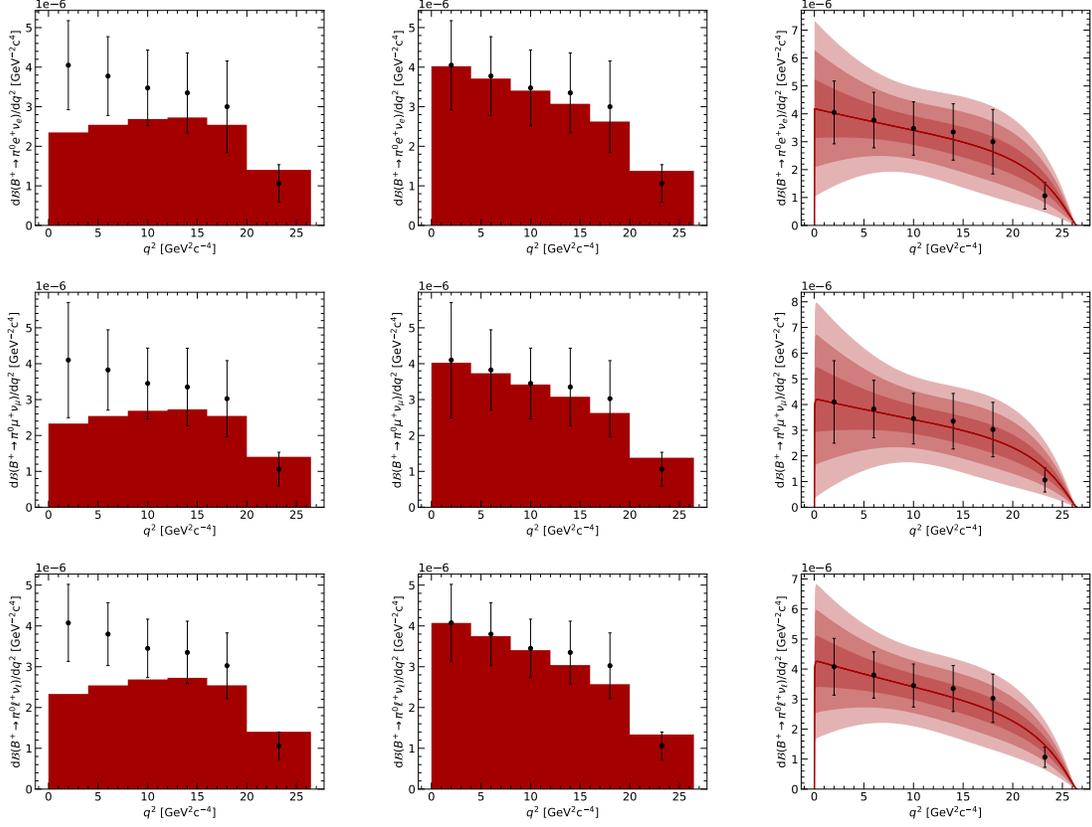
Figure 7.15: Partial branching fractions as a function of q^2 derived from $B^+ \rightarrow \pi^0 e^+ \nu_e$ (top), $B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$ (middle) and all $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Asimov data against predictions from lattice QCD using Fermilab/MILC 2015 constraints [96]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.

partial branching fraction distributions are included in Figures 7.23 and 7.24, with all post-fit $|V_{ub}|$ and χ^2/ndf values listed in Table 7.17, and all post-fit BCL parameters listed in Table 7.19.

The agreement between the measured and predicted partial branching fractions similarly improved post-fit for real data overall, though the post-fit agreement was unsurprisingly poorer than that observed for the fits to Asimov data. The measured partial branching fractions exhibited a higher degree of variation with respect to the predictions, with central values occasionally falling within the 2σ and 3σ uncertainty bands on the post-fit predicted distributions, both for the Fermilab/MILC and FLAG LQCD determinations. The post-fit BCL parameters also deviated further from the pre-fit values than for the equivalent Asimov data results, though these still lay within the scope of the LQCD constraints. Given the greater fluctuation in the measured partial branching fractions in real data with respect to
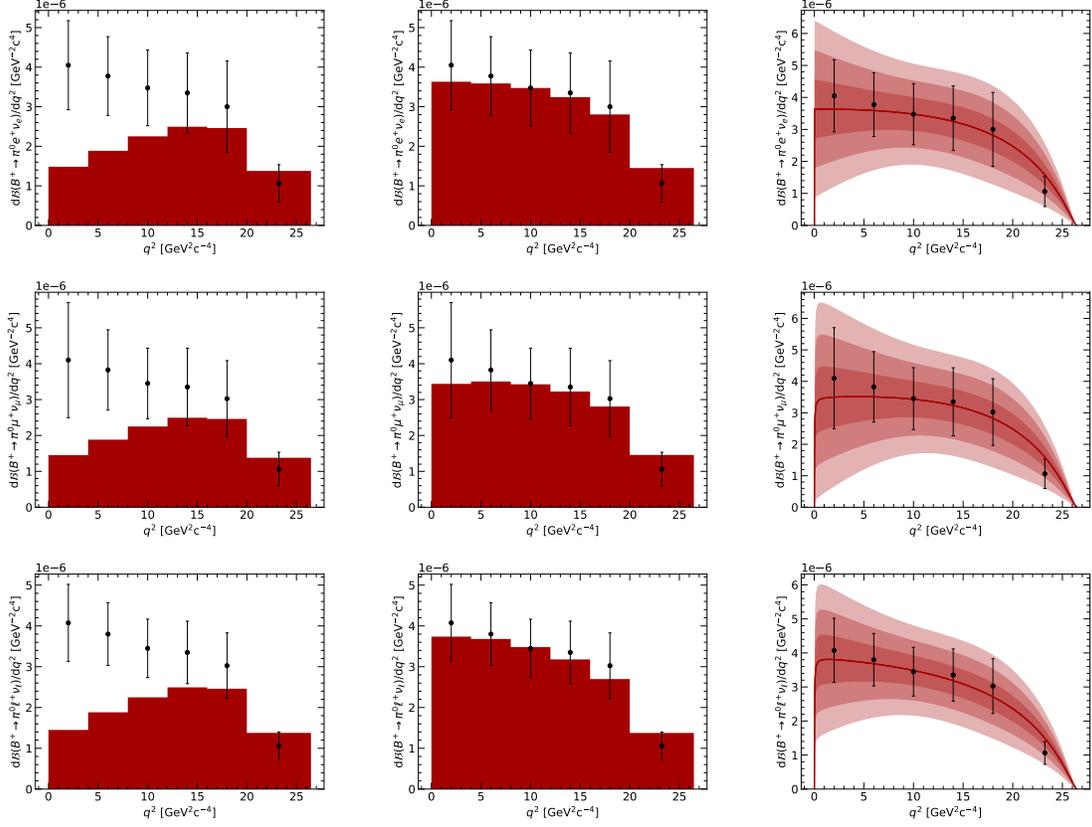
Figure 7.16: Partial branching fractions as a function of q^2 derived from $B^+ \rightarrow \pi^0 e^+ \nu_e$ (top), $B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$ (middle) and all $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Asimov data against predictions from lattice QCD using FLAG 2021 constraints [97]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.

the LQCD predictions, the post-fit χ^2/ndf values were fairly larger than the equivalent Asimov values, further reinforcing the difference in the fit quality. Quantitatively, these values corresponded to p-values of $\geqslant 11\%$ for the Fermilab/MILC LQCD predictions, and $\geqslant 7\%$ for the FLAG predictions.

In considering the values of $|V_{ub}|$ extracted from the χ^2 fits to real data, it can be clearly observed that the phenomenon of the low measured branching fractions was similarly manifested in the given results, with many of the values extracted proving inconsistent with the world average result within uncertainties. The $|V_{ub}|$ values extracted from the combined χ^2 fits to the $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ partial branching fraction distributions, the flagship results corresponding to the highest precision possible in the given analysis, were found to lie below the world average result with a total deviation of $\sim 2.1\sigma$ for both the Fermilab/MILC and FLAG determinations. At the given data integrated luminosity, these
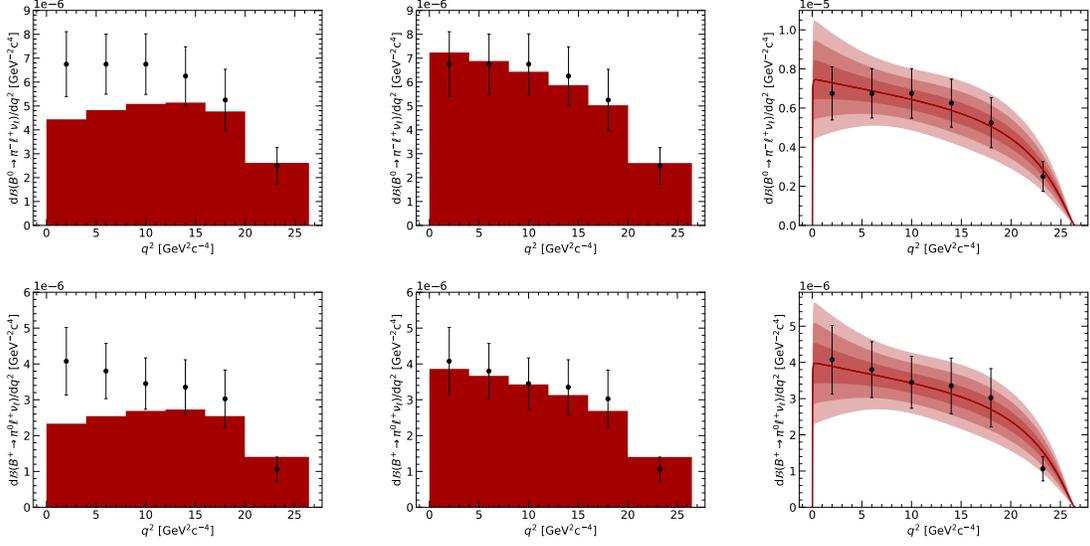
Figure 7.17: Partial branching fractions as a function of q^2 derived from $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ (top) and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Asimov data against predictions from lattice QCD using Fermilab/MILC 2015 constraints [96]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for a simultaneous χ^2 fit performed to both sets of partial branching fractions.

$|V_{\mathrm{ub}}|$ measurements corresponded to a precision of $\sim 8\%$ and $\sim 7\%$, respectively. Despite the differences in the fit quality and magnitude of the extracted $|V_{\mathrm{ub}}|$ values, the real data results reflected similar trends as those observed for Asimov data, with the FLAG predictions resulting in larger central values of $|V_{\mathrm{ub}}|$ compared with those from Fermilab/MILC alone.

With the determination of $|V_{\mathrm{ub}}|$ from the measured partial branching fraction distributions, the final and most significant measurement presented in this report, the analysis on the semi-leptonic $B \rightarrow \pi \ell \nu_\ell$ decay is concluded. After summarising the methodology and findings detailed throughout this study, the following chapter will offer some concluding remarks on the significance of the given results and the future outlook of the analysis.

Figure 7.18: Partial branching fractions as a function of q^2 derived from $B^0 \to \pi^- \ell^+ \nu_\ell$ (top) and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Asimov data against predictions from lattice QCD using FLAG 2021 constraints [97]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for a simultaneous χ^2 fit performed to both sets of partial branching fractions.

Decay mode	Asimov data		Belle II data	
	Fitted $|V_{\mathrm{ub}}|$	Fit χ^2/ndf	Fitted $|V_{\mathrm{ub}}|$	Fit χ^2/ndf
Fermilab/MILC 2015 [96]				
$B^0 \to \pi^- e^+ \nu_e$	$(3.88 \pm 0.43) \times 10^{-3}$	0.28/5	$(3.54 \pm 0.43) \times 10^{-3}$	1.92/5
$B^0 \to \pi^- \mu^+ \nu_\mu$	$(3.85 \pm 0.41) \times 10^{-3}$	0.27/5	$(2.86 \pm 0.34) \times 10^{-3}$	4.55/5
$B^+ \to \pi^0 e^+ \nu_e$	$(3.66 \pm 0.43) \times 10^{-3}$	0.97/5	$(3.36 \pm 0.44) \times 10^{-3}$	6.77/5
$B^+ \to \pi^0 \mu^+ \nu_\mu$	$(3.66 \pm 0.43) \times 10^{-3}$	1.03/5	$(3.58 \pm 0.45) \times 10^{-3}$	1.22/5
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(3.83 \pm 0.33) \times 10^{-3}$	0.41/5	$(3.16 \pm 0.30) \times 10^{-3}$	3.57/5
$B^+ \to \pi^0 \ell^+ \nu_\ell$	$(3.61 \pm 0.35) \times 10^{-3}$	1.70/5	$(3.22 \pm 0.34) \times 10^{-3}$	8.92/5
Combined fit	$(3.69 \pm 0.26) \times 10^{-3}$	2.26/11	$(3.17 \pm 0.25) \times 10^{-3}$	12.67/11
FLAG 2021 [97]				
$B^0 \to \pi^- e^+ \nu_e$	$(3.97 \pm 0.40) \times 10^{-3}$	0.99/5	$(3.65 \pm 0.40) \times 10^{-3}$	2.96/5
$B^0 \to \pi^- \mu^+ \nu_\mu$	$(3.94 \pm 0.38) \times 10^{-3}$	1.10/5	$(3.02 \pm 0.31) \times 10^{-3}$	6.72/5
$B^+ \to \pi^0 e^+ \nu_e$	$(3.79 \pm 0.40) \times 10^{-3}$	2.24/5	$(3.49 \pm 0.40) \times 10^{-3}$	7.96/5
$B^+ \to \pi^0 \mu^+ \nu_\mu$	$(3.80 \pm 0.40) \times 10^{-3}$	2.26/5	$(3.65 \pm 0.43) \times 10^{-3}$	1.30/5
$B^0 \to \pi^- \ell^+ \nu_\ell$	$(3.88 \pm 0.31) \times 10^{-3}$	1.31/5	$(3.24 \pm 0.28) \times 10^{-3}$	5.33/5
$B^+ \to \pi^0 \ell^+ \nu_\ell$	$(3.69 \pm 0.33) \times 10^{-3}$	3.16/5	$(3.29 \pm 0.32) \times 10^{-3}$	9.92/5
Combined fit	$(3.72 \pm 0.25) \times 10^{-3}$	3.31/11	$(3.21 \pm 0.23) \times 10^{-3}$	13.94/11
World avg. (excl.) [20]	$(3.70 \pm 0.10 \pm 0.12) \times 10^{-3}$			

Table 7.17: Fitted $|V_{\mathrm{ub}}|$ values from χ^2 fits to the partial branching fraction distributions of $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov and real Belle II data. The χ^2 returned from the fit divided by the number of degrees of freedom is also listed. For comparison, the world average $|V_{\mathrm{ub}}|$ measurement is also included, where the first uncertainty is experimental and the second theoretical.
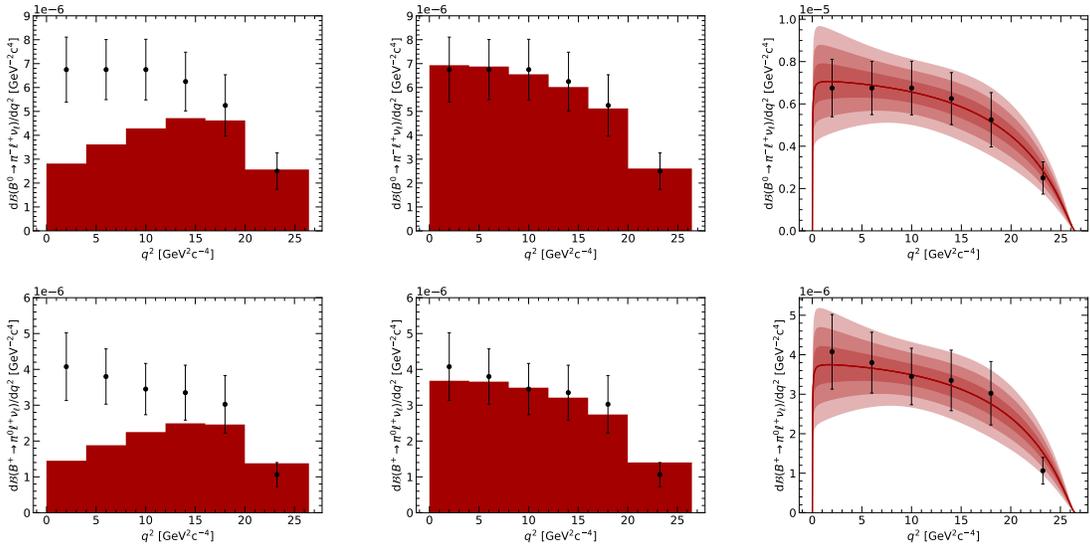
Figure 7.19: Partial branching fractions as a function of q^2 derived from $B^0 \to \pi^- e^+ \nu_e$ (top), $B^0 \to \pi^- \mu^+ \nu_\mu$ (middle) and all $B^0 \to \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Belle II data against predictions from lattice QCD using Fermilab/MILC 2015 constraints [96]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.
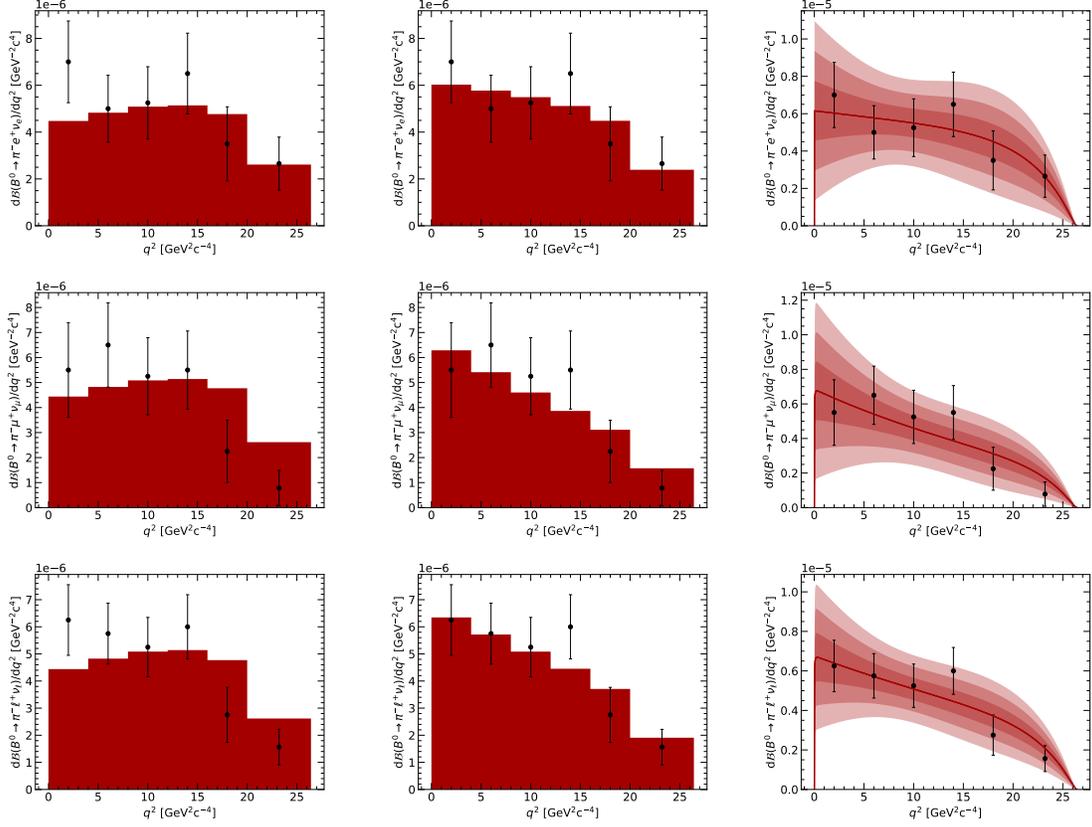
Figure 7.20: Partial branching fractions as a function of q^2 derived from $B^0 \rightarrow \pi^- e^+ \nu_e$ (top), $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (middle) and all $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Belle II data against predictions from lattice QCD using FLAG 2021 constraints [97]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.

Figure 7.21: Partial branching fractions as a function of q^2 derived from $B^+ \to \pi^0 e^+ \nu_e$ (top), $B^+ \to \pi^0 \mu^+ \nu_\mu$ (middle) and all $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Belle II data against predictions from lattice QCD using Fermilab/MILC 2015 constraints [96]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.
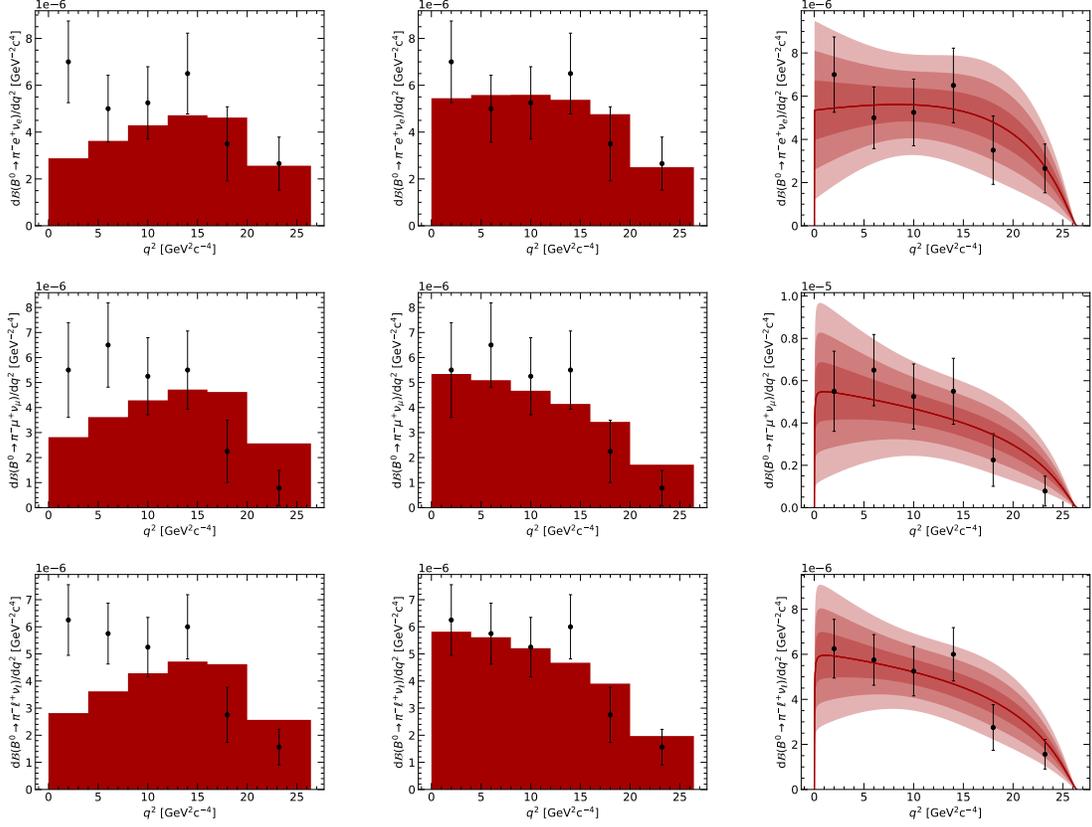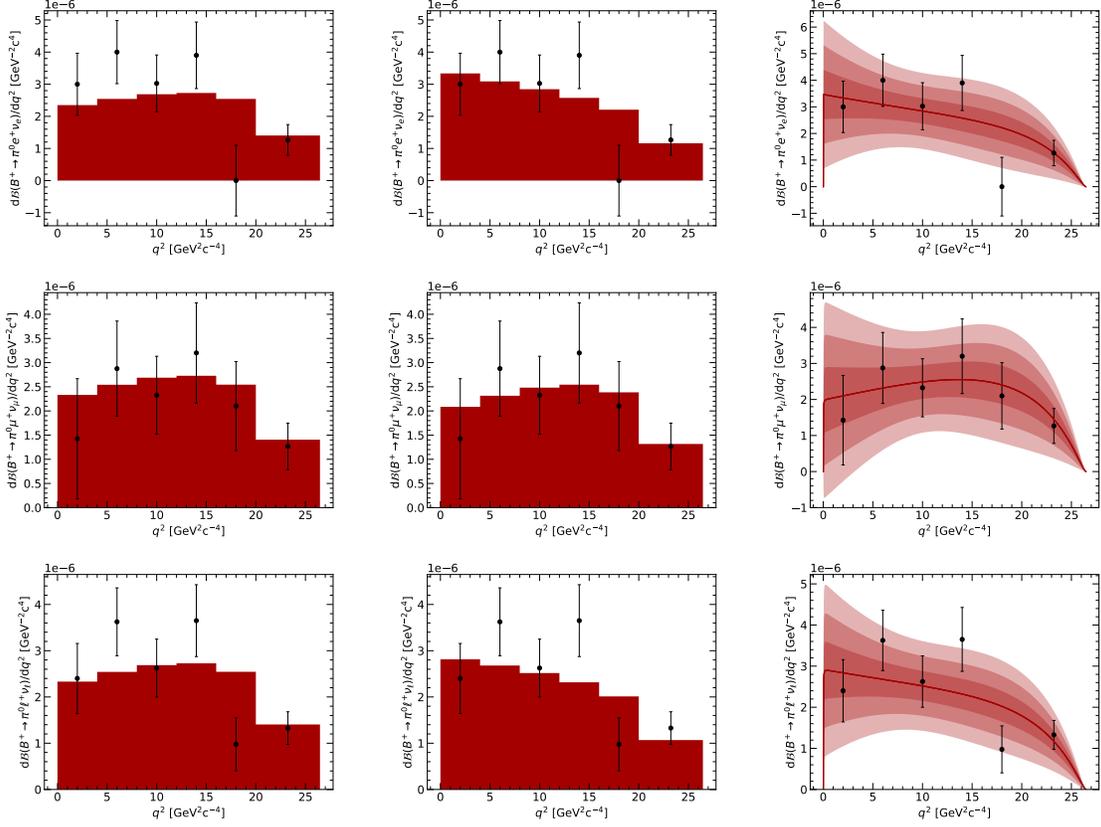
Figure 7.22: Partial branching fractions as a function of q^2 derived from $B^+ \rightarrow \pi^0 e^+ \nu_e$ (top), $B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$ (middle) and all $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Belle II data against predictions from lattice QCD using FLAG 2021 constraints [97]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for individual χ^2 fits performed to the partial branching fraction distributions of each decay mode.
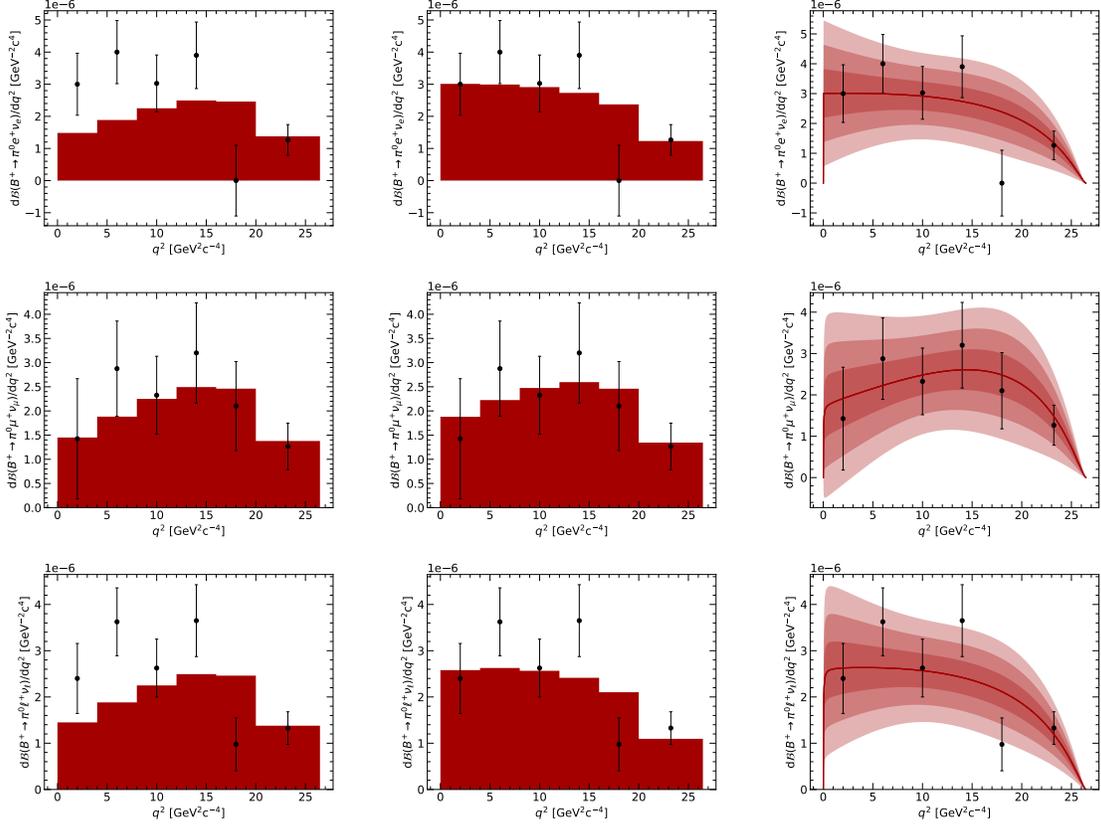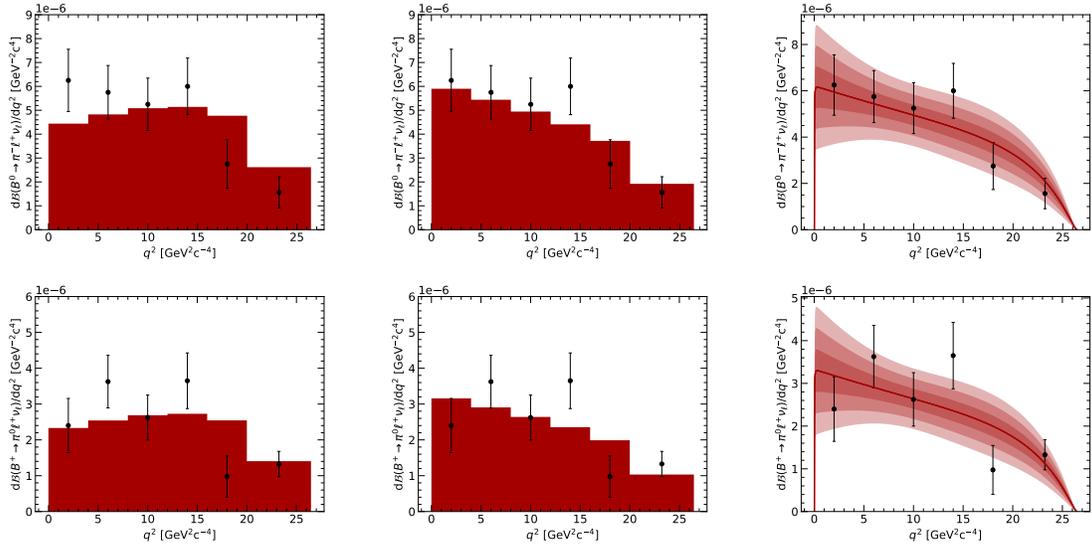
Figure 7.23: Partial branching fractions as a function of q^2 derived from $B^0 \to \pi^- \ell^+ \nu_\ell$ (top) and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Belle II data against predictions from lattice QCD using Fermilab/MILC 2015 constraints [96]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for a simultaneous χ^2 fit performed to both sets of partial branching fractions.
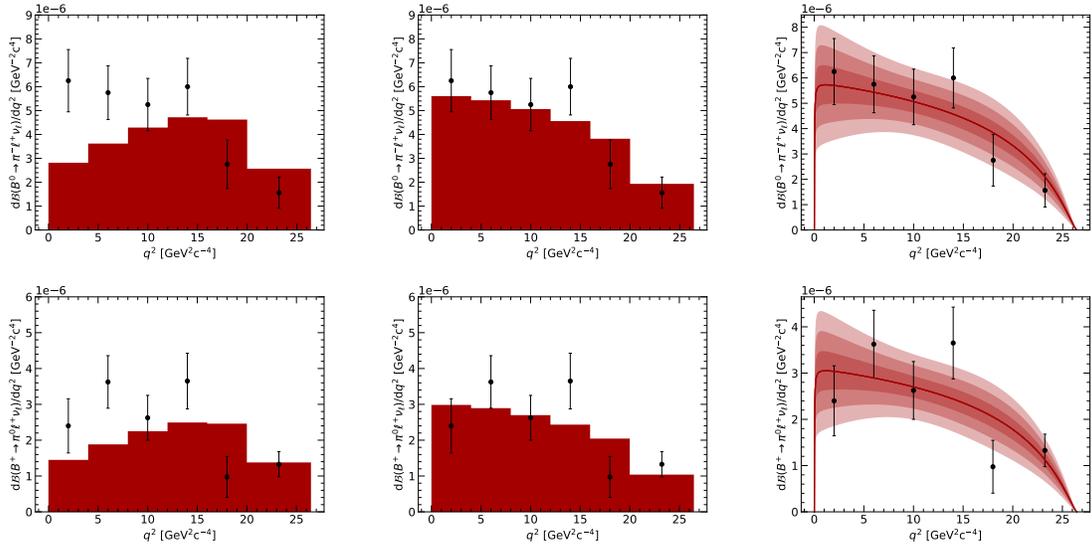
Figure 7.24: Partial branching fractions as a function of q^2 derived from $B^0 \to \pi^- \ell^+ \nu_\ell$ (top) and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (bottom) decays, where $\ell = e$ or μ, reconstructed from 362 fb^{-1} of Belle II data against predictions from lattice QCD using FLAG 2021 constraints [97]. Each distribution is shown pre-fit (left), with post-fit projections overlaid (middle), and with post-fit projections showing the 1–3σ uncertainty bands (right), for a simultaneous χ^2 fit performed to both sets of partial branching fractions.

Fermilab/MILC 2015 [96]								
p_{BCL}	b_0^+	b_1^+	b_2^+	b_3^+	b_0^0	b_1^0	b_2^0	b_3^0
Pre-fit	0.41 ± 0.02	-0.65 ± 0.16	-0.46 ± 0.88	0.40 ± 1.30	0.51 ± 0.02	-1.77 ± 0.18	1.27 ± 0.81	4.20 ± 1.40
$B^0 \to \pi^- e^+ \nu_e$	0.41 ± 0.01	-0.60 ± 0.09	-0.24 ± 0.34	0.70 ± 0.60	0.51 ± 0.02	-1.72 ± 0.10	1.49 ± 0.32	4.52 ± 0.92
$B^0 \to \pi^- \mu^+ \nu_\mu$	0.41 ± 0.01	-0.60 ± 0.09	-0.19 ± 0.36	0.76 ± 0.63	0.51 ± 0.02	-1.72 ± 0.10	1.53 ± 0.33	4.58 ± 0.93
$B^+ \to \pi^0 e^+ \nu_e$	0.41 ± 0.01	-0.57 ± 0.09	-0.03 ± 0.37	0.99 ± 0.64	0.51 ± 0.02	-1.68 ± 0.10	1.69 ± 0.34	4.79 ± 0.94
$B^+ \to \pi^0 \mu^+ \nu_\mu$	0.41 ± 0.01	-0.56 ± 0.09	-0.02 ± 0.41	1.00 ± 0.69	0.51 ± 0.02	-1.68 ± 0.10	1.69 ± 0.37	4.80 ± 0.96
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.41 ± 0.01	-0.59 ± 0.08	-0.19 ± 0.29	0.75 ± 0.54	0.51 ± 0.02	-1.71 ± 0.09	1.54 ± 0.27	4.59 ± 0.89
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.41 ± 0.01	-0.55 ± 0.08	-0.01 ± 0.32	1.01 ± 0.58	0.51 ± 0.02	-1.67 ± 0.09	1.73 ± 0.30	4.84 ± 0.91
Combined fit	0.41 ± 0.01	-0.56 ± 0.07	-0.13 ± 0.26	0.80 ± 0.50	0.51 ± 0.02	-1.69 ± 0.09	1.64 ± 0.25	4.71 ± 0.88

FLAG 2021 [97]					
p_{BCL}	a_0^+	a_1^+	a_2^+	a_0^0	a_1^0
Pre-fit	0.40 ± 0.01	-0.68 ± 0.13	-0.86 ± 0.61	0.49 ± 0.02	-1.61 ± 0.16
$B^0 \to \pi^- e^+ \nu_e$	0.41 ± 0.01	-0.58 ± 0.07	-0.39 ± 0.35	0.50 ± 0.02	-1.48 ± 0.09
$B^0 \to \pi^- \mu^+ \nu_\mu$	0.41 ± 0.01	-0.57 ± 0.08	-0.37 ± 0.36	0.50 ± 0.02	-1.47 ± 0.09
$B^+ \to \pi^0 e^+ \nu_e$	0.41 ± 0.01	-0.54 ± 0.08	-0.23 ± 0.36	0.50 ± 0.02	-1.43 ± 0.09
$B^+ \to \pi^0 \mu^+ \nu_\mu$	0.41 ± 0.01	-0.54 ± 0.08	-0.26 ± 0.39	0.50 ± 0.02	-1.44 ± 0.10
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.41 ± 0.01	-0.56 ± 0.07	-0.30 ± 0.31	0.50 ± 0.02	-1.45 ± 0.07
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.41 ± 0.01	-0.51 ± 0.07	-0.14 ± 0.33	0.50 ± 0.02	-1.41 ± 0.08
Combined fit	0.41 ± 0.01	-0.52 ± 0.06	-0.18 ± 0.28	0.50 ± 0.02	-1.42 ± 0.07

Table 7.18: Fitted BCL parameter values from χ^2 fits to the partial branching fraction distributions of $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of Asimov data. The pre-fit values are also listed for comparison.

Fermilab/MILC 2015 [96]								
p_{BCL}	b_0^+	b_1^+	b_2^+	b_3^+	b_0^0	b_1^0	b_2^0	b_3^0
Pre-fit	0.41 ± 0.02	-0.65 ± 0.16	-0.46 ± 0.88	0.40 ± 1.30	0.51 ± 0.02	-1.77 ± 0.18	1.27 ± 0.81	4.20 ± 1.40
$B^0 \to \pi^- e^+ \nu_e$	0.41 ± 0.01	-0.60 ± 0.09	-0.13 ± 0.37	0.87 ± 0.64	0.51 ± 0.02	-1.71 ± 0.10	1.56 ± 0.35	4.61 ± 0.94
$B^0 \to \pi^- \mu^+ \nu_\mu$	0.41 ± 0.01	-0.50 ± 0.10	0.28 ± 0.44	1.41 ± 0.72	0.52 ± 0.02	-1.62 ± 0.11	1.99 ± 0.40	5.22 ± 0.99
$B^+ \to \pi^0 e^+ \nu_e$	0.41 ± 0.01	-0.57 ± 0.09	-0.04 ± 0.39	0.98 ± 0.66	0.51 ± 0.02	-1.69 ± 0.10	1.67 ± 0.36	4.77 ± 0.95
$B^+ \to \pi^0 \mu^+ \nu_\mu$	0.41 ± 0.01	-0.65 ± 0.09	-0.51 ± 0.41	0.32 ± 0.69	0.51 ± 0.02	-1.78 ± 0.10	1.24 ± 0.37	4.15 ± 0.96
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.41 ± 0.01	-0.54 ± 0.09	0.09 ± 0.33	1.14 ± 0.59	0.51 ± 0.02	-1.66 ± 0.10	1.81 ± 0.31	4.96 ± 0.92
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.41 ± 0.01	-0.58 ± 0.09	-0.11 ± 0.33	0.88 ± 0.60	0.51 ± 0.02	-1.70 ± 0.10	1.61 ± 0.31	4.69 ± 0.92
Combined fit	0.41 ± 0.01	-0.55 ± 0.08	-0.01 ± 0.28	0.99 ± 0.53	0.51 ± 0.02	-1.67 ± 0.09	1.73 ± 0.27	4.86 ± 0.89

FLAG 2021 [97]					
p_{BCL}	a_0^+	a_1^+	a_2^+	a_0^0	a_1^0
Pre-fit	0.40 ± 0.01	-0.68 ± 0.13	-0.86 ± 0.61	0.49 ± 0.02	-1.61 ± 0.16
$B^0 \to \pi^- e^+ \nu_e$	0.41 ± 0.01	-0.57 ± 0.08	-0.34 ± 0.37	0.50 ± 0.02	-1.47 ± 0.09
$B^0 \to \pi^- \mu^+ \nu_\mu$	0.41 ± 0.01	-0.49 ± 0.08	-0.00 ± 0.40	0.50 ± 0.02	-1.37 ± 0.10
$B^+ \to \pi^0 e^+ \nu_e$	0.41 ± 0.01	-0.55 ± 0.08	-0.24 ± 0.37	0.50 ± 0.02	-1.44 ± 0.09
$B^+ \to \pi^0 \mu^+ \nu_\mu$	0.41 ± 0.01	-0.64 ± 0.08	-0.68 ± 0.40	0.49 ± 0.02	-1.56 ± 0.10
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.41 ± 0.01	-0.50 ± 0.07	-0.05 ± 0.34	0.50 ± 0.02	-1.39 ± 0.08
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.41 ± 0.01	-0.55 ± 0.07	-0.24 ± 0.34	0.50 ± 0.02	-1.44 ± 0.08
Combined fit	0.41 ± 0.01	-0.50 ± 0.06	-0.07 ± 0.29	0.50 ± 0.02	-1.39 ± 0.07

Table 7.19: Fitted BCL parameter values from χ^2 fits to the partial branching fraction distributions of $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays reconstructed from 362 fb^{-1} of real Belle II data. The pre-fit values are also listed for comparison.

Chapter 8

Conclusions and Outlook

Semi-leptonic B-meson decays provide a unique opportunity to test several core elements of the Standard Model (SM) of particle physics. As prime examples of flavour-changing quark transitions mediated by the weak force, these decays have proven invaluable in measuring the magnitudes of the Cabibbo-Kobayashi-Maskawa (CKM)-matrix elements characterising the strength of these interactions. With the magnitude of several matrix elements such as $|V_{ud}|$ and $|V_{tb}|$ already known to a high degree of precision given the frequency of their respective quark transitions, the element with the smallest magnitude, $|V_{ub}|$, currently possesses the largest relative uncertainty based on recent measurements [20]. The current precision of $|V_{ub}|$ constitutes a significant uncertainty in the global fits designed to test the unitarity of the CKM-matrix, a cornerstone feature of the SM, where any violation of this unitarity would constitute direct evidence for physics Beyond the Standard Model (BSM). Improving the precision on measurements of $|V_{ub}|$ is thus highly sought after within the particle physics community, with semi-leptonic B-meson decays involving the $b \to u$ quark transition forming the preferred method for such extractions. In particular, the $B \to \pi \ell \nu_\ell$ decay is considered the golden mode for $|V_{ub}|$ determinations due to the simplicity of its theoretical description, with the decay to the scalar pion characterised by a decay rate dependent on a single analytical form factor to describe the interference effects of the strong interaction.

With the $B \to \pi \ell \nu_\ell$ decay being somewhat rare, corresponding to world average branching fractions of $(1.50 \pm 0.06) \times 10^{-4}$ and $(7.80 \pm 0.27) \times 10^{-5}$ for the two B-meson flavours [20], recent measurements of $|V_{ub}|$ have been limited both by experimental sample size as well as the precision on current theoretical determinations of the form factor derived from lattice quantum chromodynamics (LQCD). The Belle II experiment, a B-factory designed to produce and detect the decay products of billions of pairs of B-mesons, aims to address the first of these limitations with a target goal of recording a total integrated luminosity of 50 ab^{-1} by the completion of the experiment, a data-set of the order of 50 billion B-meson pairs. In achieving these targets, Belle II would collect over 50 times the data of both its predecessor, Belle, and the BaBar B-factory experiment of the same generation. With advances in LQCD techniques similarly expected to increase with time, improving the theoretical precision, Belle II stands to produce $|V_{ub}|$ measurements at world-leading precision as the experiment progresses. At present, Belle II has collected a preliminary analysis sample equivalent to an integrated luminosity of 362 fb^{-1}, or $(387 \pm 6) \times 10^6$ B-meson pairs.

As described in Chapter 3, several methods exist for the reconstruction of $B \to \pi \ell \nu_\ell$ decays at B-factory experiments such as Belle II, with these grouped into two main categories, namely tagged and untagged analyses. Differentiated by their treatment of the second, non-signal B-meson in the full $\Upsilon(4S) \to B\bar{B}$ event, both methods reconstruct the signal B-meson decay with varying degrees of purity and efficiency. In an untagged analysis, only the signal B-meson is explicitly reconstructed, resulting in a high efficiency of reconstruction at the cost of significant background levels that must be mitigated through the careful application of multiple event selections. In a tagged reconstruction, however, the second B-meson in the event, the B_{tag}, is reconstructed alongside the signal B-meson in one of many hadronic or semi-leptonic decay modes, significantly improving the purity of the sample at the loss of those signal events in which no B_{tag} could be reconstructed. With the large projected data-set of Belle II, highly pure tagged analyses of rare decays such as $B \to \pi \ell \nu_\ell$ have become inherently more desirable due to the increased statistics countering the lower efficiency of such methods. When considering the semi-leptonic $B \to \pi \ell \nu_\ell$ decay in particular, hadronic tagged methods are especially powerful as the four-momentum of the only missing particle in the event, the signal-side neutrino, can be explicitly determined via four-momentum conservation.

All tagging at Belle II is conducted using the Full-Event-Interpretation (FEI) [13], a state-of-the-art machine learning algorithm where hadronic or semi-leptonic B-meson tags are reconstructed hierarchically from detector information. The algorithm is trained using simulated data, or Monte Carlo (MC), with a set of multi-variate classifiers trained at each step of the B-meson decay chain to distinguish between correctly and incorrectly reconstructed particles. The final classifier output resulting from the training, the B_{tag} `SignalProbability`, can then be used to select a single B_{tag} candidate per event with the highest likelihood of being correctly reconstructed. With simulated data only ever able to imitate the underlying physics of nature according to the imperfect collective understanding available at any one time, discrepancies between MC and data are unavoidable and necessitate that tools such as the FEI must be calibrated before applying them to real data. Within the scope of the analysis detailed in this thesis, an illustrative study of a potential method for calibrating the hadronic FEI was demonstrated, whereby a signal B-meson decay with a large branching fraction, the semi-leptonic $B \to X\ell \nu_\ell$ decay, was reconstructed alongside a hadronic B_{tag} built using the FEI. In reconstructing this decay from both samples of simulated and real Belle II data, the ratio between the number of signal events remaining in each sample was able to be determined, with this value representing a calibration factor with which the MC in independent hadronic tagged analyses could be rescaled to account for the data-MC discrepancy.

In considering each of the above points, the major analysis presented throughout this report aimed at extracting a measurement of $|V_{\text{ub}}|$ using a hadronic tagged analysis of the semi-leptonic $B \to \pi \ell \nu_\ell$ decay. After an in-depth selection optimisation procedure designed to reject the various backgrounds present after reconstruction, the distribution of the square of the missing mass in the full $\Upsilon(4S) \to B_{\text{tag}}\bar{B}_{\text{sig}}$ event was used to extract the number of signal events in data. Given the kinematically motivated signal peak at $M_{\text{miss}}^2 = 0$ corresponding

to the lone missing neutrino in the event, clear discrimination between signal and background events was possible in the reconstruction of both $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays. In each case, the signal yields were estimated through a process of fitting the observed data to a set of signal and background template probability density functions (pdfs) built from simulation, by establishing those yields for which the likelihood of the given pdfs resulting in the observed data was a maximum. For the purpose of later extracting $|V_{ub}|$ from the distributions of the partial branching fractions in six bins of the square of the momentum transfer to the leptonic system, q^2, the signal yields in each q^2 bin were determined via individual fits to the M^2_{miss} distributions restricted within these q^2 ranges. At the current data integrated luminosity, a total of six q^2 bins were defined, the maximum number possible whilst retaining sufficient statistics in each bin. After unfolding the measured signal yields to account for q^2 resolution effects, the partial branching fractions were determined and summed to extract the total branching fractions for each decay, with secondary measurements for these total branching fractions also computed from the signal yields returned from M^2_{miss} fits evaluated over the full q^2 kinematic range.

The distributions of the partial branching fractions as a function of q^2 were ultimately used to extract measurements of $|V_{ub}|$ via a second set of fits in which the value of $|V_{ub}|$ corresponding to the smallest χ^2 defined between the measured and predicted partial branching fractions was determined. Two sets of predictions for the partial branching fraction distributions were derived using the form factor determinations provided by the Fermilab/MILC collaborations in 2015 [96] and the 2021 FLAG Review [97], each using a BCL parameterisation for the $f_+(q^2)$ form factor [34]. Whilst the size of the available data sample was both far below the stated long-term target and sufficiently small that the resultant $|V_{ub}|$ determination was not expected to be competitive with respect to other recent measurements, there was still significant merit to be gained in performing the analysis on this preliminary data-set. Not only was the given analysis a direct test of the chosen procedure, including the effectiveness of the FEI at tagging rare semi-leptonic $b \to u$ decays, but it represented the first tagged measurement of $|V_{ub}|$ produced by Belle II, in turn demonstrating the capability of the experiment to the broader scientific community. In the following sections, a number of conclusions reached during the undertaking of the analysis will be expanded upon in detail.

8.1 Conclusions

8.1.1 Calibrating the FEI

Prior to the founding of the official FEI task force within the Belle II collaboration, a private calibration of the hadronic FEI was performed using the semi-leptonic $B \to X \ell \nu_\ell$ decays as the chosen signal modes, with the various details pertaining to the event selections, reconstruction, and signal extraction strategy provided in Chapter 4. Calibration factors were determined from the ratio of signal events in data and MC post-selections, the former determined via fits to the center-of-mass (CMS) momentum of the signal lepton. In practice, the $B \to X e \nu_e$ and $B \to X \mu \nu_\mu$ modes were reconstructed separately, with calibration factors determined for each mode together with an average factor. Despite the application of the in-

dependent electron and muon identification correction weights to the respective samples, the calibration factors measured from the muon channels, $B^0 \to X^- \mu^+ \nu_\mu$ and $B^+ \to X^0 \mu^+ \nu_\mu$, were significantly lower than the equivalent factors determined via the electron channels, $B^0 \to X^- e^+ \nu_e$ and $B^+ \to X^0 e^+ \nu_e$, with these values being inconsistent within uncertainties. Whilst the exact source of these inconsistencies was unclear, several potential reasons for the observed discrepancies were put forward as a result of the differences in which the electron and muon modes were treated throughout the analysis. As one major example, the momentum of the reconstructed electrons, the very variable used for the signal extraction, was corrected to account for bremsstrahlung radiation, a treatment not performed for the reconstructed muons. In addition, the number of continuum background events remaining post-selections was significantly higher for the muon channels. The relative size of the continuum background had a direct influence on the signal extraction, as not only was the continuum designated as an independent background component in the fits, but a direct constraint was applied to this component based on predictions made using off-resonance data. In the possibility that the discrimination between the various fit components was not sufficiently clear, a difference in the background levels between both channels could manifest as a discrepancy in the resultant calibration factors. To investigate these discrepancies further, future studies could evaluate any potential differences between the calibration factors obtained using a different variable for the signal extraction, one with well-established shape differences between the signal and background components and one that is not directly dependent on reweighting factors such as the bremsstrahlung corrections. In addition, with the collection of more data, higher off-resonance data statistics will allow for a deeper understanding of the continuum background and more precise constraints.

Ultimately, official hadronic FEI calibration factors produced by the FEI task force were used to correct the MC in the subsequent $B \to \pi \ell \nu_\ell$ analysis, as per Belle II recommendations. These factors, determined from a weighted average of two independent calibration studies using signal-side decays of $B \to X \ell \nu_\ell$ and $B \to D^{(*)} \pi$, constituted a more tailored approach in which several B_{tag} decay channels were calibrated individually. As secondary measurements, both studies also quoted global calibration factors averaged over all tagging modes, analogous to those determined in the presented analysis. Several inconsistencies were observed in comparing the global factors resulting from the given analysis and the official calibration, with the current study measuring larger factors overall, implying a better data-MC agreement post-fit. Whilst utilising the same signal decay for the extraction, the $B \to X \ell \nu_\ell$ analyses described in Chapter 4 and those used in the official calibration possessed a number of key differences in the event selections and signal extraction variables, making a direct comparison impossible. Nevertheless, under ideal circumstances, independent analyses should be expected to produce consistent results. In reality, however, some tension exists even between the two sets of factors determined by the official studies, particularly in the calibration of the charged B_{tag} modes, where the global calibration factor computed using the $B^+ \to \bar{D}^0 \pi^+$ signal mode for the reconstruction differed from those derived from the $B^+ \to X^0 e^+ \nu_e$ and $B^+ \to X^0 \mu^+ \nu_\mu$ channels by a value of $\sim 5\sigma$. Isolating a lone reason for this discrepancy is not feasible given the vastly different analyses and their approaches, with deficiencies in MC modelling, poor data-MC agreement in selection variables and statistical fluctuations forming some of several possible explanations. In addition, the official $B \to X \ell \nu_\ell$ study did

not observe any significant discrepancy between the electron and muon channels as opposed to the analysis presented in this report. In addition to the checks proposed above, one point of potential future investigation could focus on adapting the current analysis to better imitate the official $B \to X\ell\nu_\ell$ study in all aspects, in order to pinpoint any clear source of the observed discrepancy.

8.1.2   Analysis Validation

A common tool used throughout the entirety of the $B \to \pi\ell\nu_\ell$ analysis detailed in this thesis was the use of mock data-sets generated from MC for the purpose of validating various aspects of the undertaken study. In following the collaboration guidelines outlined during the peer-review process for the analysis, these data-sets, including both Asimov data and MC toys, were used both to test the robustness and validity of certain analysis techniques such as the fitting algorithms before applying them to real data, and to evaluate the expected performance on this data post-unblinding. The maximum likelihood fitting technique used to extract the signal yields from the M_{miss}^2 distributions was meticulously tested prior to its application to real data. For those distributions evaluated in bins of q^2 as well as over the full q^2 kinematic range, Asimov data-sets were generated directly from the MC expectation at the current data integrated luminosity. As the underlying truth information of the MC could be directly accessed, it was found that fitting these distributions returned signal and background yields that were strictly consistent with the expected yields derived from counting the true numbers of respective events. With the results of these fits validating the fitting algorithm for a single, ideal representation of the expected data-set, further validation studies were performed using MC toys, each generated from Poisson fluctuations of the MC expectation at the same integrated luminosity. Upon fitting the M_{miss}^2 distributions of the order of 1000 toys for each respective q^2 range, the resultant signal yields were found to be consistent in each case with a Gaussian distribution centered around a mean value that was in turn consistent with the expected number of signal events. Such behaviour demonstrated that, in considering a wide range of possible data-sets within the realm of the given MC expectation, the fitting was robust, determining the true underlying signal yields to a high degree of accuracy.

Asimov data was likewise instrumental in validating the chosen q^2 unfolding procedure implemented in order to correct for q^2 resolution effects in which signal events were reconstructed in a different q^2 bin to the underlying reality. By unfolding the fitted Asimov signal yields in each of the six q^2 bins studied and accessing the corresponding truth information, it was able to be directly verified that the unfolding was very successful, recovering the true yields in every case. The unfolded Asimov signal yields were subsequently used to determine a set of Asimov partial and total branching fractions for each reconstructed $B \to \pi\ell\nu_\ell$ mode. With the MC used to derive the Asimov data-set having been generated and reweighted to match the world average measurements of the respective branching fractions, this served as an additional test of the method of computing the $B \to \pi\ell\nu_\ell$ branching fractions. Each Asimov branching fraction measured was consistent with the world average measurements,

lending confidence and validity to the given procedure.

Finally, the χ^2 minimisation fitting procedure used for the extraction of $|V_{\mathrm{ub}}|$ from the partial branching fraction distributions was similarly validated using Asimov data before its application to real data. In fitting the Asimov partial branching fraction distributions alongside both sets of constraints from LQCD, values consistent with the world average measurement were likewise extracted. The goodness-of-fit measures for each fit, the values of the post-fit χ^2 divided by the number of degrees of freedom, were well within reasonable limits. Thus, simulated Asimov and toy data-sets were utilised for the successful validation of each aspect of the given analysis, each yielding positive results consistent with expectations.

8.1.3 Signal and Background Discrimination

In order to effectively isolate signal events in real data and extract estimates of the signal yields, optimising the discrimination between signal and background events was a significant component of the analysis. The event selections imposed during the reconstruction were chosen specifically for the rejection of the various backgrounds whilst retaining as many signal events as possible. A dedicated optimisation process was implemented for the tuning of the event selections in Chapter 5, with the $S/\sqrt{S+B}$ metric used to find the optimal value of each selection corresponding to the largest signal significance. As part of this process, a set of boosted decision trees (BDTs) were trained to preferentially reject the sizeable continuum backgrounds remaining after the application of the rest of the analysis selections. As judged by the distributions of the resultant classifier outputs in Figures 5.26 – 5.29, this training was especially effective, with a clear discrimination between continuum and other $\Upsilon(4S) \rightarrow B\bar{B}$ events observed across all cases.

In addition to the careful consideration of the event selections, the M^2_{miss} variable used for the signal extraction was similarly chosen for its high discriminating power. A manifestation of the unique advantage of the use of hadronic tagging in the reconstruction of a semi-leptonic signal decay, signal events were expected to peak at a value of $M^2_{\mathrm{miss}} = 0$ due to the kinematic constraints of the lone missing neutrino, unlike background events under no such constraints. In observing the relevant distributions in MC after all analysis selections (Figures 5.30 – 5.37), the discrimination in M^2_{miss} between signal events and the various backgrounds was excellent. Clear signal peaks were observed for both $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays, even in the latter case where the resultant peak was broader due to the poorer resolution of reconstructing the neutral pion from its two photon daughters. The invariant mass selection imposed on the π^0 meson in Section 5.5.1 was also instrumental in constraining the width of this signal M^2_{miss} peak for reconstructed $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays.

The clear discrimination between signal and background events was further reinforced during a test performed in Section 6.5.1, in which the robustness of the fits to the M^2_{miss} distributions of reconstructed $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ decays was evaluated in the case of splitting the background template into two components. For these decays in particular, the $B^0 \rightarrow \rho^- \ell^+ \nu_\ell$ background, though small, was sizeable enough that a handful of events fell within the signal region of $-1.0 \leqslant M^2_{\mathrm{miss}} \leqslant 1.0 \, \mathrm{GeV}^2$ for the distributions in each q^2 bin. In order to investigate

whether any signal events were being wrongfully assigned as $B^0 \to \rho^- \ell^+ \nu_\ell$ background events in real data, or vice versa, the $B^0 \to \rho^- \ell^+ \nu_\ell$ background was isolated from the remaining backgrounds and designated as a separate component in an additional set of maximum likelihood fits. The yields of the $B^0 \to \rho^- \ell^+ \nu_\ell$ backgrounds in each of these fits were then fixed to the expectation derived from an independent Belle II measurement of the $B^0 \to \rho^- \ell^+ \nu_\ell$ branching fraction. The fitted signal yields returned from these three-component fits were found to not only be consistent with, but to barely shift from those yields returned from the nominal two-component fits. A similar test was performed for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, with the shift in the fitted yields found to be negligible. Thus, as fixing the size of a dominant background within the signal region to its expected value had no significant effect on the signal yields, it was clear that the fitting algorithm was able to distinguish well between signal and background events, even within the signal-rich region of the distributions.

8.1.4 The q^2 Spectrum

For the given $B \to \pi \ell \nu_\ell$ analysis, the total branching fractions were determined from the sum of the partial branching fractions extracted in each of six q^2 bins. In this way, the q^2 dependence of the measurements could be taken into account, particularly when considering the subsequent extraction of $|V_{\mathrm{ub}}|$ from the the partial branching fraction distributions. As a general principle, when performing such determinations in distinct q^2 bins, resolution effects arising from detector limitations should be corrected for via a q^2 unfolding procedure. In order to determine the relative size and influence of the resolution effects on the analysis, q^2 bin migration matrices were constructed from MC (Figures 7.1 and 7.2), illustrating the proportion of signal events reconstructed in each q^2 bin compared with the q^2 bins to which the events actually belonged in the underlying MC truth. As only six q^2 bins were chosen for the analysis at the current data integrated luminosity, the width of each bin was relatively large. As a result, minimal bin migration was observed, with the vast majority of events being reconstructed in the same q^2 bin as they belonged to in truth. Higher levels of bin migration were observed for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays than for $B^0 \to \pi^- \ell^+ \nu_\ell$ decays due to the reconstructed π^0 resolution, though the magnitude of the observed effect was still acceptable.

After the subsequent unfolding of the q^2 spectrum using the event distribution given in the migration matrices, it was found that the unfolded data signal yields were overwhelmingly consistent within uncertainties with the raw yields returned from the M_{miss}^2 fits, a direct consequence of the minimal bin migration and the large statistical uncertainties. Thus, it was concluded that the q^2 unfolding at the current data integrated luminosity had no significant impact on the analysis. At higher integrated luminosities with a finer granularity in q^2, the implementation of the q^2 unfolding would be expected to become especially important, having a significantly higher impact on the resultant partial branching fractions.

As a secondary measurement and a cross-check, the total $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions were also determined via M_{miss}^2 fits performed over the full q^2 kinematic range. In each case, these branching fractions were found to be consistent with those deter-

mined from the respective sums of the partial branching fractions, lending further confidence to the given treatment of the q^2 dependency in the derivation of the partial branching fractions.

8.1.5 Low Measured Branching Fractions and $|V_{\mathrm{ub}}|$

Upon unblinding the real Belle II data following the analysis validation, maximum likelihood fits to the M_{miss}^2 distributions were performed and estimates of the data signal yields were extracted, both for those distributions separated into bins of q^2 and those evaluated over the full q^2 range. In each case, an expected signal yield was also derived from the number of signal MC events remaining post-selections, reweighted by all relevant data-MC corrections and normalised to the data integrated luminosity. In comparing the magnitude of the fitted yields from data with the MC expectations, a general trend was observed whereby the data yields fell below the expected yields in the majority of cases. In considering the overall discrepancies averaged over q^2, the signal yields returned from the fits over the full q^2 range resulted in total statistical deviations of $\sim 2.6\sigma$ and $\sim 2.4\sigma$ for all reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, respectively.

Naturally, when using these signal yields post-unfolding to extract measurements of the partial and total $B \to \pi \ell \nu_\ell$ branching fractions, these discrepancies were similarly manifested in the final results. With overall consistency observed between the total branching fractions evaluated from the sum of the partial branching fractions and those extracted from the M_{miss}^2 fits performed over the full q^2 range, both the measurements of the branching fractions $\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell)$ and $\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell)$ were found to be below the world average results [20]. These results are summarised in Table 8.1, with the main branching fraction and $|V_{\mathrm{ub}}|$ measurements presented alongside the current world averages. A larger discrepancy was observed for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, $\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell) = (1.14 \pm 0.11_{\mathrm{stat}} \pm 0.06_{\mathrm{sys}}) \times 10^{-4}$ (sum of partial branching fractions), consistent with a total deviation of $\sim 2.7\sigma$ below the world average result, $\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell) = (1.50 \pm 0.06) \times 10^{-4}$. For reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, a total deviation of $\sim 2.0\sigma$ was observed between the sum of the measured partial branching fractions, $\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (6.15 \pm 0.62_{\mathrm{stat}} \pm 0.55_{\mathrm{sys}}) \times 10^{-5}$, and the world average result, $\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (7.80 \pm 0.27) \times 10^{-5}$.

Whilst the observed discrepancies are not outside the realm of potential statistical fluctuations, it is also possible that these were not the sole cause of the results obtained, particularly given the fact that these discrepancies were found across multiple q^2 bins for both the charged and neutral pion channels. It is possible that the observed behaviour was related to another yet unknown source of deficiency in the MC modelling or the analysis techniques and tools. As mentioned in Section 8.1.3, an additional test was performed to investigate whether the current fitting algorithm was capable of distinguishing well between signal and background $B^0 \to \rho^- \ell^+ \nu_\ell$ events, particularly in the signal region in which the size of the latter component was not negligible. However, when modifying the fitting procedure to fix the yield of the $B^0 \to \rho^- \ell^+ \nu_\ell$ component to its expectation, no significant change in the signal yields was observed, countering the hypothesis for which some proportion of true signal

events in data may have been assigned as background events by the fit, returning lower signal yields than expected. With the results of this test, together with the extensive fit validation and the good post-fit data-MC agreement observed in the relevant M^2_{miss} distributions, it was concluded that the low measured branching fractions were likely not caused by any issue related to the signal extraction procedure.

	Measured value	World average [20]
$\mathcal{B}(B^0 \rightarrow \pi^- e^+ \nu_e)$	$(1.24 \pm 0.16_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$	
$\mathcal{B}(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)$	$(1.05 \pm 0.15_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$	$(1.50 \pm 0.06) \times 10^{-4}$
$\mathcal{B}(B^0 \rightarrow \pi^- \ell^+ \nu_\ell)$	$(1.14 \pm 0.11_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$	
$\mathcal{B}(B^+ \rightarrow \pi^0 e^+ \nu_e)$	$(6.37 \pm 0.91_{\text{stat}} \pm 0.56_{\text{sys}}) \times 10^{-5}$	
$\mathcal{B}(B^+ \rightarrow \pi^0 \mu^+ \nu_\mu)$	$(5.58 \pm 0.93_{\text{stat}} \pm 0.52_{\text{sys}}) \times 10^{-5}$	$(7.80 \pm 0.27) \times 10^{-5}$
$\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$	$(6.15 \pm 0.62_{\text{stat}} \pm 0.55_{\text{sys}}) \times 10^{-5}$	
$|V_{\text{ub}}|$	$(3.17 \pm 0.25) \times 10^{-3}$ (Fermilab/MILC 2015 [96])	$(3.70 \pm 0.10 \pm 0.12) \times 10^{-3}$
	$(3.21 \pm 0.23) \times 10^{-3}$ (FLAG 2021 [97])	(Exclusive world. avg.)

Table 8.1: Summary of the main results of the given analysis measured using 362 fb^{-1} of Belle II data, compared with the current world average results. All branching fractions listed refer to those derived from the sums of the partial branching fractions measured in six q^2 bins, and include both results from the individual and combined lepton channels, where $\ell = e$ or μ. The quoted $|V_{\text{ub}}|$ values refer to those derived from the simultaneous χ^2 fits to the $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ partial branching fraction distributions, for both sets of LQCD constraints studied.

Given this conclusion, the only other component of the analysis with the potential to really produce discrepancies of this magnitude was the FEI. Of the various MC weighting factors affecting the computation of the signal reconstruction efficiencies used as input in the branching fraction formulae, the hadronic FEI calibration factors were the most significant, representing some of the largest deviations from unity. The FEI calibration also formed a dominant source of systematic uncertainty for the analysis. The effects of any potential deficiencies in the training of the FEI algorithm would in principle be carried through to the FEI calibration and in turn, the present analysis. Previous Belle II measurements made using the hadronic FEI, including the ratio $R(X_{e/\mu}) = \mathcal{B}(B \rightarrow X e \nu_e)/\mathcal{B}(B \rightarrow X \mu \nu_\mu)$ [102] and $|V_{\text{cb}}|$ from $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ decays [103], utilised the algorithm to tag semi-leptonic signal-side decays consisting of largely $B \rightarrow X_c \ell \nu_\ell$ contributions. The analysis presented within this thesis thus represented one of the first tests of the FEI in tagging a $B \rightarrow X_u \ell \nu_\ell$ signal mode. One hypothesis formed in light of the low measured partial branching fractions was that the current FEI training may not be appropriate for use in such an application for several reasons. Firstly, the training was performed using the order of 200 million $\Upsilon(4S) \rightarrow B\bar{B}$ events. Given the rarity of the $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays, with branching fractions of the order of 10^{-4}, it was possible that too few events were present in the training sample to accurately represent the tagging of these $B \rightarrow X_u \ell \nu_\ell$ signal-side decays, particularly compared with the abundance of $B \rightarrow X_c \ell \nu_\ell$ events. As an additional differ-

ence between the current $B \to X_u \ell \nu_\ell$ and $B \to X_c \ell \nu_\ell$ analyses, the multiplicity, defined as the number of daughter particle tracks, is substantially lower for $B^0 \to \pi^- \ell^+ \nu_\ell$ (two) and $B^+ \to \pi^0 \ell^+ \nu_\ell$ (one) decays when compared with the various $B \to X_c \ell \nu_\ell$ modes that typically involve multiple charged pion and kaon tracks resulting from D and D^* meson decays. One point of potential future investigation could be the recreation of the present analysis using a version of the FEI that has been trained on a $B \to X_u \ell \nu_\ell$ rich sample such as the hybrid MC described in Section 5.1.1.

As an additional point, when considering the branching fractions measured independently from the individual electron and muon channels, it was observed that for both the charged and neutral pion modes, the central values of the branching fractions measured from the electron channels were higher than those measured from the muon channels. For reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays, the $B^0 \to \pi^- \mu^+ \nu_\mu$ branching fraction, $\mathcal{B}(B^0 \to \pi^- \mu^+ \nu_\mu) = (1.05 \pm 0.15_{\text{stat}} \pm 0.06_{\text{sys}}) \times 10^{-4}$ was found to be slightly inconsistent with the $B^0 \to \pi^- e^+ \nu_e$ branching fraction within uncertainties, $\mathcal{B}(B^0 \to \pi^- e^+ \nu_e) = (1.24 \pm 0.16_{\text{stat}} \pm 0.08_{\text{sys}}) \times 10^{-4}$. A similar trend was observed for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, though the branching fractions remained consistent within uncertainties in this case, where $\mathcal{B}(B^+ \to \pi^0 e^+ \nu_e) = (6.37 \pm 0.91_{\text{stat}} \pm 0.56_{\text{sys}}) \times 10^{-5}$ and $\mathcal{B}(B^+ \to \pi^0 \mu^+ \nu_\mu) = (5.58 \pm 0.93_{\text{stat}} \pm 0.52_{\text{sys}}) \times 10^{-5}$. Though the magnitudes of the respective uncertainties were substantially larger in this case, these observations mirrored the discrepancies observed between the calibration factors determined for the hadronic FEI in Chapter 4. As described in Section 8.1.1, the calibration factors derived using signal-side $B \to X e \nu_e$ decays were larger than those derived from $B \to X \mu \nu_\mu$ decays, implying a larger gap between the number of predicted and measured $B \to X \mu \nu_\mu$ events than observed for $B \to X e \nu_e$. Given that no such disagreement can be seen between the branching fractions derived from the two lepton channels in Asimov data, it is possible that the observed discrepancies in data are a result of the same underlying cause as that responsible for the disagreement observed in the measured calibration factors. However, statistical fluctuations cannot be discounted as the reason behind the differences considering the relative size of the uncertainties on the branching fraction measurements.

Due to the discrepancies observed in the partial branching fractions with respect to the MC expectation and the world averages, the values of $|V_{\text{ub}}|$ extracted from the χ^2 fits to these distributions were likewise found to be lower than predicted, with the world average measured from exclusive decays corresponding to $|V_{\text{ub}}| = (3.70 \pm 0.10_{\text{exp.}} \pm 0.12_{\text{theo.}}) \times 10^{-3}$. The values extracted from the individual χ^2 fits to the partial branching fraction distributions of the single lepton channels, with the exception of those measured from $B^0 \to \pi^- \mu^+ \nu_\mu$ decays, were consistent with the world average within the relatively large uncertainties. However, in reducing the size of these uncertainties with the larger samples corresponding to the combined lepton cases, the resultant values of $|V_{\text{ub}}|$ were no longer consistent with the world average, with $|V_{\text{ub}}|(B^0 \to \pi^- \ell^+ \nu_\ell) = (3.16 \pm 0.30) \times 10^{-3}$ and $|V_{\text{ub}}|(B^+ \to \pi^0 \ell^+ \nu_\ell) = (3.22 \pm 0.34) \times 10^{-3}$ for the Fermilab/MILC LQCD determinations, and $|V_{\text{ub}}|(B^0 \to \pi^- \ell^+ \nu_\ell) = (3.24 \pm 0.28) \times 10^{-3}$ and $|V_{\text{ub}}|(B^+ \to \pi^0 \ell^+ \nu_\ell) = (3.29 \pm 0.32) \times 10^{-3}$ for the FLAG LQCD determinations. As observed, the LQCD constraints taken from the 2021 FLAG Review [97], which combined the Fermilab/MILC predictions [96] with additional constraints from RBC/UKQCD [98], corresponded to measured central values that were closer to the stated

world average. The values determined from the neutral pion mode were likewise in better agreement with the world average than those from the charged pion mode, a direct consequence of the larger deviation observed in the measured $B^0 \to \pi^- \ell^+ \nu_\ell$ branching fractions.

The primary measurements of $|V_{\mathrm{ub}}|$ extracted during the analysis, however, were those that combined the partial branching fraction distributions of the reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in a simultaneous χ^2 fit for maximum statistical power. Using this fitting framework, the precision on $|V_{\mathrm{ub}}|$ was able to be reduced to $\sim 8\%$ and $\sim 7\%$ for the Fermilab/MILC and FLAG LQCD determinations, respectively, compared with the precision of the individual χ^2 fits ranging between $9 - 13\%$. The values obtained from these fits as listed in Table 8.1, $|V_{\mathrm{ub}}|$ (Fermilab/MILC) $= (3.17 \pm 0.25) \times 10^{-3}$ and $|V_{\mathrm{ub}}|$ (FLAG) $= (3.21 \pm 0.23) \times 10^{-3}$, were likewise found to be below the world average, each with a total deviation of $\sim 2.1\sigma$. This discrepancy will be able to be investigated further with the collection of more data, resulting in smaller statistical uncertainties and a higher precision on the measurement. Ultimately, global averages for $|V_{\mathrm{ub}}|$ are typically determined via fits to experimental data from multiple sources, and including various inputs such as those described in Section 1.1.1.

8.2   Future Directions

In addition to the points of potential further investigation detailed in the previous sections, a set of long-term goals and aspirations can be set for the analysis moving forward. With the collection of more data and the opportunity and time to further develop analysis techniques and methods, the future outlook for the hadronic tagged analysis of the semi-leptonic $B \to \pi \ell \nu_\ell$ decay is both positive and highly anticipated. In the following sections, the estimated future projections for the branching fraction and $|V_{\mathrm{ub}}|$ measurements as well as ideas for new techniques involving combined studies with $B \to \rho \ell \nu_\ell$ will be discussed.

8.2.1   Projections for Higher Integrated Luminosities

The measurements of the $B \to \pi \ell \nu_\ell$ branching fractions and the magnitude of the matrix element $|V_{\mathrm{ub}}|$ presented throughout this report were determined using a preliminary subset of Belle II data corresponding to an integrated luminosity of 362 fb^{-1}. As stated in the opening of this chapter, by the completion of the experiment, Belle II aims to record a total of 50 ab^{-1} of data, an amount over 130 times the size of the current data-set. Figure 8.1 depicts the expected targets for the peak luminosity and integrated luminosity of Belle II over the course of the next decade, with the end of the experiment expected around the year 2035. With the running of the experiment expected to resume in early 2024 after the end of the current shut-down period, Belle II is on track to exceed the size of the full Belle data-set by roughly the end of 2026, with an integrated luminosity of greater than 1 ab^{-1} expected by the end of 2027. Another long shut-down is planned on the timescale of 2027–2028 to work on machine improvements, as illustrated by the drop in the peak luminosity observed

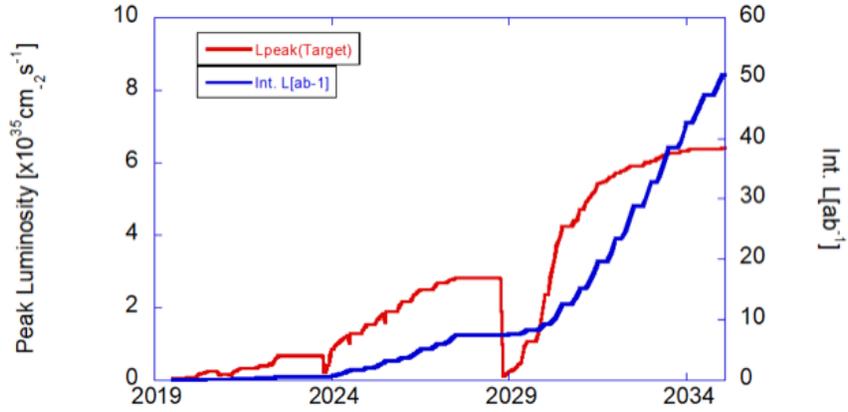in Figure 8.1 just before the planned resumption of data-taking.

Figure 8.1: Projections for the peak luminosity and integrated luminosity recorded by Belle II over the course of the experiment [104].

Using the uncertainties on the current measurements of the $B \to \pi \ell \nu_\ell$ branching fractions, rough projections can be made for the expected precision at various data integrated luminosities including 1 ab^{-1} and 50 ab^{-1}. For simplicity, given that the statistical and systematic uncertainties on the total branching fractions derived from the two independent methods were largely equivalent, the total branching fractions determined from the M^2_{miss} fits evaluated over the full q^2 range will be used for the projections. The statistical precision, or the statistical uncertainty quoted as a percentage of the measured branching fraction, can be extrapolated to larger sample sizes by considering the relative size of the Poisson error on the signal yields returned by the M^2_{miss} fits. At a value of N times the size of the current data-set, corresponding to signal yields a factor of N times larger than the nominal yields, the size of the resultant statistical uncertainty can be estimated via Poisson statistics as $\sigma_{\mathrm{stat,\ proj.}} = \sqrt{N \times S_{\mathrm{current}}} = \sqrt{N} \times \sqrt{S_{\mathrm{current}}}$, where S_{current} refers to the signal yields at the current data integrated luminosity, and thus $\sqrt{S_{\mathrm{current}}}$ represents the Poisson statistical uncertainty on these yields. Since the statistical uncertainties on the branching fractions are directly proportional to the statistical uncertainties on the yields, it directly follows that the expected statistical precision at a data-set N times the size of the current sample can be determined by scaling the current statistical precision by a factor of $1/\sqrt{N}$. In this way, using the statistical uncertainties quoted as percentages of the total $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions in Tables 7.10 and 7.13, estimates for the statistical precision at the integrated luminosity of 1 ab^{-1} expected within the next few years can be determined. Given that this integrated luminosity corresponds to a data-set 2.8 times the size of the present sample, the statistical precision can be expected to improve from 9.6% to 5.8% for the $B^0 \to \pi^- \ell^+ \nu_\ell$ branching fraction, and from 10.1% to 6.1% for the $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fraction on this timescale.

With regards to the systematic precision, similar projections can be made for higher integrated luminosities, though their determination is not as straightforward as those made

Integrated luminosity (ab^{-1})	$N\times$ current data-set	Statistical precision (%)	Systematic precision (%)	Total precision (%)
$B^0 \to \pi^- \ell^+ \nu_\ell$				
0.36	1.0	9.6	5.6	$<\sim11.1$
0.57	1.6	7.6	< 5.6	$<\sim9.4$
1.00	2.8	5.8	< 5.6	$<\sim8.0$
1.33	3.7	5.0	< 5.6	$<\sim7.5$
50.00	138	0.81	< 5.6	$<\sim5.6$
$B^+ \to \pi^0 \ell^+ \nu_\ell$				
0.36	1.0	10.1	8.6	$<\sim13.2$
0.57	1.6	8.0	< 8.6	$<\sim11.7$
1.00	2.8	6.1	< 8.6	$<\sim10.5$
1.33	3.7	5.3	< 8.6	$<\sim10.1$
50.00	138	0.86	< 8.6	$<\sim8.6$

Table 8.2: Projections of the statistical, systematic and total uncertainties quoted as percentages of the measured $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions, for various data integrated luminosities.

using the statistical uncertainties. In reality, the various sources of systematic uncertainty contributing to the branching fraction calculations were the result of the uncertainties associated with several quantities relevant to the analysis including multiple correction factors, branching fractions and form factor parameters. The uncertainties on these quantities were themselves comprised of both statistical and systematic contributions, the former expected to scale with increasing sample size in a similar manner to described above, and the latter expected to remain at roughly the same magnitude in the absence of improvements to analysis techniques, detector specifications or theoretical modelling. In order to derive a conservative estimate for the systematic precision at 1 ab^{-1}, the expected reduction in the systematic uncertainties corresponding to the most dominant sources for the current analysis were investigated. As mentioned in Section 7.1.3, these were identified as the FEI calibration and the π^0 reconstruction efficiency, with the latter only relevant to the $B^+ \to \pi^0 \ell^+ \nu_\ell$ case. At the current data integrated luminosity, the statistical and systematic uncertainties for the FEI calibration factors were found to be of the same order, with the systematic uncertainties already overwhelming the statistical uncertainties for the π^0 efficiency corrections. As these respective systematic, irreducible uncertainties were expected in both cases to dominate over the statistical uncertainties at 1 ab^{-1}, the total systematic precision of the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions measured at the current data integrated luminosity was kept the same for the projections as a rough, conservative upper limit on the expected values. As quoted in Tables 7.10 and 7.13, these values corresponded to 5.6% and 8.6% for the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions, respectively. Combining the percentage uncertainties of the statistical and systematic components in quadrature to derive the total precision expected at 1 ab^{-1}, values of 8.0% and 10.5% were obtained for $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$, respectively, compared with the precision obtained by the respective current

results, 11.1% and 13.2%. The results of these projections are summarised in Table 8.2.

Given the projected size of the statistical and systematic uncertainties at 1 ab^{-1}, it can be seen that at this data integrated luminosity, the total uncertainty on the $B^0 \to \pi^- \ell^+ \nu_\ell$ branching fraction is still expected to be dominated by the statistical uncertainty, whilst a larger systematic component is expected for the projected $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fraction. For each case, using the same estimation procedure for the projected statistical precision as above, the value of the data integrated luminosity for which the systematic uncertainty is expected to become dominant can be directly evaluated. Such a point would roughly occur at statistical uncertainties of $\sim 5\%$ and $\sim 8\%$ of the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions, respectively, corresponding to samples with integrated luminosities of 1.33 ab^{-1} and 570 fb^{-1}.

Finally, projections for the full Belle II data-set can be made for the 50 ab^{-1} of data expected by the completion of the experiment, a data-set equivalent to ~ 138 times the current sample size. Scaling the statistical precision in the same way as described above, a dramatic improvement can be anticipated at this integrated luminosity, with projected values of 0.81% and 0.86% for the $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions, respectively. Assuming roughly the same systematic uncertainties as currently observed, projections for the total percentage uncertainties on the respective branching fractions predict values of $\sim 5.6\%$ and $\sim 8.6\%$ at most. However, it should be noted that in addition to the slight reduction in the magnitude of the systematic uncertainties expected at higher integrated luminosities, improvements to analysis techniques, detector components and theoretical determinations over the course of the next decade will likely also contribute to reducing the size of the current systematic uncertainties substantially. Thus, given the fact that the uncertainties on the current $B^0 \to \pi^- \ell^+ \nu_\ell$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fractions measured during this study were limited by statistics, significant improvements to the precision of these measurements can be anticipated with more data, with reductions of the order of $\sim 28\%$ and $\sim 20\%$, respectively, already expected by the collection of 1 ab^{-1} of data. With the full projected Belle II data-set, these respective uncertainties have the potential to be reduced by the order of $\sim 50\%$ and $\sim 35\%$.

With the improvements to the precision of the measured total, and by extension, partial, branching fractions expected at higher data integrated luminosities, the total uncertainties on future measurements of $|V_{\mathrm{ub}}|$ are likewise expected to reduce in size. However, in addition to the statistical and systematic uncertainties on the partial branching fractions, the extraction of $|V_{\mathrm{ub}}|$ is also highly dependent on the magnitude of the theoretical uncertainties on the BCL parameters used in the determination of the $f_+(q^2)$ form factor. Whilst the $|V_{\mathrm{ub}}|$ values extracted during the present analysis were largely statistically limited, at the integrated luminosity thresholds for which statistical contributions no longer dominate the total uncertainties, these theoretical uncertainties are expected to become the dominant component of the total uncertainty on $|V_{\mathrm{ub}}|$. After reaching these thresholds, further increasing the data sample size will no longer result in a significant reduction in the total uncertainty. However, LQCD determinations are also expected to increase in precision with time as larger scale simulations become increasingly more possible with the development of more powerful
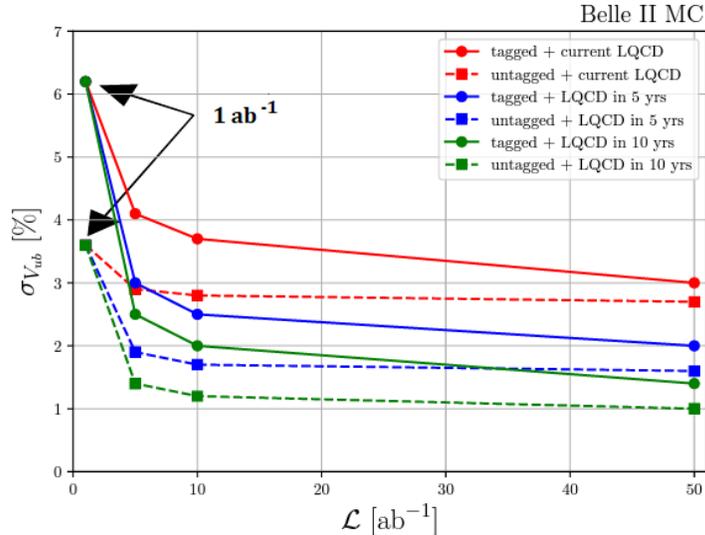
computing resources [28].

Figure 8.2: Projections of the uncertainty on measurements of $|V_{\mathrm{ub}}|$ extracted from exclusive $B \to \pi \ell \nu_\ell$ decays, for various data integrated luminosities. Both projections for tagged and untagged analyses are included, using several different predictions for the size of LQCD uncertainties over a ten year period [28].

Figure 8.2 taken from Ref. [28] illustrates a projection based on Belle II MC for the expected precision on $|V_{\mathrm{ub}}|$ as a function of the data integrated luminosity, with $|V_{\mathrm{ub}}|$ extracted solely from exclusive $B \to \pi \ell \nu_\ell$ decays. Both the projections for tagged and untagged analyses are included, for various different assumptions on the magnitude of the theoretical uncertainties derived from LQCD over the course of the next five to ten years. Whilst untagged analyses are expected to correspond to the largest reduction in the $|V_{\mathrm{ub}}|$ uncertainty due to their higher reconstruction efficiencies, significant improvements to the precision of tagged extractions can also be achieved at higher integrated luminosities. Tagged and untagged determinations are ultimately complementary in nature given the trade-off between efficiency and purity between the two methods, with tagged analyses typically resulting in a much purer reconstruction. At 1 ab^{-1}, using present-day LQCD constraints, a total precision of $\sim 6\%$ is predicted for the tagged $B \to \pi \ell \nu_\ell$ extraction, compared with the $\sim 7 - 8\%$ achieved by the current analysis. Taking projected improvements to LQCD simulations over the next decade into account, $|V_{\mathrm{ub}}|$ measurements extracted from hadronic tagged $B \to \pi \ell \nu_\ell$ decays reconstructed using the full expected Belle II data-set have the potential to report uncertainties of the order of 1%. Thus, the preliminary data-set used for the branching fraction and $|V_{\mathrm{ub}}|$ determinations detailed throughout this study demonstrates a first investigation into the types of precise measurements capable by Belle II, with world-leading results expected to emerge in the coming years.

8.2.2 Combined Analyses

Whilst several semi-leptonic $B \to X_u \ell \nu_\ell$ decays exist, each involving a $b \to u$ quark transition, measurements of $|V_{ub}|$ from exclusive semi-leptonic decays have historically largely been extracted solely from the $B \to \pi \ell \nu_\ell$ channel. This has long been the preferred choice for the extraction from both an experimental and theoretical standpoint. The $B \to \pi \ell \nu_\ell$ decay is both straightforward to reconstruct experimentally due to its low multiplicity, and benefits from a relatively clean reconstruction whereby signal and background events can be discriminated to a high degree as demonstrated by the present analysis. The $B \to \pi \ell \nu_\ell$ decay is also theoretically the most simple $B \to X_u \ell \nu_\ell$ decay to model given the scalar nature of the pion involved in the decay, with the differential decay rate dependent on the determination of a single analytical form factor. This contrasts the $B \to \pi \ell \nu_\ell$ decay from other $B \to X_u \ell \nu_\ell$ modes such as $B \to \rho \ell \nu_\ell$, where the vector ρ-meson corresponds to a much broader width in the distribution of the invariant mass, as illustrated in Figures 5.1 and 5.2. Studies on the $B \to \rho \ell \nu_\ell$ decay are also being undertaken at Belle II [1], with first measurements of the $B \to \rho \ell \nu_\ell$ branching fractions having been determined.

Measurements of $|V_{ub}|$ can in theory also be extracted from other semi-leptonic $B \to X_u \ell \nu_\ell$ modes such as $B \to \rho \ell \nu_\ell$. Whilst performing such an extraction may not be an optimal or desirable choice in lieu of a $B \to \pi \ell \nu_\ell$ extraction, there exists the potential to perform a combined analysis of multiple $B \to X_u \ell \nu_\ell$ modes including $B \to \pi \ell \nu_\ell$. If a set of partial branching fractions in bins of q^2 could be determined for $B \to \rho \ell \nu_\ell$ decays, a value of $|V_{ub}|$ could potentially be extracted from a simultaneous χ^2 fit to the partial branching fraction distributions of $B \to \pi \ell \nu_\ell$ and $B \to \rho \ell \nu_\ell$. Such a task would introduce a fair amount of technical and theoretical complexity as the $B \to \rho \ell \nu_\ell$ differential decay rate would have to be incorporated into the framework, corresponding to an additional three form factors $(A_1(q^2), A_2(q^2)$ and $V(q^2))$ which would require a suitable parameterisation [32]. Great care would also need to be taken to ensure that all correlations between the systematic uncertainties of the $B \to \pi \ell \nu_\ell$ and $B \to \rho \ell \nu_\ell$ partial branching fractions are taken into account. With all of these elements considered, however, performing an extraction of $|V_{ub}|$ using this combined approach has the potential to not only increase the statistical precision but to constrain the measurement of $|V_{ub}|$ using input from both $B \to \pi \ell \nu_\ell$ and $B \to \rho \ell \nu_\ell$ decays.

To conclude, it is clear that the prospective growth of the $B \to \pi \ell \nu_\ell$ analysis presented throughout this thesis stands to contribute significantly to future exclusive measurements of $|V_{ub}|$, ultimately leading to the improved precision of a key Standard Model parameter. With the collecting and processing of more data together with the potential evolution and innovation of analysis techniques, new and exciting developments from Belle II in the field of semi-leptonic B-meson decays can be anticipated in future.

---

[1] Whilst the two analyses were conducted independently, the hadronic tagged $B \to \rho \ell \nu_\ell$ analysis performed by Moritz Bauer of Karlsruhe Institute of Technology and the $B \to \pi \ell \nu_\ell$ analysis presented in this thesis are currently being peer reviewed in tandem within the Belle II Collaboration for the purpose of publishing a combined journal paper.

Bibliography

[1] N. Toutounji, Reconstruction Methods for Semi-leptonic Decays of B-mesons with the Belle II Experiment, Masters thesis, The University of Sydney, 2018.

[2] F. Abudinen et al., Belle II Collaboration, Exclusive $B^0 \to \pi^- \ell^+ \nu_\ell$ Decays with Hadronic Full Event Interpretation Tagging in 34.6 fb^{-1} of Belle II Data, arXiv:2008.08819 [hep-ex], 2020.

[3] F. Abudinen et al., Belle II Collaboration, Exclusive $B \to X_u \ell \nu_\ell$ Decays with Hadronic Full-event-interpretation Tagging in 62.8 fb^{-1} of Belle II Data, arXiv:2111.00710 [hep-ex], 2021.

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

[5] M. Kobayashi, T. Maskawa, CP-Violation in the Renormalizable Theory of Weak Interaction, Prog. of Theo. Phys. 49, 2, 652–657, 1973.

[6] ATLAS Collaboration, Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC, Phys. Lett. B716, 1-29, 2012.

[7] A. D. Sakharov, Violation of CP Invariance, C Asymmetry and Baryon Asymmetry of the Universe, Journal of Exp. and Theo. Phys. 5, 24-27, 1967.

[8] A. Bilal, Introduction to Supersymmetry, NEIP-01-001, arXiv:hep-th/0101055, 2001.

[9] T. G. Rizzo, Pedagogical Introduction to Extra Dimensions, SLAC-PUB-10753, arxiv-hep-ph/0409309, 2004.

[10] R. D. Peccei, The Strong CP Problem and Axions, Lect. Notes in Phys. 741, 3-17, 2008.

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

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

[13] T. Keck et al., The Full Event Interpretation – An Exclusive Tagging Algorithm for the Belle II Experiment, Comput. Softw. Big Sci 3, 6, 2019.

[14] Public domain image.

[15] S. S. Lee, A Model of Quantum Gravity with Emergent Spacetime. J. High Energ. Phys. 70, 2020.

[16] P. W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, Phys. Rev. Lett. 13, 508, 1964.

[17] M. Spira, P. M. Zerwas, Electroweak Symmetry Breaking and Higgs Physics, CERN-TH/97-379, DESY 97-261, arXiv:hep-ph/9803257, 1997.

[18] S. Raby, Grand Unified Theories, arXiv:hep-ph/0608183, 2006.

[19] F. De Aquino, Theory of Everything, arXiv:gr-qc/9910036, 2013.

[20] R.L. Workman et al., Particle Data Group, Prog. Theor. Exp. Phys. 083C01, 2022 and 2023 Update.

[21] E. Majorana, Teoria Simmetrica dell'Elettrone e del Positrone, Il Nuovo Cimento 14, 171-184, 1937.

[22] Y. Fukuda et al., Super-Kamiokande Collaboration, Evidence for Oscillation of Atmospheric Neutrinos, Phys. Rev. Lett. 81, 8, 1562–1567, 1998.

[23] L. Wolfenstein, Parametrization of the Kobayashi-Maskawa Matrix, Phys. Rev. Lett. 51, 1945.

[24] J. Charles et al., CKMfitter Group, Eur. Phys. J. C41, 1-131, arXiv:hep-ph/0406184, 2005.

[25] M. Bona et al., UTfit Collaboration, The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of δms and BR($B \to \tau\nu$), J. High Energ. Phys. 10, 081, arXiv:hep-ph/0606167, 2006.

[26] CKMfitter Group, Updated Results on the CKM Matrix (Preliminary), 2022, http:ckmfitter.in2p3.fr/

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

[28] E. Kou et al., The Belle II Physics Book, Prog. of Theo. and Exp. Phys. 12, 2019.

[29] O. Eberhardt et al., Status of the Two-Higgs-Doublet Model of Type II, J. High Energ. Phys. 07, 118, 2013.

[30] M. Schmaltz, Y-M. Zhong, The Leptoquark Hunter's Guide: Large Coupling, J. High Energ. Phys. 132, 2019.

[31] Y. Amhis et al., Averages of b-hadron, c-hadron, and τ-lepton Properties as of 2021, Phys. Rev. D 107, 052008, 2023.

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

[33] C. G. Boyd, B. Grinstein, R. F. Lebed, Constraints on Form Factors for Exclusive Semi-leptonic Heavy to Light Meson Decays, Phys. Rev. Lett. 74, 23, 1995.

[34] C. Bourrely, L. Lellouch, I. Caprini, Model-independent Description of $B \to \pi l \nu$ Decays and a Determination of $|V_{ub}|$, Phys. Rev. D 79, 013008, 2009.

[35] B. Aubert et al., BaBar Collaboration, $B \to \pi \ell \nu$ and $B \to \rho \ell \nu$ Decays and Determination of $|V_{ub}|$, Phys. Rev. D 72, 051102, 2005.

[36] P. del Amo Sanchez et al., BaBar Collaboration, Study of $B \to \pi \ell \nu$ and $B \to \rho \ell \nu$ Decays and Determination of $|V_{ub}|$, Phys. Rev. D 83, 032007, 2011.

[37] J. P. Lees et al., BaBar Collaboration, Branching Fraction and Form-factor Shape Measurements of Exclusive Charmless Semi-leptonic B Decays, and Determination of $|V_{ub}|$, Phys. Rev. D 86, 092004, 2012.

[38] S. B. Athar et al., CLEO Collaboration, Study of the q^2-dependence of $B \to \pi \ell \nu$ and $B \to \rho(\omega) \ell \nu$ Decay and Extraction of $|V_{ub}|$, Phys. Rev. D 68, 072003, 2003.

[39] N. E. Adam et al., CLEO Collaboration, A Study of Exclusive Charmless Semi-leptonic B Decay and $|V_{ub}|$, Phys. Rev. Lett. 99, 041802, 2007.

[40] T. Hokuue et al., Belle Collaboration, 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, Phys. Lett. B 648, 139-148, 2007.

[41] F. Abudinen et al., Belle II Collaboration, Study of Exclusive $B \to \pi e \nu_e$ Decays with Hadronic Full-event-interpretation Tagging in 189.3 fb^{-1} of Belle II Data, BELLE2-CONF-PH-2022-008, arXiv:2206.08102 [hep-ex], 2022.

[42] M. Okamoto et al., Semi-leptonic $D \to \pi/K$ and $B \to \pi/D$ Decays in 2+1 Flavor Lattice QCD, Nucl. Phys. Proc. Suppl. 140, 461-463, 2005.

[43] E. Gulez et al., B Meson Semi-leptonic Form Factors from Unquenched Lattice QCD, Phys. Rev. D 73, 074502, 2006; Erratum- Phys. Rev. D 75, 119906, 2007.

[44] KEK Press Release, Kick-off of the Belle II Phase 3 Physics Run, 2019.

[45] N. Toge et al., KEKB B-Factory Design Report, KEK Report 95-7, KEK, 1995.

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

[47] Belle II Software Documentation on Sphinx,
https://software.belle2.org/development/sphinx/index.html

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

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

[50] Belle II Collaboration, Snowmass Whitepaper: The Belle II Detector Upgrade Program, arXiv:2203.11349v1 [hep-ex], 2022.

[51] C. Z. Yuan, Belle II Collaboration, The Belle II Experiment at the SuperKEKB, Presented at Flavor Physics and CP Violation, Hefei, China, arXiv:1208.3813 [physics.ins-det], 2012.

[52] B. Wang, Belle II PID Group, The Assembly of the Belle II TOP Counter, Nucl. Inst. and Meth. in Phys. Res. A 766, 204-207, 2014.

[53] S. Sandilya, Belle II Collaboration, Particle Identification with the TOP and ARICH Detectors at Belle II, UCHEP-17-04, arXiv:1706.08515 [physics.ins-det], 2017.

[54] T. Auschev et al., A scintillator Based Endcap K_L and Muon Detector for the Belle II Experiment, Nucl. Inst. and Meth. in Phys. Res. A 789, 134-142, 2015.

[55] A. Abashian et al., Muon Identification in the Belle Experiment at KEKB, Nucl. Inst. and Meth. in Phys. Res. A 491, 69–82, 2002.

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

[57] T. Kuhr et al., The Belle II Core Software, Comput. Softw. Big Sci. 3, 1, 2019.

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

[59] R. Brun, F. Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. and Meth. in Phys. Res. A 389, 81-86, 1997.

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

[61] S. Agostinelli et al., Geant4 — a Simulation Toolkit, Nucl. Inst. and Meth. in Phy. Res. A 506, 250-303, 2003.

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

[63] M. Feindt et al., The NeuroBayes Neural Network Package, Nucl. Inst. and Meth. in Phy. Res. A 559, 1, 190-194, 2006.

[64] A. Hoecker et al., TMVA: Toolkit for Multivariate Data Analysis, CERN-OPEN-2007-007, arXiv:physics/0703039 [physics.data-an], 2007.

[65] S. Neuhaus et al., A Neural Network z-vertex Trigger for Belle II, J. of Phys.: Conf. Ser. 608, 012052, 2015.

[66] T. Keck, FastBDT: A Speed-optimized and Cache-friendly Implementation of Stochastic Gradient-boosted Decision Trees for Multivariate Classification, arXiv:1609.06119 [cs.LG], 2016.

[67] J. Friedman, Stochastic Gradient Boosting, Comp. Stat. and Data Ana. 38, 4, 367-378, 1999.

[68] T. Keck, Machine Learning Algorithms for the Belle II Experiment and their Validation on Belle Data, EKP-2017-00067, 2017.

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

[70] F. Abudinen et al., Belle II Collaboration, A Calibration of the Belle II Hadronic Tag-side Reconstruction Algorithm with $B \to X\ell\nu_\ell$ Decays, arXiv:2008.06096 [hep-ex], 2020.

[71] J. Stypula, M. Rozanska et al., Belle Collaboration, Evidence for $B^- \to D_S^+ K^- \ell^- \bar{\nu}_\ell$ and Search for $B^- \to D_S^{*+} K^- \ell^- \bar{\nu}_\ell$, Phys. Rev. D 86, 072007, 2012.

[72] M. Prim, b2-hive/eFFORT v0.1.0, Zenodo, 2020.

[73] F. U. Bernlochner et al., $B \to \rho\ell\nu$ and $B \to \omega\ell\nu$ in and Beyond the Standard Model: Improved Predictions and $|V_{ub}|$, Phys. Rev. D 104, 034032, 2021.

[74] D. Scora, N. Isgur, Semi-leptonic Meson Decays in the Quark Model: An Update, Phys. Rev. D 52, 2783–2812, 1995.

[75] G. Duplancic, B. Melic, Form Factors of B, $B_s \to \eta, \eta'$ and $D, D_s \to \eta, \eta'$ Transitions from QCD Light-cone Sum Rules, J. High Energ. Phys. 138, 2015.

[76] C. G. Boyd, B. Grinstein, R. F. Lebed, Model-Independent Determinations of $B \to D\ell\nu$, $B \to D^*\ell\nu$ Form Factors, Nucl. Phys. B 461, 493-511, 1996.

[77] R. Glattauer et al., Measurement of the Decay $B \to D\ell\nu$ in Fully Reconstructed Events and Determination of the Cabibbo-Kobayashi-Maskawa Matrix Element $|V_{cb}|$, Phys. Rev. D 93, 032006, 2016.

[78] D. Ferlewicz et al., Revisiting Fits to $B^0 \to D^{*-}\ell^+\nu_\ell$ to Measure $|V_{cb}|$ with Novel Methods and Preliminary LQCD Data at Nonzero Recoil, Phys. Rev. D 103, 073005, 2021.

[79] A. Abdesselam et al., Belle Collaboration, Precise Determination of the CKM Matrix Element $|V_{cb}|$ with $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ Decays with Hadronic Tagging at Belle, BELLE-CONF-1612, arXiv:1702.01521 [hep-ex], 2017.

[80] `pidvar`, private software package developed by Dr. William Sutcliffe of the University of Bonn, Belle II Collaboration.

[81] N. Tan et al., Magnetic Field Mapping of the Belle Solenoid, IEEE Transactions on Nuclear Science 48, 3, 900-907, 2001.

[82] Q.-D. Zhou, Correction for Tracking Momentum Bias Based on Invariant Mass Peak Studies, BELLE2-NOTE-PH-2020-030, 2020.

[83] S. Schmitt, Data Unfolding Methods in High Energy Physics, Proceedings of "XII Quark Confinement and the Hadron Spectrum", EPJ Web of Conferences, arXiv:1611.01927 [physics.data-an], 2016.

[84] R. Barlow, C. Beeston, Fitting Using Finite Monte Carlo Samples, Comput. Phys. Commun. 77, 219–228, 1993.

[85] W. Verkerke, D. P. Kirkby, The RooFit Toolkit for Data Modeling, Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05): Proceedings, Oxford, UK, eConf C0303241, 2003.

[86] C. Ramirez, J.F. Donoghue, G. Burdman, Semi-leptonic $b \to u$ Decay, Phys. Rev. D 41, 5, 1990.

[87] B. O. Lange, M. Neubert, G. Paz, Theory of Charmless Inclusive B Decays and the Extraction of V_{ub}, Phys. Rev. D 72, 073006, 2005.

[88] Modified from image by William Sutcliffe of the University of Bonn, Belle II Collaboration.

[89] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic Formulae for Likelihood-based Tests of New Physics, Eur. Phys. J. C 71, 1554, 2011, Erratum: Eur. Phys. J. C 73, 2501, 2013.

[90] D. M. Asner et al., CLEO Collaboration, Search for Exclusive Charmless Hadronic B Decays, Phys. Rev. D 53, 3, 1039-1050, 1996.

[91] E. L. Lehmann, J. P. Romano, Testing Statistical Hypotheses, Springer Texts in Statistics, 513–517, 2005.

[92] K. Adamczyk et al., Belle II Collaboration, Determination of V_{ub} from Untagged $B^0 \to \pi^- \ell^+ \nu_\ell$ Decays Using 2019–2021 Belle II Data, BELLE2-CONF-PH-2022-017, arXiv:2210.04224 [hep-ex], 2022.

[93] S. Choudhury et al., Belle Collaboration, Measurement of the B^+/B^0 Production Ratio in e^+e^- Collisions at the $\Upsilon(4S)$ Resonance Using $B \to J/\psi(\ell\ell)K$ Decays at Belle, Phys. Rev. D 107, 3, 2023.

[94] H. Dembinski, M. Schmelling, Bias, Variance, and Confidence Intervals for Efficiency Estimators in Particle Physics Experiments, arXiv:2110.00294 [stat.AP], 2022.

[95] T. Adye, Unfolding Algorithms and Tests using RooUnfold, Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, arXiv:1105.1160 [physics.data-an], 2011.

[96] J. A. Bailey et al., $|V_{\mathrm{ub}}|$ from $B \to \pi\ell\nu$ Decays and 2 + 1-flavor Lattice QCD, Phys. Rev. D 92, 1, 2015.

[97] Y. Aoki et al., FLAG Review 2021, Eur. Phys. J. C 82, 10, 2022.

[98] J. M. Flynn et al., $B \to \pi \ell \nu$ and $B_s \to K \ell \nu$ Form Factors and $|V_{ub}|$ from 2+1-flavor Lattice QCD with Domain-wall Light Quarks and Relativistic Heavy Quarks, Phys. Rev. D 91, 7, 2015.

[99] `VubFitter`, private software package developed by Professor Florian Bernlochner of the University of Bonn, Belle II Collaboration.

[100] K. L. Wuensch, Chi-Square Tests, International Encyclopedia of Statistical Science, Springer, 2014.

[101] F. James, MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1, CERN-D-506, 1994.

[102] L. Aggarwal et al., Belle II Collaboration, Test of Light-Lepton Universality in the Rates of Inclusive Semi-leptonic B-Meson Decays at Belle II, Phys. Rev. Lett. 131, 5, 051804, 2023.

[103] F. Abudinen et al., Belle II Collaboration, Measurement of the $B^0 \to D^{*-} \ell^+ \nu_\ell$ Branching Ratio and $|V_{cb}|$ with a Fully Reconstructed Accompanying B-meson in 2019–2021 Belle II Data, arXiv:2301.04716 [hep-ex], 2023.

[104] Belle II public luminosity projection plot, https://confluence.desy.de/display/BI/Belle+II+Luminosity

[105] `WG1Template`, private software package developed by Dr. William Sutcliffe and Henrik Junkerkalefeld of the University of Bonn, Belle II Collaboration.

Appendix A

Decay Modes of Intermediate Particles Supported by the Full Event Interpretation Algorithm

The following table lists each of the decay modes supported by the hadronic Full Event Interpretation (FEI) algorithm, together with their designated `decayModeID`s, for all intermediate particles reconstructed prior to the reconstruction of the parent B-mesons. The hadronic B-meson decay channels are listed in the main matter, in Table 3.3.

decayModeID	Channel	decayModeID	Channel
	π^0		D_S^+
0	$\pi^0 \to \gamma\gamma$	0	$D_S^+ \to K^+ K_S^0$
	K_S^0	1	$D_S^+ \to K^+ \pi^+ \pi^-$
0	$K_S^0 \to \pi^+\pi^-$	2	$D_S^+ \to K^+ K^- \pi^+$
1	$K_S^0 \to \pi^0\pi^0$	3	$D_S^+ \to K^+ K^- \pi^+ \pi^0$
	J/ψ	4	$D_S^+ \to K^+ K_S^0 \pi^+ \pi^-$
0	$J/\psi \to e^+ e^-$	5	$D_S^+ \to K^- K_S^0 \pi^+ \pi^+$
1	$J/\psi \to \mu^+ \mu^-$	6	$D_S^+ \to K^+ K^- \pi^+ \pi^+ \pi^-$
	D^0	7	$D_S^+ \to \pi^+ \pi^+ \pi^-$
0	$D^0 \to K^- \pi^+$	8	$D_S^+ \to K_S^0 \pi^+$
1	$D^0 \to K^- \pi^+ \pi^0$	9	$D_S^+ \to K_S^0 \pi^+ \pi^0$
2	$D^0 \to K^- \pi^+ \pi^0 \pi^0$		D_S^{*+}
3	$D^0 \to K^- \pi^+ \pi^+ \pi^-$	0	$D_S^{*+} \to D_S^+ \gamma$
4	$D^0 \to K^- \pi^+ \pi^+ \pi^- \pi^0$	1	$D_S^{*+} \to D_S^+ \pi^0$
5	$D^0 \to \pi^- \pi^+$		Λ^0
6	$D^0 \to \pi^- \pi^+ \pi^+ \pi^-$	0	$\Lambda^0 \to p^+ \pi^-$
7	$D^0 \to \pi^- \pi^+ \pi^0$		Σ^+
8	$D^0 \to \pi^- \pi^+ \pi^0 \pi^0$	0	$\Sigma^+ \to p^+ \pi^0$
9	$D^0 \to K_S^0 \pi^0$		Λ_c^+
10	$D^0 \to K_S^0 \pi^+ \pi^-$	0	$\Lambda_c^+ \to p^+ K^- \pi^+$
11	$D^0 \to K_S^0 \pi^+ \pi^- \pi^0$	1	$\Lambda_c^+ \to p^+ \pi^- \pi^+$
12	$D^0 \to K^- K^+$	2	$\Lambda_c^+ \to p^+ K^- K^+$
13	$D^0 \to K^- K^+ \pi^0$	3	$\Lambda_c^+ \to p^+ K^- \pi^+ \pi^0$
14	$D^0 \to K^- K^+ K_S^0$	4	$\Lambda_c^+ \to p^+ K^- \pi^+ \pi^0 \pi^0$
	D^+	5	$\Lambda_c^+ \to p^+ \pi^+ \pi^+ \pi^- \pi^-$
0	$D^+ \to K^- \pi^+ \pi^+$	6	$\Lambda_c^+ \to p^+ K_S^0$
1	$D^+ \to K^- \pi^+ \pi^+ \pi^0$	7	$\Lambda_c^+ \to p^+ K_S^0 \pi^0$
2	$D^+ \to K^- K^+ \pi^+$	8	$\Lambda_c^+ \to p^+ K_S^0 \pi^+ \pi^-$
3	$D^+ \to K^- K^+ \pi^+ \pi^0$	9	$\Lambda_c^+ \to \Lambda^0 \pi^+$
4	$D^+ \to \pi^+ \pi^0$	10	$\Lambda_c^+ \to \Lambda^0 \pi^+ \pi^0$
5	$D^+ \to \pi^+ \pi^+ \pi^-$	11	$\Lambda_c^+ \to \Lambda^0 \pi^+ \pi^- \pi^+$
6	$D^+ \to \pi^+ \pi^+ \pi^- \pi^0$	12	$\Lambda_c^+ \to \Lambda^0 \pi^+ \gamma$
7	$D^+ \to K_S^0 \pi^+$	13	$\Lambda_c^+ \to \Lambda^0 \pi^+ \pi^0 \gamma$
8	$D^+ \to K_S^0 \pi^+ \pi^0$	14	$\Lambda_c^+ \to \Lambda^0 \pi^+ \pi^- \pi^+ \gamma$
9	$D^+ \to K_S^0 \pi^+ \pi^+ \pi^-$	15	$\Lambda_c^+ \to \Sigma^+ \pi^+ \pi^-$
10	$D^+ \to K^+ K_S^0 K_S^0$	16	$\Lambda_c^+ \to \Sigma^+ \pi^+ \pi^- \pi^0$
	D^{*0}	17	$\Lambda_c^+ \to \Sigma^+ \pi^0$
0	$D^{*0} \to D^0 \pi^0$		
1	$D^{*0} \to D^0 \gamma$		
	D^{*+}		
0	$D^{*+} \to D^0 \pi^+$		
1	$D^{*+} \to D^+ \pi^0$		
2	$D^{*+} \to D^+ \gamma$		

Table A.1: The decay modes of intermediate particles supported by the hadronic FEI algorithm, along with their `decayModeID` flags.

Appendix B

Additional Post-Fit Distributions for $B \to \pi \ell \nu_\ell$ Decays Reconstructed From 362 fb^{-1} of Belle II Data, in Six q^2 Bins

The following figures include additional distributions of key kinematic quantities reconstructed in Belle II data against predictions from Monte Carlo (MC). The MC has been fitted to the data using the binned maximum likelihood procedure described in Section 6.1, with each distribution analogous to the post-fit M_{miss}^2 distributions included in Figures 6.35 and 6.39.
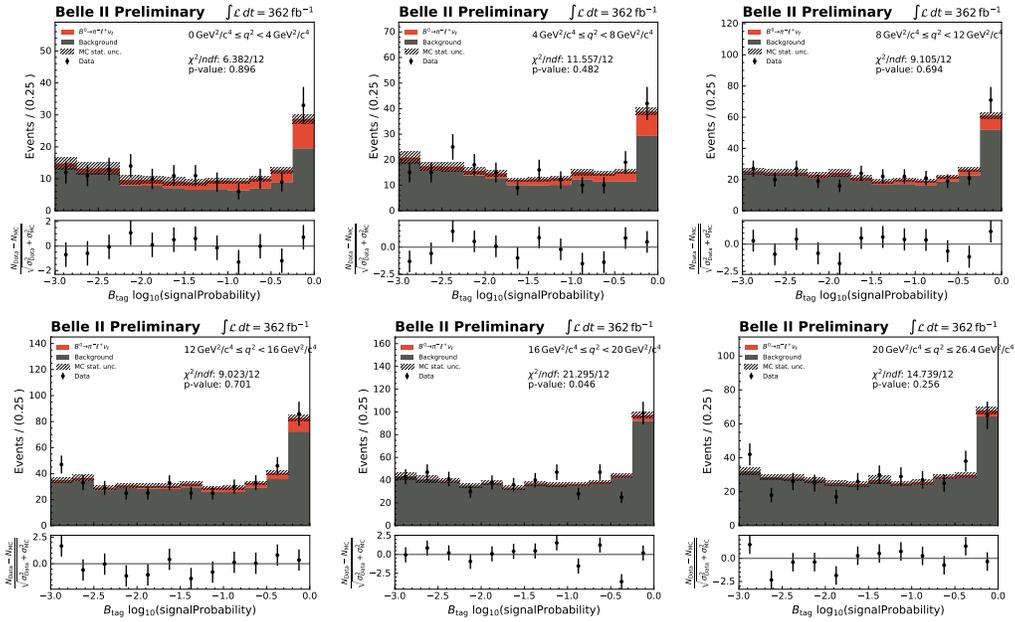
Figure B.1: Distributions of the log of the FEI classifier output, $\log_{10}(\texttt{SignalProbability})$, built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.
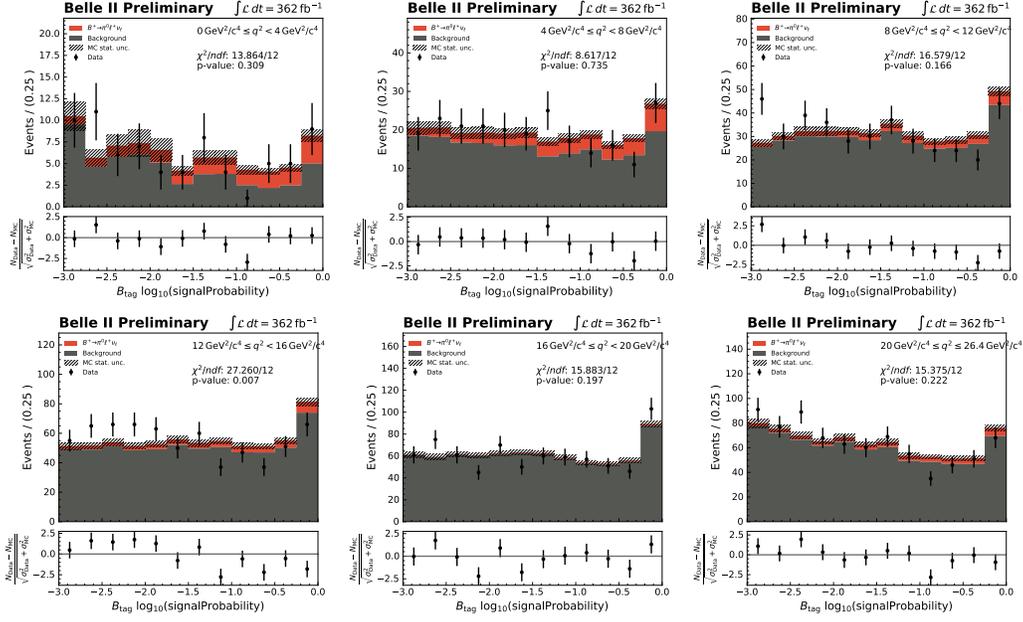
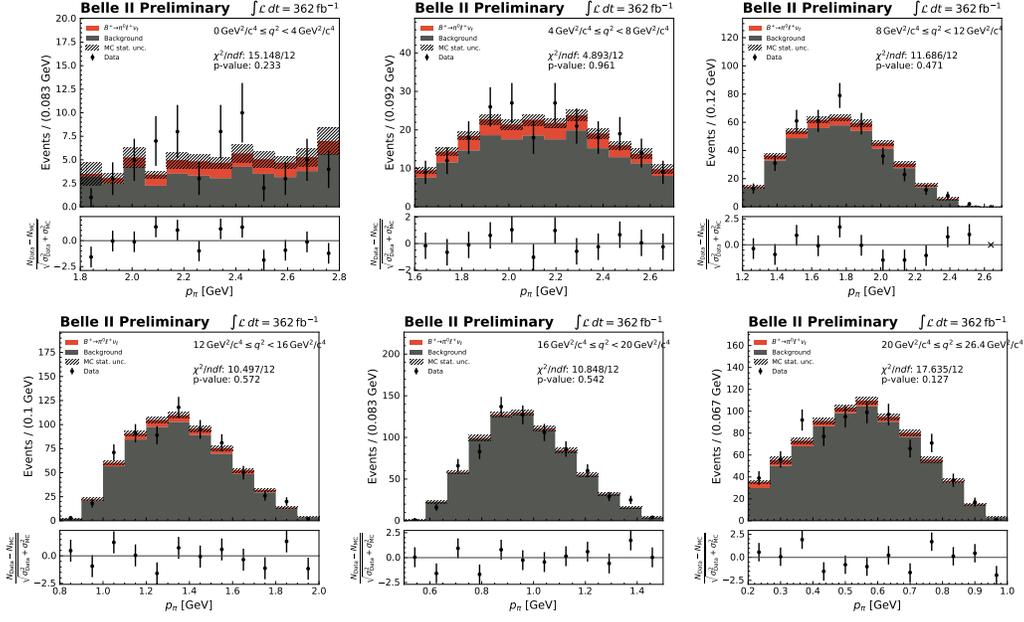Figure B.2: Distributions of the log of the FEI classifier output, $\log_{10}(\texttt{SignalProbability})$, built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.
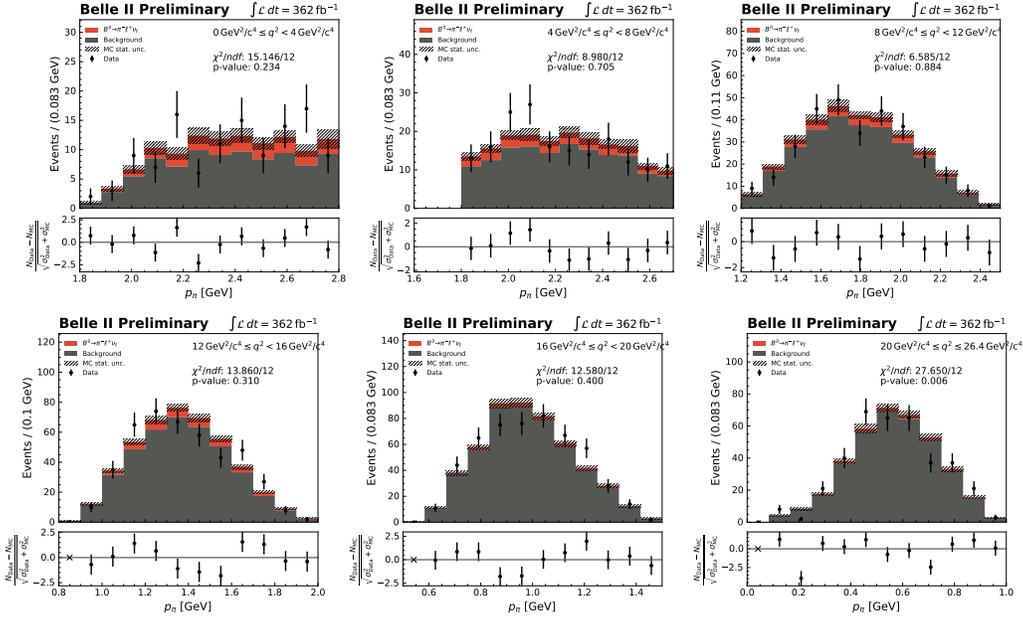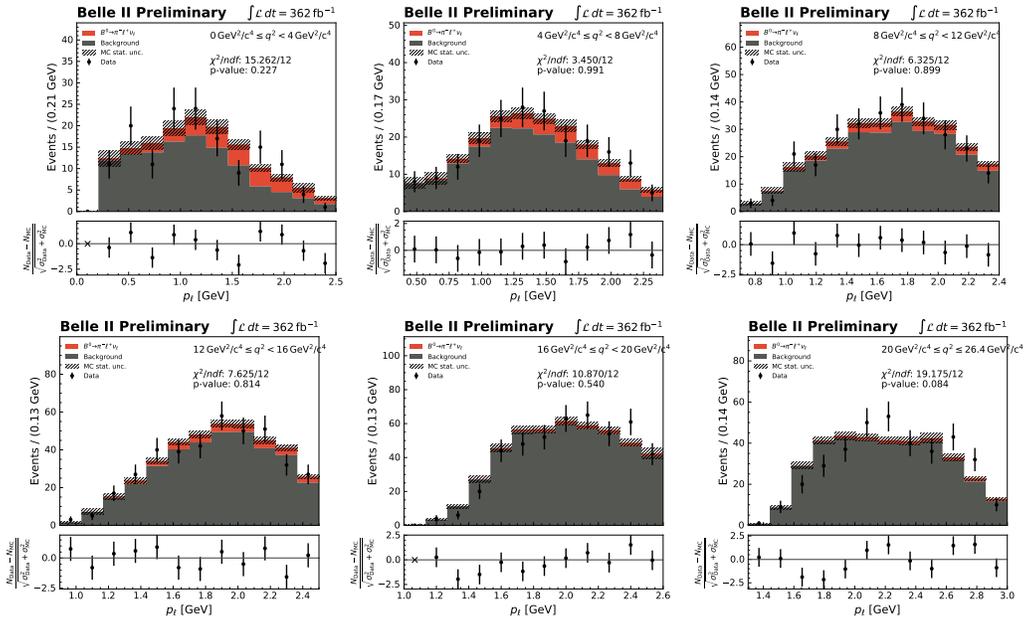

Figure B.3: Distributions of the lab-frame momentum p_π built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.

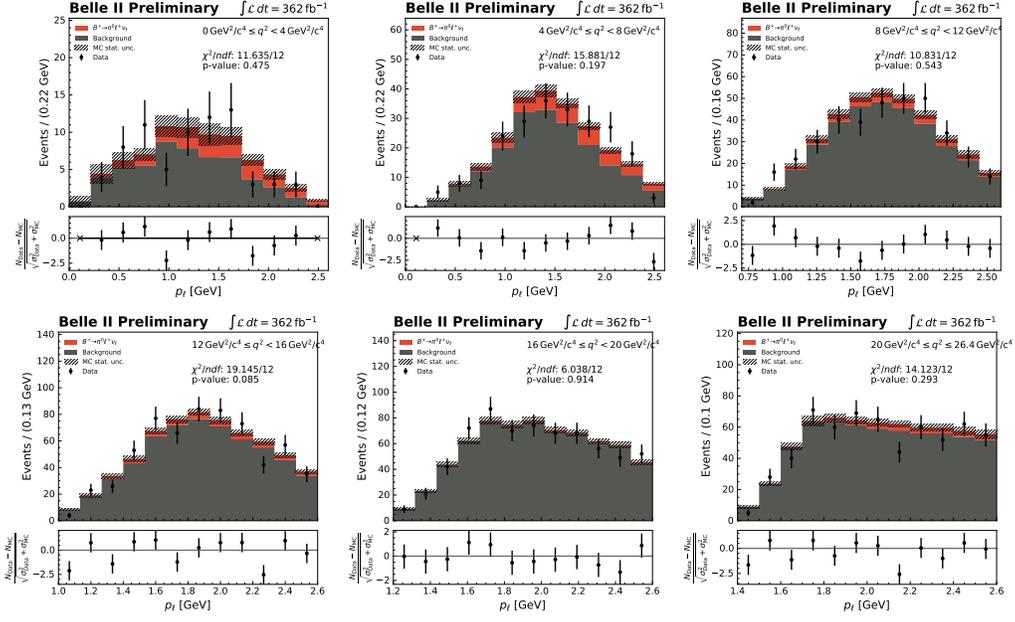Figure B.4: Distributions of the lab-frame momentum p_π built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.

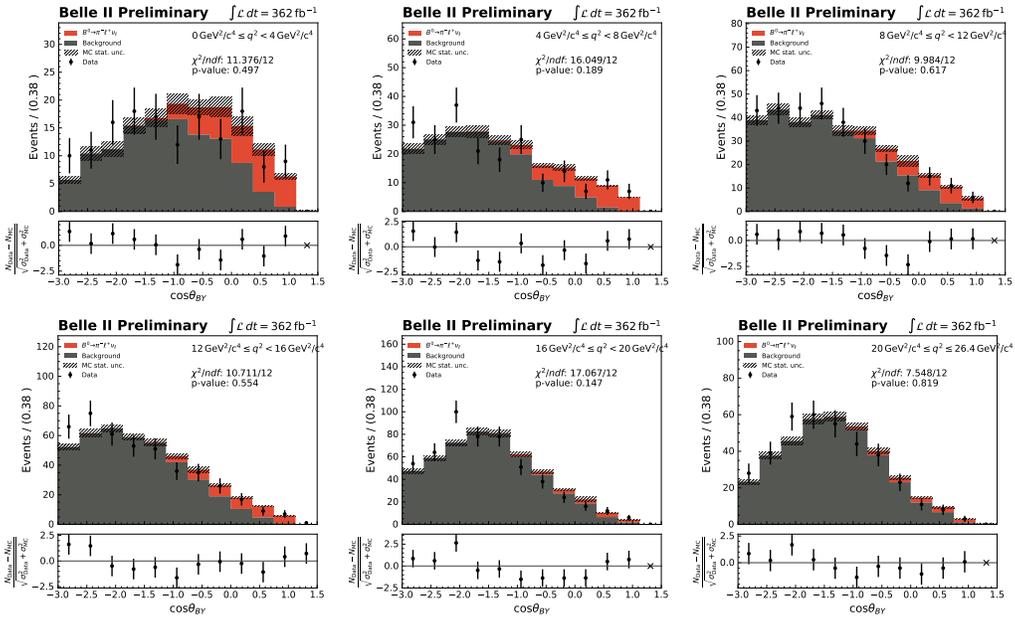Figure B.5: Distributions of the lab-frame momentum p_ℓ built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.
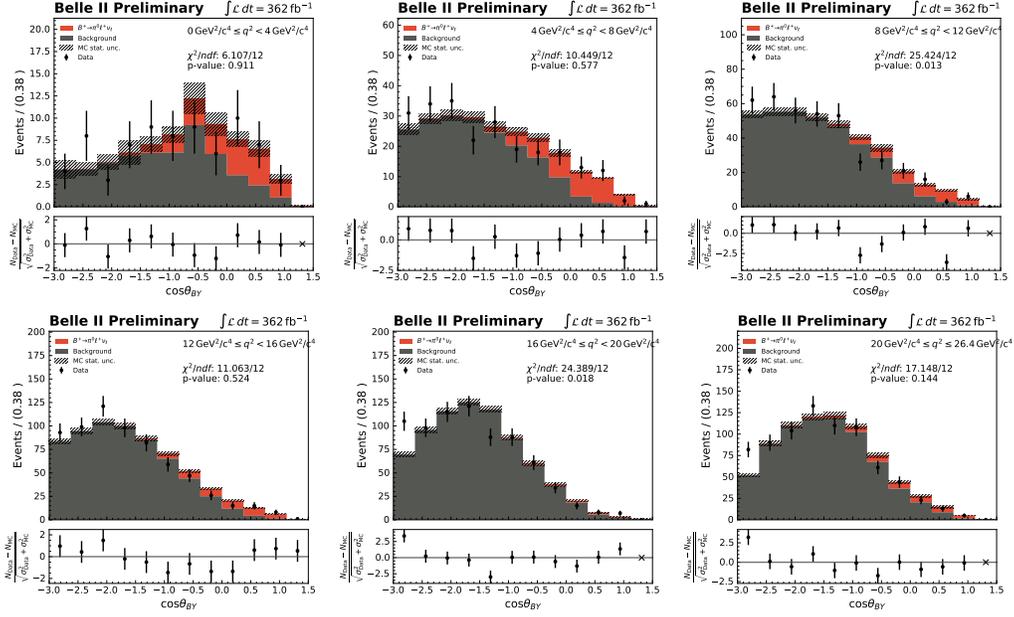
Figure B.6: Distributions of the lab-frame momentum p_ℓ built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.

Figure B.7: Distributions of B_{sig} $\cos\theta_{BY}$ built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.
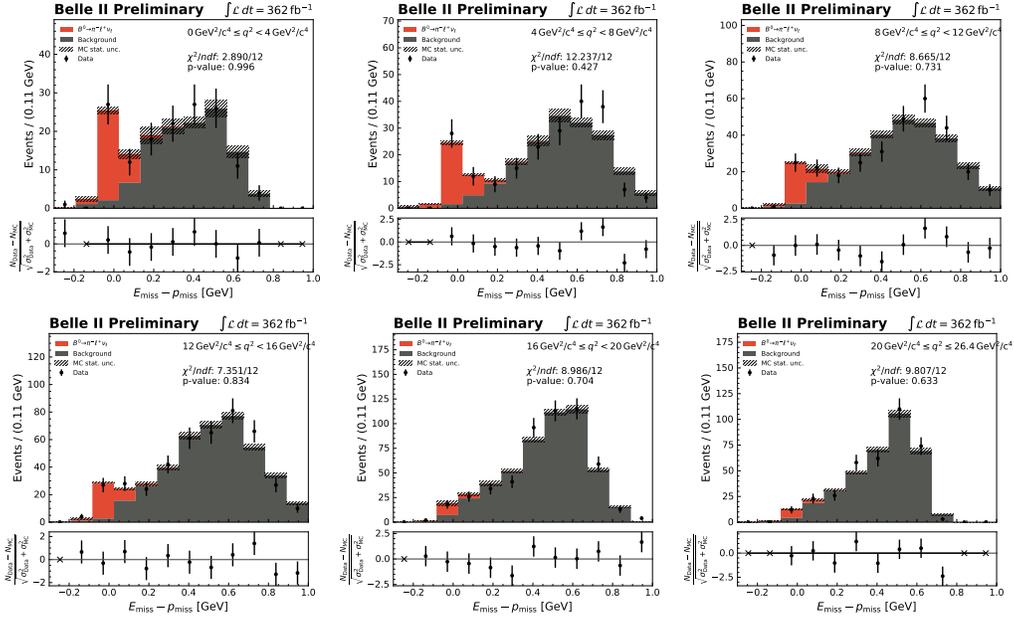
Figure B.8: Distributions of B_{sig} $\cos\theta_{BY}$ built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.

Figure B.9: Distributions of $E_{\mathrm{miss}} - |\vec{p}_{\mathrm{miss}}|$ built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^0 \to \pi^- \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.
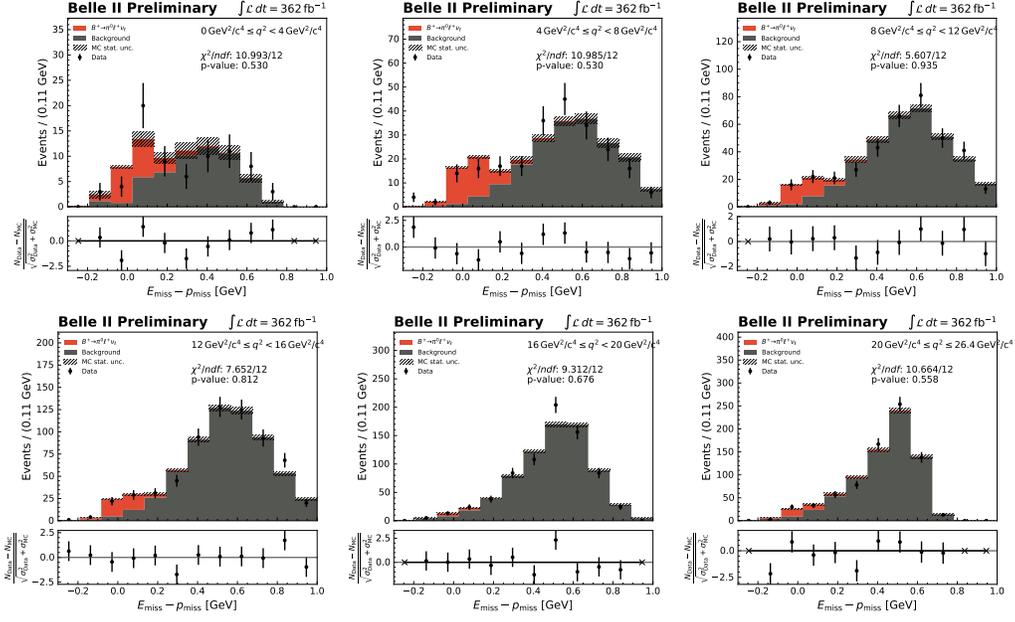
Figure B.10: Distributions of $E_{\text{miss}} - |\vec{p}_{\text{miss}}|$ built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.
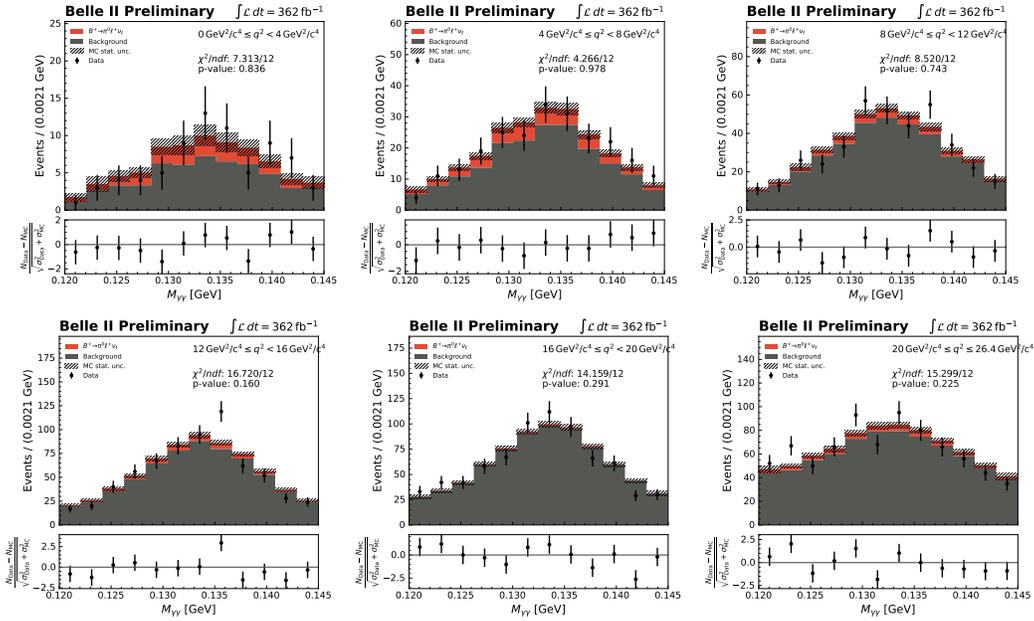

Figure B.11: Distributions of the π^0 invariant mass, $M_{\gamma\gamma}$, built from 362 fb^{-1} of Belle II data with MC fit projections overlaid, for reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays in six q^2 bins, where $\ell = e$ or μ.