TECHNISCHE
UNIVERSITÄT
WIEN

DISSERTATION

Searches for New Physics using Missing Energy at the Belle II Experiment

zur Erlangung des akademischen Grades
Doktor der Technischen Wissenschaften

ausgeführt an der Technischen Universität Wien
in Zusammenarbeit mit dem Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften

unter der Anleitung von
Privatdoz. Dipl.-Ing. Dr.techn. Christoph Schwanda
&
Univ.Lektor Dr. Gianluca Inguglia

durch

Huw Haigh, MPhys.
Matrikelnummer: 12106355

Österreichische Postsparkasse, 3. Stock,
Georg-Coch-Platz 2, 1010 Wien

August 7, 2024

Christoph Schwanda	Huw Haigh

Abstract

This thesis presents a search for new physics processes using 362fb^{-1} of collected data produced at a centre-of-mass energy of 10.58 GeV at the Belle II experiment in Tsukuba, Japan. The studies described herein focus on events that feature missing energy, which could indicate the presence of an as yet undiscovered particle. Such an observation could shed light on gaps in current understanding of the universe, such as in the nature of dark matter. This work specifically investigates events in which a pair of muons are produced along with some undetected particle that induces a recoil in the muon pair. Searches are conducted for invisibly decaying Z′ and muonphilic scalar particles, both of which are expected to appear as resonant structures in the recoil mass distribution (M_{recoil}). Data are interpreted under multiple theoretical models including various couplings to dark matter. Upper limits are calculated on the cross section and coupling strength for each model that is investigated, providing results that are competitive with, and expand on, previous experimental efforts. The study makes use of a neural network trained using the novel Punzi-loss function for rejection of background processes, the development and implementation of which is outlined. Furthermore a phenomenological reinterpretation of recent measurements related to the branching fraction of the B→K$\nu\bar{\nu}$ decay process is conducted. Previous results from multiple sources are studied in conjunction under a light new physics model with an associated two body B→KX decay process, and constraints are thereby set on such a model.

Kurzfassung

Diese Dissertation präsentiert eine Suche nach neuen fundamentalen Wechselwirkungen von Elementarteilchen unter Verwendung von Daten, die bei einer Schwerpunktsenergie von 10.58 GeV im Belle II Experiment in Tsukuba, Japan, gesammelt wurden und einer integrierten Luminosität von 362 fb^{-1} entsprechen. Die hierin beschriebenen Studien konzentrieren sich auf Ereignisse mit fehlender Energie, was auf das Vorhandensein eines noch unentdeckten Teilchens hinweisen könnte. Die Beobachtung neuer Physik könnte Licht auf Lücken im aktuellen Verständnis des Universums werfen, wie etwa die Natur der Dunklen Materie. Diese Arbeit untersucht speziell Ereignisse, bei denen ein Paar Myonen zusammen mit einem nicht nachgewiesenen zurückstoßenden Teilchen produziert wird. Es werden Suchen nach unsichtbar zerfallenden Z′- und myonphilen Skalarbosonen durchgeführt, die beide als resonante Strukturen in der Rückstoßmassenverteilung (M_{recoil}) erscheinen sollen. Die Daten werden mit mehreren theoretischen Modellen interpretiert, einschließlich verschiedener Kopplungsstärken mit Dunkler Materie. Konfidenzintervalle werden für den Wirkungsquerschnitt und die Kopplungsstärke für jedes untersuchte Modell berechnet, wobei Ergebnisse erzielt werden, die mit früheren experimentellen Resultaten konkurrieren. Diese Arbeit nutzt ein neuronales Netzwerk, das unter Verwendung der neuartigen Punzi-Loss Funktion zur Trennung von Hintergrundprozessen trainiert wurde. Die Entwicklung und Implementierung dessen wird beschrieben. Darüber hinaus wird eine phänomenologische Neuinterpretation aktueller Messungen im Zusammenhang mit dem Verzweigungsverhältnis im $B \to K \nu \bar{\nu}$ Zerfallsprozess durchgeführt. Frühere Ergebnisse aus mehreren Quellen werden in Verbindung mit einem Modell leichter neuer Physik untersucht, das einen damit verbundenen zweikörperigen $B \to K X$ Zerfallsprozess beinhaltet, und es werden dadurch Restriktionen für ein solches Modell bestimmt.

Acknowledgements

I would like to thank my supervisors, Gianluca and Christoph, for the support and guidance that they have provided me through these three years. Specifically I would like to thank Gianluca for his tireless efforts to ensure that I was afforded every opportunity to develop my abilities as a research scientist, and get the best out of my PhD studies. I would like to further thank Enrico Graziani, Marcello Campajola and Giacomo de Pietro for their valuable collaboration in the $Z' \to inv$ analysis.

Special thanks to the rest of the Belle II group at HEPHY. Michel, Rajesh, Paul, Nadia, Daniel, Géraldine, Phillip, Petar and Abdul, together you created a friendly and supportive working environment, and have been an absolute joy to work with.

I would like to further extend my thanks to the Belle II collaboration as a whole. None of the work outlined in this thesis could have been done without their mammoth efforts. It has been a pleasure to work within the collaboration, which has provided an inclusive and cooperative environment. I am also immensely grateful for the hard work of those at HEPHY who helped to facilitate my move to Vienna during a global pandemic.

Contents

Introduction

The standard model of particle physics provides an elegant description of the fundamental particles of nature and the means by which they interact with one another. There are, however, areas in which it has thus far failed to produce a conclusive understanding. For instance, the nature of dark matter and energy, which together comprise the majority of the known universe, remains completely elusive. In fact the force of gravity itself, which reveals the very existence of that dark matter, is unexplained in the standard model (SM). There are fundamental questions to be answered, and a variety of approaches have been adopted by physicists today to address them. From astronomers probing the deepest reaches of the skies, to particle physicists scrutinising those tiniest of objects that comprise the universe - a clue that might shed light on such unknowns will surely reveal itself somewhere.

The search for new physics at particle collider experiments, however, is not a straightforward task. Experimental searches have been ongoing for many years now and have yet to provide any such observations. Thus, one must look for more subtle signatures in the data that might indicate the presence of interaction processes beyond those of the SM. One such avenue is to focus not on what one *can* detect, but on what one fails to. By conservation of momentum, one can deduce if an observed event contains some entity that has escaped the detector without leaving a trace. Indeed the neutrinos of the SM do exactly this, due to their exceedingly small chance of interacting with the material through which they pass.

This thesis work focuses on such undetectable particles, specifically those produced in the process $e^+e^- \to \mu^+\mu^-(\text{inv.})$, where the final state is characterised by two oppositely charged muons with some third *recoil* entity. The primary candidate that is studied herein is the Z' boson, which can appear in such an invisible manner through decays to neutrinos, or perhaps even through coupling to some dark matter candidate. In addition to this the search for a muonphilic scalar particle is included. Due to the similar final state this can provide a further result with little change to the overall analysis procedure.

Part I seeks to first establish the motivation behind these searches. Comprehensive discussion of the theoretical grounding is provided on both the current limitations of understanding, and also the introduction of these new proposed particles to the framework of the SM. Previous experimental efforts to constrain the models are considered, and the Belle II experimental apparatus is described. A phenomenological reinterpretation of the measurement of a SM process, which was conducted as part of this thesis work, is then introduced. The $B^+ \to K\nu\bar{\nu}$ decay that is studied provides an interesting avenue in the

search for new physics due to precise theoretical predictions and also the missing energy due to neutrinos. Multiple experimental measurements of the branching fraction are collectively analysed and interpreted under a light new physics model which can thereby be constrained.

Part II then introduces two studies aimed at quantifying aspects the performance of the Belle II detector. The first of these looks at the cluster-track matching efficiency of the detector using radiative bhabha decays. A tag-and-probe method is employed, with a single electron and photon tagged, and then a search for the other electron is conducted in the recoil direction. Both real and simulated events are analysed and the rate at which the second electron is found in the expected region is then used to define a per track cluster-track match uncertainty, which is then to be propagated in further physics studies. The second of these studies measures the efficiency of the trigger lines that are selected for the following physics analysis. This is again calculated both in real and simulated data, and necessary correction factors to bring the two into agreement are defined. The third chapter then outlines the development of a new loss function, dubbed 'Punzi-loss', that is designed specifically for applications in searches for new particles. This is based on the Punzi figure-of-merit and the subsequent trained neural network (labelled the 'Punzi-Net') provides a classifier with notable improvements in sensitivity to the desired new physics signal over more commonly used methods.

Finally in part III the search for invisibly decaying Z' and muonphilic scalar bosons is presented. The event selection criteria are outlined: both preselection cuts applied to events and then further discussion of the implementation of the Punzi-Net. The expected resulting distributions are discussed thoroughly, with particular attention paid to those background sources that present a notable risk of mis-modelling. Control channels consisting of decay processes that are similar in topology to that for which is searched are analysed, and from these systematic uncertainties associated with different aspects of the analysis are quantified. The method of hypothesis testing by binned maximum likelihood fits is introduced with the particular methods outlined for the different models investigated as part of this work. Finally the projected upper limits on the associated cross-sections and coupling strengths of these models are then shown, along with comparison to other experimental measurements.

Notes on this work

There are a few important points to make clear about the work presented in this thesis. Firstly, the primary analysis discussed is the search for an invisibly decaying Z' boson. The work outlined in this document refers to a current iteration of this study that has not yet reached its conclusion and the data remains blind, meaning that full internal review of the methods has not yet finished and so only results from simulated data can be shown. However, during the full PhD study an earlier analysis was conducted and unblinded. The general framework of that analysis was similar, but with less comprehensive coverage of dark matter models and no inclusion of the muonphilic dark scalar search. The results of the study are discussed along with other previous experimental searches.

It is important for one to acknowledge the highly collaborative nature of experimental particle physics. The Belle II collaboration brings together over 1000 members from 27 countries around the globe, and any given physics analysis often combines the efforts of multiple members. Furthermore many tasks, such as the generation of simulated MC data or the development of necessary corrections, are carried out centrally by a specific task force with the results then utilised by all members. In this thesis each study that is discussed is preceded by a small note to highlight those collaborators who had some direct involvement, be that in an advisory role or in carrying out some analysis work.

Part I

Motivation

1. Theory and experimental background

1.1 The Standard Model of particle physics

The Standard Model (SM) of particle physics is a theoretical framework that describes the known fundamental particles and the mechanisms by which they interact with one another. These interactions are governed by three of the four forces: electromagnetism, the strong force, and the weak force. The last of these, gravity, is as yet unexplained in the SM. Electromagnetism is described by the theory of Quantum Electrodynamics (QED), which underpins the interaction between electrically charged particles and is mediated by the photon, which itself is both massless and chargeless. Electromagnetism is responsible for such natural phenomena as the orbital nature of electrons around the atomic nucleus. Similarly, the strong force is underpinned by the theory of Quantum Chromodynamics (QCD), which, mediated by gluons, creates the neutron-proton structure that can be observed within that nucleus itself. Finally there is the weak force, which is mediated by the $W^{\pm}$ and Z bosons, is responsible for the likes of radioactive β-decays that occur within the Sun.

The Standard Model contains two main types of particles: fermions and bosons. Fermions can be broken down further into leptons and quarks. They have half integer spin, are defined by Fermi-Dirac statistics and obey the Pauli exclusion principle. The bosons, however, have integer spin and are commonly referred to as force carriers, owing to the fact that they mediate the previously described fundamental forces. The exception to this being the spin zero Higgs boson, which instead accounts for the mass of the fermions and massive bosons of the SM.

The leptons and quarks come in 3 generations, each containing 2 particles respectively. All of these in turn also has an oppositely charged anti-particle to accompany. There exists a generational mass hierarchy within the SM and, owing to the inherent instability of heavier particles, it is the lower mass first generation that makes up the majority of matter naturally occurring in the universe. Atoms contain electrons, which belong to the first generation of leptons, orbiting around a nucleus that is comprised of protons and neutrons, which in turn consist of up and down quarks: the two quarks belonging to the first generation. Figure 1.1 summarises these various particles, their attributes, how they are classified and the mechanisms by which they interact with one another.

The Standard Model is quantum field theory in which a Lagrangian function describes the dynamics of the fields and their respective particles, that exist as excitations in

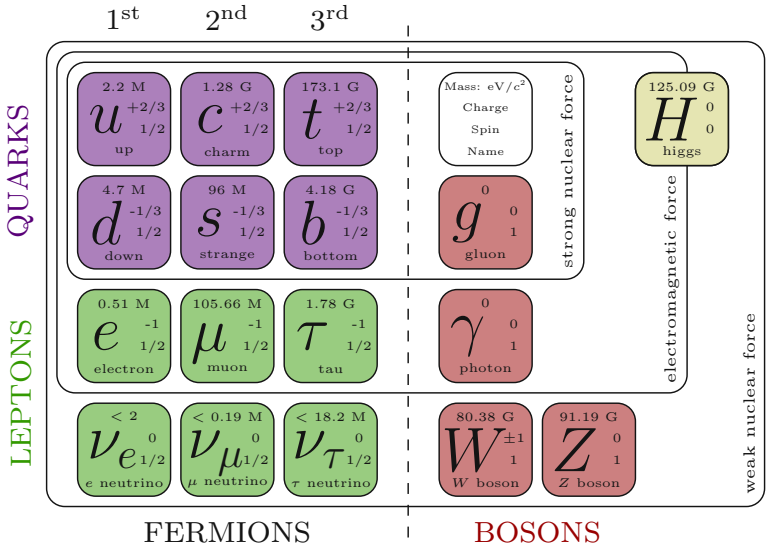

Figure 1.1: The particles of the Standard model shown divided into fermions and bosons, with the former also divided into the leptons and quarks. Each is shown with its mass, charge and spin.

those fields. The fundamental pillar of this mathematical description is known as *gauge invariance*, which dictates that the Lagrangian must be locally invariant under a set of specific transformations. This leads to the set of symmetries defined as $SU(3)_c \times SU(2)_L \times U(1)_Y$, the terms of which essentially give rise to underlying interactions that occur within the SM. As dictated by Noether's theorem, each of these symmetries comes with some quantity that must be conserved during the interactions.

Quantum chromodynamics is brought about in the standard model by the inclusion of the $SU(3)$ symmetry. This introduces the conserved *color charge* quantity, meaning that each quark flavour has with it three colour fields: red, green and blue. This results in a total of 8 gluon fields. Only those particles with non-zero colour charge couple to gluons, and it is for this reason that leptons do not interact strongly; they posses a neutral color charge.

The further inclusion of the $U(1)_Y \otimes SU(2)_L$ gauge symmetry introduces the electroweak interaction that describes the unified electromagnetic and weak forces. The subscript L refers to the fact that only left-chiral fermions transform under the $SU(2)$ group, thereby dictating that only they, and not right-chiral fermions, interact via the weak force. With these two lie groups, U(1) and SU(2), come the quantities weak hypercharge (Y_W) and weak isospin (T) respectively. Individually neither of these quantities is absolutely conserved [1], however the electric charge, which comes about in a linear combination of the two, is a conserved quantity.

The breaking of this electroweak symmetry is necessary to explain the large masses

[1] Weak hypercharge and weak isospin are not conserved in interactions with the Higgs field.

of the $W^{\pm}$ and Z bosons, and indeed is the reason as to why neither Y_W nor T are conserved, and why the electromagnetic and weak forces exist individually and not as a single combined force. It is brought about with the introduction of the scalar Higgs field, which has a non-zero vacuum expectation value. Excitation of this new field brings with it the famous Higgs boson and the mass of the fermions then arises due to coupling between their respective fields and the Higgs field.

This introduction utilises material found in Modern Particle Physics [1], in which one can find a comprehensive introduction to the current theoretical and experimental landscape of particle physics today.

Particle Interactions

The scale and strength of these fundamental forces in the SM vary greatly. For instance, as discussed, the strong force responsible for holding together the protons and neutrons within an atomic nucleus. This encapsulates the stark difference in strength between it and the electromagnetic force, which is at the same time trying push the positively charged protons away from one another. The strengths of the forces are defined by the respective coupling constant: the strong force with coupling constant g_s, the electromagnetic with e and the weak force with charged and neutral constant g_W and g_Z respectively.

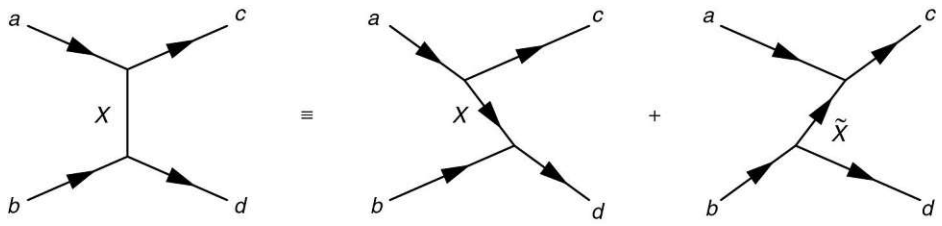

Figure 1.2: The scattering process $a + b \rightarrow c + d$ as the sum of two processes with slightly different time-orderings, with one involving the exchange of some particle, X, and the other instead showing the anti-particle, $\bar{X}$.

Particle interactions are commonly shown in the form of *Feynman diagrams*, which are not only a useful means by way to visualise a process, but also a powerful tool for calculating the probability of the evolution of a system from one quantum state to another. The feynman diagram of a given process represents the sum over all possible slight variations in the process defined by the in-going and out-going set of particles. There can be great deals of complexity, with so-called *virtual* particles being created and decaying over the course of the interaction. Feyman diagrams represent the passage of time in the x-direction, and space in the y. Figure 1.2 demonstrates the scattering process $a + b \rightarrow c + d$ as the sum of two processes with slightly different time-orderings, with one involving the exchange of some particle, X, and the other instead showing the anti-particle, $\bar{X}$.

1.2 Beyond the Standard Model

1.2.1 Dark Matter

Observational evidence

Perhaps the most ubiquitous of the SM's shortfalls is that of the unexplained nature of so-called dark matter (DM). Named as such for its apparent inability to interact via any of the three fundamental forces understood through the SM, DM is thought to make up around 85% of the physical matter within the universe, thereby dwarfing the normal baryonic matter that one can actually observe. Over the years there have been a plethora of observations that indicate the existence of this dark matter, invariably through its gravitational interaction with SM particles. The first of such hints came with the observation of galaxy clusters that were conducted in the 1930s, with Fritz Zwicky noting that the galaxies within the Coma cluster appeared to be moving with greater velocity than the gravitational pull of the cluster could feasibly contain [2]. This would then suggest that some invisible matter must be present to provide this extra gravitational pull.

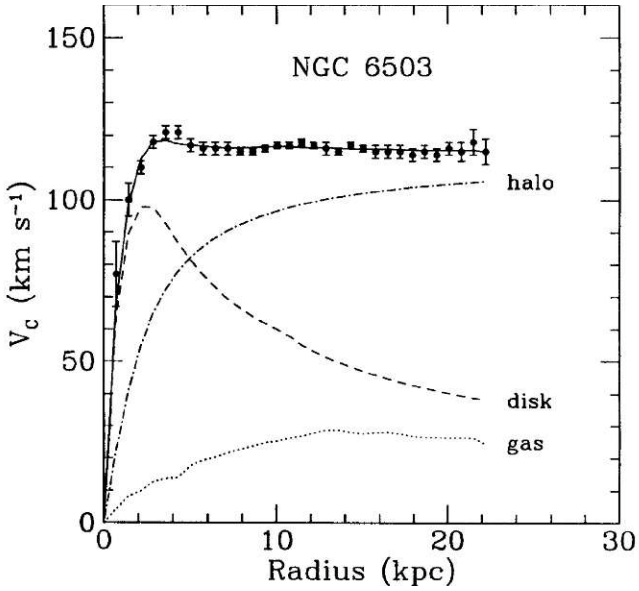

Figure 1.3: The observed galaxy rotation curve of the NGC 6503 galaxy, with the estimated contributions from the baryonic matter disk and gas, and the dark matter halo that is necessary to explain the observed data [3].

More concrete evidence came with similar measurements of the stars contained within galaxies themselves. It was found that the distribution of star velocities as a function of distance from the centre of the galaxy was quite different to that which one would expect given the visible matter present. This finding can be seen in fig. 1.3, which shows the comparison of expected and observed velocities as a function of radius in the dwarf

spiral galaxy NGC 6503 [3].

The figure demonstrates clearly that the expected contributions due to the disk and gas, both comprised of baryonic matter, do not suffice to explain the observed data points. One must add a large dark matter *halo* component to the equation, called such as it must continue well beyond the radial extent of the visible galaxy to produce the plateaued distribution. It is estimated that such DM would comprise as much as 80 to 90% of the total galactic mass in order to explain this observation [4].

In cosmology, simulation of the large scale structure of the universe is found to rely heavily on the slight density fluctuations that were present in the early cosmos. Various large scale simulations under different dark matter models have been conducted, with some fitting the observational data almost perfectly [5].

There have of course been varied efforts to explain such observations without the need to introduce new particles to the standard model. Modified Newtonian Dynamics (MOND) is perhaps the most well known attempt at this. MOND, specifically developed in response to the previously discussed galaxy rotation curves, puts forward the notion that Newtonian dynamics does not hold true at the limit of small accelerations [6]. This allows the model to then explain the galaxy rotation curves purely with the baryonic matter that is present.

MOND, and other attempts to explain away the particle nature of dark matter, tend to come undone when one considers the observation of *the Bullet Cluster* [7]. The name of which comes from the unique situation in which the cluster finds itself: moving away from another, larger cluster through which it has passed during a collision. In such, the galaxies that occupy the clusters are expected to pass through one another largely unaffected, with direct collisions between them being unlikely. One can deduce the mass distribution of the galaxy clusters by way of observing the gravitational lensing they induce. This is shown in the left hand image of figure 1.4, in which the green contours demarcate the mass density, which is calculated using the lensing. The two regions of high mass density as one might expect coincide with the two post-collision clusters.

During the collision, the intracluster medium of gas and dust that comprise the large bulk of baryonic material within each cluster will interact electromagnetically with that of the other, and thereby both will be slowed during the process. These can then be observed in the x-ray spectrum and thus their distributions mapped. This is shown in the right-hand image of fig.1.4, where they are superimposed over the mass density contours observed via gravitational lensing.

It is in the comparison of these that the ground-breaking observation lies: The large majority of baryonic matter has been effectively kicked out from the centre of both clusters, and yet, in those centres remains the center of mass. This strongly implies that observations of dark matter are unlikely to be a mere misunderstanding of gravity, in which case the lensing would still coincide with the distribution of the intracluster medium. Instead, one can surmise that dark matter is surely particle-like in nature.

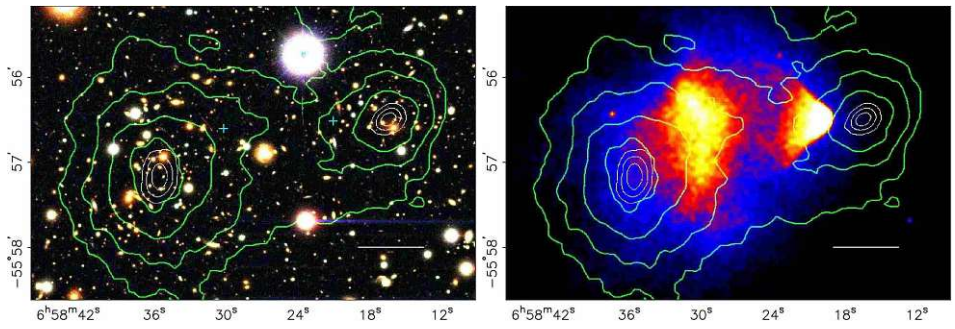

Figure 1.4: Two images of the Bullet Cluster, both with green contours that represent the mass density observed by gravitational lensing. The left image shows a visible spectrum image, the right hand image instead shows an image in the x-ray spectrum which highlights the intracluster medium of gas and dust [7].

Searching for dark matter

The ΛCDM model of the cosmology, commonly referred to as the *standard model of cosmology*, is the current accepted framework that provides a description of the structure of the universe [8]. From this model, the current estimate of the dark matter contribution to the total energy-matter density of the universe, Ω, is at Ω_c=0.265(7). The vast bulk of the density comes from *dark energy*, giving Ω_Λ=0.685(7), while standard baryonic matter comprises the remain 4.9% (Ω_b=0.0493(6)) [5].

This model specifies that the dark matter component of the universe is comprised of *cold dark matter* (indeed this is what "CDM" stands for) which refers to it being non-relativistic. The observed *relic density* of this DM is commonly attributed to the *freeze-out* scenario, in which DM was previously in thermal equilibrium with the plasma of the early universe. This means that the relative abundances of DM and SM particles would remain coupled as annihilation of pairs of one would produce a pair of the other. As the universe expanded and cooled the thermal annihilation rate $\langle\sigma v\rangle$ would decrease until some critical *freeze-out temperature* is reached at which point the DM decouples and the relic density that is observed today is reached [10].

This process is encapsulated in figure 1.5, which shows the expected relic density (Ω_X) (Y is instead the number density scaled by the entropy density) for a dark matter particle of mass m_X = 100 GeV/c^2 as a function of time, t, and temperature, T. Ω_X is found to be inversely proportional to the annihilation rate $\langle\sigma v\rangle$, which is in turn inversely proportional to the square of the DM mass.

The solid line defines Ω_X at an annihilation rate required to explain the observed relic abundance, and the contours show variation around this by factors of 10, 100 and 1000. From this one can infer that a DM particle in the mass range $m_X \sim$ 100GeV/c^2 - 1 TeV/c^2 could explain the observed relic density. This finding is known as the *WIMP miracle*, which refers to the fact that the mass range coincides with already hypothesised

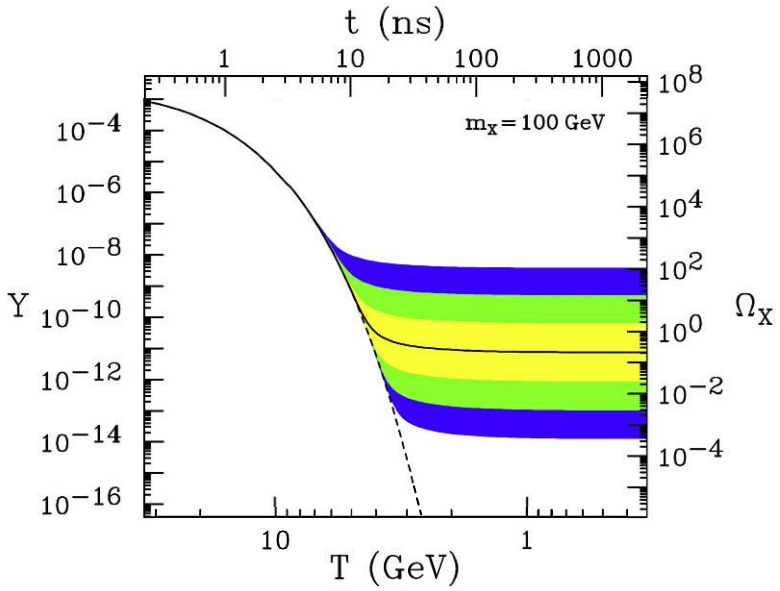

Figure 1.5: The expected relic density Ω_X and number density scaled by entropy density Y, for a dark matter particle of mass 100GeV/c^2 as a function of time t and temperature T. [9]

weakly interacting massive particles (WIMPs), commonly originating in *supersymmetry* [9].

For some time WIMPs were seen as an excellent candidate to explain the DM puzzle, however, optimism has somewhat faded in recent years with a lack of experimental evidence [11]. There may yet be hope in the high mass (multi-TeV+) end of the range which future experiments such as the proposed muon collider could one day probe [12].

One can still seek to explain the observed relic density of dark matter in the universe via other methods however. Of particular interest in this thesis is the case in which some new light (less than a few GeV) mediator particle is introduced that couples to both fermions and DM. Such an addition can alter the annihilation rate such that this mediator along with a light dark matter candidate can still produce the observed relic density [13] [2].

Experimental searches for dark matter can be conducted via a few main methods:

- *Direct searches*: Direct searches involve directly observing the process of some dark matter interaction occurring within an experimental detector [15]. Often these are highly sensitive detectors that hope to detect the vanishingly rare interaction of a dark matter particle passing passing as Earth moves around the galaxy. The Xenon dark matter project for instance searches for the tiny nuclear recoil that is expected to be produced upon interaction between a WIMP and a Xenon nucleus

[2] Of course there exists a huge range of proposed DM candidates, a full summary is beyond the scope of this work, however comprehensive discussion can be found in [9, 14]

[16].

- *Indirect searches*: Indirect searches for dark matter instead hope to observe telltale signs of dark matter decay or annihilation occurring in the cosmos, specifically in heavy objects such as galaxies, where one expects a concentration of dark matter. These methods often utilise x-ray/radio telescopes or neutrino detectors [17].

- *Collider searches*: Finally collider searches generally hope to observe the effect of dark matter's contribution to the events produced at the interaction point. Of course due to the fact DM must interact exceedingly weakly with SM particles (or not at all), these searches invariably involve events with some missing, undetected energy [18].

1.2.2 The Muon g-2

One notable experimental observation that appears to differ from the SM prediction is that of the *magnetic moment* of the muon, which leads to the so-called muon g-2 anomaly. A particle with spin $\vec{S}$, mass m and charge q will have a magnetic moment $\vec{\mu}$ define as:

$$\vec{\mu} = g\frac{q}{2m}\vec{S} \tag{1.1}$$

where g is what is known as the g-factor and is defined as the ratio of the magnetic moment to the angular momentum. In 1928 Dirac predicted that all spin-$\frac{1}{2}$ particle would have a value of $g = 2$. Further treatment via relativistic quantum field theory demonstrated that loop corrections due to interactions with other SM particles would push this value fractionally away from the tree-level value of just 2, thereby giving rise to the *anomalous magnetic moment* of the muon, a_μ. These come from various SM sources [19]:

$$a_\mu^{SM} = a_\mu^{QED} + a_\mu^{EW} + a_\mu^{QCD} \tag{1.2}$$

Where contributions from quantum electrodynamic (QED), electroweak (EW) and quantum chromodynamics (QCD) are included. Example feynman diagrams of processes associated with the leading order contributions for each of these three are shown in figure 1.6.

The first two of these contributions (QED and EW) can be calculated to high precision by perturbation theory. The QCD component, however, is less well defined and thereby dominates the uncertainty associated with the theoretical prediction of a_μ^{SM} [20].

The Fermilab Muon g-2 experiment began collecting data in 2018 and ran until 2020. The experiment used a storage ring containing muons with aligned spins, and observed the precession of these as they orbit the ring. From this, a measurement of the anomalous magnetic moment can be inferred. The latest publication from the collaboration provided the most accurate experimental measurement yet with $a_\mu(\text{exp}) = 116592055(24)\times10^{-11}$ [21]. The experimental average, which includes previous measurements by Fermilab and BNL, has a value of,

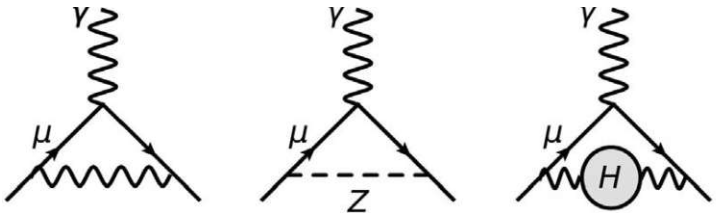

Figure 1.6: Feynman diagrams showing leading order contributions to the anomalous magnetic moment of the muon from multiple sources: QED (left), EW (centre) and QCD (right).

$$a_\mu^{exp} = 116592059(22) \times 10^{-11} \tag{1.3}$$

The current accepted SM prediction for the value, defined in 2020 by the Muon g-2 Theory Initiative [20], has a value of,

$$a_\mu^{SM} = 116591810(43) \times 10^{-11} \tag{1.4}$$

And herein lies the anomaly: there is a substantial difference between these measured and calculated values. This could imply the existence of some new particle that would further add to the loop corrections and thereby explain the difference.

It is important to note, however, that the ground is not completely stable under this standard model calculation. A recent publication by the BMW collaboration, who investigate QCD contributions via lattice calculations, brings with it a theoretical prediction that is in better agreement with the observed values [25][3]. Furthermore, recent results from the CMD-3 e^+e^- collider experiment, which can evaluate hadronic contributions to a_μ via measurement of the $e^+e^- \to \pi^+\pi^-$ cross-section, appear to lend credence to this updated theory prediction [24].

The various results discussed here are depicted in figure 1.7. The experimental measurements from BNL and Fermilab, and the resulting average, are shown in green. Previous results by the KLOE, BaBar and previously mentioned CMD-3 e^+e^- collider experiments are shown in blue, along with the current white paper theory prediction depicted by the blue shaded region. Finally the new prediction from the BMW collaboration is shown in purple. One can see that this brings the tension between theory prediction and experimental observation from 5.2σ to just 1.8σ.

1.3 The Z′ boson

One possible extension that seeks to explain some of these shortcomings in the Standard Model is the addition of a $U(1)'$ gauge symmetry. This then gives rise to an associated gauge boson, dubbed the Z' boson, which would couple to SM particles and perhaps

[3] A further calculation was presented very recently that agrees with the experimental average to 0.9σ [26], however this result is still in review.

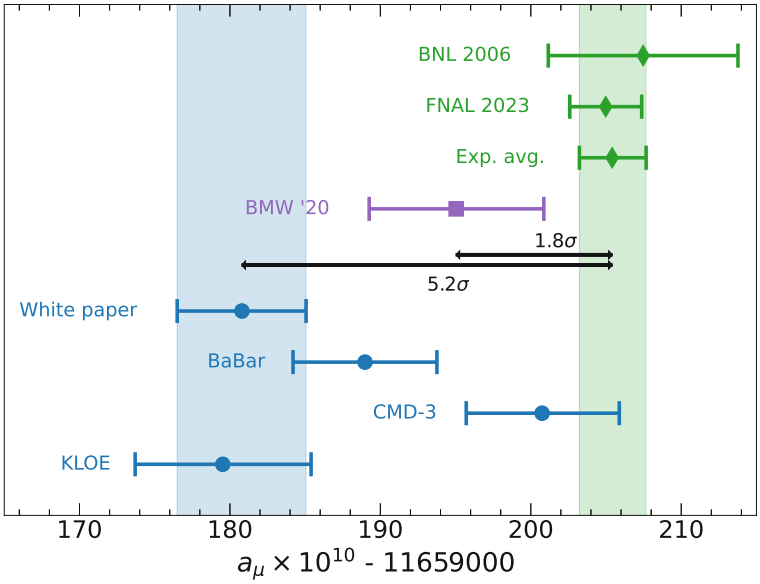

Figure 1.7: The direct experimental measurements by BNL and then more recently at Fermilab [21], with the final experimental average. The bottom three blue lines show results from measurements of the two pion spectrum observed at electron-positron colliders KLOE [22], BaBar [23] and CMD-3 [24]. The blue shaded region depicts the current theory expectation define by the Muon g-2 Theory initiative [20]. Finally the newer theory prediction by the BMW collaboration [25] is shown in purple. Notably this theory prediction reduces the tension with the experimental average from 5.2σ to just 1.8σ. Figure modified from [26].

to others beyond the Standard Model such as dark matter. The work described herein focuses mostly on one specific theoretical model, known as the L_μ-L_τ framework [27, 28, 29], which outlines a Z' boson that couples only to the muon and tau leptons, and their respective neutrino partners, ν_μ and ν_τ.

This proposed model provides possible resolutions to multiple questions that are as yet unaddressed in the SM. The additional loop corrections to the muon anomalous magnetic moment provided by a Z' (similar in form to that shown in the central Feynman diagram of 1.6) could go some way to explaining the observed discrepancy with respect to the SM theory calculation [30, 31]. It may resolve a number of tensions related to measurement of flavour observables conducted by the LHCb, Belle and BaBar [32, 33, 34]. Such a Z' can also be used to explain the observed relic density of dark matter, assuming that such dark matter is charged under the L_μ-L_τ framework. For such an explanation there are two possible dark matter candidates: sterile neutrinos [28] or light Dirac fermions [35].

The L_μ-L_τ model does of course provide the possibility to search for such a Z' boson by studying decays to those products that would be directly observable in the detector, and indeed searches for the $Z' \to \mu^+\mu^-$ process are conducted at the Belle II experiment [36]

and others. However, here the search is conducted specifically for the *invisible* decay of the Z′. In what shall be henceforth referred to as the *vanilla* L_μ-L_τ model, such invisible decays would be the result of Z′ decay to neutrinos, which cannot be reconstructed by direct detection methods and so their presence is inferred by conservation of momentum, or, the presence of *missing energy* in the event.

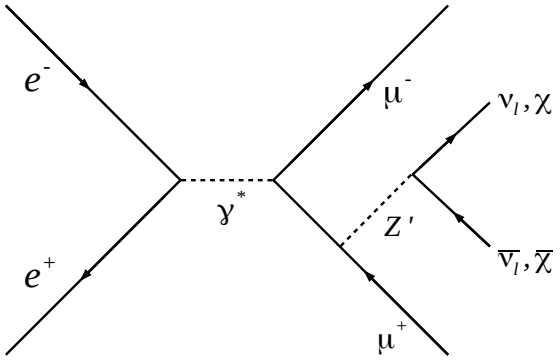

Figure 1.8: A Feynman diagram depicting the production of two muons and a Z′ boson which subsequently decays invisibly to either neutrinos or some dark matter particle (χ).

A Feynman diagram depicting the process of this Z′ creation and subsequent invisible decay is shown in figure 1.8, and the interaction Lagrangian underpinning this model is given by,

$$\mathcal{L} = \sum_{\ell} \theta g' \bar{\ell} \gamma^\mu Z'_\mu \ell \tag{1.5}$$

where the sum runs over the heavy leptons and their respective (left-handed) neutrino species $\ell = \mu,\ \tau,\ \nu_{\mu,L},\ \nu_{\tau,L}$, and θ takes the value -1 for $\ell = \mu,\ \nu_{\mu,L}$ and 1 for $\ell = \tau,\ \nu_{\tau,L}$. The strength of the interaction is defined by the g' coupling constant. The partial widths for the Z′ decay to leptons and neutrinos respectively are then defined as:

$$\Gamma(Z' \to l^+ l^-) = \frac{g'^2 M_{Z'}}{12\pi}\left(1 + \frac{2M_l^2}{M_{Z'}^2}\right)\sqrt{1 - \frac{4M_l^2}{M_{Z'}^2}}, \tag{1.6}$$

$$\Gamma(Z' \to \nu_l \bar{\nu}_l) = \frac{g'^2 M_{Z'}}{24\pi} \tag{1.7}$$

The total branching fraction to invisible decays in the vanilla L_μ-L_τ model is thereby defined as:

$$BF[Z' \to invisible] = \frac{2\Gamma(Z' \to \nu_l \bar{\nu}_l)}{2\Gamma(Z' \to \nu_l \bar{\nu}_l) + \Gamma(Z' \to \mu^+\mu^-) + \Gamma(Z' \to \mu^+\mu^-)} \tag{1.8}$$

One can see that the BF thus is dependent on the mass of the Z′ boson, $M_{Z'}$, and also the lepton mass in the case of visible decays to muons or tau leptons. This leaves 3 distinct regimes in the branching fraction of invisible decay,

- $M_{Z'} < 2M_\mu$: Here one expects a BF$[Z' \to invisible] = 1$, as only decays to neutrinos are kinematically possible from a Z′ boson produced at rest.
- $2M_\mu < M_{Z'} < 2M_\tau$: With decays to a pair of muons now kinematically available, but still not to tau leptons, the branching fraction is reduced to an expectation of BF$[Z' \to invisible] \sim 1/2$.
- $M_{Z'} > 2M_\tau$: With a Z′ mass greater than twice the tau mass, all final states are then kinematically available and the expected branching fraction is further reduced to BF$[Z' \to invisible] \sim 1/3$.

In addition to this vanilla L_μ-L_τ, where coupling exists only to these SM particles, one can also study separately a case in which there is the addition of possible dark matter coupling. This interaction would be defined with the coupling constant,

$$\alpha_D = \frac{g_D'^2}{4\pi} \tag{1.9}$$

with the newly defined dark matter coupling, g'_D. There is no current limit set on α_D and so there is no *a priori* reason to assume it is small. Given this, one can reasonably assume $g'_D \gg g'$ and thereby BF$[Z' \to \chi\bar{\chi})] \approx 1$. This model shall be referred to as the *dark* Z′ model and will be studied in conjunction with the vanilla model. In this study, the mass of the dark matter is assumed to be one third of the Z′ mass, $M_\chi = M_{Z'}$.

The branching fraction of the invisible decay mode for the Z′ is shown as a function of $M_{Z'}$ in figure 1.9. This shows the expected branching fraction for both the vanilla L_μ-L_τ model, with the three regions of behaviour outlined previously, and also in the dark Z' model with various values of α_D [4].

This higher possible coupling in the dark Z' model does lead to an important change in the analysis method, despite the fact both cases involve the inference of some invisible

[4] The branching fractions under the dark Z' model are calculated here using the g′ limits set in the previous iteration of this analysis. These results are discussed in section 1.5.

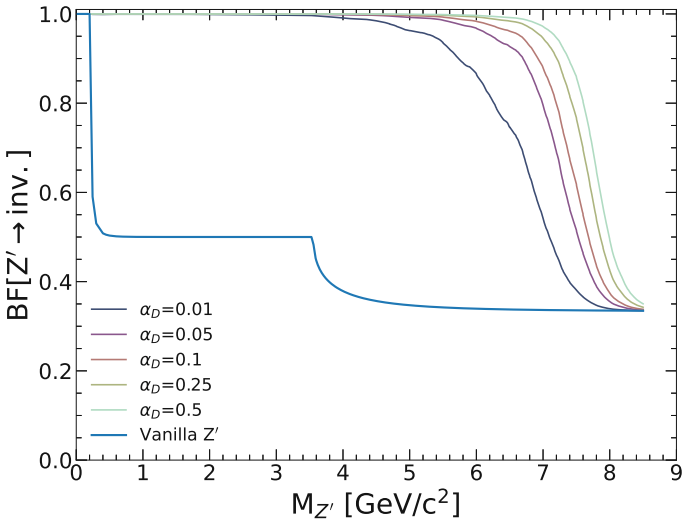

Figure 1.9: The branching fraction of the invisible decay mode for the Z′ as a function of $M_{Z'}$. The L_μ-L_τ model is shown, along with the dark Z′ model for which the distribution under multiple values of α_D are shown.

particles produced in the decay. The width of a Z′ decay to a pair of dark matter particles, $\chi\bar{\chi}$, can be calculated by:

$$\Gamma(Z' \to \chi\bar{\chi}) = \frac{\alpha_D M_{Z'}}{12}(1 - \frac{M_\chi}{M_{Z'}})^{\frac{3}{2}} \tag{1.10}$$

Due the dependence of α_D on g'_D, and the relatively large value of g'_D, one must begin to consider the width of the decay to dark matter with respect to the detector resolution. In the vanilla model, g' is small and thereby the width (eq. 1.8) can be considered negligible as the achievable resolution of the detector limits the observable width. Figure 1.10 shows the width of Z′ decay to a pair of dark matter particles, $\chi\bar{\chi}$, for a set of different α_D values along with the measured detector resolution shown in a black dashed line.

One can see that in fact for values of α_D above approximately 0.01, the assumption of negligible width no longer holds as the physical width surpasses the detector resolution. This effect is more prominent in higher recoil masses, and for larger values of α_D the range over which the width is no longer negligible grows.

For study of this fully invisible model no change is made to the reconstruction or selection of events (discussed in section part III), however the variation in signal shapes is accounted for in the statistical analysis of data, where a set of α_D values are investigated and parameter limits are subsequently set with this varied interpretation.

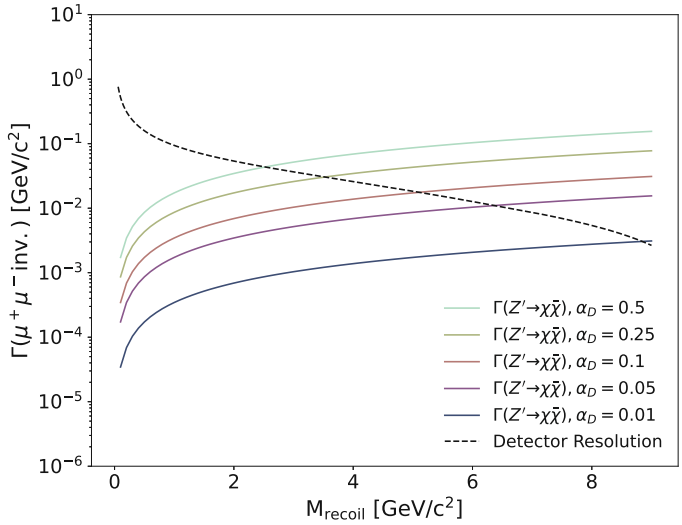

Figure 1.10: The width of Z′ decay to a pair of dark matter particles, $\chi\bar{\chi}$, for a set of different α_D values. The measured detector resolution is shown by a black dashed line.

1.4 Muonphilic Dark Scalar

Due to the identical final state of a pair of muons produced with some some missing energy, a *muonphilic dark scalar* particle [37, 38], can easily be studied in conjunction with the invisibly decaying Z′, with little adjustment to the analysis process. Such a particle is theorised to couple exclusively to muons via Yukawa-like interactions, defined by the interaction Lagragian:

$$\mathcal{L} \supset g_S S\bar{\mu}\mu \tag{1.11}$$

where g_S is the coupling constant of the dark scalar, S. This particle is often proposed as a solution to the muon g-2, with a coupling constant introducing a shift to the value of Δa_μ defined with the approximation:

$$\Delta a_\mu^2 \approx 2 \times 10^{-9} (\frac{g_S}{10^{-3}})^2 (\frac{700\text{MeV}}{m_S})^2 \tag{1.12}$$

Which holds on in the mass limit $m_S \gg m_\mu$, thereby meaning a strong coupling would be required to explain the anomaly. Furthermore, with the requirement of a scalar mass larger than twice the muon mass, the only tree-level decay channel would be $S \to \mu^+\mu^-$, with $S \to \gamma\gamma$ and $S \to \nu\bar{\nu}$ occurring at the one loop level. These are highly suppressed however and so decay to a muon pair dominates with a branching fraction close to one.

This would of course then form a signal topology of 4 muon tracks in the detector and so is not of interest in this study that focuses on invisibly decaying particles. Attention is therefore limited to the small mass region below the muon limits, $m_S < 2m_\mu$ where S either decays to neutrinos or has a long lifetime allowing it to escape the detector before decaying to observable products.

1.5 Previous experimental searches

Searches for a Z′ boson decaying visibly to a pair of muons have been conducted previously by the *BABAR* [39], Belle [40], Belle II [36], CMS [41] and ATLAS [42] experiments. These of course can only constrain the vanilla L_μ-L_τ model.

Searches for purely invisibly decaying Z′ bosons (i.e. the dark Z′ model) have been much more limited. The NA64-e experiment [43] performed a search constraining only the very low mass regions (below the muon mass threshold $2M_\mu$), and more recently the NA64-μ experiment provided a new measurement [44]. These searches can constrain both the vanilla and dark matter Z′ models.

The first publication produced by the Belle II collaboration was in fact an initial search for the invisibly decaying Z′ boson, using just 0.276fb^{-1} of data collected during the initial commissioning runs in 2018 [45]. More recently this was updated in 2023 using 79.7fb^{-1} [46]. This second analysis was undertaken as part of the doctoral studies summarised within this thesis and thereby followed a similar structure to that which is outlined in this thesis, such as using an earlier iteration of the Punzi-Net (see section 3) for background rejection, and similar methods of statistical analysis. The 90% confidence level upper limits set by this study on the cross-section of an invisibly decaying Z′ boson are shown in figure 1.11. The figure shows the upper limits under the assumption of a negligible decay width in black. The green dashed line then shows a basic interpretation of the cross section limits where the width is then non-negligible at a value of 10% of the mass of the Z′.

Figure 1.12 shows the 90% CL upper limits on the g′ coupling constant under the assumption of a vanilla L_μ-L_τ model Z′. In this scenario the analysis was able to set world leading limits in the very low mass region between 11.5 and 211 MeV/c^2. Figure 1.13 shows the limits on the g′ coupling constant under the fully-invisible Z′ boson, in which the BF to dark matter was assumed to be 1. Again the limits are shown under the assumption of a negligible decay width in black and then with a decay width at 10% of the Z′ mass in green. The grey dashed line displays the point above which the BF assumption is not valid in the former of these two cases. These provided the first direct-search results fully excluding an invisibly decaying Z′ boson as the explanation for the $(g\text{-}2)_\mu$ anomaly in the mass range $0.8 < M_{Z'} < 5.0\ \text{GeV}/c^2$.

As for the search for a muonphilic dark scalar boson, there had been no experimental constraint set until recently with a publication by the BES-III collaboration who, as in this study, incorporated the search into their study of an invisibly decaying Z′ boson [47].

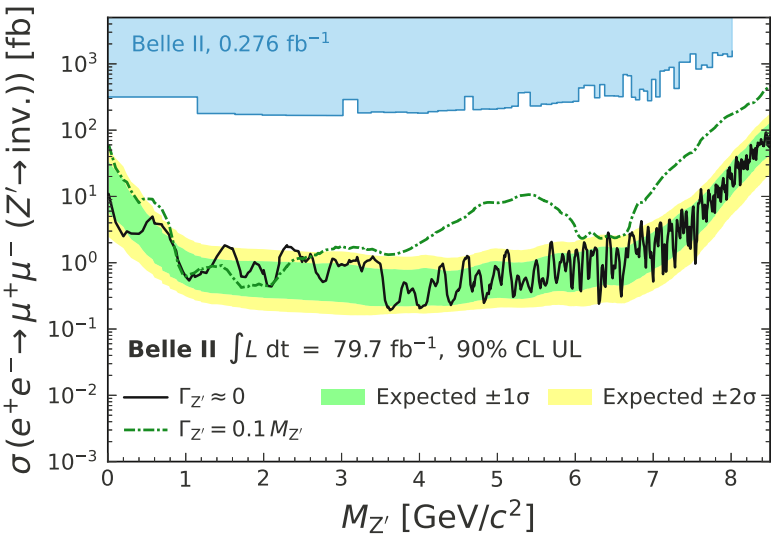

Figure 1.11: The 90% confidence level upper limits set by the 2023 study on the cross-section of an invisibly decaying Z′ boson. In black is the limit under assumption of negligible decay width. The green dashed line shows the limit under the assumption of a non-negligible width with a value at 10%

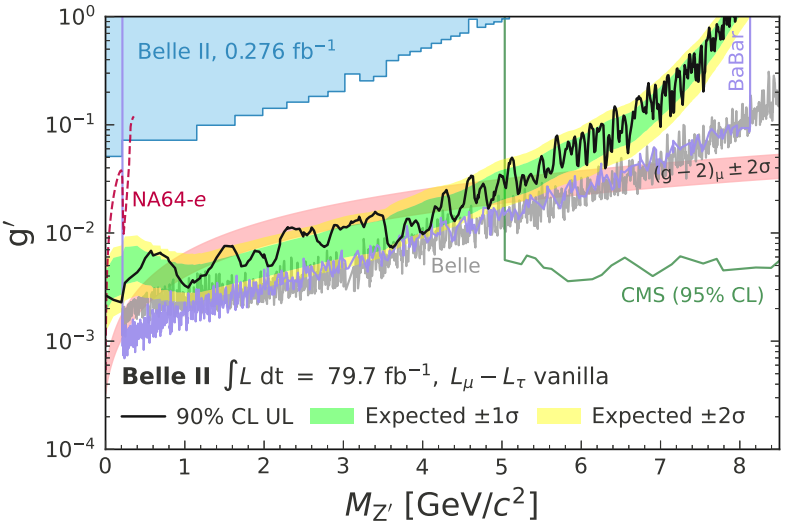

Figure 1.12: The 90% confidence level upper limits set by the 2023 study on the cross-section of an invisibly decaying Z′ boson. In black is the limit under assumption of negligible decay width. The green dashed line shows the limit under the assumption of a non-negligible width with a value at 10%

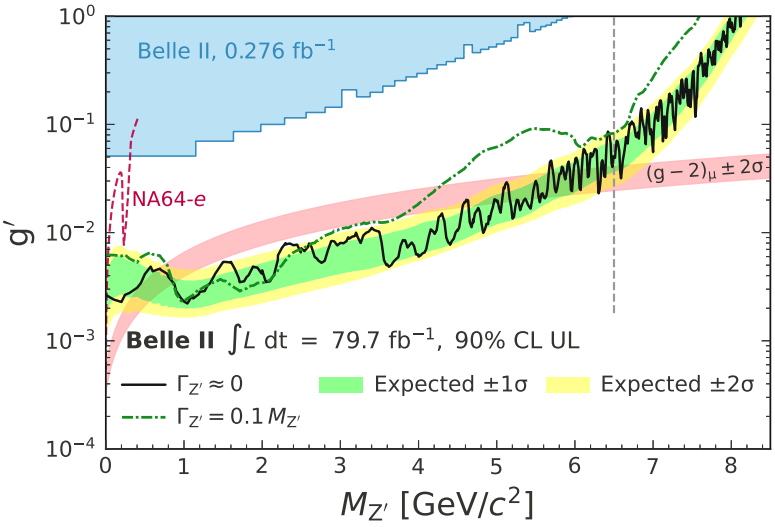

Figure 1.13: The 90% confidence level upper limits set by the 2023 study on the cross-section of an invisibly decaying Z′ boson. In black is the limit under assumption of negligible decay width. The green dashed line shows the limit under the assumption of a non-negligible width with a value at 10%

2. Experimental Particle Physics

2.1 Collider Experiments

2.1.1 Collider Physics

Collider experiments in particle physics utilise beams of particles, typically electrons or protons, that have been accelerated to relativistic velocities (i.e. at some large fraction of the speed of light, usually denoted as $\beta = v/c$). In such cases, the energy and momentum are modified by the *Lorentz factor*, defined as:

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{2.1}$$

thus giving:

$$E = (\gamma - 1)m_0c^2 \quad \text{and} \quad \overrightarrow{p} = \gamma m_0 \overrightarrow{v} \tag{2.2}$$

Where m_0 is the rest mass of the particle. As a result of this, the particles posses massive kinetic energies and thereby, upon collision with one another, open the door to explore processes and states that only exist in such high energy regimes. As discussed previously, the universe is predominantly made of the lightest particles in the standard model. On the other hand the top quark has a mass approximately 170 times greater than that of the proton. To produce these high mass particles at the collision point, a collider must posses a *centre-of-mass* energy at least equal to the rest mass of the particles produced. In a collider experiment this is given by square root of the Lorentz invariant quantity s, defined as:

$$s = \left(\sum_i^2 E_i\right)^2 - \left(\sum_i^2 p_i\right)^2 \tag{2.3}$$

Where E and p are the energy and momentum of particle i involved in the collision. Typically the beams used in collider experiments actually consist of many tightly packed bunches of particles. In addition to the centre-of-mass energy, the other key parameter often considered when discussing colliders is the *instantaneous luminosity*, defined as:

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{2.4}$$

where f is the frequency of bunch crossings, n_1 and n_2 are the number of particles in the two colliding bunches, and, σ_x and σ_y are the root-mean-square of the horizontal and vertical beam sizes. Typically at collider experiments one studies the cross-section, σ, which is essentially a measure of the rate of some interaction process occurring. This quantity has units of area and, as one might expect when dealing with particles, invariably takes very small values. In nuclear and particle physics the unit of *barns* are used, with 1 barn equal to 10^{-28}m^2. One can seek to measure the cross-section of a given process with the relation given by:

$$N = \sigma \int \mathcal{L}(t)dt \tag{2.5}$$

Where the luminosity is now integrated over the run time of the experiment, and the number of observed occurrences of the process, also referred to as *events*, is given by N. Of course in order to count these events one must devise a way to observe the particles that are produced as a result of the collisions.

2.1.2 Detecting Particles

During high energy collision, many of the particles that are produced are inherently unstable and will quickly decay to more stable states. These unstable particles will travel a distance that is dependent on the mean rest frame lifetime of the particle, τ, and the Lorentz factor, $\gamma = 1/\sqrt{1 - v^2/c^2}$, which accounts for time dilation due to special relativity. Depending on τ, the particle may decay at a distance of a few metres from the interaction point (IP) [1], in which case one might observe both the initial and final states, or it may decay effectively immediately and so one can only ever hope to observe the final decay products. Commonly these decay products will come in the form of photons, light leptons or more stable hadrons such as pions. The methods by which different particles may be detected depends on the nature by which they tend to interact with matter.

Charged particles travelling at relativistic speeds through a medium will ionise atoms in the process and thereby lose energy as they traverse the material. The rate of this loss per distance travelled for a particle travelling at velocity v $= \beta c$, through a medium with number density n and atomic number Z, is given by the *Bethe-Bloch* equation, defined as:

$$\frac{dE}{dx} \sim -4\pi\hbar^2 c^2 \alpha^2 \frac{nZ}{m_e v^2} \left(ln(\frac{2\beta^2\gamma^2 c^2 m_e}{I_e}) - \beta^2 \right) \tag{2.6}$$

where m_e is the mass of the electron, I_e is the ionisation potential of the material and *alpha* is the fine-structure constant. In simple terms, the rate at which a particle will deposit energy is therefore predominantly dependant on the density, ρ, of the material through which it is travelling, and the velocity of the particle itself. This is shown in figure 2.1, where the function has been scaled by the density, and the curves show the

[1] Various beyond standard model theories also put forward Long Lived Particles (LLPs) that decay well outside any current experimental setups [48].

energy loss of a particle with unit charge travelling through some common elements. Notably one can see that dE/dx is maximal for low velocity particles, and only rises logarithmically in the relativistic regime.

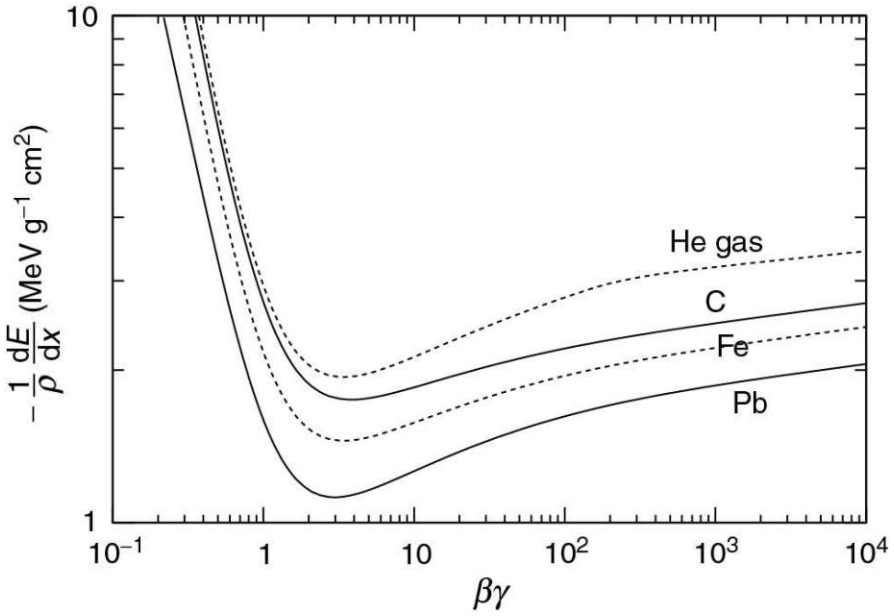

Figure 2.1: The rate of energy deposition in different materials for a particle of unit charge as a function of the particles velocity (boost) [1]

A typical experimental detector setup will include various sub-systems each with their own specific aspect of the event that they might detect. Immediately around the interaction point one commonly finds a tracking sub-system. Today these are often comprised of semiconductor materials arranged in radial layers around the IP. A charged particle passing through this will then ionise atoms and thereby create a measurable electric signal. Such detectors are typically segmented in some manner. By then considering the signals created in given segments across the individual layers one can begin to construct a picture of the track described by the particle as it travels radially out from the IP. The addition of a strong magnetic field then forces the charged particle to describe a curved trajectory, the radius and direction of which can by used to infer both the momentum and charge of the particle.

Continuing radially outward from the tracking system, one arrives at the calorimetry sub-detectors, coming in the form of electromagnetic or hadronic calorimeters, the job of which being to entirely stop particles and thereby measure their energy. The former of these will observe *electromagenetic showers* created when either an electron or a photon interacts with the medium. This is the process by which a high energy electron will emit a *bremsstrahlung* photon, which in turn decays to an electron/positron pair which themselves carry on to emit further photons. Typically an electromagnetic calorimetry

system is comprised of an array of segmented scintillating crystals, like in the tracking system this allows one to resolve the angle at which the shower occurs.

Hadronic showers are somewhat more complicated processes that involve decay chains of hadrons, with multiple decay and interaction processes taking part. These tend to be far larger than their electromagnetic counterparts. To account for this, hadronic calorimeters are comprised of layers of dense absorber materials which induce the showers, with active detector material spaced between in order to sample the energy deposition.

Figure 2.2 shows a depiction of this layered radial structure of sub-detectors that make up a full collider-based experiment. It shows the layered tracking subsystem with slightly curved tracks left by the charged particles passing through. One can see then how the electron and photon (e^- and γ) are stopped completely in the electromagnetic calorimeter (ECAL), and likewise a neutron and pion (n and π^+) traverse this before creating showers in the hadronic calorimeter (HCAL). The muon is shown traversing all of these sub-systems, with a small amount of energy being deposited in the ECAL before then hitting the muon system which is commonly found on the outermost layer of an experiment. Located as such it can additionally serve as a veto system for those produced outside of the experiment. Finally the diagram shows a neutrino (ν) passing through all sub-systems with no interaction. One commonly has to infer the presence of neutrinos by conservation of momentum: the sum of momenta of all detected particles should come to zero, if not then one can infer that some particle has escaped undetected.

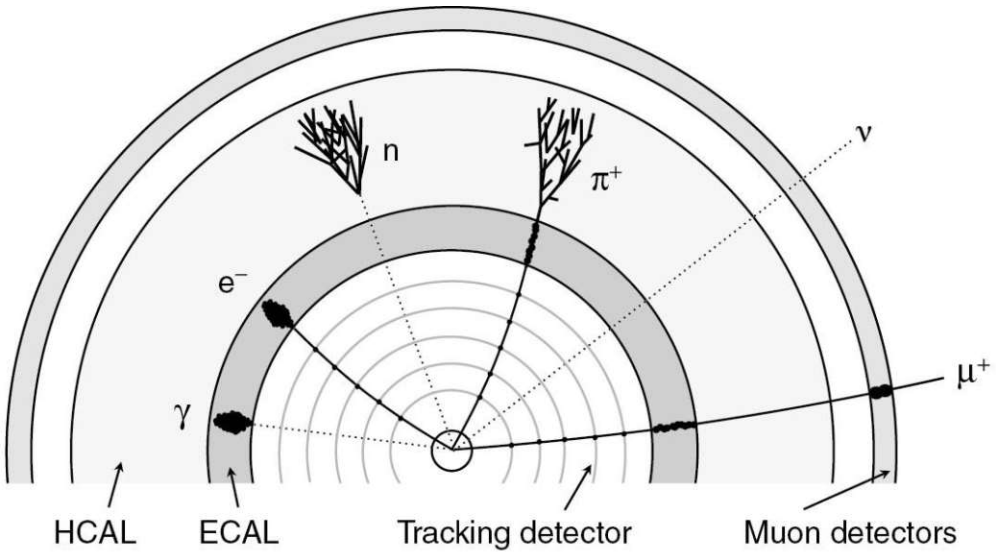

Figure 2.2: A depiction of a simple collider experiment detector set up. This shows the radial structure of subsystems and the types of particles they might be expected to detect [1].

2.2 The Belle II Experiment

There are two main approaches to probing the standard model (and what may lay beyond it) using particle colliders. The more well know being the approach of exploring the *energy frontier*, as is done with the Large Hadron Collider (LHC) at CERN. Protons beams are accelerated to an energy of 6.8TeV while travelling around the 27km circumference beam pipe before smashing into one-another at a *centre-of-mass* energy of 13TeV. This allows for physicists working on the 4 experiments there (ATLAS, CMS, ALICE, and LHCb) to probe even the most massive particles of the standard model, and led to the first observation of the Higgs boson by ATLAS [49] and CMS [50] with a mass of around 125GeV.

The Belle II experiment on the other hand conducts searches at the *intensity frontier*, which utilises sheer volumes of data to carry out accurate measurements of rare particle interactions with low statistical uncertainty. Belle II is often referred to as a *B factory*, due to it operating at a *centre-of-mass* energy equal to that of the $\Upsilon(4S)$ resonance (10.58 GeV). The result of this is a greatly enhanced rate of production of B-meson pairs, opening the door to many new ways to probe the SM. Indeed this allowed the Belle experiment (Belle II's predecessor), along with the BaBar experiment, to observe charge parity (CP) violation in neutral B-meson decays [51].

2.2.1 SuperKEKB

The Belle II detector lines at the interaction point of the SuperKEKB electron-positron collider. This is comprised of a *high energy ring* (HER) that circulates electrons at 7 GeV around the 3km circumference of the collider, and a *low energy ring* (LER) that circulates positrons at 4 GeV in the opposite direction. Electrons and positrons are initially accelerated and introduced to the beam line by the injector linac (linear accelerator), with a 1.1GeV positron damping ring[2] [52]. This structure is shown in figure 2.3

The asymmetric beam energies are purposefully chosen to introduce a Lorentz boost of $\beta\gamma = 0.28$, the result of which is a boosted centre-of-mass system. Due to the collider operating at the centre-of-mass energy equal to the $\Upsilon(4S)$ (10.58 GeV) resonance, the $B\bar{B}$ mesons, with masses of 5.279 GeV/c^2, are produced essentially at rest. Due to their very short life time ($\tau = 1.519 \times 10^{-12}$s [53]) they would therefore travel an imperceptible distance and thus make measuring the decay vertices, which is fundamental to the study of time-dependent CP violation, impossible.

As the only current member of a second generation of B-factory accelerators, the target integrated luminosity of SuperKEKB aims to dwarf that of its predecessor, and with that open the door to measurements of SM parameters with unprecedented precision. In June 2022 SuperKEKB achieved a world record instantaneous luminosity at a value of $4.7 \times 10^{35} cm^{-2} s^{-1}$ [54]. This was achieved by introducing the *Nano-beam* scheme, which essentially focuses the beam down to approximately 10μm and 0.06μm in the x

[2] This reduces the *beam emittance* - essentially regulating the spread and momenta of particle bunches in order to improve luminosity.

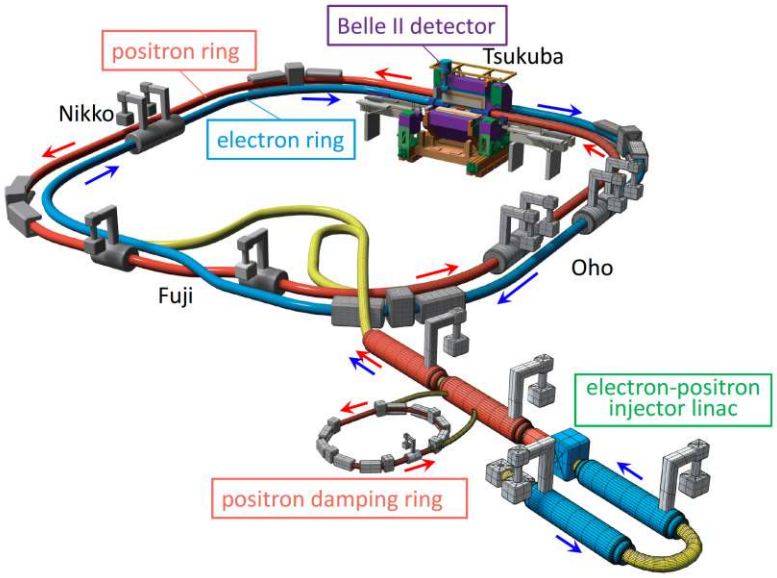

Figure 2.3: A depiction of the SuperKEKB collider and Belle II detector [52]. The elecron and positron rings are shown in blue and red respectively.

and y directions (perpendicular to the beam axis). Additionally to this the beam current is increased to around double that of KEKB (the predecessor to superKEKB) [55].

2.2.2 Beam Backgrounds

With this increased luminosity comes a large boost in the amount of *beam backgrounds*. These are comprised of a set of different processes:

- *Touschek scattering*: This is a result of coulomb scattering between particles within bunches as they travel around the beam and is exacerbated by the Nano-beam scheme. When two particles within a bunch interact via coulomb scattering, an amount of energy is imparted from one onto the other and thereby both deviate from the nominal bunch energy. The particles are then commonly lost into the beam pipe inner wall which, when occurring close to the detector, can produce a shower of particles. The rate of this background relies on multiple factors and is estimated to happen at a rate 20 times higher than in the KEKB accelerator. It is mitigated using collimators and shields around the ring that serve to catch such particles before they get near to the IP.

- *Beam-gas scattering*: Air is of course evacuated from the beam line however a perfect vacuum can never be achieved and so there is a contribution to the beam backgrounds coming from direct interaction between the beam and what few particles may still remain. This again commonly occurs via coulomb scattering or by

Bremsstrahlung. The result is much the same as that of Touschek and so the same countermeasures are found to be effective.

- *Synchotron radiation*: Synchotron radiation is an inherent problem in the nature of accelerating charged particles around a collider ring. It is produced when any relativistic charged particle feels an accelerative force perpendicular to the direction of motion, just as one orbiting a collider does. The result is electromagnetic radiation that is emitted tangential to the collider. This background is a particular danger to the silicon tracking detectors near to the IP which can be damaged by the radiation produced. A gold absorber is applied round the inner surface of the beam pipe to mitigate this.

- *Radiative Bhabha events*: Photons produced in radiative Bhabha events ($e^+e^- \rightarrow e^+e^-\gamma$) travel along the direction of the beam axis and strike the iron magnets which in turn produces large numbers of neutrons. This occurs by way of the photo-nuclear effect in which the photon will excite a nucleus which then kicks out a neutron during the subsequent de-excitation. These constitute a large source of backgrounds for the KLM detector and are combated by shielding in the tunnel housing the beamline.

- *Two photon process*: Finally the two photon process $e^+e^- \rightarrow e^+e^-e^+e^-$ will produce electron positron pairs with very low momentum which results in the tracks curving within the CDC and producing many hits from just a single particle.

2.2.3 Detector Subsystems

Figure 2.7 highlights the position of each of the following detector sub-systems in the Belle II experiment.

Vertex Detector - VXD (PXD/SVD)

It is paramount that a B-factory experiment achieves accurate measurement of decay vertices, which invariably occur at fractions of a centimeter from the interaction point. The Belle II experiment makes use of two devices working in unison that comprise the VXD subsytem. These are the Silicon Vertex Detector (SVD) and the Pixel Detector (PXD), both of which are arranged in radial layers around the interaction point. The former of these comprises four layers of double-sided silicon strip sensors at radii ranging from 38mm to 140mm. The latter contains two layers of pixelated DEPFET sensors at radii of just 14mm and 22mm. This layered structure can be seen in figure 2.4, where the SVD is shown in red and the PXD is shown in light blue [56]. These tight dimensions not only provide excellent determination of decay vertices near to the IP, but provide the ability to reconstruct tracks of low momentum particles that might otherwise curve within the inner CDC due to the magnetic field.

There are 20 diamond detectors (each containing a single 4.5×4.5× diamond crystal) arranged isotropically around the beam pipe and residing within the PXD/SVD support

structure. These are designed to monitor the beam backgrounds, which could cause radiation damage in the VXD system, and facilitate an immediate shut-down response if these are to spike.

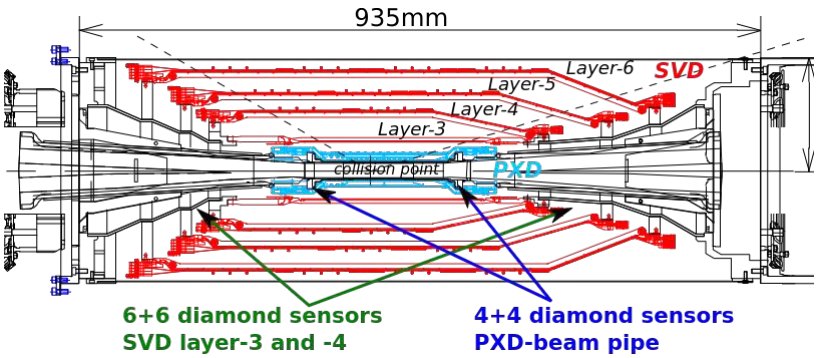

Figure 2.4: The VXD subsystem, with the SVD shown in red and the PXD in light blue. Additionally the positions of the diamond sensors are indicated [56].

Electronic Calorimiter - ECL

The *electronic calorimiter* (ECL) is comprised of 8736 thallium-doped caesium iodide CsI(Tl) crystals, separated into 3 sub-sections covering the barrel and forward/backwards end caps. This covers around 90% of the solid angle around the interaction point and gives excellent angular resolution for detection of ECL showers. As discussed earlier, the main purpose of the ECL is for the detection of photons but also for the separation of electrons from other particles such as pions.

Central Drift Chamber - CDC

With a radius of 1130mm, the *central drift chamber* (CDC) is the largest of the sub-detectors that make up the full Belle II experimental apparatus. The CDC encases the PXD and SVD, and makes up part of the tracking system of Belle II. It is comprised of approximated 14000 wires arranged in 56 layers inside a chamber that is filled with an equal mixture of He and C_2H_6 gasses. The layers of wire are grouped into 9 super-layers orientated in two distinct ways: *axial*, meaning the wires of the layer are aligned with the magnetic field produced by the solenoid, and *stereo* where a slight angular skew of either 45.5 or 74 mrad from the magnetic field is added. This structure is shown in figure 2.5.

Charged particles moving through the CDC will cause ionisation of the gas mixture, the resulting ions are then accelerated towards the wires due to an induced electric potential difference. Upon reaching these wires, the cascade of ions then produces a pulse which can be used to infer position along the wire. The layers of wires with altered orientation thus allows for the charged particle's path to the reconstructed in 3 dimensions [57].

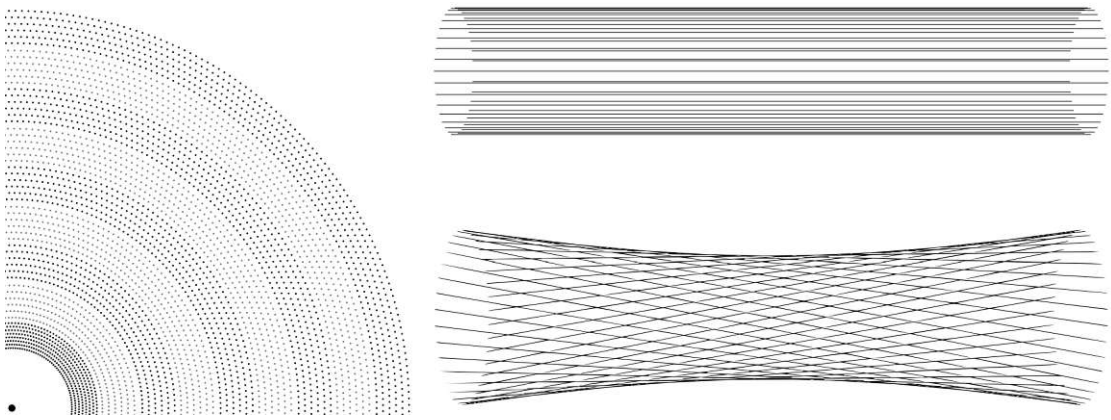

Figure 2.5: The left figure shows a cross section in the r-ϕ plane, with wires indicated by dots and super-layers shown with alternating shades. [58]

KLM

The K_L-Muon detector (KLM) resides at the outermost layer of the Belle II experiment, beyond the superconducting solenoid that provides the magnetic field. It is comprised of a layered structure of 4.7cm thick iron plates alternated with resistive plate chamber elements for sampling of deposited energy. The iron further serve as a magnetic flux return for the solenoid. Due to the expected high beam backgrounds, the active material has been upgraded from the glass-electrode resistive plate chambers that were used in Belle, with Belle II now using layers of scintillating material with silicon photomultipliers (SiPMs).

PID system (TOP and ARICH)

The final two subsystems of Belle II are the time-of-propagation counter (TOP) and aerogel ring-imaging detector (ARICH), which together comprise the particle identification system (PID). The TOP resides in the barrel region and is comprised of sixteen modules encompassing the CDC, each of which consists of a 2.6m quartz bar, with a second smaller prism that expands to cover a photo-detector at the end as shown in figure 2.6. These photo-detectors have a single photon time resolution of just 100ps. The ARICH detector is situated at the forward endcap region of the detector and is primarily designed for the seperation of low momentum pions and kaons. Both of these systems utilise the Cherenkov radiation that is produced when a charged particle travels through a dialectric medium (in the case of ARICH this is aerogel) faster than the phase velocity of light in that medium. This creates a ring of electromagnetic radiation, the geometry and clarity of which in the detector can be used to infer the type of particle and it's momentum.

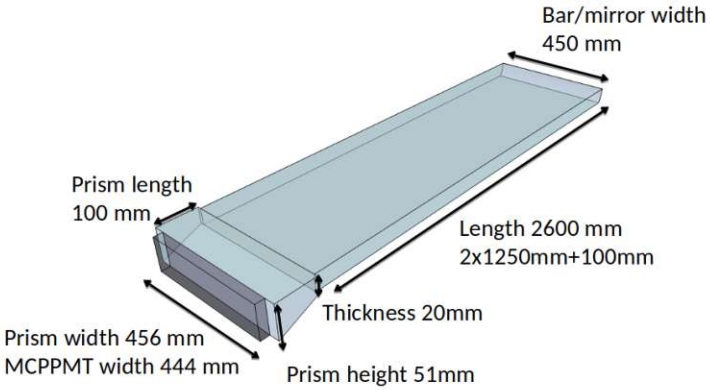

Figure 2.6: One of the sixteen modules of the TOP detector, comprised of two quartz crystals (a cuboid and a prism) with a photo-detector at the end [59].

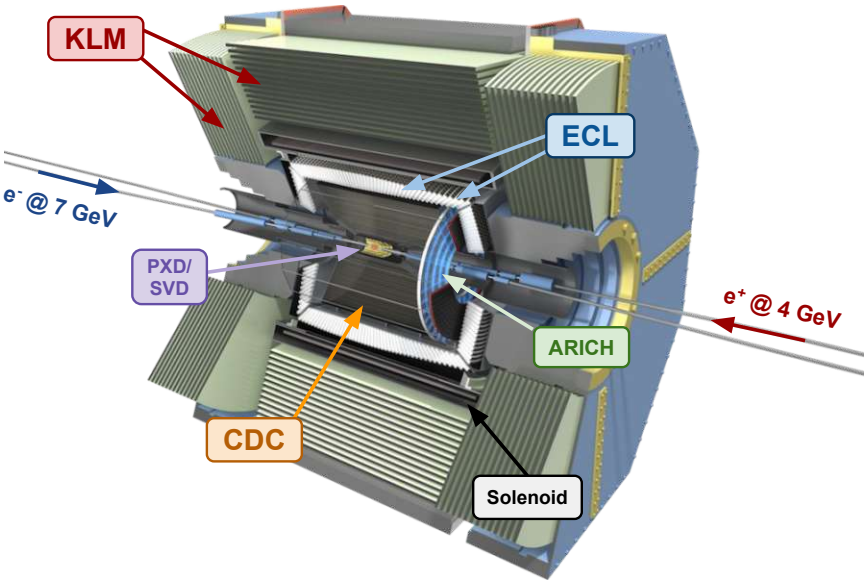

Figure 2.7: A depiction of the Belle II detector with the various subsytems indicated along with the solenoid.

2.2.4 Charged Particle Identification

The identification of charged particles at Belle II is done by way of using a combined likelihood where information from each of the previously discussed sub-systems is considered to provide discrimination between the likes of electrons, muons, pions and kaons.

Each of them provides an individual likelihood, $\mathcal{L}_i^{det}$, for the given particle hypothesis i. A global likelihood can then be defined as the product of this likelihood across all sub-detectors [60],

$$\mathcal{L}_i = \prod_{det}^{CDC,ECL,...} \mathcal{L}_i^{det} \tag{2.7}$$

The ECL for instance provides measurement of the energy deposited by a charge particle, while CDC provides the momentum. The ratio of these quantities can be quite consistent for some particles, for instance in the case of electrons E/p is invariably often close to 1. The KLM instead can differentiate between kaons and muons (and also photons) by providing information about the depth to which the particle may penetrate into the layers.

With a set of measurements x, and given all the possible particle candidates, $A_j = e, \mu, \pi, ...$, the likelihood ratio can then be defined as,

$$P(X)_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j} \tag{2.8}$$

With this, each charged track in the event then also comes with the probability $P(X)_i$ for each possible particle hypothesis, and thus one can apply some selection cuts to achieve pure samples containing only the desired candidates.

2.2.5 Trigger system

The trigger system serves the role of identifying individual events that may be of physics interest. Different *trigger lines* are devised with varied requirements such as number of tracks or energy of clusters in a given event. As discussed a large number of beam-induced backgrounds are observed during data taking. Most of these can be characterised by events containing just one or two tracks and ECL clusters, meaning they can look quite similar to various low-multiplicity or dark sector searches. Thus an extensive and robust array of trigger lines provides a strong advantage to Belle II in these fields, opening doors to studies that previous B-factor experiments were unable to undertake.

The trigger is comprised of two separate sub-systems, the *low level trigger* (L1) and the *high level trigger* (HLT). The L1 trigger is hardware based and allows for low latency, high frequency triggering with a maximum rate of 30kHz at which it is limited by the read-in rate of the data acquisition system [55]. It takes event data from the CDC, ECL, TOP and KLM sub-detectors, each of of which has its own sub-trigger system monitoring specific aspects of the events: the ECL monitors number and energy of clusters, the CDC reconstructs charged particle tracks, the KLM can observe muons independently from the CDC, and finally the TOP provides event timing information. This information is then propagated to the Global Reconstruction Logic (GRL) which must then decide to pass it on to the HLT or discard the event in just a $5\mu s$ window. This is a new addition to the trigger system with respect to Belle and is fundamental to operating at the increased luminosity.

The purpose of the HLT is then to further analyse these events passed on from the L1 trigger, with the main task of reducing the event rate to 10kHz for storage. In addition to this, the HLT selects tracks of interest in the PXD to then be saved, thus greatly reducing the data readout from the sub-system by vetoing beam backgrounds.

2.2.6 Belle II software

The Belle II Analysis Software Framework (basf2 [61]) is used for for the purposes of event *reconstruction*. That is to say, it is the means by which one can process low-level detector information such as individual hits in the CDC or ECL and interpret them as higher level objects such as full charged tracks with matched clusters. Basf2 is a framework containing individual modules, mostly written in C++, that each perform specific sequential tasks in some order defined in a *steering file*, which are instead commonly written in Python.

3. Phenomenological Search for New Physics

The work described herein, specifically the statistical analysis of data, was conducted by the author, in collaboration with W. Altmannshofer, A. Crivellin, G. Inguglia and J. Camalich. The full results are published in Physical Review D [62].

3.1 Reinterpretation of $B \to K^{(*)}\nu\bar{\nu}$

In addition to direct searches for new physics, such as the searches for BSM particles described within this thesis, one can also look to probe specific standard model processes for possible NP-induced enhancement. This could take the form of perhaps an observed increase in branching fraction of some known process beyond that which is theorised in the standard model. Indeed the Belle II experiment recently measured, using 362fb^{-1} of data, the branching fraction of the $B^+ \to K^+\nu\bar{\nu}$ decay at a value of $[2.3 \pm 0.5(\text{stat.})^{+0.5}_{-0.4}(\text{syst.})]\times10^{-5}$, giving an excess above the standard model prediction of 2.7 standard deviations [63].

This value was derived via combination of two separate analyses within the collaboration that used either a *hadronic tagging* method of event selection, or an *inclusive tagging* method, with each measuring branching fractions of $[2.7 \pm 0.5(\text{stat.}) \pm 0.5(\text{syst.})]\times10^{-5}$ and $[1.1\ ^{+0.9}_{-0.8}(\text{stat.})^{+0.8}_{-0.5}(\text{syst.})]\times10^{-5}$ respectively. The former method involves the complete reconstruction of the hadronic decay of one of the B mesons produced at the interaction point, with the other B meson being then used to study the $B^+ \to K^+\nu\bar{\nu}$ decay. The latter instead merely requires the reconstruction of the decay of interest, and all other detected tracks and clusters are subsequently attributed to what is dubbed as the *rest-of-event*. This inclusive method thereby achieves a considerably higher signal efficiency and, in this Belle II study, provides a measurement ~ 3 standard deviations above the SM prediction.

One can reasonably surmise then that perhaps there is indeed some other process contributing to this decay, thereby causing this apparent excess. As discussed previously, neutrinos cannot feasibly be detected in collider experiments and their existence must be inferred via the presence of missing energy, much like in the search for a Z' boson (see section 1.3). One cannot necessarily discern if that missing energy is due to the presence of neutrinos, or some new BSM particle.

With that, the measured excess in BF could thereby by reinterpreted as being artifi-

cially enhanced by the two body decay B→KX, where X is some long lived or invisibly decaying particle. This puts some constraint on the possible mass of the particle, with the requirement $m_X \leq m_B - m_k$, and also thereby dictates that the presence of this decay process would produce some resonance-like enhancement in the q^2 distribution (in the nominal B→K$\nu\bar{\nu}$ process this quantity would equate to the squared invariant di-neutrino mass).

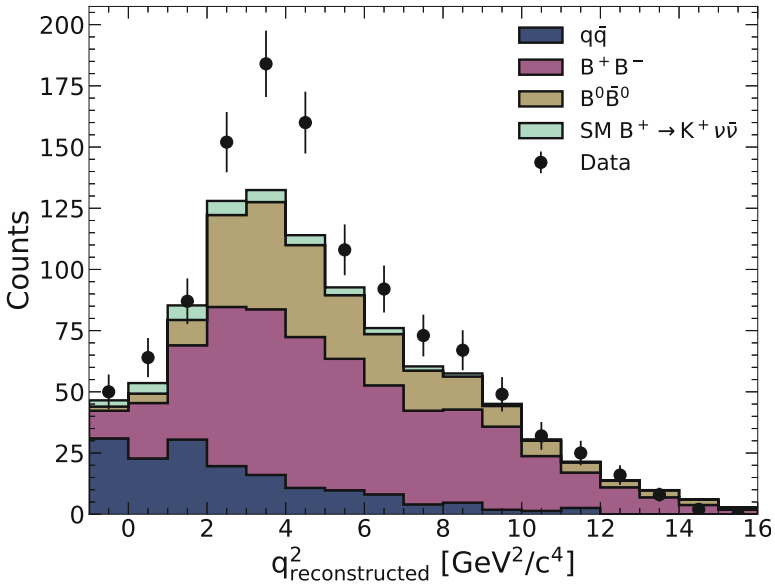

Figure 3.1: Reconstructed q^2 distribution provided by the Belle II collaboration. The standard model MC generated $B^+ \to K^+\nu\bar{\nu}$ and backgrounds are shown, along with the observed data counts. This shows a clear peaking excess in data around 3 to 4 GeV^2/c^4.

Figure 3.1 shows the published reconstructed q^2 distribution found by the Belle II analysis [64], both with MC predicted counts and observed data. This distribution is given only for the inclusive tagged method, and so only it and not the hadronically tagged analysis can be included in this study. The $B^+ \to K^+\nu\bar{\nu}$ histogram is normalised to just the expected standard model contribution. One can see a clear resonant-like structure in the data, peaking in the 3 to 4 GeV^2/c^4, well above the distribution due to SM predictions.

This sample makes up one of the four *signal efficiency quantiles* that define the signal region from which the $B^+ \to K^+\nu\bar{\nu}$ BF is extracted in the Belle II study. This full signal region is shown in fig. 3.2, where the four efficiency quantiles, each containing 2% signal efficiency, are denoted by the *signal discriminator* variable[1]. Each of these quantiles then contains an expected 2% signal efficiency. While the top efficiency quantile depicted in figure 3.1 displays the resonant-like excess very well, it is this full signal region sample that is utilised in the fitting strategy here.

[1]This is defined as 1 - signal eff., meaning the last quantile covers the region 0.98 - 1.00

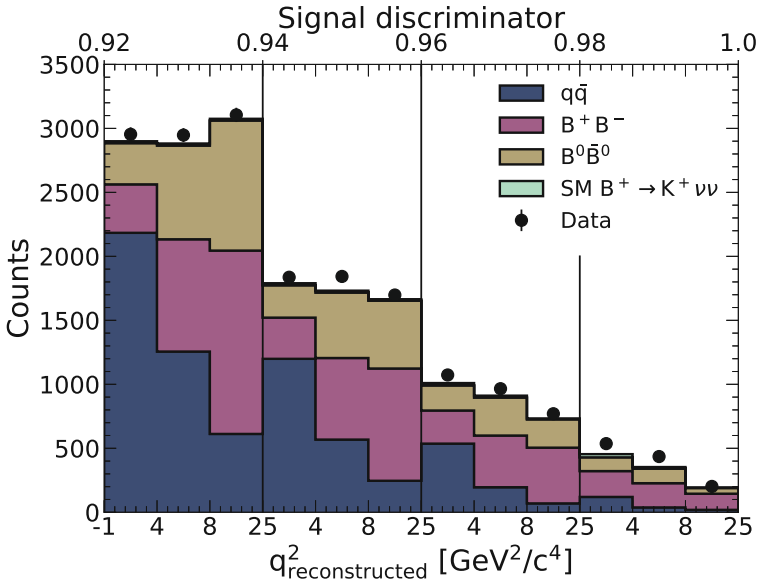

Figure 3.2: Reconstructed q^2 distribution provided by the Belle II collaboration. The data are binned in 4 *signal efficiency quantiles*, each of which containing 2% signal efficiency. These are then each binned in 3 q^2 bins, thus providing the 12 bins shown.

Of course, to properly constrain any NP contribution that may be present in such decay processes, it is important to consider all measurements that have been made. The BaBar collaboration conducted a similar search, this time using the hadronic tagging method and including also the decay to neutral kaons ($B^0 \rightarrow K^0 \nu\bar{\nu}$), and also charged and neutral decays to excited kaon states ($B^{+/0} \rightarrow K^{*+/0} \nu\bar{\nu}$)[65]. Figure 3.3 shows the resulting reconstructed q^2 distributions for the charged and neutral decay channels [2]. Analyses using the semi-leptonic tagging method (whereby $B \rightarrow D^* l \nu_l$ decays are reconstructed from one of the B-mesons produced at the interaction point and the other is used to probe the decay of interest) have been conducted by the Belle [66] and BaBar collaborations [67], however the reconstructed q^2 is provided in neither and so cannot be included in this study.

Contrary to the observed data-to-simulation agreement of the Belle II study, there appears no clear excess of data above standard model expectations in either the charged or neutral channels studied by the BaBar collaboration. Thus one can expect that the inclusion of this study shall constrain heavily the possible contribution of the 2-body decay $B \rightarrow KX$. However, it is important to note that the signal efficiency in this analysis is considerably lower than that of the Belle II study, at approximately 0.75% and 0.37% for the charged and neutral channels respectively. Furthermore the region between 3 and 4 GeV^2/c^4 in which the Belle II excess appears has severely limited statistical precision

[2]The BaBar study provides these distributions in a variable, S_B, defined as q^2 scaled by the squared B meson mass. This scaling is reversed so as to have all studied distributions defined in q^2

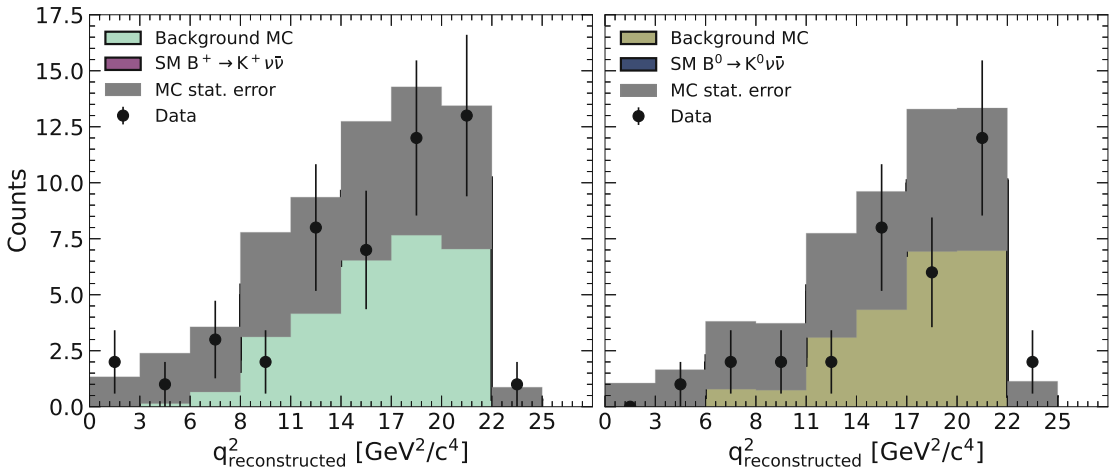

Figure 3.3: Reconstructed q^2 distributions of the studied $B^+ \to K^+ \nu\bar{\nu}$ (left) and $B^0 \to K^0 \nu\bar{\nu}$ (right) decays, provided by the BaBar collaboration. The expected standard model MC signal and backgrounds are shown, along with the observed data counts.

due to the near complete lack of events.

Figure 3.4 shows the reconstructed q^2 distributions for the charged and neutral decay to excited Kaon states found by the BaBar analysis. These suffer from not just larger backgrounds than those in the non-excited channels, but also signal efficiencies approximately an order of magnitude lower. This results in the SM $B^+ \to K^{*(+/0)} \nu\bar{\nu}$ occupying a barely visible portion of the histograms, and would lead one to expect that this is unlikely to constrain heavily any possible NP contribution.

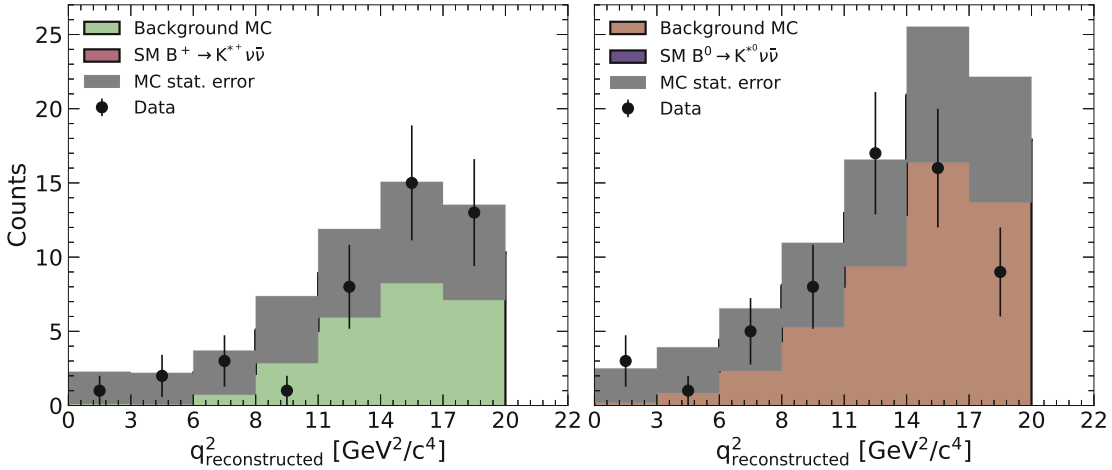

Figure 3.4: Reconstructed q^2 distributions of the studied $B^+ \to K^{*+} \nu\bar{\nu}$ (left) and $B^0 \to K^{*0} \nu\bar{\nu}$ (right) decays, provided by the BaBar collaboration. The expected standard model MC signal and backgrounds are shown, along with the observed data counts.

The three q^2 distributions associated with decays to ground state kaon systems (figures 3.2, 3.3) are used, both individually and combined, in a binned maximum-likelihood fit to extract a branching fraction, Br[B→KX], and preferred mass, M_X, associated with some new physics contribution. This is carried out with the `pyhf` software package[3] [68, 69]. Three separate fits are conducted, the first of which using only Belle II data and MC. The second is a combined fit where both the charged and neutral BaBar distributions are fitted simultaneously, and finally all three are used together in a single simultaneous combined fit. A fourth fit is conducted separately using the combined channels of charged and neutral decays to excited kaon states (figure 3.4).

In each of these, one must define templates that are then fit to the observed data distributions. For the fits to BaBar data, the templates included are the respective homogeneous backgrounds of the two channels, whereas in the fit to Belle II data the individual backgrounds ($q\bar{q}$, $B^0\bar{B}^0$, B^+B^-) are included. The expected standard model distribution of $B^{(*)} \to K^{(*)}\nu\bar{\nu}$ is included in each, with the normalisation of this template correlated across all channels.

A template to represent the proposed new physics contribution of the B→K$^{(*)}X$ decay is included, and the parameter-of-interest of the fit is then the normalisation, or *signal strength*, of this. The resonance is modelled by a simple Gaussian distribution with a standard deviation of 1.5 GeV^2, while the decay itself is expected to have negligible width, this value is chosen so as to approximate the detector resolution of Belle II[4]. This was estimated through study of the q^2 distribution in simulated $B^+ \to K^+\bar{D}^0$ decays, where the subsequent $\bar{D}^0$ decay is omitted from calculation of q^2, thereby roughly approximating the kinematics of $B^+ \to K^+X$ (under the assumption that $M_X \sim M_{\bar{D}^0}$).

The individual background templates of the Belle II channel are each assigned a pre-fit 10% normalisation uncertainty, and the $q\bar{q}$ template is additionally assigned a 10% shape uncertainty, allowing each bin to fluctuate independently by 10%. These are chosen to roughly reflect values defined in the Belle II study itself. Only statistical uncertainties are considered in the BaBar analysis. This is of course a very simplistic application of experimental uncertainties that are present in these measurements of the $B^{(*)} \to K^{(*)}\nu\bar{\nu}$ processes however for the purposes of this phenomenological reinterpretation, where no real measurement is being done, they are deemed sufficient. Furthermore, due to the lack of access to the full Monte-Carlo samples that would be used in the original analyses, this fitting procedure is heavily limited by statistical uncertainty.

As a basic check, a fit to extrapolate the $B^+ \to K^+\nu\bar{\nu}$ branching fraction is conducted to the Belle II signal region (omitting of course any NP contribution template), with the discussed template normalisation and shape uncertainties applied. From this a best fit value of $\text{Br}[B^+ \to K^+\nu\bar{\nu}] = (2.8 \pm 0.7)\times10^{-5}$ is found, showing good agreement with that which is derived in the Belle II inclusive analysis and thereby serving to confirm that the simple model describe herein is of good approximation. Furthermore, a similar

[3] The use of pyhf is discussed in detail in section 3.1, where it is also applied for statistical analysis in the search for an invisibly decaying Z′

[4] This of course does not necessarily accurately reflect the BaBar detector resolution but for such a phenomenological re-interpretation it is deemed a sufficient assumption.

fit to the combined charged and neutral BaBar distributions provides an upper limit of $Br[B\to K\nu\bar{\nu}] \leq 1.88\times 10^{-5}$. This is a slight excess of that which they quote ($Br[B\to K\nu\bar{\nu}] \leq 1.7\times 10^{-5}$), however, this is not unexpected and likely a result of the aforementioned statistical constraints due to lack of access to the full MC-simulated data.

Fits are conducted at mass values in the range 0.1 to 3.0 GeV/c^2, with the assumption that, in the two body decay $B\to KX$, $q^2 \sim M_X^2$, and so probing the q^2 range 0.01 to 9.0 GeV$^2/c^4$. The resulting best fit values of Br[B→KX] and Br[B→K*X], along with the respective $\pm 1\sigma$ error bands, are shown in the left and right subplots of figure 3.5 respectively. The former is split into the three previously discussed separate fits conducted: Belle II only (red), combined charged and neutral BaBar (light green), and the combined fit of Belle II and BaBar distributions (dark blue). As expected the combined BaBar charged and neutral channels shows the lowest fitted BF across all masses. Due to the lack of statistics available and low signal efficiency this also has largest error, giving a result that is compatible with a branching fraction of zero, meaning no significant indication of NP contribution.

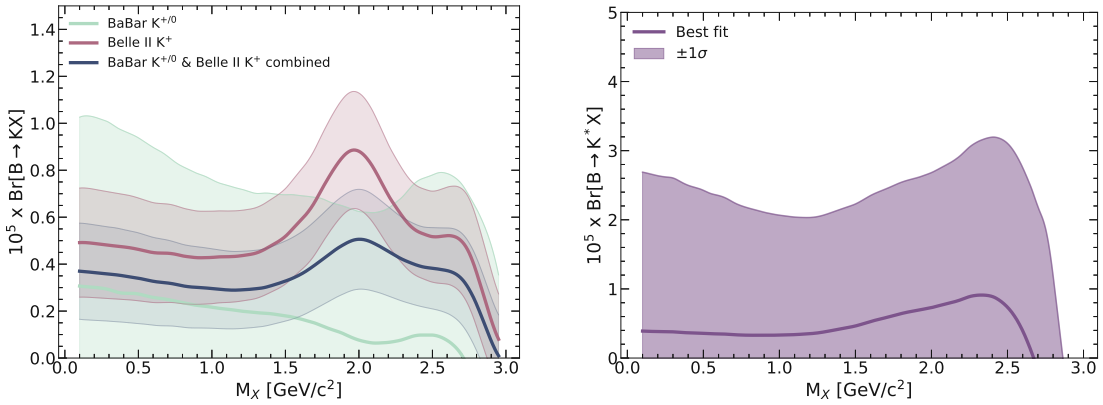

Figure 3.5: Left: Best fit values of Br[B→KX] for the three fits conducted on decay channels to ground state kaon systems. Belle II only (red), combined charged and neutral BaBar (light green), and the combined fit of Belle II and BaBar distributions (dark blue) Right: Best fit values of Br[B→K*X] for the combined fit conducted to the BaBar charged and neutral distributions.

Conversely the fit to the Belle II data shows a clear peaking structure in the fitted BF at $M_X \sim 2$ GeV/c^2, with a value of $Br[B\to KX] = (8.8 \pm 2.5) \times 10^{-6}$ at a significance of $\sim 3.6\sigma$. As one would expect this coincides with the observed resonant-like structure in data shown in figure 3.1. The combined fit of both Belle II and BaBar results then shows a similar structure with the expected stronger constraint enforced by the BaBar dataset. This results in a BF at $M_X \sim 2$ GeV/c^2 of $Br[B\to KX] = (5.1 \pm 2.1) \times 10^{-6}$ at a more reduced significance of just $\sim 2.4\sigma$. Shown in the right hand sub-plot, much like in the decay to ground state kaon systems, best fit values Br[B→K*X] are accompanied with large uncertainties due to statistical limitations and low signal efficiencies, leaving

the result compatible with a branching fraction of zero.

Figure 3.6 shows again the reconstructed q^2 distribution, this time with the addition of the Gaussian resonance representing the $B^+ \to K^+X$ process, at the best fit value for the shown data of $Br[B \to KX] = (8.8 \pm 2.5) \times 10^{-6}$. This demonstrates quite clearly the ability for such a resonance to account for the excess observed in data.

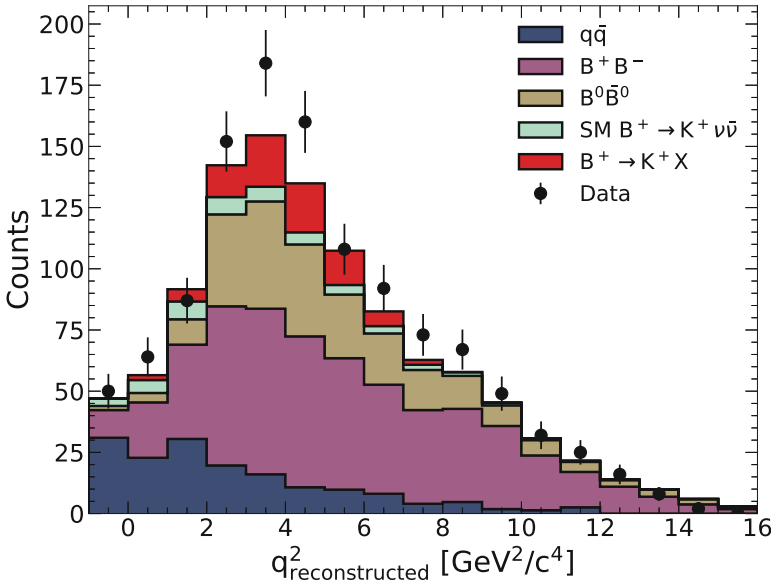

Figure 3.6: Reconstructed q^2 distribution provided by the Belle II collaboration. The standard model MC generated $B^+ \to K^+\nu\bar{\nu}$ and backgrounds are shown, along with the observed data counts, and also the contribution of a Gaussian resonance representing the contribution of $B^+ \to K^+X$ at the (Belle II only) best fit value of $Br[B \to KX] \sim (8.8 \pm 2.5) \times 10^{-6}$.

3.2 Light physics explanation of excess

With this statistical re-interpretation of the data concluded, and best fit particle mass and branching fraction associated with the suggested two body decay defined, one can now look for some theoretical framework that could explain the observation. Furthermore the results of the combined fit can be used to constrain parameters of such a model.

The SM $B^+ \to K^+\nu\bar{\nu}$ decay process involves the weak transition of a b quark to an s quark with the production of a neutrino/anti-neutrino pair. This is what is known as a flavour changing neutral current (FCNC) transition, and occurs at loop level, where contributions are highly suppressed both due to loop factors and also the associated elements of the CKM matrix [70]. Two of the leading order feynman diagrams describing this process are shown in figure 3.7. The left hand depicts a so-called *penguin diagram*, while the right hand shows a box diagram.

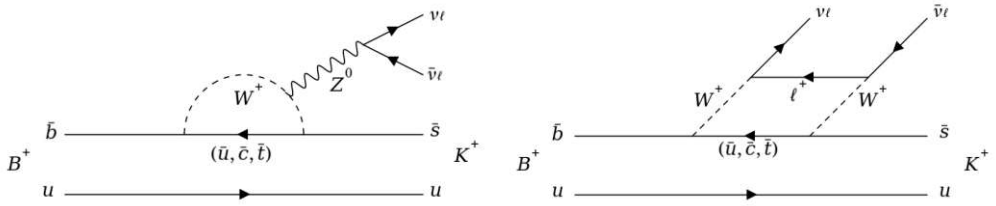

Figure 3.7: Leading order Feynman diagrams, an electroweak penguin (left) and a box diagram (right), for the quark level transition $b \to s\nu\bar{\nu}$, responsible for the decay $B^+ \to K^+\nu\bar{\nu}$.

Two beyond standard model particles that could cause the proposed B→KX are a flavoured axion or axion-like particle (ALP), and a light neutral Z′ boson. Both come with some basic caveats in order to fulfill this role, the first of which being that they must not decay to charged standard model fermions. If this were the case, such particles would have been observed in the study of decays involving $b \to sl^+l^-$ transitions, however recent measurement of which by the LHCb collaboration has shown no such enhancement to the SM expectations [32]. Any coupling to light leptons or quarks must be exceedingly weak such that the decay occurs outside of the detector, or the decay must be completely invisible. These criteria do then rule out the L_μ-L_τ Z′ model.

Full theoretical treatise describing the inclusion of such particles into the standard model is beyond the scope of this experimentally focused thesis work. However, the constraints that can be set on this Z′ model can be briefly outlined to highlight the impact of the study. The Lagrangian describing the interaction of this Z′ is defined (up to dimension 6) as:

$$\mathcal{L}_{Z'} \supset \left\{ g_L^{(4)} Z'_\mu (\bar{s}\gamma^\mu P_L b) + \frac{g_L^{(5)}}{\Lambda} Z'_{\mu\nu} (\bar{s}\sigma^{\mu\nu} P_R b) + \frac{g_L^{(6)}}{\Lambda^2} \partial^\nu Z'_{\mu\nu} (\bar{s}\gamma^\mu P_L b) \; + \text{h.c.} \right\} + \{L \leftrightarrow R\} \tag{3.1}$$

where $Z'_{\mu\nu} = \partial_\mu Z'_\nu - \partial_\nu Z'_\mu$ is know as the Z' field strength tensor. The Lagrangian is fully discussed in [71] but for the purposes here the important thing to note is that it contains vector and axial couplings, defined as $g_V^d = g_R^{(d)} + g_L^{(d)}$ and $g_A^d = g_R^{(d)} - g_L^{(d)}$ where L and R refer to the left or right handed *chirality* of the interaction.

The left hand plot in figure 3.8 shows coloured bands that define the correlations between the branching ratios of the B→K*Z′ and B→KZ′ for the different operators that are defined within the Lagrangian. These are shown in comparison with the upper limits on BR(B→K*Z′) set from Belle (hatched), the study of the more recent Belle II measurement on the non-excited decay (BR(B→KZ′)) in blue, and finally the full results from the analysis of the combined measurements in red. This plot uses values defined at the best fit mass of 2 GeV/c^2. Notably one can see that couplings to left or right-handed quarks ($g_{L/R}^{(4/6)}$, $g_{L/R}^{(5)}$) would produce larger B→K*Z′ branching ratios that exceed the experimental observations. Thus vector couplings ($g_V^{(4/6)}$) must dominate in such a

Z' model. The right-hand plot in fig. 3.8 shows the preferred regions in a 2D plane of the left and right-handed couplings, with the results for B→K*Z′ and B→KZ′ shown separately in yellow and blue respectively, and again the final combined results shown in red. From this one can see that couplings on the order of approximately just 1×10^{-8} are suggested by the data.

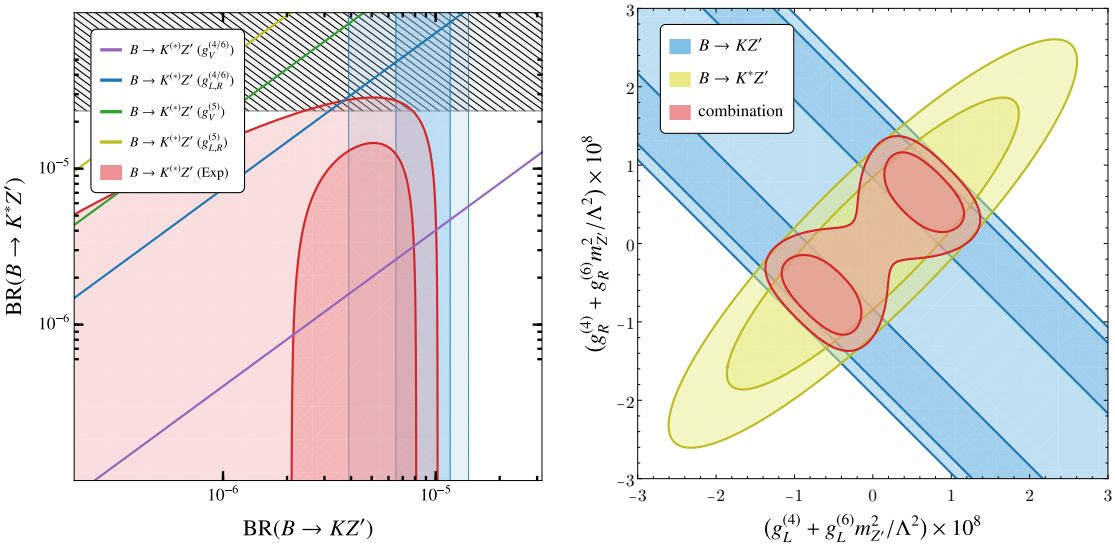

Figure 3.8: Left: Coloured bands showing the correlations between the branching ratios of the B→K*Z′ and B→KZ′ for the different operators that are defined within the Lagrangian. These are shown in comparison with the upper limits on BR(B→K*Z′) set from Belle (hatched), the study of the more recent Belle II measurement on the non-excited decay (BR(B→KZ′)) in blue, and finally the full results from the analysis of the combined measurements in red. Right: The preferred regions in a 2D plane of the left and right-handed couplings, with the results for B→K*Z′ and B→KZ′ shown separately in yellow and blue respectively, and again the final combined results shown in red.

This study of course is only a reinterpretation of the experimental measurement of the $B^+ \to K^+\nu\bar{\nu}$ decay process and, as such, in no way seeks claim confirmation that the observed excess BF is due to such a Z′. However, the study does serve as a means of exploring how new physics models can be probed by measurement of SM interactions. The results here can thus provide motivation for more dedicated searches for such particles.

Part II

Physics Performance and Analysis Tools

1. Charged Particle Tracking Performance

The work described herein was implemented solely by the author based on work previously conducted within the Belle II collaboration, and with advice provided by P. Rados and G. Inguglia. This work was carried out well before other studies discussed within this thesis and so results here have been superseded, nevertheless it is included here for completeness.

1.1 Measuring tracking efficiency with Bhabha events

As discussed in section 2.2, charged particles that are created during collisions will describe curved trajectories through the detector material due to the presence of the magnetic field. The path of such a particle can be reconstructed as a trail of consecutive hits through the VXD and CDC. The momentum can be calculated from the track curvature, and tracks can be associated with hits in other sub-detector systems such as the KLM or ECL. As such, a charged particle will have some track with associated clusters, whereas a neutral particle will appear only as the clusters with no track.

In the event of a mis-reconstruction by the Belle II reconstruction software, one may find a charged particle being mis-identified as a neutral. This study investigates the case in which software event reconstruction fails to properly associate an electron's ECL cluster with its track in the CDC/VXD and thus attributes it to a non-existant photon.

To identify such occurrences one must define some event criteria in which an electron will land in a specific, expected region of the detector. As such one would expect when looking at this region to find a reconstructed electron with its associated track from the CDC/VXD and cluster in the ECL. If, on the other hand, not an electron but a photon is found then one can conclude that there has been some failure in the track reconstruction software.

The study utilises a *tag-and-probe* method with radiative bhabha scattering events ($e^+e^- \rightarrow e^+e^-\gamma$). A tag side consisting of one electron and the photon is explicitly reconstructed, along with a probe side consisting of the other electron (if present) and any other reconstructed photons, the full reconstruction criteria is shown in table 1.1. This defines the angular and energy requirements for the tag electron and photon, and the probe side. The sum of cluster energies of both tag and probe sides is required to be greater than 10.4 GeV, such that only well captured events with closed kinematics are kept. Subsequently a recoil momentum vector is calculated with respect to the electron-

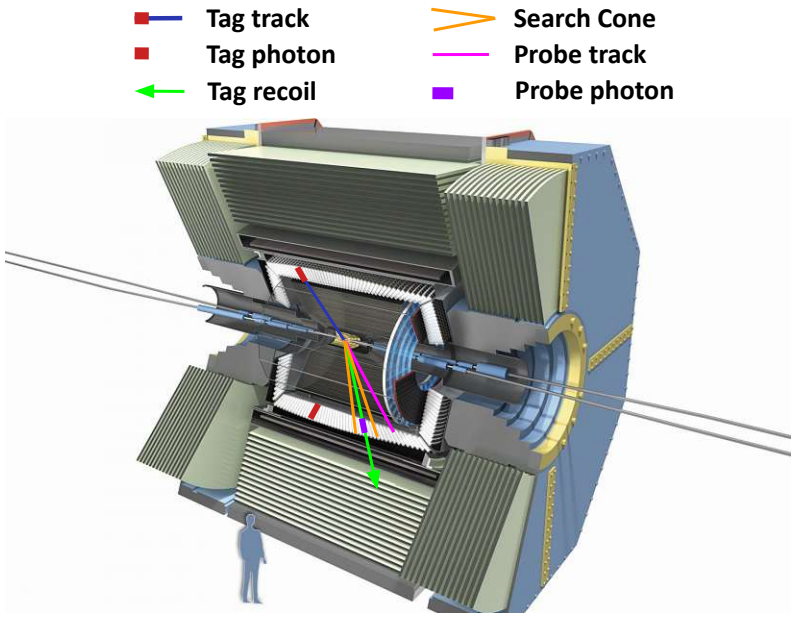

Figure 1.1: A depiction of the track or gamma matching process in the belle II detector. The search cone is shown along with the tag and probe tracks and photons. This shows the *Gamma match* case, with the probe track residing outside of the search cone.

photon tag side which then defines the direction in which the probe electron should be found.

Figure 1.1 shows a diagram of this process. A 20° search cone is defined in the recoil direction, within which one may observe one of four outcomes:

- *Track match*: The second electron track is found within the cone and so the track-cluster reconstruction was successful. The track must be of opposite charge to the tag electron, and the energy must be greater than the tag photon.
- *Gamma match*: No charged track is found, but an ECL cluster that has been associated to a photon is present. This would indicate instead that the algorithm failed to match the track of the electron to its ECL cluster, and thus falsely attributed it instead to a photon.
- *Track and gamma match*: Both a charged track and a photon are reconstructed within the search cone. This is likely a result of FSR from the electron.
- *No match*: No cluster or track is found in the search cone. This is rare (<0.1% of events) and likely due to the electron completely escaping ECL coverage (for instance through a gap between crystals).

The cluster-track matching efficiency can then be defined as the ratio of events with successfully matched tracks to the total number of events,

$$\epsilon = \frac{N_{\text{good}}}{N_{\text{good}} + N_{\text{bad}}} \tag{1.1}$$

Where N_{good} refers to events with a track match or track and gamma match, and N_{bad} refers to events with a gamma match, or no track or gamma match at all. This quantity is measured in both MC and data, and the cluster-track match inefficiency ratio is calculated as,

$$\delta^* = 1 - \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}} \tag{1.2}$$

This is of course the important quantity to define, as it provides a level of uncertainty one must have in the simulated MC. Therefore it can then be used in future physics analyses as a systematic uncertainty on track reconstruction.

Event selection criteria

Cut	**Value**
Triggers	hie or bha3d or bhabha
Good tracks*	<3
***Good track definition**	
|dz|	<2cm
dr	<0.5cm
Tag side selection	
tag track quality	Good track
E/P$_{e^{tag}}$	∈[0.95, 1.05]
clusterE$_{\gamma^{tag}}$	<clusterE$_{e^{tag}}$
clusterE$_{\gamma^{tag}}$	1st or 2nd highest in event
$\theta_{e^{tag}}$	∈[0.6435, 2.0944] rad
$\theta_{\gamma^{tag}}$	∈[0.6435, 2.0944] rad
Probe side selection	
clusterE$_{probe}$	>0.3 GeV
clusterE$_{probe}$	$>clusterE_{\gamma^{tag}}$
$\lvert\theta_{recoil} - \theta_{probe}\rvert$	<0.2 rad
clusterE^{tag} + clusterE^{probe}	>10.4 GeV
Matched track selection	
Probe track quality	Good track
$\lvert\theta_{recoil} - \theta_{e^{probe}}\rvert$	<0.2 rad
Probe charge	Opposite tag charge
clusterE$_{\gamma^{tag}}$	Most energetic in event
Matched gamma selection	
$\lvert\theta_{recoil} - \theta_{\gamma^{probe}}\rvert$	<0.2 rad

This study was conducted in 2021/2022 and used only run independent MC14 simulated samples (all other analysis discussed within this thesis used the newer MC15 run dependent samples). The study covers data produced in experiments 7 to 18 (outlined in

table 2.2) which totals 189.6fb^{-1} of integrated luminosity. Only the $e^+e^- \to e^+e^-$ simulated MC sample is used, other channels (such as $e^+e^- \to e^+e^-e^+e^-$, $e^+e^- \to e^+e^-\mu^+\mu^-$ and $e^+e^- \to \tau^+\tau^-$) where found to have little-to-no contribution due to the stringent event selection criteria.

The data and MC distributions of the transverse momentum and polar angle of the recoil vector ($\mathrm{pt}_{\mathrm{recoil}}$ and θ_{recoil}) are shown in fig. 1.2. Both of these show generally good agreement. The polar angle agreement is flat through the barrel region with slight data excesses at the extreme forward and backward directions. A mild data excess does grow towards lower recoil transverse momentum. The observed agreement is deemed to be sufficient for the purposes of this study.

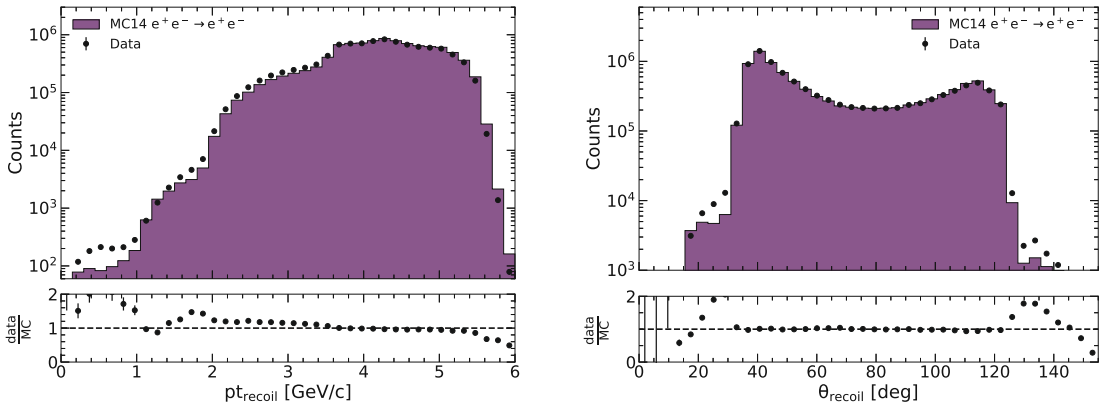

Figure 1.2: Data and MC distributions after all selection criteria is applied. The transverse recoil momentum is shown on the left, and the recoil polar angle on the right.

1.2 Results

The calculated cluster-track matching efficiency found in both data (black) and MC (purple) is shown in figure 1.3, where the results are given in binned distributions of the transverse momentum and polar angle of the recoil vector (left and right respectively). As one would hope, the efficiencies are in general very high with values remaining above 99.5% for the vast majority of events. The simulated MC events have consistently higher measured efficiency, particularly at high transverse recoil momentum. Efficiency drops rapidly below $\mathrm{pt}_{\mathrm{recoil}} \approx 2$ GeV/c, however this region is quite statistically limited.

Charge asymmetry

During further investigation of the observed difference between data and MC efficiency, the calculations were repeated with events separated by the measured charge of the tag electron. This would of course then dictate that the probe side electron has the opposite charge. Figure 1.4 shows the measured efficiency in events with a positive or negative

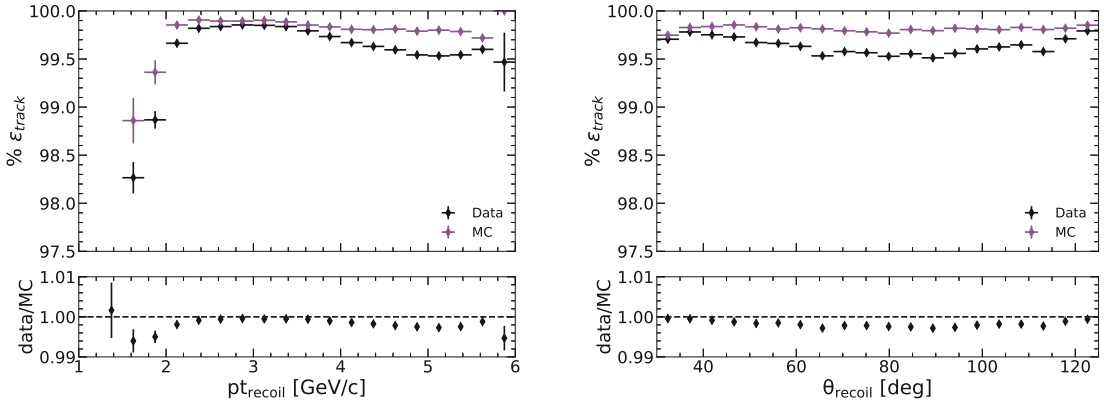

Figure 1.3: The cluster-track match efficiency in both data (black) and MC (purple). The left hand figure shows the efficiency calculated as a function of the transverse recoil momentum, the right hand figure shows it instead calculated across the recoil polar angle.

tag electron (green and blue respectively). The left hand figure shows the results from MC, and the right hand shows those from data. One can see a stark difference between these two, with no difference between opposite charge tagged events in MC contrasted by a clear angular dependence for the two in data. The lower sections of each plot show the ratio of positive to negative charge tagged events.

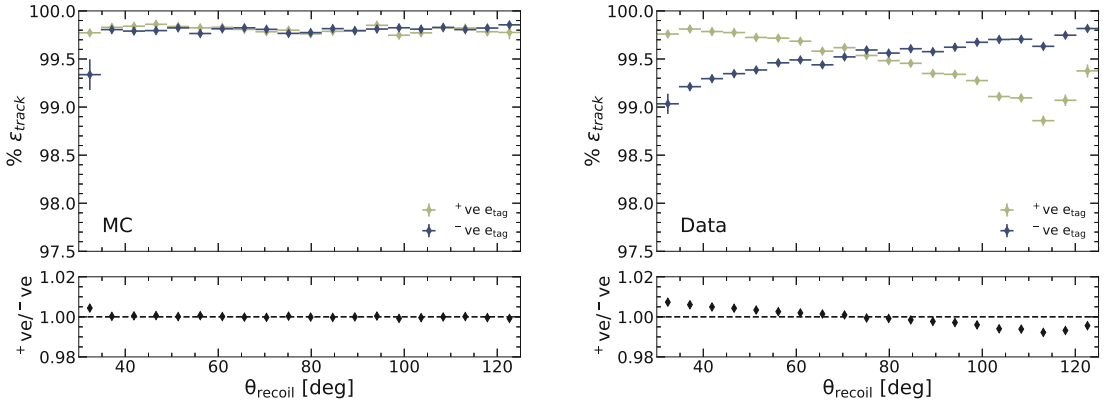

Figure 1.4: The calculated cluster-track match efficiency as a function of the recoil polar angle. The left hand shows the result for MC events and the right hand for data events, with each of these separated by the charge of the tag electron.

In data events with a negatively charged tag electron the efficiency rises from 99% in the forwards direction (low θ) to 99.8% in the backwards. Conversely those with positive charges decrease from 99.8% to a minimum of at around 110° approximately 98.8% before rising again to 99.5%. Figure 1.5 shows the result of this splitting procedure when the efficiency is instead calculated in bins of transverse recoil momentum. In this case again

there is good agreement between the opposite charge tagged events in MC, and this time similarly good agreement in data. There is a very slight deviation between the two in the region between approximately 3.5 and 4.5 GeV/c.

There is currently no full understanding of the origin of this discrepancy and the study is still ongoing, now with more recently processed MC and data with initial results appearing to show improvement.

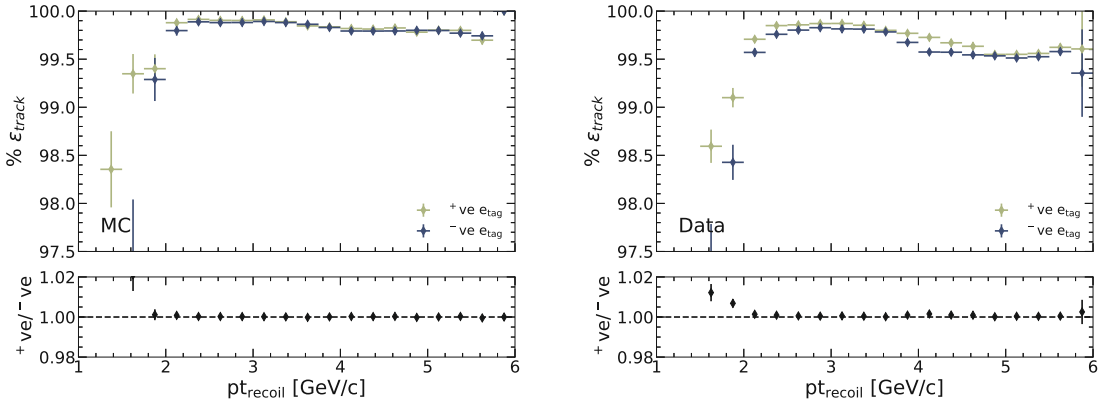

Figure 1.5: The calculated cluster-track match efficiency as a function of the transverse recoil momentum. The left hand shows the result for MC events and the right hand for data events, with each of these separated by the charge of the tag electron.

Data and MC distributions

Figure 1.6 shows the recoil polar angle distribution of data and MC events in which a track is found in the search cone as expected (left) or both a track and photon are found (right). The former of these shows little difference with respect to the full data-to-MC agreement shown in fig. 1.2. This is expected as of course the events constitute the vast bulk of those studied, in which all tracks and clusters are reconstructed properly. The latter (right hand) figure appears to show reasonable agreement between data and MC, however one may note that there is a growing data excess towards the forward detector region.

Figure 1.7 instead shows those events in which only a photon is found in the search cone (left) and no track or cluster is found (right). The agreement between data and MC in this latter case appears to be good, with a notable tendency for the events to reside in the far forward region of the detector barrel. One can see large disagreement in the data-to-MC ratio of events in which only a photon is reconstructed in the search cone. This appears most notable in the central region of the detector barrel. These events are, of course, those that make up the observed disagreement in cluster-track matching efficiency between data and MC.

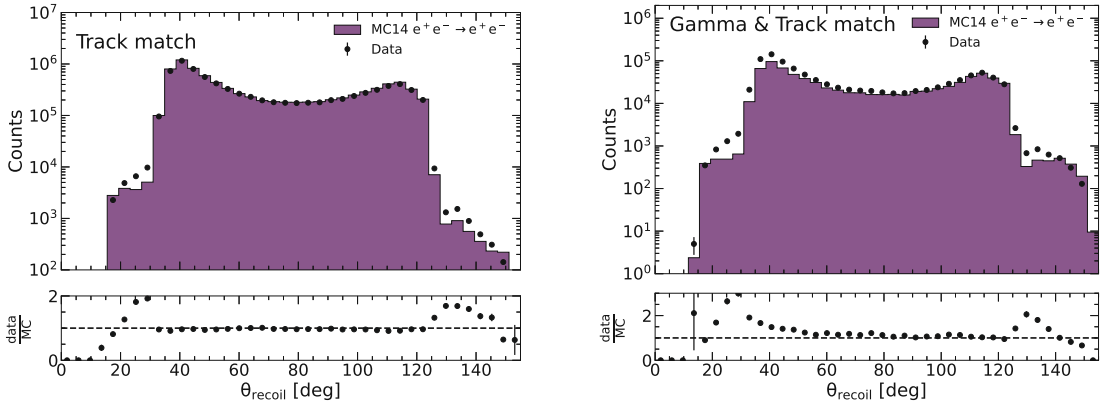

Figure 1.6: The recoil polar angle distribution for data and MC events. The left hand plot shows only events in which a track is found in the search cone (track match). The right hand plot shows instead only events in which both a track and cluster are found in the search cone (track & gamma match).

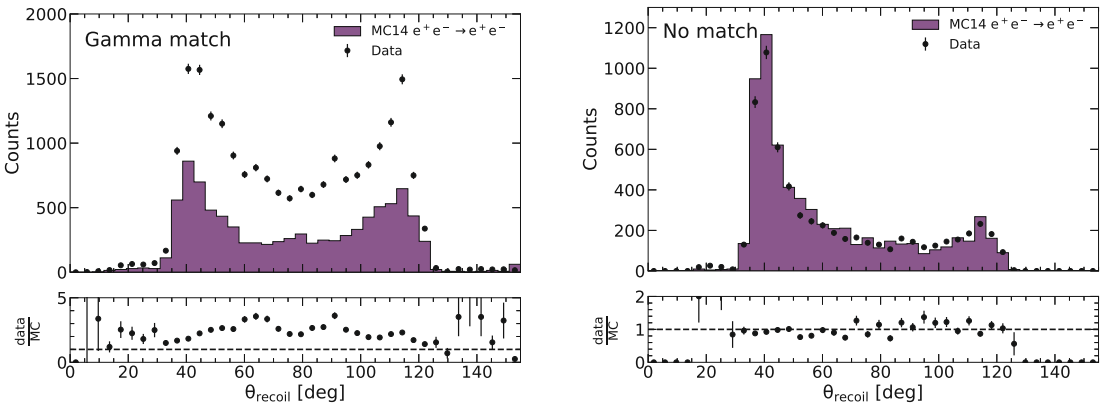

Figure 1.7: The recoil polar angle distribution for data and MC events. The left hand plot shows only events in which a photon is found in the search cone (gamma match). The right hand plot shows instead only events in which neither a track nor cluster is found in the search cone (no match).

Total cluster-track match uncertainty

Full treatment of systematic uncertainties associated with this measurement is not yet implemented, and indeed the observed charge asymmetry would appear to dominate over initial investigation of sources such as trigger usage or momentum corrections. Furthermore, this study itself is intended to merely define a track-associated uncertainty for further use in physics analyses, and not any correction factors. It is deemed sufficient to not investigate such sources further.

However, a rudimentary systematic uncertainty is defined by measurement of the

efficiency as a function of the tag electrons polar angle $\theta_{e^{tag}}$, which has good agreement between data and MC. The standard deviation of this measurement in both data and MC is then attributed as a systematic uncertainty to each. The data and MC distributions of this variable are shown on the left of fig. 1.8 and the resulting calculated efficiencies in data and MC are shown on the right.

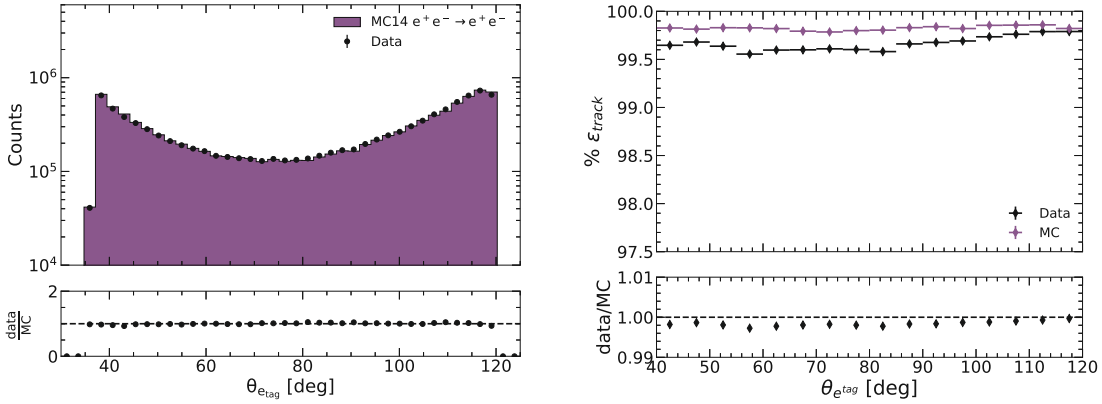

Figure 1.8: Left: the θ angle distribution of the tag electron in data and MC events. Right: the calculated cluster-track match efficiency as a function of the tag electron's θ angle .

Due to the fact that the scale of the cluster-track match efficiency is low ($<$1%), it is deemed unnecessary to quantify uncertainty values binned in some variable. Rather a single value is defined and recommended as a per-track uncertainty in any given analysis. The total efficiency values for both data and MC are calculated as:

$$\epsilon_{data} = 99.71 \pm 0.002 \pm 0.07\%$$

$$\epsilon_{MC} = (99.83 \pm 0.002 \pm 0.02)\%$$

Where the given uncertainties are statistical and systematic respectively. The total cluster-track match inefficiency ratio is thus:

$$\delta^* = 0.13 \pm 0.07\% \tag{1.3}$$

This therefore can be applied as the per-track reconstruction uncertainty. This was found to be in good agreement with a similar study that was conducted using tau decays to pions, where the a pion track could then be used in much the same way as the electron probe is used in this analysis. Again it should be noted that this result is now outdated as the study was conducted early in this thesis work.

2. Trigger Efficiency study

The work described herein was conducted solely by the author, and with advice provided by M. Bertemes, E. Graziani, M. Campajola and supervisor G. Inguglia.

2.1 Measuring trigger efficiency

As discussed in section 2.2, low multiplicity event searches at the Belle II experiment are greatly enhanced by the implementation of a varied set of L1 trigger lines. This section will investigate the efficiency of a set of such trigger lines that are of particular use in the search for an invisibly decaying Z′. These efficiencies are evaluated with respect to certain key variables relating to the kinematics of the physics process being utilised in the given sample. Both real data (collected between 2019 and 2022) and run-dependant MC samples are studied, the latter of these utilising the basf2 trigger simulation framework, *TSIM*, the accuracy of which can thereby be evaluated.

Trigger efficiency can be calculated either relative to some *orthogonal* trigger line, meaning that it has totally independent requirements, or in absolute terms. An example of orthogonality in triggers would be their use of independent sub-detectors: one may fire with some energy being deposited in the ECL, while the other would require only a track to be present in the CDC. These two can of course be fired simultaneously, for instance in a $e^+e^- \to \mu^+\mu^-\gamma$ event, the muons would create tracks in the CDC while the photon would create a cluster in the ECL. The trigger efficiency can thus be defined as,

$$\epsilon_{test} = \frac{N_{test} \cap N_{ref}}{N_{ref}} \tag{2.1}$$

Where N_{test} and N_{ref} are the sets of events that fire the trigger of interest and the orthogonal reference trigger respectively. In the case of calculating the absolute efficiency, N_{ref} is amended to just the total number of events utilised in the study.

Over the data-taking period, the selection of trigger lines was expanded. The trigger lines investigated here have the following requirements:

- *ffo*: Available for full data-taking period, superseded by *fyo* in exp. 12. Requires

two full tracks in CDC with a transverse plane opening angle, in the LAB frame, of $\Delta\phi_{LAB} > 90^{\circ}$

- *ff30*: Available from exp. 12, superseded by *fy30* in exp. 14. Requires two full tracks in CDC with a transverse plane opening angle, in the LAB frame, of $\Delta\phi_{LAB} > 30^{\circ}$

- *fyo*: Superseded *ffo* from exp. 12 of data taking period. Requires two full tracks in CDC with a transverse plane opening angle, in the LAB frame, of $\Delta\phi_{LAB} > 90^{\circ}$ with added requirement of at least one *Neuro 3D track*.

- *fy30*: Superseded *ff30* from exp. 14 of data taking period. Requires two full tracks in CDC with a transverse plane opening angle, in the LAB frame, of $\Delta\phi_{LAB} > 30^{\circ}$ with added requirement of at least one *Neuro 3D track*.

- *cdcklm1(2)*: Available from exp. 12 of data taking period. Requires 1(2) CDC track(s) with a matched KLM cluster.

- *stt*: Available from exp. 16 of data taking period. Requires just a single *Neuro 3D track* with a momentum $p > 0.7\text{GeV/c}$.

The *fyo* and *fy30* trigger lines are analogous to the *ffo* and *ff30* lines respectively, but with the added requirement that one of the two tracks must be a *Neuro 3D track*. These are characterised by the use of a neural network to interpret 2D reconstructed tracks in three dimensions [72].

The trigger efficiencies are measured as functions of four variables, each providing an avenue for the study of different aspects of the trigger performance:

- $\Delta\phi^{LAB}$: The previously mentioned transverse plane opening angle between the two muon tracks, measured in the LAB reference frame. This variable is used to study the activation threshold of those triggers lines that incorporate some requirement in the variable.

- $p_{t,min}^{LAB}$: The lowest transverse component of momentum of the two muons in the event, measured in the LAB frame. Trigger lines often have a transverse momentum threshold requirement and so this can be used to measure the momentum at which trigger efficiency plateaus. Taking the minimum value of those measured from the two tracks in the event allows study of a single variable that limits efficiency.

- $\theta_{\mu_{\pm}}^{LAB}$: The polar angle of the two muons (μ^{+}, μ^{-}) in the event, measured in the LAB frame. This is used to check for trigger efficiency effects imparted by the detector instrumentation coverage.

- $p_{t,\mu_{\pm}}^{LAB}$: The transverse component of momentum of the two muons in the event, measured in the LAB frame.

The *hie* trigger line is used as an orthogonal reference trigger line. This is available through the whole data-taking period and requires a total energy of 1 GeV is deposited in the ECL during the given event, in addition to the event not being flagged as bhabha scattering. The absolute efficiency of this trigger is calculated at a value of ∼99.7%, which remains reasonably stable across the different variables investigated herein.

2.2 Data and Monte Carlo samples

This study makes use of two distinct physics processes to study trigger efficiencies: $e^+e^- \rightarrow \mu^+\mu^-\gamma$ and $e^+e^- \rightarrow e^\pm\mu^\mp$. Events in these channels are expected to generate 2 CDC tracks, and at least one KLM cluster due to the presence of a muon in each. The latter of the two is only expected to reliably trigger *cdcklm1* due to the lack of a second muon. Both are expected to trigger the orthogonal reference trigger line, *hie*, by the respective presence of the photon or the electron. These processes are indeed also used for validation and analysis of systematic uncertainties in the search for an invisibly decaying Z′ boson and are discussed in greater detail later in sections 3.1 and 3.2 respectively.

The full data set taken before long shutdown 1, spanning the years 2019 to 2022, is studied. This comprises a total integrated luminosity of $362 fb^{-1}$ divided into numbered *experiments*, defining the experimental environment in which the collisions are conducted with perhaps a combination of hardware, software or beam parameter changes implemented between. These are subsequently divided into *runs* defining periods of beam runtime. The availability through the data-taking period of the trigger lines studied herein is shown in table 2.1 below.

	Experiment											
Trigger	7	8	10	12	14	16	17	18	20	22	24	26
ffo	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-
ff30	-	-	-	✓	✓	✓	-	-	-	-	-	-
fyo	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
fy30	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓
cdcklm1(2)	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
stt	-	-	-	-	-	✓	✓	✓	✓	✓	✓	✓

Table 2.1: Availability by experiment of the trigger lines studied herein.

	Integrated Luminosity [fb^{-1}]		
Experiment	On-resonance	Off-resonance	Other
7	0.510 ± 0.002	-	-
8	4.459 ± 0.003	0.813 ± 0.001	0.0038 ± 0.001
10	3.635 ± 0.001	-	-
12	54.388 ± 0.004	8.716 ± 0.002	-
14	16.385 ± 0.005	-	-
16	10.321 ± 0.004	-	-
17	10.714 ± 0.004	-	-
18	89.176 ± 0.010	8.424 ± 0.003	-
20	3.779 ± 0.002	-	-
21	-	-	19.661 ± 0.004
22	31.884 ± 0.006	-	0.040 ± 0.001
24	85.130 ± 0.010	-	-
25	-	24.377 ± 0.006	-
26	54.055 ± 0.008	-	-
Total	364.463 ± 0.02	42.329 ± 0.007	19.705 ± 0.004

Table 2.2: Integrated luminosity gathered during each experiment across the 2019 - 2022 period before LS1. On-resonance refers to running at the $\sqrt(s) = \Upsilon(4S)$ resonance (10.58 GeV/c). Off-resonance refers to running at 60 MeV below this, and final category contains any other centre-of-mass energy used.

2.3 Trigger Efficiency

2.3.1 CDC-based early trigger lines

The trigger efficiency found in experiments 7 to 10, while comprising only a small fraction of the total luminosity, is measured here separately due to the fact that only the *ffo* trigger line was available during this period. The *ff30* trigger was then activated in experiment 12, and the effect of then including this trigger in an OR logic is studied. Figure 2.1 shows the calculated efficiency as functions of the opening angle and minimum transverse momentum respectively.

The addition of the ff30 trigger line, and the associated reduction in opening angle requirement from 90°to 30°, can be see in the left hand side of fig. 2.1. In both cases the trigger efficiency is seen to reach, and plateau, at a constant level above the respective required opening angle. So as to measure events only comprising this plateau region, any further calculations of trigger efficiency for these experiments is done with only those events passing these opening angle requirements. Similarly, the right hand plot of fig. 2.1 shows trigger efficiency dropping to 0 below a $p^{LAB}_{t,min}$ of approximately 0.5 GeV. This is expected and is likely a result particles in this region failing to traverse the CDC chamber completely due to the magnetic field inducing strong curvature in the particles

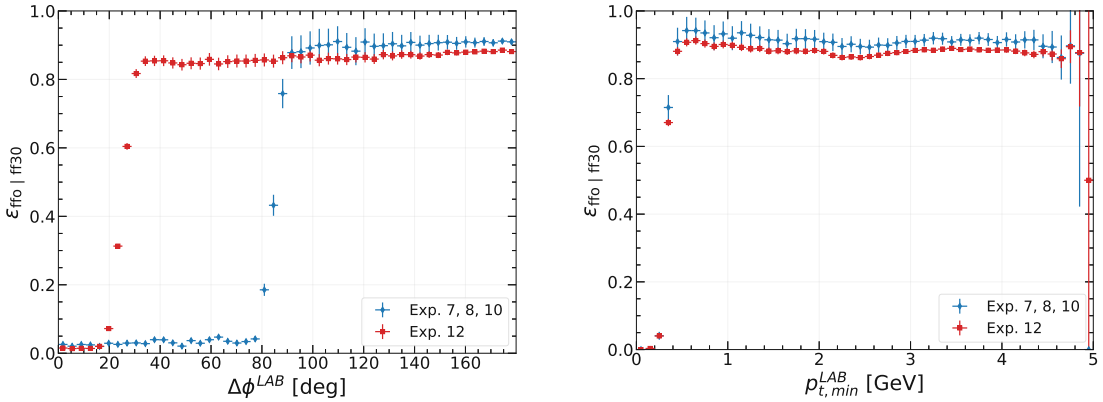

Figure 2.1: Measured efficiency for the OR logic combination of *ffo* or *ff30* trigger lines as a function of the opening angle (left) and the minimum muon transverse momentum (right).

trajectory. Again this can be accounted for with a requirement of $p^{LAB}_{t,min} > 0.5$ applied to events. It should be noted however that the number of such low momentum events is low, comprising approximately 3% of the total events studied.

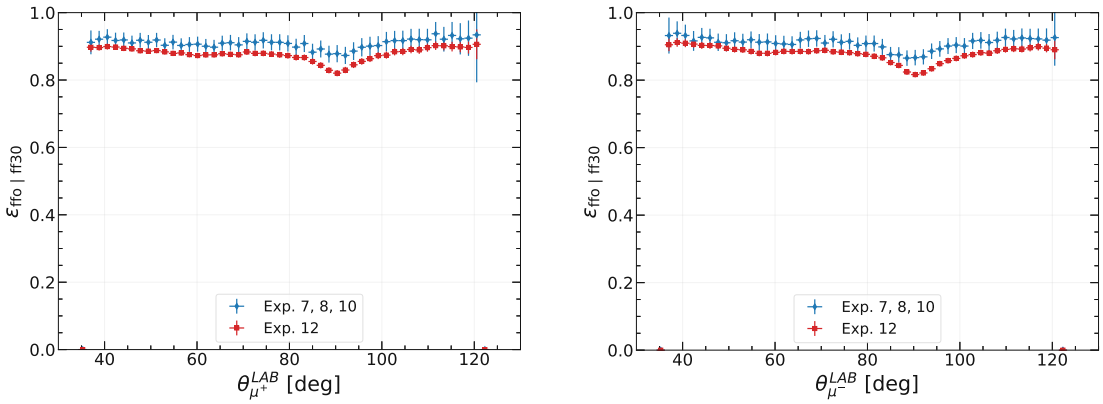

Figure 2.2: Measured efficiency for the OR logic combination of *ffo* or *ff30* trigger lines as a function of the polar angle of the two muons.

Figure 2.2 shows the calculated trigger efficiencies as functions of the respective muon polar angles, $\theta^{LAB}_{\mu_1}$ and $\theta^{LAB}_{\mu_2}$. The studied range covers the full CDC barrel region and shows an approximately flat trend with the exception of a dip in the region around 90°. This is a commonly observed feature and, while there is no current understanding of the origin, it is thought to be associated with cases in which the particle tracks lie perpendicular to CDC wires. This can maximise the charge accumulated on the wire, which in turn can reduce the dead-time of the given wire and thereby slightly reduce the trigger efficiency.

2.3.2 Full trigger selection

Above this first set of experiments the available trigger selection is improved with the shift to neural network based CDC triggers (*fyo* and *fy30*), the addition of KLM based triggers, and finally a single-track trigger. The overall trigger efficiency is studied across the full data-taking period using all triggers, where available, shown in table 2.1 in an OR logic, and with the previously studied activation requirements considered in the early experiments. The CDCKLM and single-track trigger have no such angular requirements and so these are neglected in experiments 12 and above.

Figure 2.3 shows the trigger efficiency calculated for all individual runs across the twelve experiments that make up the data-taking period. This shows good consistency of the trigger performance across the experiments. Efficiency in the early experiments is at the lower level expected due to the smaller selection of trigger lines. One can see the period early in exp. 12 during which further trigger lines were added. Importantly, this serves as a check that trigger efficiency is not appreciably impacted by changes in the experimental environment, for instance in the later experiments where higher instantaneous luminosity leads to a large increase in beam-backgrounds.

The efficiency found in MC simulated events is also calculated and compared with those results from data. Correction factors and systematic uncertainties can then derived from this. Figures in 2.4 show the calculated total trigger efficiency in both data (black) and run-dependant MC (red) in the previously discussed opening angle and minimum transverse momentum respectively. Across the full range in each variable there is a disagreement showing higher trigger efficiency in the MC simulated samples.

There appears 3 clear regions of distinct behaviour, mirrored in both data and MC, across the opening angle distribution (fig. 2.4): below 25° where data and MC disagreement is maximal and the efficiency appears to fluctuate mildly, between 25° and 90° where agreement between data and MC is closest and flat, and finally above 90° where the MC efficiency rises to 100% while data remains below this with a slight trend upward from 97.0% to around 98.5%. The minimum transverse momentum shows more stable distributions (right of fig. 2.4). Below 1 GeV data and MC agreement differs by up to 9% and rapidly drops off towards lower values. As discussed previously, this is not unexpected and possibly a result of low momentum tracks curving heavily in the CDC chamber. Above 1 GeV both data and MC remain flat at approximately 98% in data and between 99% and 100% in MC

The total trigger efficiency for each experiment is calculated for both MC and data, the results are shown in fig. 2.5. One can see a fluctuation and sometimes quite large overall disagreement between the two. It does largely improve in later experiments but remains non-negligible.

These differences between modelling of trigger lines in MC, and the real-life effect in data lead one to decide whether to still utilise the simulated trigger system, TSIM, with some additional calculated correction factors, or, merely derive and apply weights to the MC such as to represent the trigger's effect. Neither is necessarily better however the latter option provides a more simplistic solution to trigger modelling and is thus adopted for use in the full study detailed in this thesis. Correction factors are calculated in bins

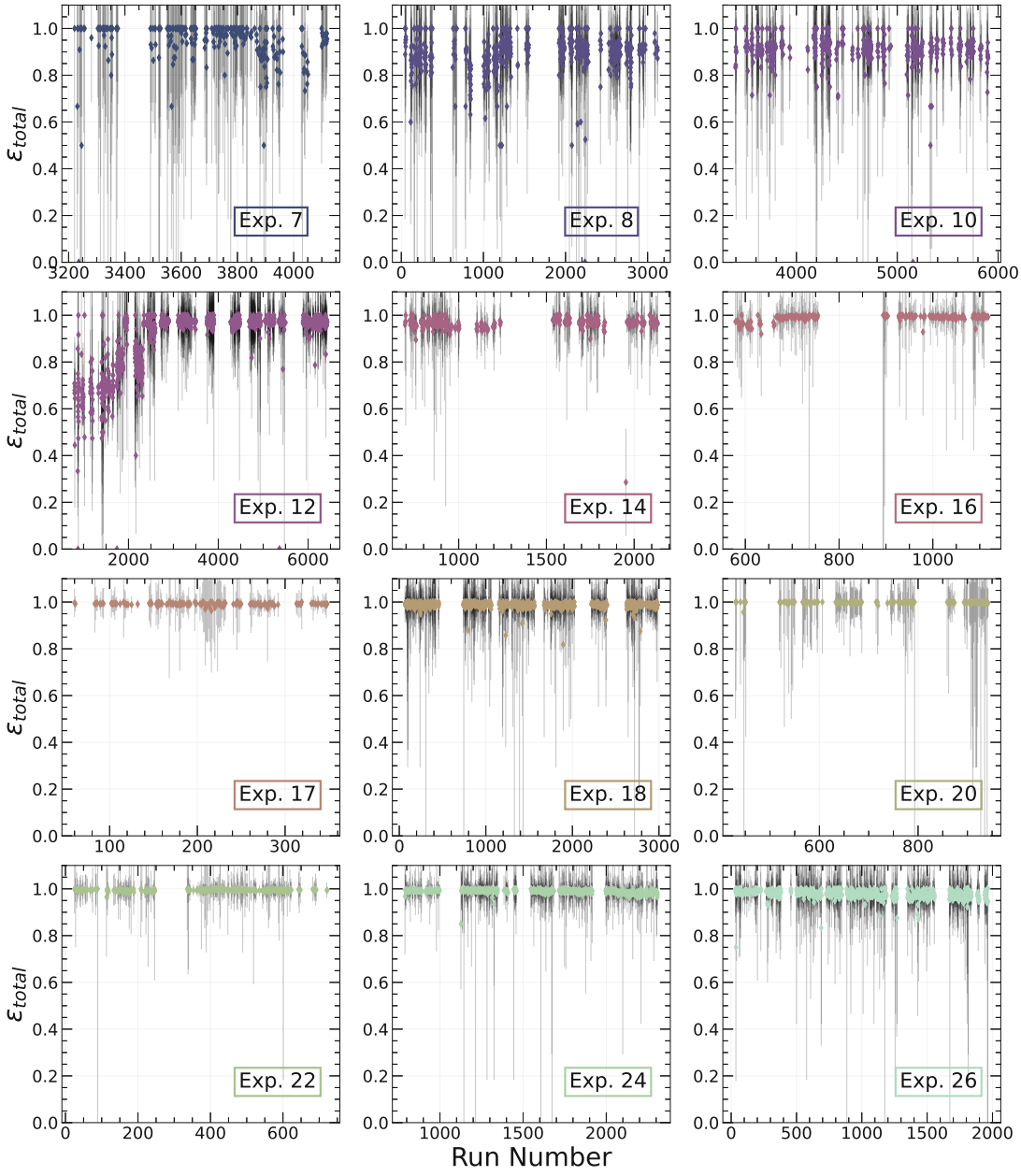

Figure 2.3: Trigger efficiency calculated for each run during each of the twelve experiments spanning the full data-taking period.

of a given variable which must be chosen.

The data trigger efficiency is studied in the two-dimensional distribution of muon polar angle ($\theta^{LAB}_{\mu^\pm}$) against transverse momentum ($p^{LAB}_{t,\mu^\pm}$). The results for the two muons in the event are shown below in figure 2.6. These two-dimensional distributions serve

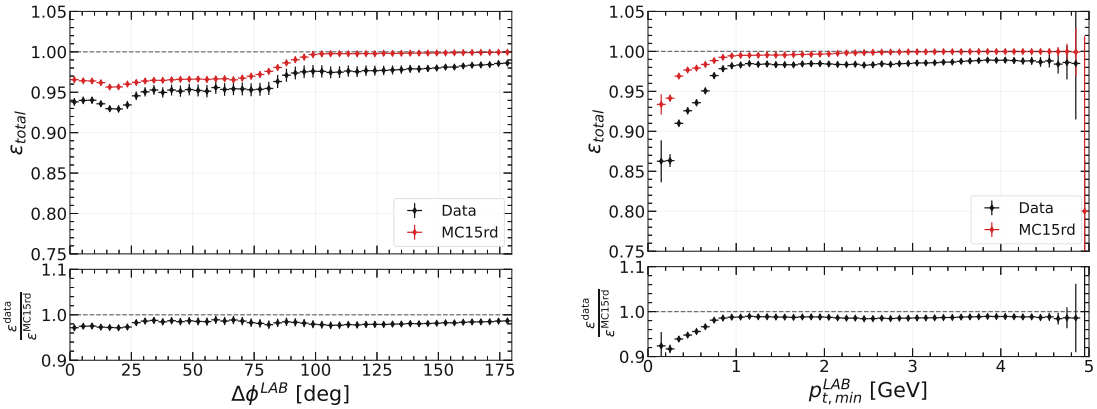

Figure 2.4: Measured efficiency for the OR logic combination of all trigger lines as a function of the opening angle (left) and the minimum muon transverse momentum (right). Measurements in both data and MC are shown (black and red respectively), and the ratio of data to MC is shown in the lower section.

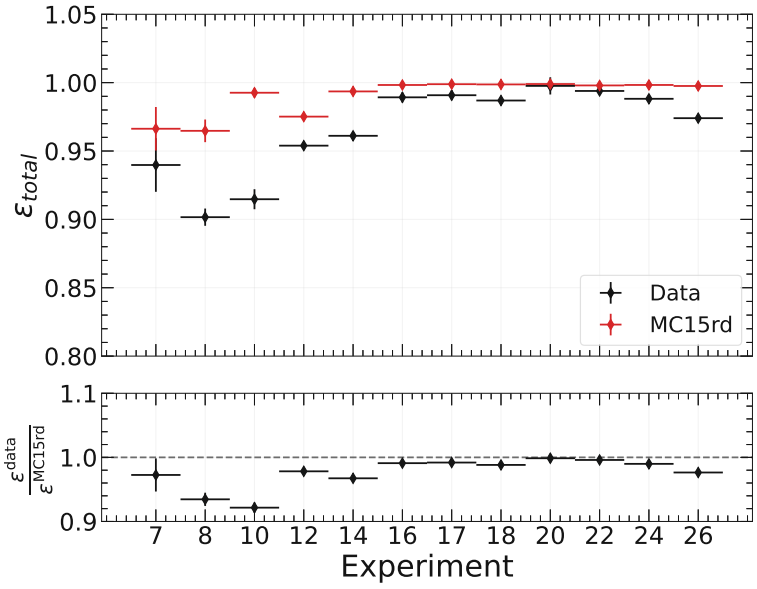

Figure 2.5: Total calculated trigger efficiency by experiment, showing results of MC (red) and data (black).

as a check for regions in the detector that may be particularly problematic and thereby require specific attention when deriving correction values.

The plots show again, as observed previously in figure 2.2, a slight drop in efficiency around the 90° region, likely due to the perpendicular intersection of tracks and CDC wires. Trigger efficiency does drop for those muons with trajectory more in the *forward*

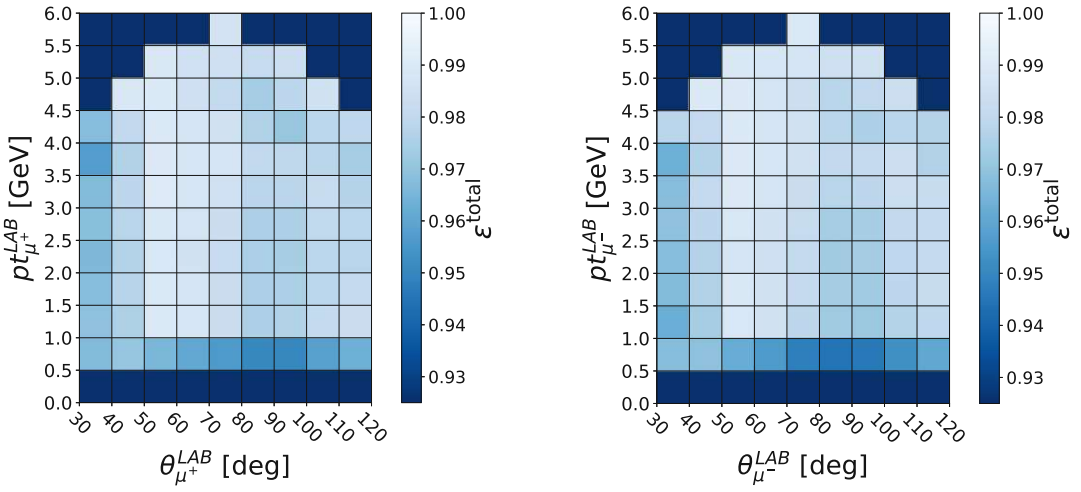

Figure 2.6: Trigger efficiency calculated in data in 2D bins of $\theta^{LAB}_{\mu^\pm}$ against transverse momentum $pt^{LAB}_{\mu^\pm}$ for the two muons in the event.

direction of the detector, however statistics here are low and so the effect is not significant. Those with transverse momenta below approximately 1 GeV also suffer a similar decrease as is already observed in the minimum transverse momentum variable shown in fig. 2.4.

Trigger efficiency corrections are chosen to be calculated in bins of the minimum transverse momentum variable, due to both the previously discussed mis-modelling that is captured therein at low values, and the relatively stable statistics found across the range. In addition, it provides a single variable describing events that is related to the function of not only those triggers relying on the detection of 2 tracks, but also those requiring just a single track. The calculated values are given with systematic and statistical uncertainties. Due to the nature of these corrections being calculated by way of a ratio, systematic effects are not considered to be of large significance as they will be to some extent cancelled out. Nevertheless, a systematic uncertainty is attributed to each bin by taking the largest variation ($\epsilon_{max} - \epsilon_{min}$) of calculated efficiency as a function of muon polar angle ($\theta^{LAB}_{\mu^\pm}$) for those events residing within the given $p^{LAB}_{t,\mu^\pm}$ bin. The correction values with statistical and systematic uncertainties are given in table 2.3 below.

Finally, these corrections can be validated by checking the independent $e^+e^- \rightarrow e^\pm\mu^\mp$ sample. Figure 2.7 shows the polar angle of the recoil vector calculated from the $e^\pm\mu^\mp$ system, in both data and MC. The left and right plots show the distributions with and without the trigger correction factors applied respectively. This variable is minimally correlated to the minimum transverse momentum used to calculate correction factors. Furthermore it is used in the 2D maximum likelihood fit of the Z′ to invisible search outlined later in section 3, and thus is an important variable to check.

$p_{t,min}^{LAB}$ [GeV]	Correction (+ stat. + sys.)
0.0 - 0.5	0.697 ± 0.003 ± 0.063
0.5 - 1.0	0.963 ± 0.002 ± 0.038
1.0 - 1.5	0.984 ± 0.002 ± 0.017
1.5 - 2.0	0.984 ± 0.002 ± 0.015
2.0 - 2.5	0.984 ± 0.002 ± 0.017
2.5 - 3.0	0.984 ± 0.002 ± 0.017
3.0 - 3.5	0.986 ± 0.001 ± 0.016
3.5 - 4.0	0.988 ± 0.001 ± 0.013
4.0 - 4.5	0.988 ± 0.002 ± 0.013
4.5 - 5.0	0.855 ± 0.006 ± 0.024

Table 2.3: Trigger correction factors, with statistical and systematic uncertainties quoted.

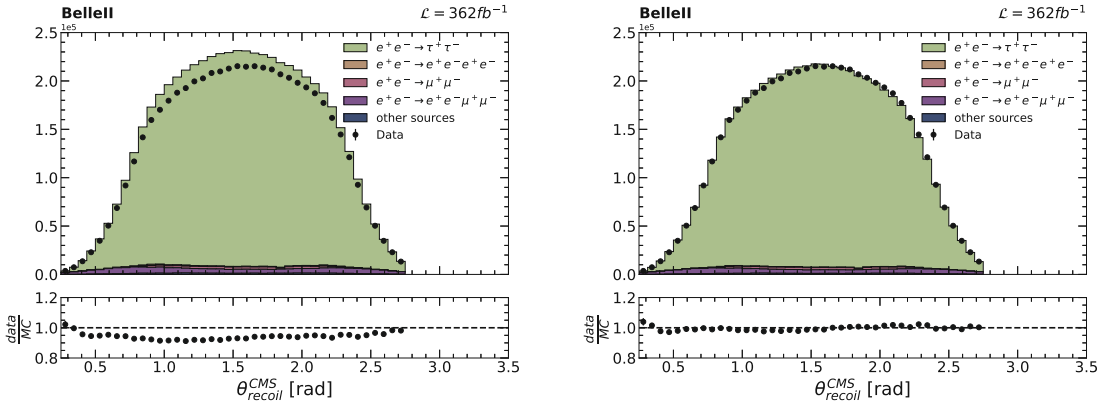

Figure 2.7: The polar angle of the recoil vector in reconstructed $e^+e^- \to e^\pm\mu^\mp$ events, both from data and MC, with the full trigger selection applied. On the right the trigger correction factors (2.3) are applied to the MC, the left hand instead has no trigger correction applied.

3. The Punzi-Net

The work described herein was conducted in collaboration with fellow PhD student P. Feichtinger and supervisor G. Inguglia, and is published in the European Physics Journal [73].

3.1 Figures of Merit

When conducting the search for a new particle, a key step is the reduction of SM interactions that may imitate and thus obscure the decay process of interest. These are referred to as *background* processes, and the process of interest is referred to as the *signal*. Often one can reduce the background by applying cuts to some given variable in the data. Perhaps a problematic background process often contains a high energy photon while the signal does not. Applying a cut to detected photon energy could thereby remove this without impacting the signal.The selection and implementation of such cuts used is discussed in section 2.2.

In the search for some physical process, one commonly carries out what is referred to as a *counting experiment*, whereby the number (n) of signal (S) and background (B) events are counted. One can then test the signal-plus-background and background only hypotheses, H_{S+B} and H_B respectively, and claim a discovery or set limits on parameters associated with the new particle accordingly. These hypotheses are describe by the Poisson distributions,

$$p(n|H_B) = \frac{B^n e^{-B}}{n!} \tag{3.1}$$

and

$$p(n|H_{S+B}) = \frac{(S+B)^n e^{-(S+B)}}{n!} \tag{3.2}$$

By employing *Monte Carlo* (MC) simulated data one can define some set of selection criteria, i.e. cuts to certain variables, that remove the maximal amount of background events and thereby allow for high confidence in rejecting one (or in some cases perhaps both) of the hypotheses. Such selection invariably comes at a cost, as it more often than

not removes not just background, but signal events as well. Thus it is important that one can optimise the criteria so as to balance this loss of signal with rejection of background. To this end, one can employ a *figure of merit* (FOM), that provides a single measure of these two aspects. The values $S(t)$ and $B(t)$ are defined as the numbers of signal and background events given some applied selection criteria t. Two FOMs, commonly employed in particle physics studies, are,

$$FOM = \frac{S(t)}{\sqrt{(B(t))}} \tag{3.3}$$

and

$$FOM = \frac{S(t)}{\sqrt{(S(t) + B(t))}} \tag{3.4}$$

In the case of a search for some new BSM particle, however, one encounters a problem with these: the number of signal events, $S(t)$, is dependant on some unknown free parameter(s), such as mass. One can therefore define $S_m(t)$ as the number of signal events with some unknown free parameter, m, given some selection criteria, t. To overcome this problem, a modified FOM was proposed in the early 2000s which has come to be known as the *Punzi FOM*, owing to its originator Giovanni Punzi. The derivation of this is fully outlined in [74], but shall be briefly stepped through here to contextualise the subsequent discussion of the Punzi-loss function.

A test of the background only hypothesis is characterised by the set of values of some observable, X, that would reject it. One can then quote a *significance level*, α, that is defined as the probability of X falling within this critical region of values when H_B is actually true (type I error). Conversely, one could fail to reject H_B when H_{S+B} is in fact true (type II error), as defined by the probability β. The probability of correctly claiming a discovery of a new particle with some free parameters, m, can be expressed with the power function: $1 - \beta(m)$. Thus the sensitivity to making a discovery with the experiment at a chosen *confidence level* (CL) is define as,

$$1 - \beta_\alpha(m) > CL \tag{3.5}$$

The critical region for a counting experiment is merely the number of events, n_{min}, one would need to observe to reject the background only hypothesis. This is, of course, dependant on the chosen significance level α. More specifically in the case of signal events, one can then define the requirement that $S_m(t)$ must be greater than some minimum number, S_{min}. As shown in [74], this quantity can then be expressed as:

$$S_m(t) = \frac{b^2}{2} + a\sqrt{B(t)} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)} \tag{3.6}$$

Where a and b are the number of sigmas corresponding to one sided Gaussian tests at significance α and β respectively. The relationship between $S_m(t)$ and the luminosity, $\mathcal{L}$, cross-section of the particle given the free parameter m, σ_m, and the efficiency of the selection criteria, $\epsilon(t)$, can be defined:

$$S_m(t) = \epsilon(t)\sigma_m\mathcal{L} \tag{3.7}$$

Combining equations 3.6 and 3.7 one can define the minimum detectable cross-section given the chosen selection criteria:

$$\sigma_m = \frac{\frac{b^2}{2} + a\sqrt{B(t)} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}}{\epsilon(t)\mathcal{L}} \tag{3.8}$$

and finally from this define the Punzi FOM,

$$FOM_{punzi} = \frac{\epsilon(t)}{b^2 + 2a\sqrt{B(t)} + b\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}} \tag{3.9}$$

Which reaches its maximum value when selection criteria, t, provides the maximum sensitivity to the new particle.

3.2 Neural networks

In recent years machine learning tools have become the gold standard in the task of event classification. Neural networks (NNs), being one such tool, can come in a variety of *architectures* that lend themselves to specific applications. This investigation makes use of a standard fully connected feed-forward NN.

A neural network is comprised of a interconnected layers of *nodes*, with each connected to all those in the layers behind and in front of it. Each node is described by mathematical function that takes input from all those nodes in the preceding layer, adds some bias factor and applies a specific *activation* function, σ,

$$a^l_j = \sigma(\sum_k w^l_{jk} a^{l-1}_k + b^l_j) \tag{3.10}$$

where a^l_j denotes the output of the j^{th} node in the current layer (l), b^l_j denotes its bias and w^l_{jk} denotes the weight of its connection to the k^{th} node of the previous layer (a^{l-1}_k).

The activation function can take many forms, with traits of each that are desirable for specific applications. Typically, one uses the likes of a rectified linear activation (ReLU) for nodes within the *hidden* layers (those that do not directly connect to the inputs or output), while output nodes might have some function with a more well defined step such as a sigmoid function. This is especially common for applications such as event classification in particle physics, where one seeks to clearly define if an event is signal (with an output of 1) or background (with an output of 0).

Thus a neural network is a tool that can map a set of input variables to some such desired output. Before this can occur, however, the various *hyperparameters* of the NN (such as the previously mentioned weights and biases) must be tuned and optimised. In this study this is achieved by carrying out *supervised training*, whereby training data, x, consisting of perhaps a set of variables derived from reconstructed particles in a collision, is passed to the network along with the correct labels, y, (i.e. signal or background). The output from the network when fed this data, $\hat{y}$, can then be compared to these correct variables and a subsequent scoring of the NN's performance can be conducted. This is done by way of a *loss function*, which numerically quantifies the NN's inaccuracy. One such commonly used loss function is the *binary cross entropy*, defined as,

$$L = -yln\hat{y} - (1-y)ln(1-\hat{y}) \tag{3.11}$$

Note that this requires $y \in \{0,1\}$ and $\hat{y} \in [0,1]$, however this is commonplace and in-fact inherent when using sigmoid activation on the output node. With a loss function chosen, one must then seek for the optimal set of hyperparmeters, which can done by way of some optimisation algorithm such as *gradient descent* that will allow them to be adjusted iteratively in the direction opposite that of the loss function's gradient:

$$w_{n+1} = w_n - \eta\frac{\delta L}{\delta w_n} \tag{3.12}$$

and

$$b_{n+1} = b_n - \eta\frac{\delta L}{\delta b_n} \tag{3.13}$$

where w_{n+1} and b_{n+1} are some weight and bias values at the $n^{th}+1$ iteration of training, η is the *learning rate* and L is the loss value. The two derivatives noted in these equations are calculated by way of the *backpropagation* algorithm, which steps backwards through the layers of the neural network iteratively calculating the derivative by using the chain rule.

3.3 Implementation and results

With the Punzi figure of merit and the basic workings of neural networks understood, discussion can now turn to how one may combine the two in order to produce an event classification NN that is trained for the precise task of new particle searches, where one seeks to achieve the highest possible sensitivity to the particle in question. To this end, a new loss function that is based on the Punzi FOM was developed. This is dubbed the *Punzi-loss*, and the NN that is subsequently trained by way of it, the *Punzi-Net*.

As mentioned previously, the task of training a neural network is defined by the minimisation of the loss function. Thus one must start from the minimum detectable cross-section (eq. 3.8), and not the Punzi FOM itself (eq. 3.9) as this would conversely be maximised at optimal selection criteria. One cannot use eq. 3.8 in its current form though, as both the number of background events, $B(t)$, and the signal efficiency, $\epsilon(t)$, are discrete functions of the selection criteria. That is to say one cannot half count an event, $B(t)$ must be some integer value.

This presents a problem when it comes to training the NN: a gradient based optimisation algorithm requires a differentiable loss function. One can remedy this by replacing these terms with sums over the NN output value for all individual events, with weightings from their respective labels, y_i, such that only true signal events are considered when calculating $\epsilon(t)$, and only true background events are considered when calculating $B(t)$. The former is defined with a label of 1 and the latter with a label of 0. The substitute terms, now dependent on the output, $\hat{y}$, given the current weights and biases of the network, w and b, can thus be defined as:

$$\epsilon(t) \to \epsilon(w,b) = \sum_{i \in x} \frac{y_i \cdot \hat{y}_i(w,b) \cdot s_{\text{sig}}}{N_{\text{gen}}} \tag{3.14}$$

and

$$B(t) \to B(w,b) = \sum_{i \in x} (1 - y_i) \cdot \hat{y}_i(w,b) \cdot s_{\text{bgd}} \tag{3.15}$$

Where the summation runs over all training events, i, in the batch, x. The signal efficiency is calculated with respect to the initial number of MC signal events generated, N_{gen}. Both calculations include scaling factors, s_{sig} and s_{bgd} , for signal and backgrounds respectively. These are dependent on the given use case and might consider scaling of samples to a desired luminosity, sample fraction in the case of batched training or detector effects such as trigger efficiencies (see section 2). In the optimal selection case where all signal and background events are correctly identified, the Punzi-Net would thus provide output values approaching 1 or 0 for all signal or background events respectively and thereby would recover the original definitions.

Finally the *Punzi-loss* function can be defined by the mean value of this continuous Punzi sensitivity,

$$\sigma_m(w,b) = \frac{\frac{b^2}{2} + a\sqrt{B(w,b)} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(w,b)} + 4B(w,b)}}{\epsilon(w,b)\mathcal{L}} \tag{3.16}$$

for all signal hypotheses, N, across a range of values for the free parameter, m, that are used during training, giving:

$$C_{\text{Punzi}} = \frac{1}{N_m}\sum_{i\in m}\sigma_m(w,b) \tag{3.17}$$

Thus far, the implementation of the Punzi-Net has been discussed in generic study-independent terms, however, the specific application conducted herein is that of the search for an invisibly decaying Z′ boson. In this case, the free parameter across that defines the signal hypotheses is the mass of such a particle, $\mathrm{M}_{Z'}$. The full details of this study are outlined in part III of this thesis and shall not be explored further here, rather just those points pertinent to the Punzi-Net training and performance alone.

The network is implemented using the `PyTorch` library [75]. The selection of input variables is discussed in section 2.3. The network uses 8 input variables connected via 3 hidden layers of 16, 24 and 8 nodes respectively to a single output node. It was found that increasing numbers of layers or nodes-per-layer further than this provided little improvement and substantial increases in training time. Nodes in the hidden layers use the tanh activation function and the architecture was chosen by way of a simple grid-search method. To mitigate possible problems introduced by scale differences in the input variables, all were scaled such that their distributions take a mean value of 0 and variance of 1.

It was found necessary to first pre-train the network with the commonly used *Binary Cross-entropy* loss function. This introduces a degree of separation power to the network, without which any training using the Punzi-loss function regularly failed. This pre-training was conducted for 70 Epochs with the LR being halved after 10 consecutive epochs showing no decrease in loss. Training data were divided into batches of 2048 events. All signal events were assigned a weight such that the weighted sum of signal events is equal to that of background events.

Following the pre-training, the network was then trained on the Punzi-loss function with a=3 and b=1.28. A learning rate (LR) of 0.05 was chosen, which was found to be small enough to not introduce instability to the minimisation of the loss function while still allowing for rapid training. Training data were again divided into batches, however, in the case of training with the Punzi-loss function, where one seeks to optimise for sensitivity at some specific luminosity, this introduces the need to assign weights, s_{bgd} and s_{sig}, to both the number of background events and the signal efficiency, previously mentioned in equations 3.15 and 3.14 respectively.

That is to say if one wishes to train the network for use on 100fb^{-1} of collected data, and does so using three separate background samples with integrated luminosities of 500, 1000 and 2000fb^{-1} of simulated events, divided into batches each equaling 10% of the full training data. A single batch would then contain the equivalent of just 50, 100 and 200fb^{-1} of each background process, and thus each of these categories would require assigned weights of $s_{bgd} = 2$, 1 and 0.5 respectively in order for the loss calculation for each batch to be accurately approximating the Punzi sensitivity. Similarly, one must take into consideration the effect batching has on the calculation of signal efficiency. A batch size of 10% would artificially reduce the signal efficiency calculation (eq. 3.14) by a factor of 10 and must be accounted for with a weight of $s_{sig} = 10$.

Hyper-parameter	BCE	Punzi-loss
Learning rate	0.1	0.05
Batch-size	2048	100000
Epochs	70	600
LR reduction	0.5	-
Patience (Epochs)	10	-

Table 3.1: The hyper-parameters used in both the initial training using the BCE loss function and subsequent training with the Punzi-loss

Figure 3.1 shows the evolution of Punzi-loss as training progresses for the first 100 epochs, with the value calculated using both the training data set (dashed) and also an independent testing data set (solid). The figure shows the effect of varying the batch size used, with training using batch sizes of 1×10^4, 2.5×10^4 and 1×10^5 shown. One may note that these are considerably larger batch sizes than one commonly might see. This is due to the inherent statistical nature of the objective, where large numbers of events are required to effectively calculate estimate the Punzi sensitivity and thereby the Punzi-loss itself.

Indeed it was found that even the lower batch sizes shown here would introduce a degree of instability to the fitting procedure, as can be seen quite clearly in the figure. Choosing anything much lower than these was found to regularly cause complete failure of the training, with the loss often jumping to a local minimum at some higher value and subsequently remaining there until the end of the training. This is thought to likely be a result of such small batch sizes containing too few signal events for the individual mass hypotheses across which the loss function is averaged, thereby causing eq. 3.16 to become unstable.

A batch size of 1×10^5 was used for the final training of the Punzi-Net. The hyper-parameters used in both the initial training using the BCE loss function and subsequent training with the Punzi-loss are defined in table 3.1. During training, the loss calculation for each individual mass hypothesis is conducted using only those background events that reside within $\pm3\sigma$ of the signal peak in M^2_{recoil}. Furthermore only every other (odd) generated mass hypothesis was used for training, with the rest then kept for validation.

Figure 3.2 shows a comparison of the achievable Punzi figure-of-merit (left) and sensi-

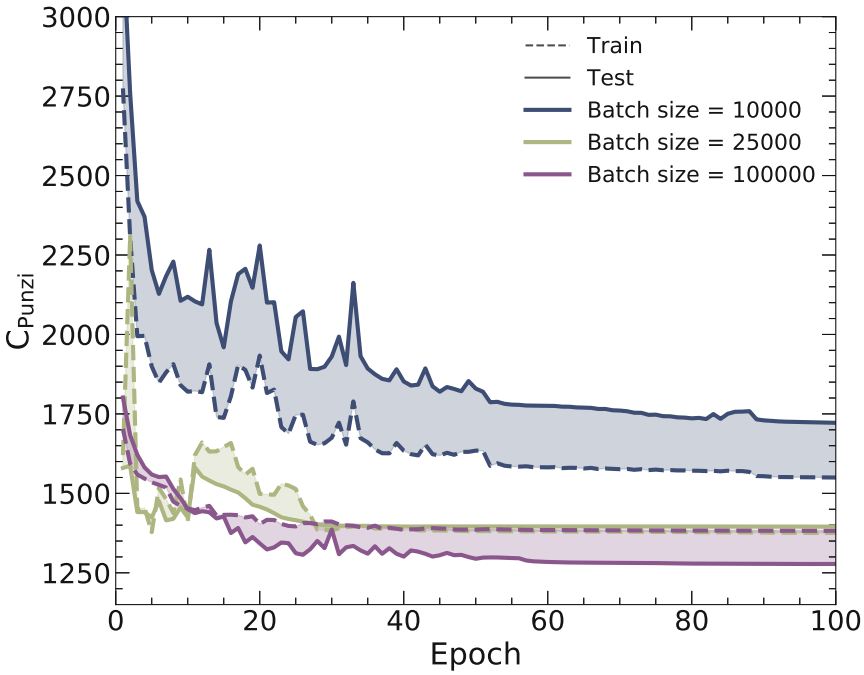

Figure 3.1: Distributions of the Punzi-loss values calculated at each epoch of training with independent train and test samples. These are shown in the case of 3 different batch sizes being used.

tivity (right) when using the BCE-trained network and fully trained Punzi-net using the Punzi-loss function. Both figures contain solid lines depicting the best achievable value when an optimal cut is applied to the output at the given Z′ mass hypothesis. The dashed lines instead show the single cut to the output that provides a best average result across all mass hypotheses. Immediately clear is the advantage provided by the Punzi-Net in the region between approximately 3 and 6 GeV/c^2, with up to a ∼75% improvement at 4 GeV/c^2. Furthermore, one can achieve almost the optimal performance when applying just a single cut to the output of the Punzi-Net, demonstrated by the close trend in both the dashed and solid lines.

Conversely, while the same may be true of the BCE-trained network for the region between approximately 1 and 6 GeV/c^2, it does not hold for the extremes of the mass spectrum. The single best cut provides a notable reduction of ∼50% below 1 GeV/c^2, while performance completely drops away in the high mass with no sensitivity to a Z′ signal in the region above ∼7 GeV/c^2. This means if one were to employ the BCE-trained network and hope to achieve the best possible performance, either a varied cut value to the output would have to be utilised across the full mass range, or multiple separately trained networks would be required for different mass ranges. Both of these options would add a further layer of complexity to the analysis, thereby complicating and likely worsening systematic uncertainties associated with the event selection process.

It is this aspect of the Punzi-loss (and trained Punzi-Net) that is perhaps the most powerful. Not only does it provide better performance than a more traditional implementation with the BCE loss function, but it is able to do so with just a single cut value applied to a single classifier for all mass hypotheses in the given analysis.

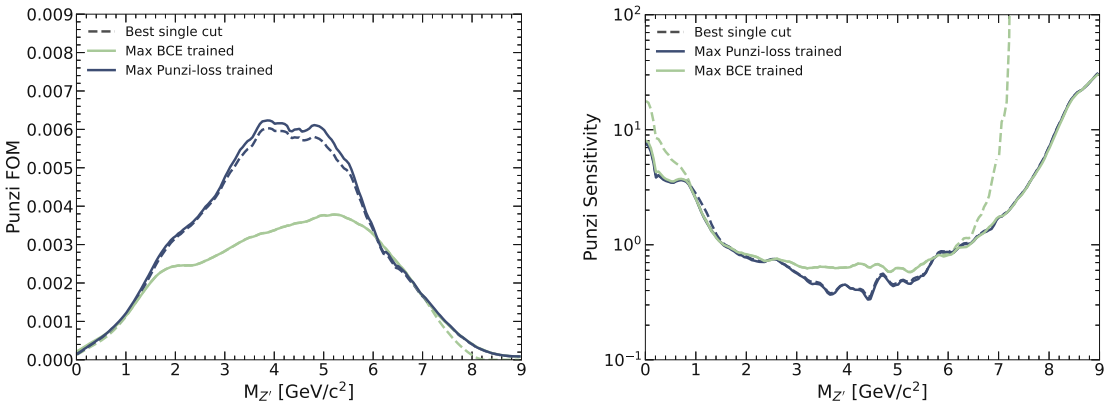

Figure 3.2: Left: The Punzi figure-of-merit across all Z′ mass hypotheses for both the Punzi-Net (blue) and BCE-trained network (green). The solid line shows the maximum achievable value with an optimal cut applied at each mass hypothesis, while the dashed line shows the maximum average across mass hypotheses achievable with a single cut. Right: The same definitions, now showing the achieved Punzi sensitivity for each mass hypothesis.

Part III

Search for an invisibly decaying Z′ boson at the Belle II experiment

1. Analysis Overview

This chapter investigates the full experimental analysis process in the search for an invisibly decaying Z′ boson using data collected with the Belle II experiment. Furthermore, the additional requirements concerned with expanding the study to include the search for a muonphilic scalar, S, are discussed. Due to the strong similarities between the two it is assumed that the systematic uncertainty studies and statistical analysis method conducted for the Z′ boson search can be applied also in the case of the muonphilic scalar study.

In this chapter the data studied (both real and simulated) is introduced, before then walking through the event selection criteria. The expected composition of the remaining background events is analysed in detail, with the main contributing processes investigated individually, and the final simulated signal distributions and efficiency is presented. *Control channels* are then employed for the validation of selection criteria, and derivation of any subsequent systematic uncertainties associated with various aspects of the analysis. These control channels are comprised of specific physics processes, similar in topology to the nominal $\mu^+\mu^-(inv)$ cases, that allow for comparison of real data and simulated Monte Carlo samples. Finally, the statistical interpretation of data is discussed, with the methods used, resulting expected limits and their interpretation presented.

The signal events, both Z′ and muonphilic scalar S, have a simple topology, with the observable particles being just two oppositely charged muons. The particles for which this study searches are, as discussed, inherently undetectable by the Belle II experiment, due to the fact they decay either to neutrinos or to some possible dark matter candidate.

The appearance of such particles therefore occurs only in the *recoil*, defined as the system recoiling from the detected muon pair in an event. The four-vector characterising such a system thereby describes the kinematics of undetected particles associated with the event. In the case of backgrounds this could be a composite object of multiple undetected SM particles, in the case of a signal event it of course describes directly the Z'/S particle itself. The invariant mass of this system thereby directly corresponds to the mass of some such particle. The search conducted herein specifically utilises the square of the recoil mass, defined as:

$$M^2_{\text{recoil}} = s + M^2_{\mu^+\mu^-} - 2\sqrt{s}E_{\mu^+\mu^-} \tag{1.1}$$

A signal would thereby appear as a peaking structure in this distribution, centred at $\mathrm{M}^2_{\text{recoil}} = \mathrm{M}^2_{Z'/S}$. Further discussion of this study will often switch between the particle

mass and the recoil mass. The polar angle of the recoiling system in the CMS frame, θ^*_{recoil}, is also used to improve separation from SM background processes.

The statistical analysis is conducted by way of a binned maximum likelihood fit in which templates (essentially histograms) corresponding to the expected 2D M^2_{recoil} vs. θ^*_{recoil} distribution (derived from MC) are fit to the observed data. Templates corresponding to both the signal and background are fitted, and the Z'/S cross section is extrapolated as the parameter of interest. The derived systematic uncertainties are introduced to the fit as either normalisation or shape uncertainties on the respective templates. The former of these corresponds to a single fitted value, or *nuisance parameter*, that corresponds to the yield of the template, while the latter can have individual values for each bin of the template that allow it to then fluctuate by the given amount during the fit process.

As discussed previously multiple Z' models are studied, the vanilla L_μ-L_τ where the decay occurs only to SM particles, and the dark Z' model in which instead decays to dark matter are allowed. For each of the dark matter couplings (α_D) studied, signal templates are created using MC samples generated with the given coupling, and new binning schema are defined for each so as to capture the wider signal peaks in M^2_{recoil}.

For each of the studied models, expected 90% confidence level upper limits are provided for both the cross section and relevant coupling strength. These define the region in which the signal hypothesis can be rejected with the stated confidence level (CL). Tests are conducted to ensure the accuracy of the fitting procedure, with signal events being injected into toy sampled data to ensure that when using real data any signal events would be properly measured.

2. Event reconstruction and selection

2.1 Data sets

This study makes use of data collected by the Belle II experiment between 2019 and 2023, providing a total integrated luminosity of 362fb^{-1} collected at a centre-of-mass energy of 10.58 GeV. Table 2.2 details the breakdown of this data set into individual experiments.

Signal samples

Monte Carlo signal samples are generated using the MadGraph 5 package [76]. Z′ signal samples are generated for mass hypotheses in steps of 10 MeV/c^2 ranging from 0.01 to 9.0 GeV/c^2. For each, 20,000 events are generated. The final binned maximum-likelihood fit (discussed in section 3) requires further samples to be generated. This is done to ensure that a mass hypothesis corresponding to each of the M^2_{recoil} bins that span the region studied herein is tested. This binning schema is defined in section 4.1.1.

A further set of signal samples is generated in steps of 100MeV/c^2 ranging from 0.1 to 9.0 GeV/c^2, this time using what is described as *run-dependent* conditions. This means that known experimental conditions that have been used to collect data, such as beam current for instance, are applied during the generation and reconstruction of the events. Again these are generated with 20,000 events for each mass hypothesis, and conditions spanning a set of runs and experiments are used: exp. 22 run 30, exp. 22 run 468, exp.24 run 985 and exp. 26 run 1485. This allows for dedicated study of the variation in signal efficiency across the full data taking period. Additionally a set of muonphilic scalar mass hypotheses are generated in steps of 10 MeV/c^2 ranging from 10 to 210 MeV/c^2, again with 20,000 events for each.

To study cases in which some non-zero coupling to dark matter is present, further Z′ signal samples are generated across the full mass range with valued of $\alpha_D = 0.05$, 0.1, 0.25 and 0.5. For each of these, specific mass hypotheses are chosen based on the binning schema that is constructed and defined in section 4.1.1.

Background samples

Background contributions from Standard model processes that could mimic the signal are studied, with various low multiplicity channels included. These are generated centrally by the Belle II collaboration, and are done so in a run dependant manner so as to

Process ($e^+e^- \to$)	$\int Ldt$ (ab^{-1})
$\mu^+\mu^-(\gamma)$	1.444
$\tau^+\tau^-(\gamma)$	1.444
$e^+e^-\mu^+\mu^-$	0.361
$e^+e^-(\gamma)$	0.036
$\pi^+\pi^-(\gamma)$	0.361
$K^+K^-(\gamma)$	0.361
$K^0\bar{K}^0(\gamma)$	0.361
$e^+e^-\pi^+\pi^-$	0.361
$e^+e^-\tau^+\tau^-$	0.361
$\mu^+\mu^-\mu^+\mu^-$	0.361
$e^+e^-e^+e^-$	0.361

Table 2.1: The process and corresponding integrated luminosity of each background sample used in this analysis, from the official MC15rd campaign.

accurately reflect the full data taking period. The luminosity of each sample is varied, table 2.1 summarises the processes studied and their respective luminosities.

2.2 Event reconstruction

The expected final state observed in the detector simply consists of two muons and some missing energy due to the undetected Z' boson. During the reconstruction process a set of basic requirements are imposed on events so as to remove any with noise in the detector (such as cosmic or beam background particles) or badly reconstructed components. Further to this, cuts are also applied to the *rest-of-event* (ROE), which contains the sum of all other ECL clusters not associated with a track. These clusters must lie within the ECL barrel with energy > 0.075 GeV. Before the use of MVA methods for signal selection, a set of preselection cuts are applied to events in order to reduce distinct background sources. These are selected so as to have a minimal effect on the signal efficiency.

1. Events are required to contain two, and only two, clean oppositely-charged tracks with *muon-ID* values greater than 0.5. Centrally generated correction factors for this variable are applied. These tracks are required to have originated in the interaction point with impact parameter requirements $|dz| < 2$ cm and $|dr| < 0.5$ cm.

2. An upper limit of 9 GeV/c^2 is applied to the recoil mass, above which SM processes dominate and become increasingly hard to separate from the signal.

3. In order to maintain a well modelled trigger efficiency, the polar angle, θ, of each muon is required to fall within the range covered by the ECL barrel region, $37° < \theta < 120°$. Outside this range the trigger efficiency drops rapidly and agreement between data and simulation worsens.

4. $e^+e^- \to \mu^+\mu^-$ events containing no recoiling quantity (i.e. lacking the presence of any undetected particle) are rejected by way of a slight cut to the 3D opening angle between the two muons, demanding that it be less than 179.5°.

5. Events must posses a dimuon transverse momentum greater than 0.5 GeV/c . This cut was chosen to reject a large portion of the main background processes, however, it does have a notable effect on the signal efficiency at high and very low masses. Regardless this is then superseded by the further MVA selection described in section 2.3, and so the cut is utilised.

6. A *photon veto*, which demands that no reconstructed photon lies within 15° of the recoil momentum vector, is employed to remove $e^+e^- \to \mu^+\mu^-\gamma$ backgrounds.

7. Events are required to have no greater than 0.5 GeV extra energy associated with clusters in the ROE.

8. The tracks associated with the two muons are required to be matched with an ECL cluster, and the ratio of ECL cluster energy to reconstructed track momentum must be equal to or less than 1.

9. To account for regions of the detector lacking good coverage with ECL instrumentation, such as gaps where cabling is fed through, cuts are applied to the polar angle of the recoil momentum vector, θ_{recoil}. $((91.15° < \theta_{\text{recoil}} < 123.19°)$ **or** $(33.8° < \theta_{\text{recoil}}) < 88.85^{\circ}))$ **or** $M_{\text{recoil}} > 2$ GeV/c^2. This is done to avoid events in which a single photon associated with an $e^+e^- \to \mu^+\mu^-(\gamma)$ process does not interact with the ECL and thus mimics a signal. For recoil masses above 2 GeV/c^2, most of the $\mu\mu(\gamma)$ background is due to two missed photons, one collinear with the beam and the second in one of the gaps. In this case, the direction of the recoil momentum has no direct connection with the direction of the missed photons, and therefore no constraint is imposed.

10. Events are required to have a dimuon mass greater than 0.5 GeV/c^2. This cut is imposed during production of $e^+e^- \to e^+e^-\mu^+\mu^-$ MC, and so is necessary to maintain agreement between data and simulation.

The signal efficiency is shown as a function of the successive application of each of these pre-selection cuts in figure 2.3, with Z′ mass hypotheses spanning the full range studied herein shown. The final signal efficiency ranges from approximately 20% for those hypotheses at the extremes of the range to 35% for those in the middle of the range. Cut 2, demanding the muon tracks reside within the ECL barrel region, has the largest effect on the efficiency, almost halving it for most mass hypotheses.

The 8.9 GeV/c^2 mass hypothesis is notably effected by the cut limiting the recoil mass to 9GeV/c^2, resulting from the spread of events due to detector resolution. Cut 5 reduces the signal efficiency particularly in the lowest mass hypothesis, owing to the fact that these signal events with low recoil often have near back-to-back muons and so the dimuon transverse momentum is low.

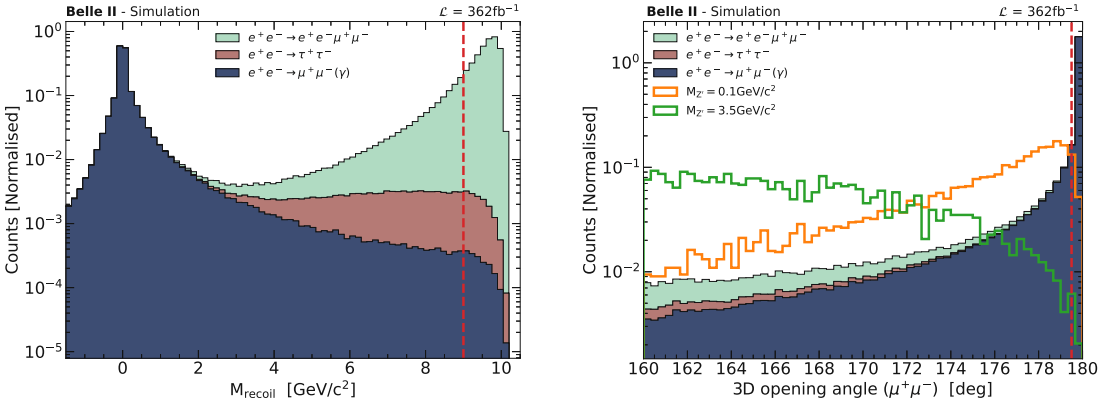

Figure 2.1: Left: The cut applied to limit M_{recoil} at 9 GeV/c^2 (cut 1). Right: Cut to the 3D opening angle between the muons less than 179.5° (cut 3), to remove the large bulk of $\mu^+\mu^-$ events where there is no recoiling particle.

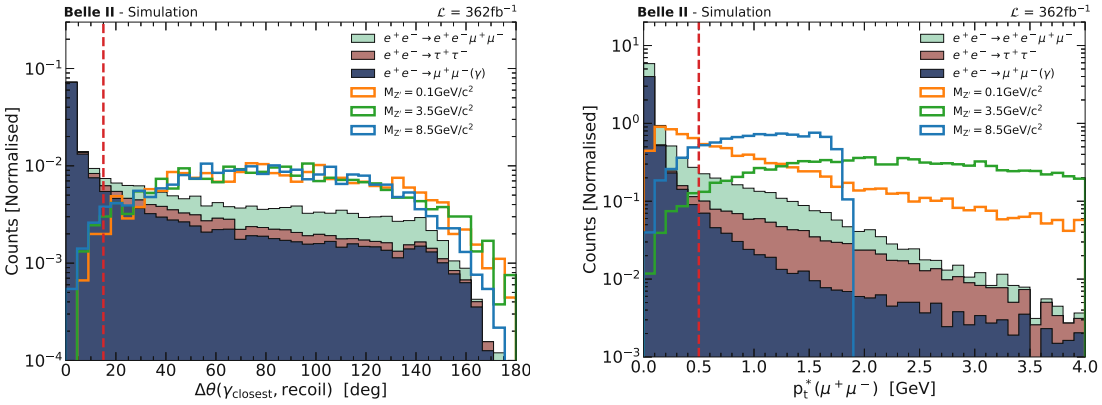

Figure 2.2: Left: The photon veto cut (cut 6), removing events where a photon lies within 15° of the recoil axis. Right: Cut to the dimuon transvserse momentum (cut 5), ensuring it to be greater than 0.5 GeV/c

The remaining distribution of background events after these pre-selection cuts are shown in fig. 2.4 above and originate from three main decay processes:

- $e^+e^- \to \mu^+\mu^-(\gamma)$: These events dominate the very low recoil mass region, with a peaking structure clear at around $M^2_{recoil} \sim 0$ GeV^2/c^4. While the majority of those pure un-radiative $\mu^+\mu^-$ events are removed by the cut the 3D opening angle (cut 3), there still remains a notable portion of radiative events. The photon veto (cut 4.) removes a large bulk of these events, however, many can still survive due to the photons residing outside detector acceptance, either through gaps in the ECL instrumentation or near collinear to the beam itself.

- $e^+e^- \to \tau^+\tau^-$: These events have a strong presence across the full M^2_{recoil} spec-

trum, comprising the dominant background source between approximately 5 and 50 $\mathrm{GeV}^2/\mathrm{c}^4$. Commonly these taus will decay leptonically to muons with associated muon neutrinos, however a fraction likely also come from 1-pronged decays to pions where the pion is then mis-identified as a muon.

- $\mathrm{e}^+\mathrm{e}^- \rightarrow \mathrm{e}^+\mathrm{e}^-\mu^+\mu^-$: The high squared recoil mass region, above approximately 50 $\mathrm{GeV}^2/\mathrm{c}^4$ is completely dominated by these events, with the two electrons produced in conjunction with the detected muons comprising the sizeable recoiling system. These events are primarily created by way of a two photon fusion process, and the electrons invariably are produced near to collinear with the beam pipe, thereby remaining outside of detector acceptance.

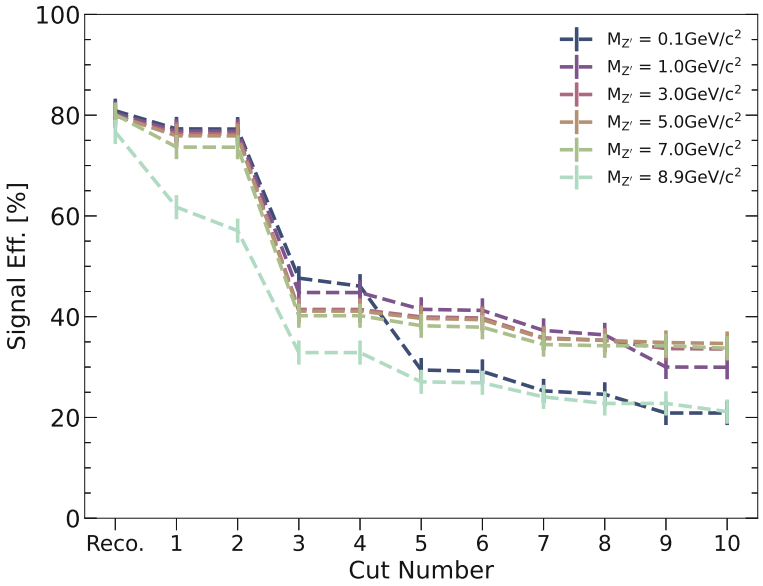

Figure 2.3: Signal efficiency as preselection cuts are applied for a set of Z′ mass hypotheses covering the full mass range. Reco. refers to the initial signal efficiency after reconstruction of events from detector data.

2.3 Background rejection with the Punzi-net

Bulk background rejection is carried out with a neural network that is trained using the *Punzi-loss* function, which was developed specifically for the purposes of searching for an invisibly decaying Z' boson. The subsequent trained neural network is dubbed the *Punzi-net*, and is optimised for the search such that it allows for just a single classifier to be used, with a single cut value applied to the output for all mass hypotheses investigated. The Punzi-net is discussed fully in section 3.

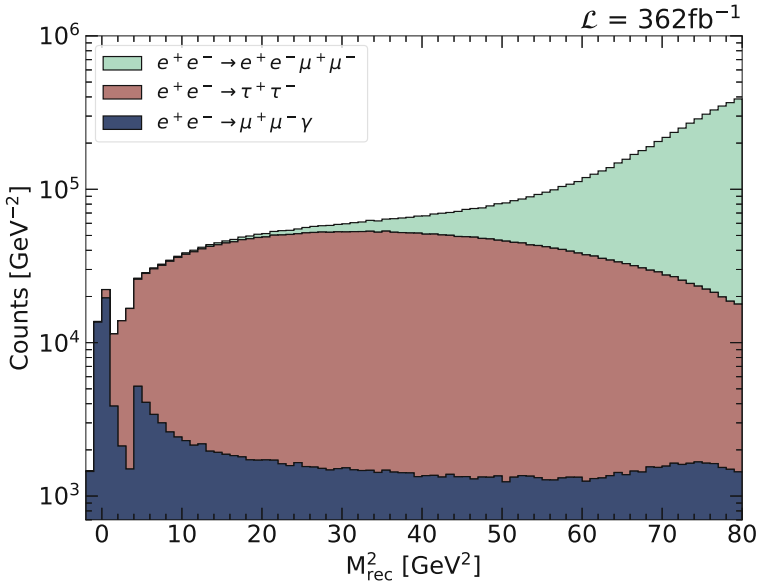

Figure 2.4: M^2_{recoil} distribution of the three main background processes after all preselection cuts are applied.

2.3.1 Choice of Variables

The Punzi-Net is trained using 8 input variables, most of which concern relationships between the two muons' momenta, the recoil momentum, and the thrust vector. The variables are chosen by way of ranking in mass sub regions of 1 GeV/c^2 width across the full Z′ mass range of [0, 9] GeV/c^2. In each mass range, a boosted decision tree is trained with a large set of event variables. This is done using the XGBoost package [77], which allows for the importance of individual variables to be extracted.

Those variables achieving highest importance in any of the mass regions are then used in the final choice of 8. This is done instead of considering the full mass range at once so as to avoid overlooking variables that have strong separation power in only small regions of the parameter space.

- $P_t^*(\mu^+\mu^-)$: Transverse momentum of the dimuon pair in the CMS frame.
- $P^*_{t,recoil}(\mu_{max})$: Transverse momentum of the higher momentum muon with respect to the recoil momentum in the CMS frame.
- $P^*_{t,recoil}(\mu_{min})$: Transverse momentum of the lower momentum muon with respect to the recoil momentum in the CMS frame.
- $P^*_{l,recoil}(\mu_{max})$: Longitudinal momentum of the higher momentum muon with respect to the recoil momentum in the CMS frame.

- $P^*_{l,\mu_{min}}(\mu_{max})$: Longitudinal momentum of the higher momentum muon with respect to the lower momentum muon in the CMS frame.
- $P^*_{t,thrust}(\mu)$: Transverse momentum of the dimuon pair with respect to the *thrust* axis in the CMS frame. The thrust vector is defined by the vector, $\hat{t}$, that maximises the relation:
$$T = \frac{\sum_i |p^*_i \cdot \hat{t}|}{\sum_i p^*_i}$$
where p^*_i is the momentum of the i^{th} muon in the event.
- $A_{\mu\mu}$: defined in the following subsection.
- $D_{\mu\mu}$: defined in the second following subsection.

The distributions of the first six of these variables are shown in figure 2.5, including the three main background contributions and three example Z′ signals at mass hypotheses of 0.5, 4.0 and 7.5 GeV^2/c^4.

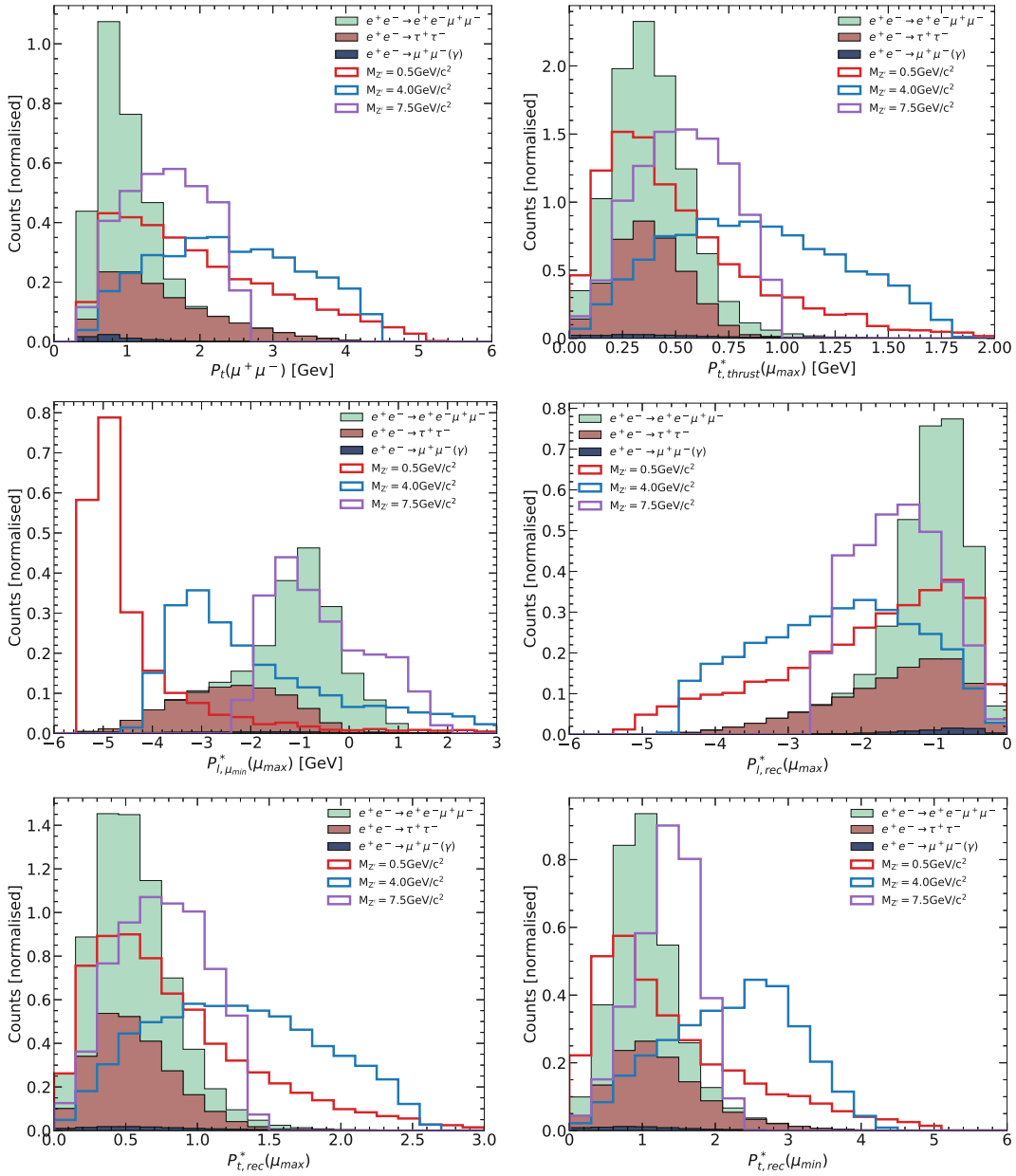

Figure 2.5: Signal and background distributions of the first 6 Punzi-net input variables, with 3 different signal mass hypotheses shown (0.5, 4.0 and 7.5GeV/c^2).

$\mathbf{A}_{\mu\mu}$ variable

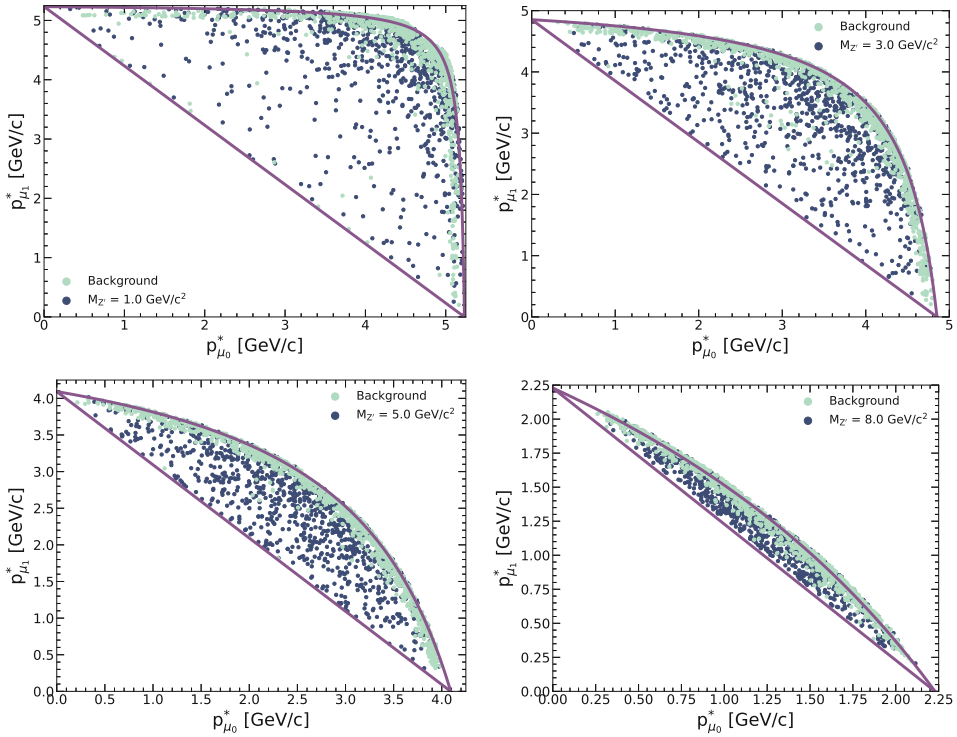

Figure 2.6: 2D distribution of the momenta of the two muons, shown are background signal events that fall within 0.5σ of the signal peak of the given mass hypothesis. The lines constraining the distributions are shown

Figure 2.6 shows the 2D distribution of the CMS-frame momenta of the two muons at four chosen Z′ mass hypotheses. Events are restricted to $\pm 0.5\sigma$ around the M^2_{recoil} peak of the given mass hypothesis, so as to include only events with similar kinematics. One can see that each of these distributions can be constrained by two lines (shown in purple), both of which can be defined analytically by considering the centre-of-mass energy, $\sqrt{s}$, and the physical limit of the recoil momentum, P, which in the case of a signal would be the momentum of the Z′ boson. It therefore corresponds to a recoiling system of mass $M_{Z'}$ recoiling against a dimuon system with the lowest possible squared invariant mass, $2m^2_\mu$. This can be defined analytically as,

$$P = \frac{\sqrt{(s + M^2_{Z'} - (2m_\mu))^2 - 4sM^2_{Z'}}}{2\sqrt{s}} \tag{2.1}$$

and the two lines that constrain the distribution can then be written,

$$p^*_{\mu_1} = P - p^*_{\mu_0} \tag{2.2}$$

$$p^*_{\mu_1} = \frac{k}{p^*_{\mu_0} - \sqrt{s}/2} + \sqrt{s}/2, \quad \text{with} \quad k = \sqrt{s}/2(\sqrt{s}/2 - P) \tag{2.3}$$

where the lower bounding straight line is given by 2.2, and the higher bound is given by the hyperbolic function 2.3. One can see a clear difference in the distributions of signal and background events within this 2D space, with the latter tending to populate a region that follows the hyperbolic edge closely and the former describing a more even distribution between the two bounds.

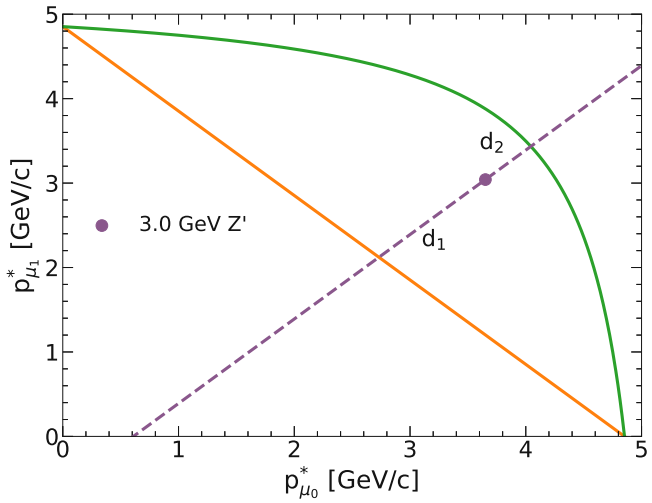

Figure 2.7: A single event shown by the purple point in the 2D muon momenta plane with the lines defined by equations 2.3 and 2.2 shown in green and yellow respectively. The two distances, d_1 and d_2, are then shown with the two portions of the dashed purple line that is perpendicular to that defined by eqn.2.2.

With this observation in mind, one can seek to define some single variable that might encapsulate this information in a single number. Figure 2.7 below shows the same 2D space with just a single signal event, at a mass of 3 GeV/c^2. One can draw a 3^{rd} line, perpendicular to eqn. 2.2, that intersects the point defined by the given event. Lengths d_1 and d_2, spanning the distances along this line from the given event to the bounding lines 2.2 and 2.3 respectively are defined, and then such a variable can be constructed from these,

$$A_{\mu\mu} = \frac{d_1 - d_2}{d_1 + d_2} \tag{2.4}$$

This provides a single value, that varies between -1 and 1, and captures the given event's distance from the hyperbolic function. Due to background events clustering near to this hyperbola, one can expect the value of $A_{\mu\mu}$ to peak at 1, whereas signal events will have a flat distribution. Furthermore, this provides a variable that does not vary in scale with the event's squared recoil mass, and so helps to mitigate any concern of biasing the search towards specific mass regions.

It was found that in the higher squared recoil mass region, a small number of events were not accurately constrained by the two defined boundaries and would sit outside of the region. This would result in a secondary peak at a lower value of 1. In such cases it was decided that the value would be modified by $A_{\mu\mu} \rightarrow 1/A_{\mu\mu}$, resulting in these events taking higher values than 1 as can be seen in fig. 2.8.

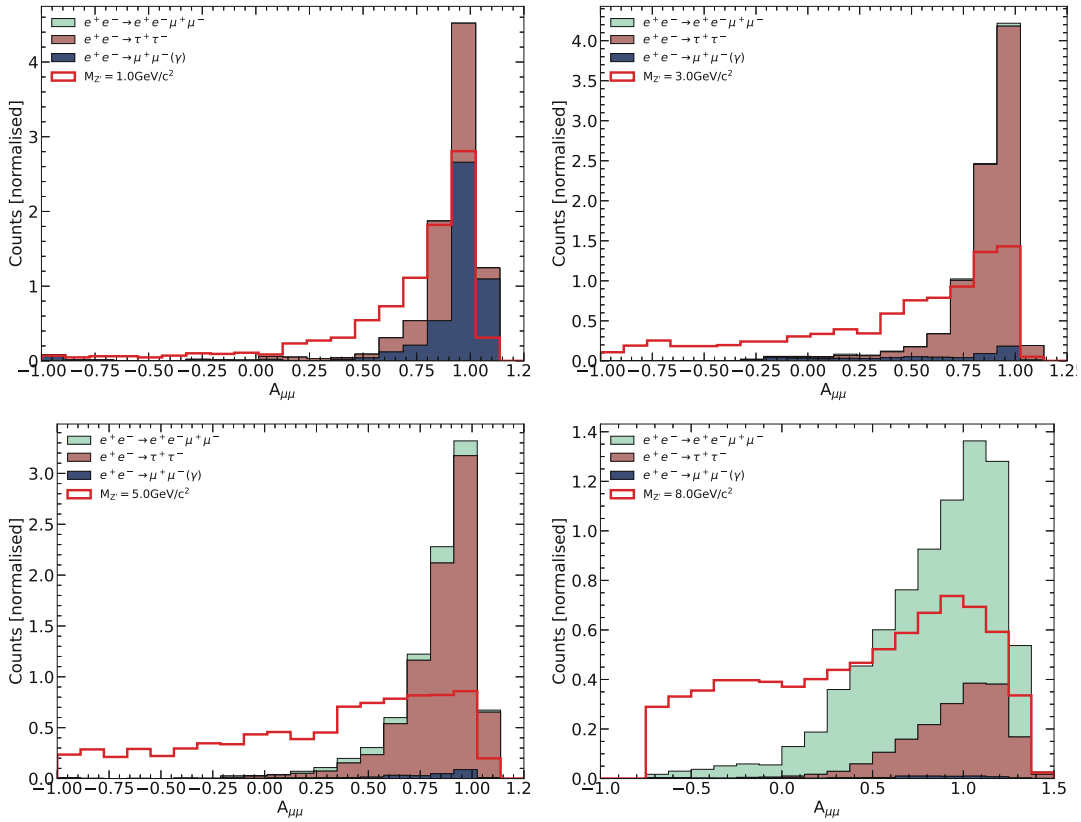

Figure 2.8: Distribution of $A_{\mu\mu}$ for signal and background events that fall within 0.5σ of the signal peak of the given mass hypothesis.

$D_{\mu\mu}$ variable

One can make use of the so-called *antler topology* [78] of the $e^+e^- \rightarrow \tau^+\tau^-$ process to define a new variable that provides strong separation power from signal events. Such

events invariably contain two tau leptons that are both on-shell and produced back-to-back. One can thereby approximate the decay process as $\tau^{\pm} \rightarrow \mu^{\pm}$N, where N is a composite object comprised of the muon and tau neutrinos that would be produced in such a process. This is shown in figure 2.9, with two tau leptons decaying to muons and the associated composite objects comprising of the two neutrinos in each decay. Note that this a rough graphical representation and not an accurate feynman diagram for such a process, which would include $W^{\pm}$ bosons mediating the tau lepton decays.

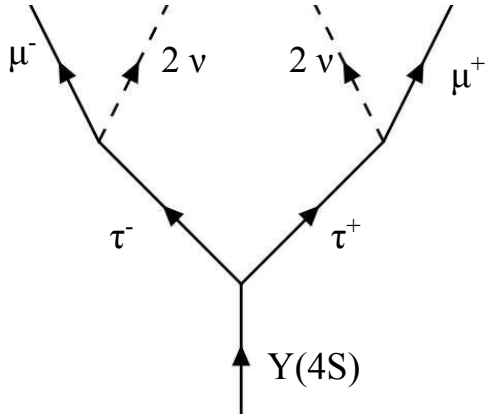

Figure 2.9: The antler decay topology, with two tau leptons decaying to produce muons and associated neutrinos. The tau lepton decay to a muon is mediated by a $W^{\pm}$ boson, which is not shown here.

One can then reasonably make the assumption that such a composite object made only of neutrinos would have a negligible mass. On top of this, if one approximates also the muon mass as negligible, the dimuon mass can then be defined as:

$$(M^{\tau}_{\mu\mu})^2 = \frac{2(m_{\tau}^4 - m_{\tau}^2\sqrt{s}(E_{\mu^+} + E_{\mu^-}) + 2(s - 2m_{\tau}^2)E_{\mu^+}E_{\mu^-})}{s - 4m_{\tau}^2} \tag{2.5}$$

Where $E_{\mu^{\pm}}$ is the measured energy of the muons, s is the centre of mass energy and m_{τ} is the mass of the tau lepton. The difference between this calculated quantity and the measured dimuon invariant mass, $M_{\mu\mu}$, can then be taken to define a new variable:

$$D_{\mu\mu} = (M^{\tau}_{\mu\mu})^2 - M^2_{\mu\mu} \tag{2.6}$$

This then has an expected value of 0 for $e^+e^- \rightarrow \tau^+\tau^-(\gamma)$ background events, thereby allowing for these to be separated from signal events. The proposal of this variable is discussed in more detail for a similar search at Belle II in [79]. The distribution of signal and background events with four example mass hypotheses are shown in figure 2.10, where only those backgrounds residing near the given signal peak in M^2_{recoil} are shown. One can see the $e^+e^- \rightarrow \tau^+\tau^-(\gamma)$ background events do mostly reside at $D_{\mu\mu} \sim 0$ as one would expect, thus providing separation power particularly in the mid masses (3, 5 GeV/c^2) where this background dominates.

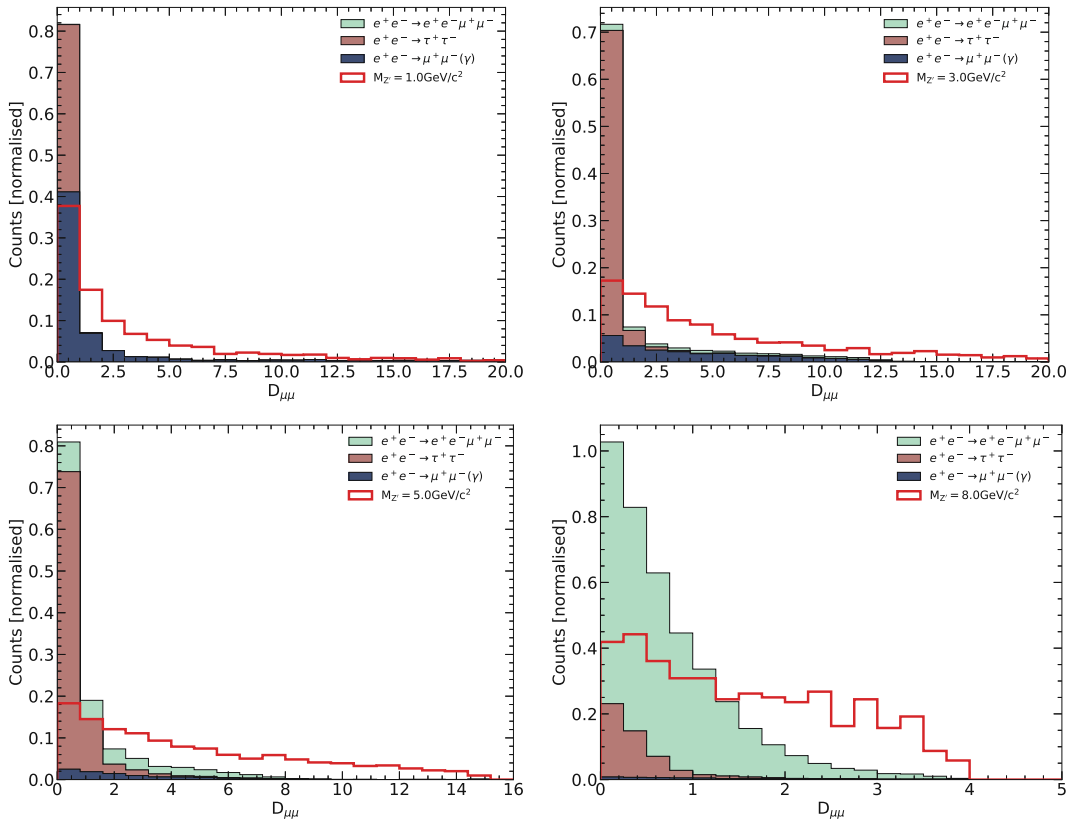

Figure 2.10: Distribution of $D_{\mu\mu}$ for signal and background events that fall within 0.5σ of the signal peak of the given mass hypothesis.

2.3.2 Rejection of backgrounds for muonphilic scalar search

As discussed in section 1.4, the search for BSM physics associated with the production of missing energy in conjunction with a muon pair is extended also to the case of a muonphilic scalar particle, S. This is limited to the low mass region, below the muon mass limit ($M_{recoil} < 2M_\mu$). The signature of such a particle is in general the same, but there are kinematic differences that must be considered. These primarily relate to the fact that, a Z′ is invariably *soft* at these low masses[1]. This is expected for such a vector particle, in a similar way a final state radiation photon emitted from the muon would also be soft, the muonphilic dark scalar is instead preferentially produced *hard*.

This results from the difference in spin between the two. Both are produced via photon-mediated e^+e^- interactions which proceed primarily through ± 1 helicity states, as 0 helicity states are suppressed by a factor $m_e/\sqrt{s}$. Due to the conservation of angular momentum, vector and scalar production occur through s-wave and p-wave processes respectively. At low M_S, p-wave suppression occurs and the scalar process grows slowly

[1]Soft and hard refer to the momentum of the radiated particle, low or high respectively.

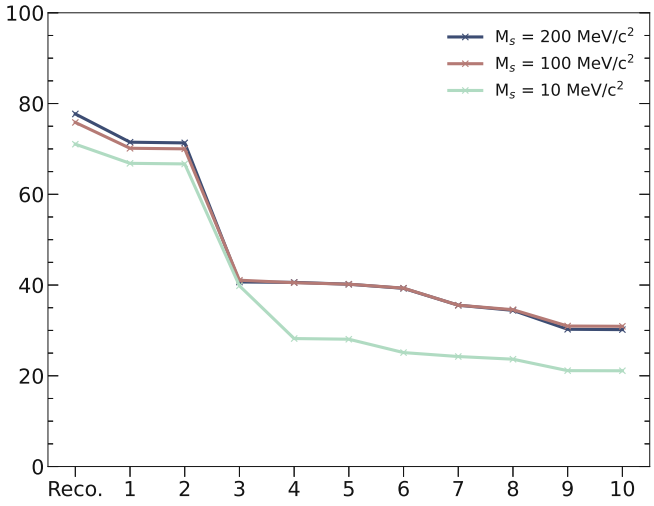

Figure 2.11: The signal efficiency as preselection cuts are applied to muonphilic scalar signal MC. 'Reco.' refers to the efficiency after initial event reconstruction via basf2.

with energy. Conversely the vector processes undergoes no such suppression and has something of a sudden increase in cross section.

The same preselection criteria are used as in the case of the Z′. The signal efficiency after each of these are applied is shown in figure 2.11, with three mass hypotheses (10, 100 and 200 MeV/c^2) spanning the range studied. The 10 MeV/c^2 appears to be heavily effected by the cut on the 3D opening angle (cut 4 in section 2.2) which drops the efficiency from approximately 40% to 25%.

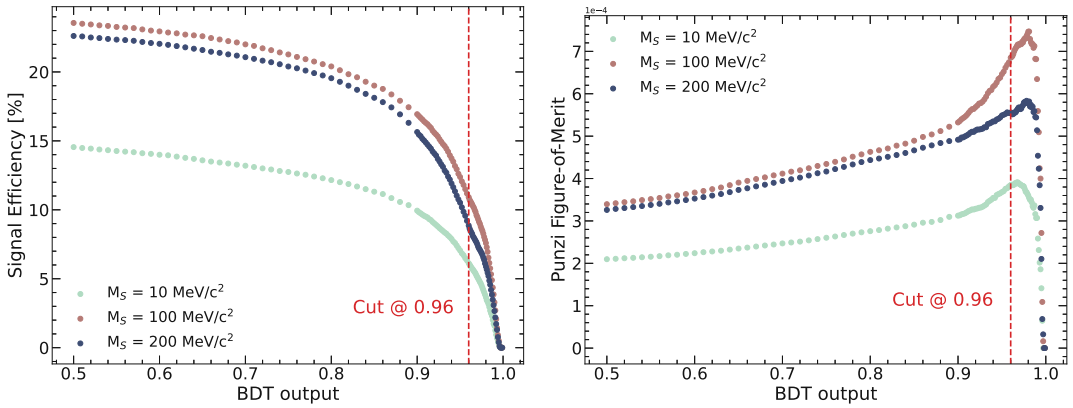

Figure 2.12: The left hand plot shows the signal efficiency for the three mass hypotheses spanning the range, while the right hand plot shows the Punzi figure-of-merit achieved when cutting at the given output.The chosen cut value of 0.96 is marked.

It was found that the Punzi-net used for the selection of Z′ candidates did not provide

optimal background rejection for this study. Due to the small mass range in which it is conducted, a simple BDT classifier is in instead employed. This is trained using the same 8 variables. The comparison of these variables can be seen in figure 2.13. The hyper-parameters were chosen by way of a simple grid search and are summarised in table 2.2.

Hyper-parameter	Value
Learning rate	0.025
N estimators	3000
Max depth	3

Table 2.2: The hyper-parameters used in the BDT.

Training is conducted using a set of generated signal MC samples, with masses from the set $M_{Z'} \in [10, 15, ..., 195, 200]$ MeV/c^2. Figure 2.12 below shows the signal efficiency (left) and Punzi figure-of-merit (right) as a function of the BDT output for a set of 3 masses spanning this range. A cut value of 0.96 is applied for final signal selection, which provides approximately the optimal Punzi figure-of-merit across the masses while still maintaining a signal efficiency of between 6 and 12%. The considerable drop in signal efficiency for the 10 MeV sample is attributed to pre-selection cuts aimed at removing $e^+e^- \to \mu^+\mu^-(\gamma)$ backgrounds, such as the cut to the 3D muon opening angle (cut 3) shown on the right in figure 2.1.

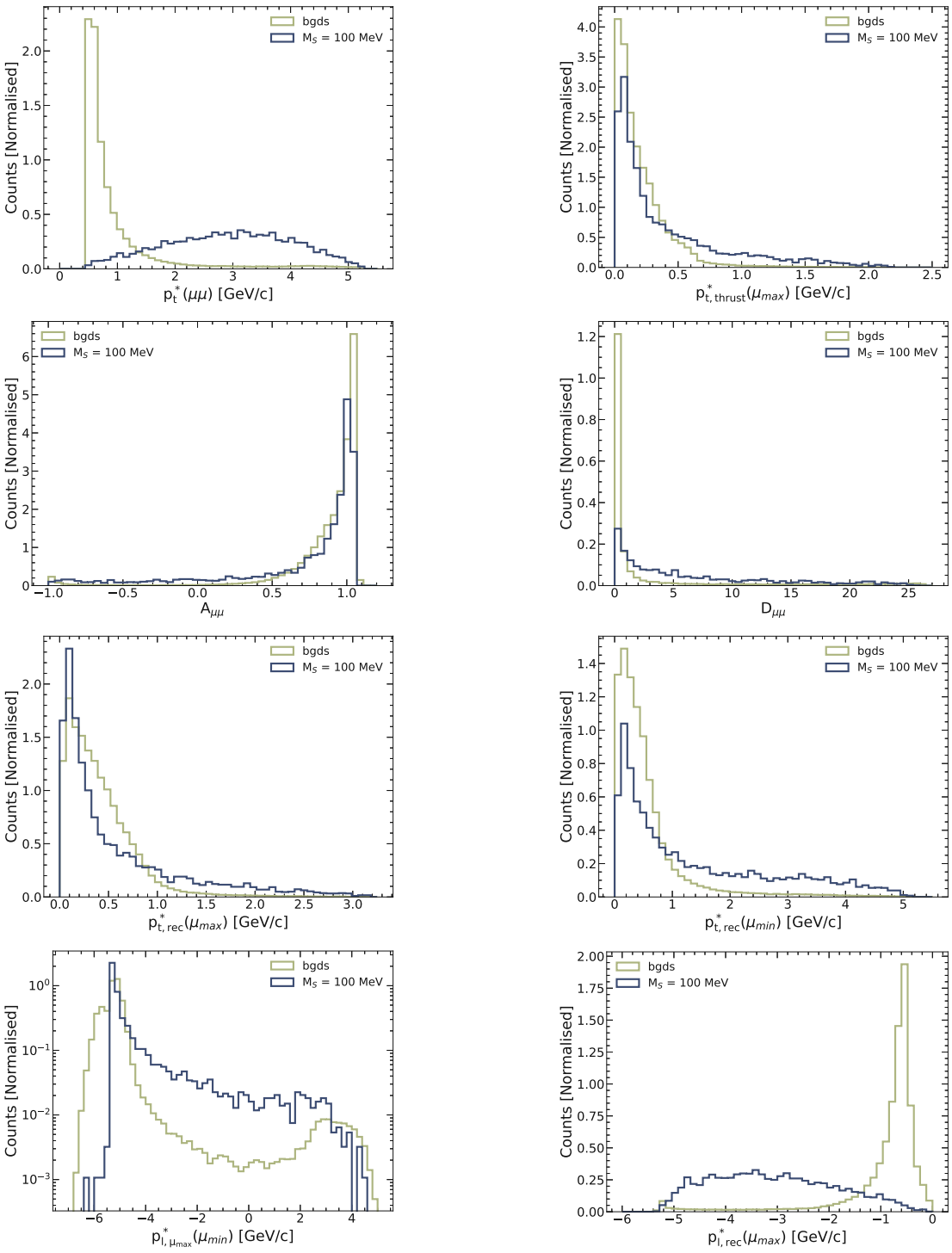

Figure 2.13: Signal and background distributions for the 8 input variables of the BDT used for selection in the muonphilic scalar search.

2.4 Final background composition

Figure 2.14 shows the squared recoil mass distribution, M^2_{recoil}, of the expected background processes after all preselection cuts and the Punzi-Net selection are applied. The low mass region contains a large peaking structure of $e^+e^- \to \mu^+\mu^-(\gamma)$ events centred at a squared recoil mass of 0 GeV^2/c^4. This is then followed by a region of near-to-no backgrounds up to approximately 30 GeV^2/c^4 where a small number of $e^+e^- \to \mu^+\mu^-(\gamma)$ appears again. Above approximately 50 GeV^2/c^4 a rapidly growing contribution of $e^+e^- \to e^+e^-\mu^+\mu^-$ events appears. As discussed previously in such occurrences the two electrons are invariably lost up the beam pipe and so completely outside of detector acceptance.

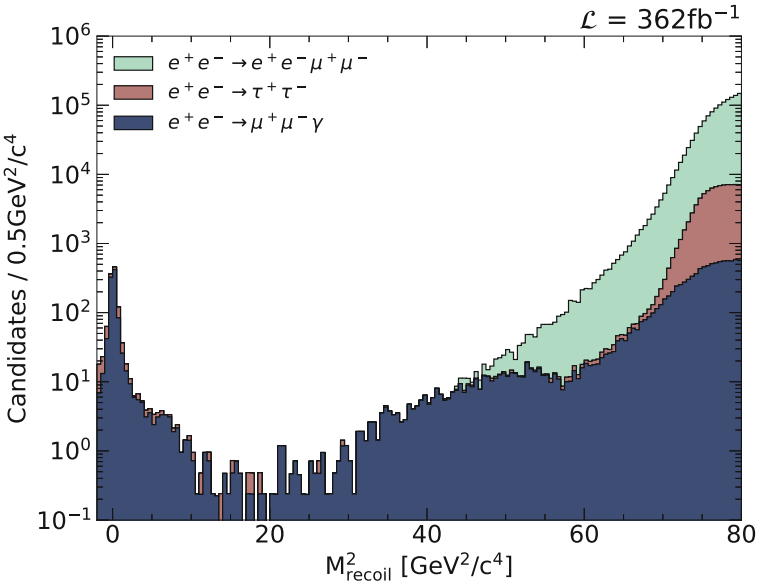

Figure 2.14: The squared recoil mass distribution after all preselections and the Punzi-Net have been applied.

The 2D distributions of recoil polar angle in CMS frame (θ^{CMS}_{recoil}) against squared recoil mass (M^2_{recoil}) for each of the three main background components along with a set of four signal mass hypotheses at $M_{Z'}$ = 0.01, 3.0, 5.5 and 8.0 GeV/c^2 are shown in figure 2.15 below. One can see a notable shaping in the polar angle for events away from the limits of the M^2_{recoil} distribution. For the majority of the range the events are constricted to between approximately 70 and 120 degrees. The $e^+e^- \to e^+e^-\mu^+\mu^-$ and $e^+e^- \to \tau^+\tau^-$ show a flat distribution across the polar angle, as do the signal samples.

The $e^+e^- \to \mu^+\mu^-(\gamma)$ instead show some clear trends across θ^{CMS}_{recoil}. It is important for one to properly understand the mechanisms behind these. The previous iteration of the analysis was faced with a large data discrepancy after unblinding that resulted from mis-modelling of these radiative $\mu^+\mu^-\gamma$ events and so here this channel is studied

comprehensively to avoid such risks.

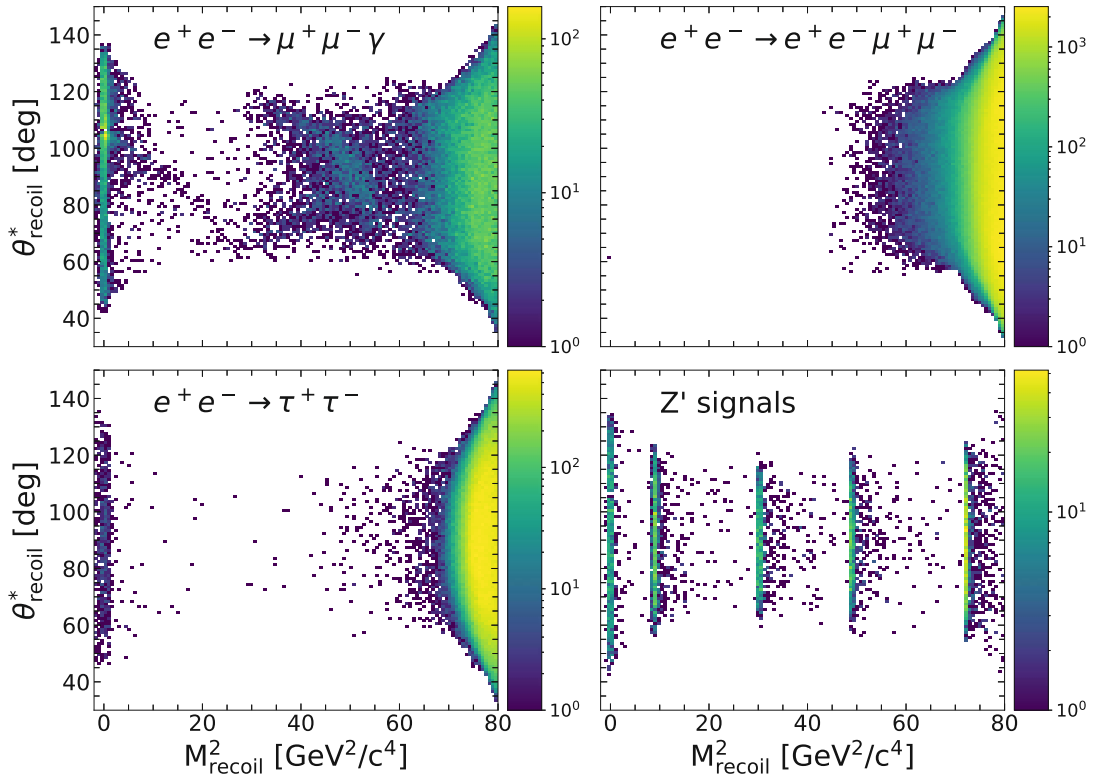

Figure 2.15: 2D distribution of recoil polar angle in CMS frame (θ^{CMS}_{rec}) against squared recoil mass (M^2_{recoil}) for each of the three main background components along with a set of four signal mass hypotheses ($M_{Z'}$ = 0.01, 3.0, 5.5 and 8.0 GeV/c^2)

2.4.1 Residual radiative backgrounds

The presence of large $e^+e^- \to \mu^+\mu^-(\gamma)$ backgrounds after the application of the Punzinet, shown in the top left panel of figure 2.15, indicates that there are a number of events containing photons that go somehow undetected by the ECL. This results in the basf2 software failing to properly reconstruct the photon and thereby allowing the event to survive the ECL-based photon veto that is applied during preselection (cut 6). Such events could easily mimic the signature of a Z′, especially around $M^2_{recoil} \sim 0 GeV^2/c^4$ where they form a peaking structure, and so must be understood better.

There are clear band-like structures in the recoil polar angle across M^2_{recoil} up to approximately 60 GeV^2/c^4, which would indicate that these photons are likely associated with specific regions of the detector. These structures can be isolated for further study of three regions in which one might expect different contributing processes: the low squared recoil mass region ($M^2_{recoil} \leq 4$ GeV^2/c^4) where events span the full polar angle range

but have clear structures within this, the $4 < M^2_{recoil} \leq 30$ GeV^2/c^4 region in which events are more constrained to a specific, varying polar angle range, and finally the $30 < M^2_{recoil} \leq 60$ GeV^2/c^4 region where one might expect multiple missed high energy photons contributing to the event to create the large recoiling system with respect to the two muons. There are, of course, $e^+e^- \to \mu^+\mu^-(\gamma)$ events present above this point, however, in this region there is no clear structure in the distribution shown in fig. 2.15.

Low squared recoil mass region ($M^2_{recoil} < 4$ GeV^2/c^4)

Figure 2.16 contains distributions showing only MC $e^+e^- \to \mu^+\mu^-$ events with $M^2_{recoil} \leq 4$ GeV^2/c^4, with all variables shown in the LAB frame. The left hand sub-figure shows the distribution of the MC-generated polar angle, $\theta^{generated}_{\gamma_1}$, of the highest energy photon in a given event against the recoil polar angle, θ^{LAB}_{recoil}. This shows a strong trend with the majority of events containing a photon correlated directly with the recoil direction. The majority of such cases show the photon landing in the 90° gap. A small number do appear to instead have a photon travelling in the forward or backward direction nearly parallel to the beam pipe where ECL coverage does not reach.

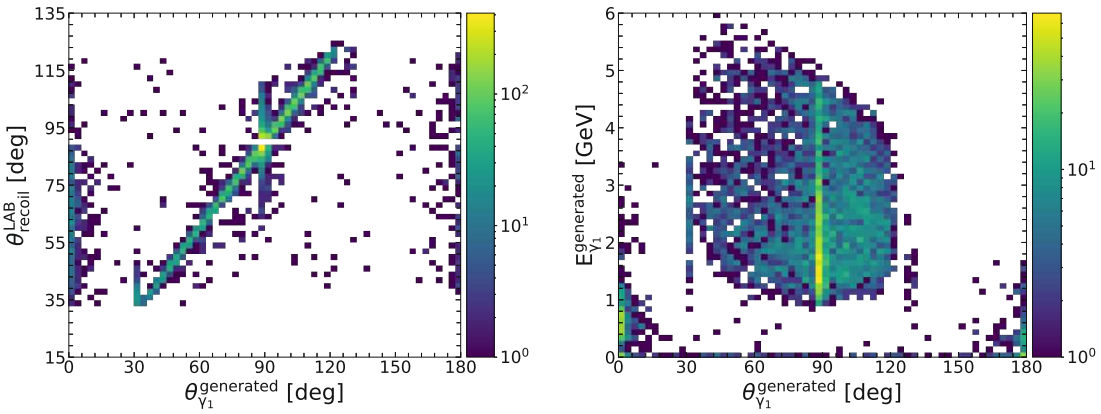

Figure 2.16: Left: distribution of the recoil polar angle and MC generated polar angle (LAB frame) of the photon in $\mu^+\mu^-\gamma$ events that survive all selections due to the photon going undetected. This figure contains only events with $M^2_{recoil} \leq 4$ GeV^2/c^4. Right: distribution of the MC generated energy and polar angle (LAB frame) of the photon in $\mu^+\mu^-\gamma$ events that survive all selections due to the photon going undetected. This figure contains only events with $M^2_{recoil} \leq 4$ GeV^2/c^4.

The right hand sub-figure shows the MC-generated polar angle of the highest energy photon in a given event against the MC-generated photon energy. Again one can see a clear structure at approximately 90°. Additionally this shows that those photons lost up or down the beam pipe invariably have low energies, below approximately 1 GeV. From these two distributions, one can conclude that those $\mu^+\mu^-$ events with $M^2_{recoil} \leq 4$ GeV^2/c^4, and that pass all selections, often contain a single *hard* photon (a photon of

energy greater than ≈1 GeV), coincident with the recoil axis, that is commonly lost in the $\sim$ 90° gap of the ECL.

Mid squared recoil mass region ($4 < M^2_{recoil} \leq 30$ GeV^2/c^4)

The left hand plot of figure 2.17 shows again the distribution of the MC-generated polar angle, $\theta^{generated}_{\gamma_1}$, of the highest energy photon in a given event against the recoil polar angle, θ^{LAB}_{recoil}, however now containing only those events that lie in the range $4 < M^2_{recoil} \leq 30$ GeV^2/c^4. With this, the correlation that previously existed between the two angles is gone, with these photons now more consistently residing either again at the $\sim$ 90° gap, or at $\sim$ 30° where the join between the ECL barrel and endcap resides, providing a small region for these photons to escape undetected.

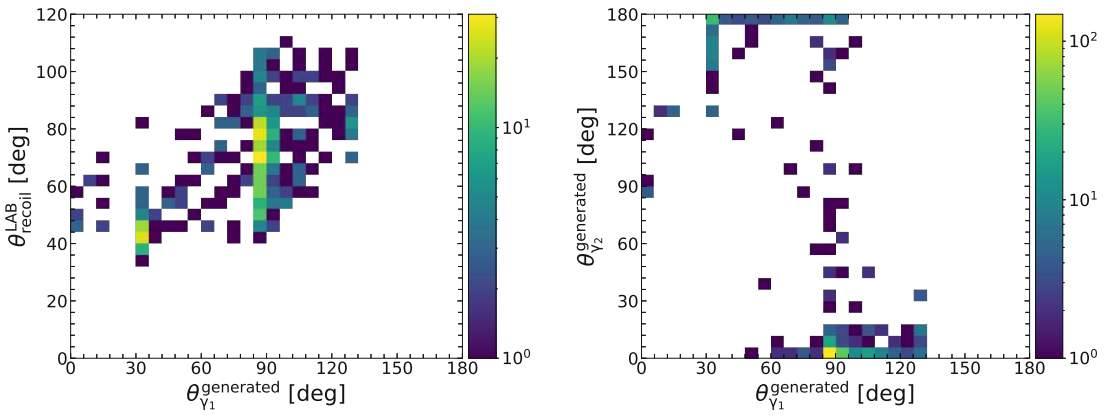

Figure 2.17: Left: distribution of the recoil polar angle and MC generated polar angle (LAB frame) of the photon in $\mu^+\mu^-\gamma$ events that survive all selections due to the photon going undetected. This figure contains only events in the range $4 < M^2_{recoil} \leq 30$ GeV^2/c^4. Right: 2D distribution of the MC-generated polar angles of the highest (x-axis) and 2nd highest (y-axis) energy photons. This figure contains only events in the range $4 < M^2_{recoil} \leq 30$ GeV^2/c^4.

Furthermore, the lack of correlation between these highest energy photon with the recoil axis would imply that other energetic photons are present and thus contribute to an overall recoil axis that does not coincide with any individual. The right hand plot of figure 2.17 shows the MC-generated polar angles of the highest (x-axis) and 2nd highest (y-axis) energy photons. From this one can see that in fact the majority of events in this squared recoil mass region contain two photons, with the leading photon lost, as discussed, to instrumentation gaps at 30° or 90°, and the secondary photon being lost along the beam pipe.

High squared recoil mass region ($30 < M^2_{recoil} \leq 60$ GeV^2/c^4)

Finally in the higher squared recoil mass region one could expect to have contributions from multiple high energy photons. The left hand plot of figure 2.18 shows the 2D distribution of the MC-generated energies of the highest (x-axis) and 2nd highest (y-axis) energy photons. Indeed this shows that events in this squared recoil mass region have have two high energy photons, of at least approximately 3 and 1 GeV respectively. The right hand plot shows the generated polar angles of these two photons, both of which now are invariably lost either in the gaps where the ECL barrel and endcaps meet ($\approx 30°$, 130°), or into the beam pipe itself.

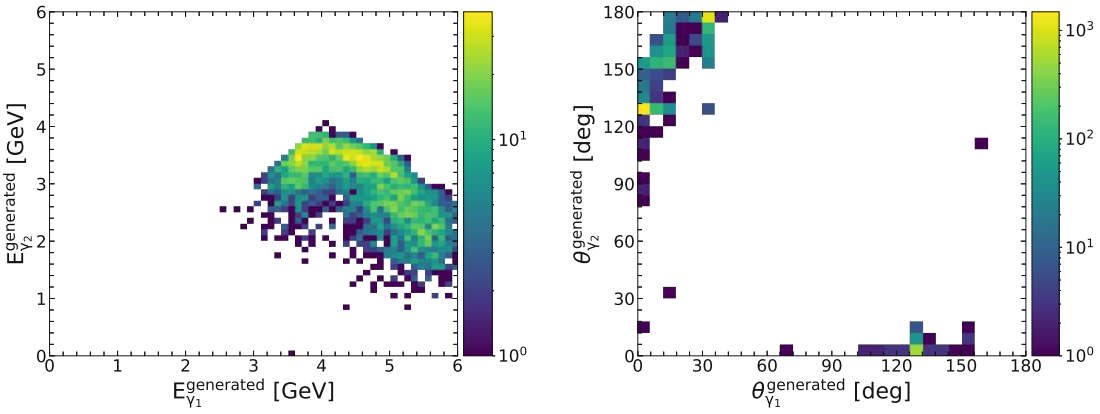

Figure 2.18: Left: 2D distribution of the MC-generated energies of the highest (x-axis) and 2nd highest (y-axis) energy photons. This figure contains only events in the range $30 < M^2_{recoil} \leq 60$ GeV^2/c^4. Right: 2D distribution of the MC-generated polar angles of the highest (x-axis) and 2nd highest (y-axis) energy photons. This figure contains only events in the range $30 < M^2_{recoil} \leq 60$ GeV^2/c^4.

Implementation of KLM-based photon veto

The remaining $\mu^+\mu^-$ that survive both the ECL-based photon veto cut and further Punzi-net selection often have photons that escape detection via regions of the ECL known to have poor coverage. The KLM sub-system, which doesn't suffer from this same problem, does provide a second line of defense against such decays. Although designed to detect primarily K_L or muons, photons can induce showers upon interacting with iron plates, just at a K_L would, and thereby be picked up by the sub-detector.

Based on the study of event topologies of the residual radiative backgrounds in distinct regions of the squared recoil mass, an additional KLM-based photon veto can be defined. This contains selections based on two different mass regimes:

- $M^2_{recoil} < 4$: No non-track matched KLM cluster with klmClusterInnermostLayer ≤ 4 and $N^{KLM}_{layers} \leq 6$ within 15° of the recoil vector.

- $M^2_{\text{recoil}} \geq 4$: No non-track matched KLM cluster with klmClusterInnermostLayer ≤ 4 and $N^{\text{KLM}}_{\text{layers}} \leq 6$ within ECL instrumentation gaps, defined as;
  - $30.0° < \theta < 34.0°$
  - $87.5° < \theta < 92.5°$
  - $125.0° < \theta < 130.0°$

Where 'klmClusterInnermostLayer' refers to the innermost layer that shows hits, a higher number would instead imply the source is possibly from outside the detector. $N^{\text{KLM}}_{\text{layers}}$ refers to the number of layers showing hits.

The effect this further veto has on the distribution of $\mu^+\mu^-$ backgrounds is shown in figure 2.19, with the initial distribution on the left and after the application of the veto on the right. One can see from the color scale that this veto reduces the peaking structure around ≈110° substantially. Furthermore the band structures in the intermediate mass range is cleaned up well. There exists still a round band between 30 and 45 GeV^2/c^4.

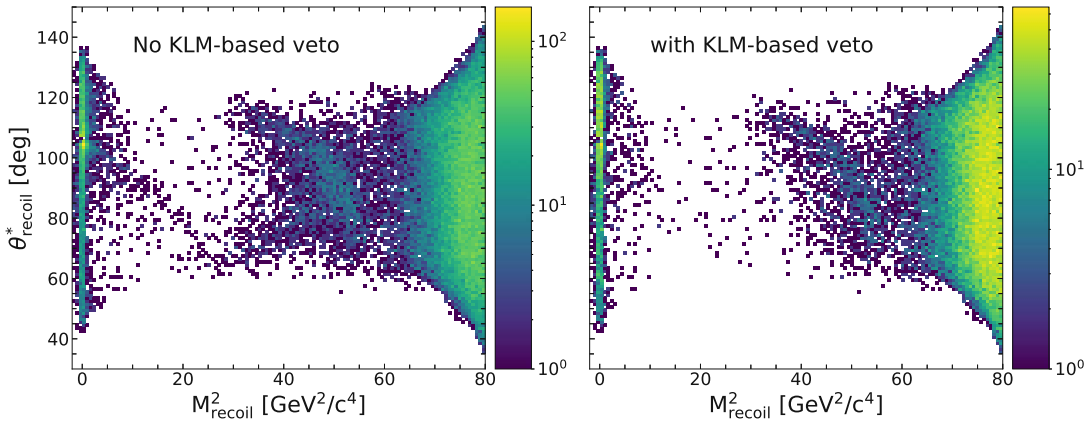

Figure 2.19: 2D distribution of recoil polar angle in CMS frame ($\theta^{\text{CMS}}_{\text{rec}}$) against squared recoil mass (M^2_{rec}) of $\mu^+\mu^-(\gamma)$ background events surviving the Punzi-Net. The left plot shows this before the application of the additional KLM-based photon veto, the right plot shows the result after this is then applied.

2.4.2 Signal Efficiency

The signal efficiency after pre-selection cuts (dashed line) and full Punzi-Net selection (solid line) is shown in figure 2.20. Both run independent and run dependent MC samples are shown, with experimental conditions corresponding to 4 different experiment and run periods shown. There is a notable decrease in signal efficiency for signal samples produced under the conditions of experiment 26, run 1485. This is due to the collider being run with a higher current and thereby instantaneous luminosity during this period, the result of which is a large increase in beam backgrounds (discussed in section 2). This leads to events in general having more activity in the ECL and CDC, and thus a larger portion of signal events are rejected with cuts to the number of tracks or energy deposited by the rest-of-event in the ECL. There is a step in signal efficiency after pre-selection at $M_{Z'}$=2 GeV/c^2. This is due to cut 9, which applies selection on events below this mass.

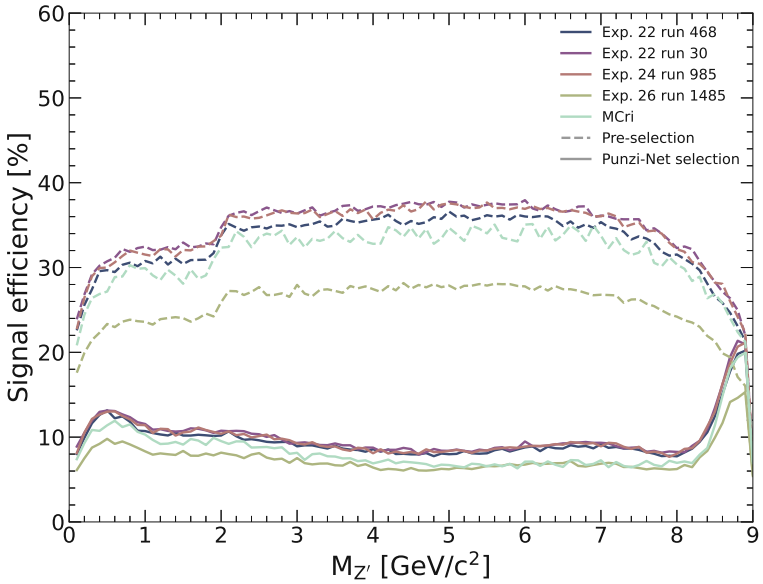

Figure 2.20: Signal efficiency across mass hypotheses both after pre-selection cuts (dashed line) and after application of the Punzi-Net (solid line). Both run-independent and run-dependent MC samples are shown, with experimental conditions from multiple experiments and run conditions used.

The final signal efficiency after all selections sits relatively stable at around 10% across the full range of masses, with a notable peak at around 14% between 0 and 1 GeV/c^2. This is due to the presence of irreducible $\mu^+\mu^-(\gamma)$ backgrounds where the photon has gone somehow undetected in both the ECL and KLM, thereby mimicking signal and thus forcing the Punzi-Net to adopt less aggressive selection in this region. Similarly there is a peak in signal efficiency towards 9 GeV/c^2, this time due to irreducible $e^+e^-\mu^+\mu^-$ backgrounds where both electrons are invariably in the direction of the beam pipe and so undetected.

The signal efficiency is also checked for those signal samples generated with some level of coupling to dark matter included. Figure 2.21 shows the signal efficiency as a function of the mass hypothesis for dark matter couplings of α_D = 0.05, 0.1, 0.25 and 0.5. Again the efficiency is shown both after preselection cuts are applied (dashed line) and then after full Punzi-Net selection (solid line). The efficiency appears to follow the same general trends as in the vanilla case (fig. 2.20), however the bump at low mass does appear to be greatly extended. It is not clear at the moment what causes this. The signal efficiency in the high mass region near 9GeV/c^2 does show a large drop for α_D=0.5 in comparison to the others. This is due to the much wider signal peak in this case which results in a large portion of the events reside above the 9GeV/c^2 limit on M_{recoil}.

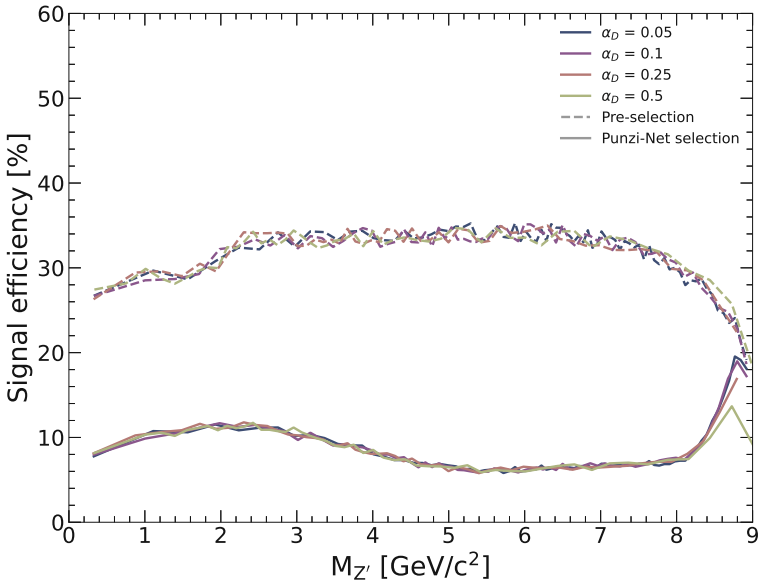

Figure 2.21: The signal efficiency after preselections (dashed) and full Punzi-Net selection (solid) for signal hypotheses with varied couplings to dark matter (α_D)

2.5 Signal modelling

For both the training of the Punzi-net (section 2) and definition of the templates used in the binned maximum-likelihood fit (section 3), a binning scheme in the M^2_{recoil} variable is required. In the latter case particularly, these bins must be chosen such that their width is not so great as to obfuscate the signal structure amongst background processes. What's more, the resolution of the detector varies with M^2_{recoil}, and thus the width of a signal peak and thereby the optimal bin width will vary across the range. To define these bins, unbinned maximum-likelihood fits are conducted to the M^2_{recoil} distribution of MC signal samples spanning the mass range of interest.

The data were fitted with a Crystal Ball function, which consists of a Gaussian and power-law tail and is defined by four fit parameters: the mean (μ), the width of the CB (σ) and the other CB parameters (n, α) which describe the power-law and transition point from Gaussian to power-law tail. The resulting distribution of Crystal Ball widths, σ_{CB}, is shown against M^2_{recoil} in figure 2.22 below. A polynomial is fitted to this, from which consecutive bins of width $\sim 1\sigma$ can be defined spanning the full range. This results in 702 bins that span the range [-2, 81]GeV^2/c^4. These bins are utilised in training of the Punzi-Net (section 2.3) and the signal extraction fit (section 3.1).

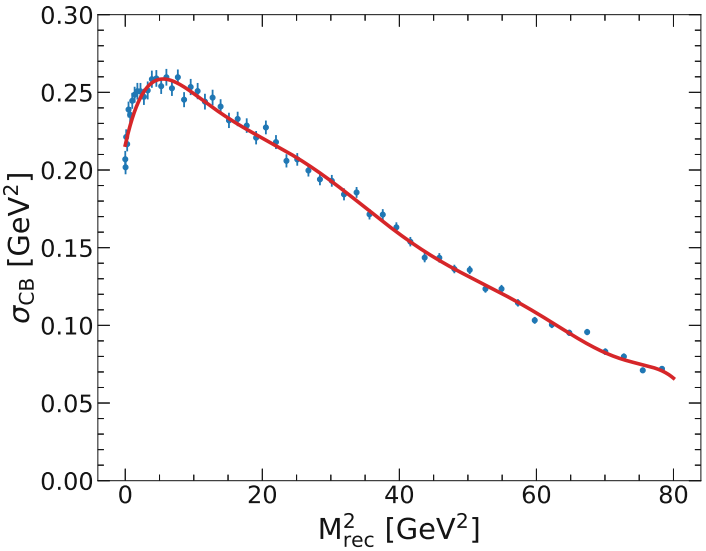

Figure 2.22: The crystal ball width as a function of M^2_{recoil}, resulting from fits to Z' signals with mass hypothesis covering the range of interest. An 8^{th}-order polynominal (shown in red) is fitted to the distribution for definition of consecutive bin widths.

3. Control channel checks

3.1 Control channel checks

In order to validate the event selection and background rejection methods, and to define systematic uncertainties associated with the search for an invisibly decaying Z′ boson, one can employ control channels with similar dynamics to that of the signal process but some distinct criteria that negate any risk of unblinding the analysis. Due to the similar nature of the analysis flow, it is assumed that the systematics defined herein can also be applied in the search for a muonphilic scalar boson too. Of course only the values defined in the mass range below $0.2\text{GeV}/c^2$ are relevant in this case.

The MC simulated events in each of these channels can be compared with real data and from this one can define systematic uncertainties associated with aspects of the main analysis itself. The following physics processes are utilised:

- $e^+e^- \to \mu^+\mu^-\gamma$: This control channel retains all of the same reconstruction and preselection procedure as in the nominal analysis (described in 2.2), however, the *photon veto* is now inverted. This means that events are now required to contain a reconstructed photon with energy of at least 1 GeV deposited in the ECL. In addition, the selection criteria concerning the rest-of-event energy is modified to $ROE_E - E_\gamma < 0.4\text{GeV}$. The rest-of-event energy is the energy attributed to any other clusters within the event not attributed to either of the muons or the photon. This channel is used to check the initial data-to-MC agreement, resolution modelling and background rejection efficiency. It is also used to study event trigger efficiencies and define correction factors for this, as discussed in section 2.

- $e^+e^- \to e^\pm\mu^\mp$: Again this control channel utilises the same event selection criteria but with the electron particleID likelihood >0.5 for one of the tracks, instead of being classified as a muon. This channel is used to check the initial data-to-MC agreement and background rejection efficiency. It is also used to validate event trigger efficiencies corrections, as discussed in section 2.

- $e^+e^- \to e^+e^-$: In this control channel, both charged tracks are required to be identified as an electron. There are further criteria applied to reduce final state radiation, the details are provided in section 3.1.4. This channel is used to study the efficiency and data-to-MC agreement of the photon veto criteria and define required corrections to MC.

This section studies the data-to-MC agreement found in key variables for each of these channels individually. After this, various studies are conducted that utilise these channels to define and quantify systematic uncertainties which can then be propogated as nuisance parameters in the binned maximum likelihood fit (discussed later in section 3)

3.1.1 $e^+e^- \to \mu^+\mu^-\gamma$

The distribution of M^2_{recoil} observed in data and MC is shown in figure 3.1 below, both before the application of the Punzi-Net (left), and after (right). The lower portion of each plot shows the ratio of the number of data and MC events found in each of the bins. Good agreement is observed in the lower squared recoil mass region, however a notable upward trend in the ratio is present from approximately $40 GeV^2/c^4$. This is expected in the given control channel due to the fact that the $e^+e^- \to e^+e^-\mu^+\mu^-$ MC sample is not generated with initial state radiation (ISR) and so does not contain the required photon to pass selection criteria. Due to this sample containing two electrons, it occupies the higher recoil mass region where this discrepancy is observed and, as a result, one cannot use this channel for validation of higher recoil mass events.

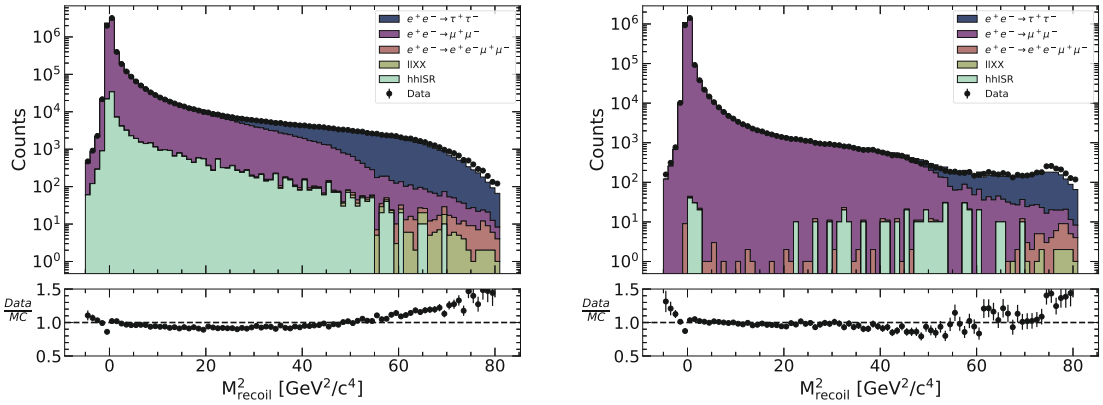

Figure 3.1: M^2_{recoil} distribution of data and MC events in the $\mu^+\mu^-\gamma$ control sample before (left) and after (right) the application of the Punzi-Net.

There is a notable mis-modelling present around 0 GeV^2/c^4. A closer look at this region is provided on the left of figure 3.2, where one can see there is a deficit of data events below 0 GeV^2/c^4, and an excess above this. This would imply that in MC the number of events with no recoiling system is overestimated in comparison to data, likely an effect of detector resolution modelling. This peaking structure that is thereby formed in data could impact the fitting procedure and so weights are defined using the $\mu^+\mu^-\gamma$ control sample to flatten this agreement. These are defined in 14 bins spanning the range $-1.5 < M^2_{recoil} < 2.0$ GeV^2/c^4. The resulting data/MC agreement is shown on the right of figure 3.2, and the weights used are shown in table 3.1. This effect is also seen in the $e^+e^- \to e^\pm\mu^\mp$ control sample, and the corrections defined here are thus validated via

this independent channel.

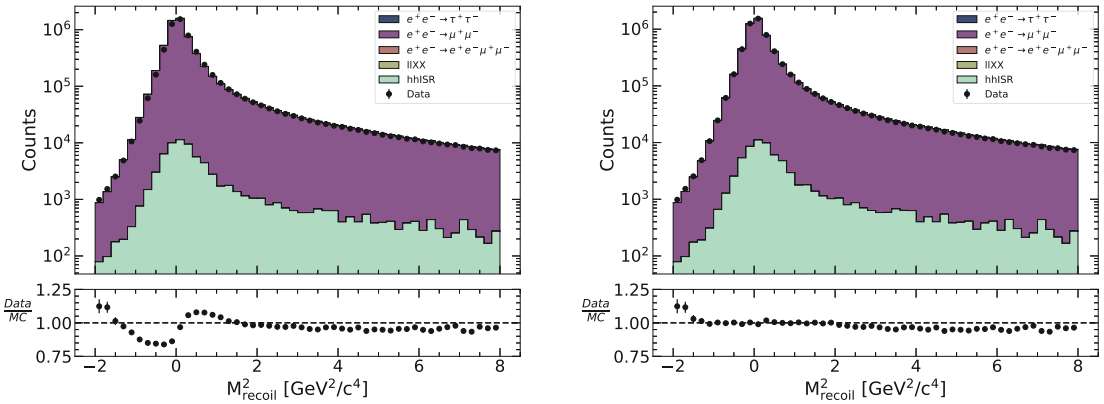

Figure 3.2: M^2_{recoil} distribution of data and MC events in the low recoil mass region of the $\mu^+\mu^-\gamma$ control sample. Left: before reweighting of MC distribution to reflect data. Right: after the reweighting of MC events.

Figure 3.3 shows the data and MC distributions of the 8 input variables of the Punzinet, before its application. In general the agreement seen in these is good, with flat data-to-MC ratios. There are slight deviations in regions that contain the hhISR MC sample ('h' here stands for hadron and this sample contains the likes of $\pi^+\pi^-\gamma$ events), which would appear to be mildly overestimated in the MC. While there is no exact understanding of the cause for this, it is not deemed to be problematic for the analysis as this MC sample is not observed after nominal signal preselection criteria. Furthermore even in the context of this control sample this MC category provides only a very small contribution and so this has a negligible effect on the calculations of systematic. Additionally there is a slight excess of data in the high end of the $A_{\mu\mu}$ variable, which coincides with events at the high end of the squared recoil mass distribution and so is again a result of the missing ISR in the $e^+e^- \to e^+e^-\mu^+\mu^-$ MC sample.

bin [GeV^2/c^4]	correction
(-1.50, -1.25)	0.97 ± 0.020
(-1.25, -1.00)	0.94 ± 0.012
(-1.00, -0.75)	0.87 ± 0.006
(-0.75, -0.50)	0.85 ± 0.004
(-0.50, -0.25)	0.84 ± 0.002
(-0.25, 0.00)	0.85 ± 0.001
(0.00, 0.25)	0.98 ± 0.001
(0.25, 0.50)	1.07 ± 0.002
(0.50, 0.75)	1.08 ± 0.003
(0.75, 1.00)	1.06 ± 0.003
(1.00, 1.25)	1.04 ± 0.004
(1.25, 1.50)	1.00 ± 0.004
(1.50, 1.75)	1.00 ± 0.005
(1.75, 2.00)	1.00 ± 0.005

Table 3.1: Correction factors defined using the $\mu^+\mu^-\gamma$ sample

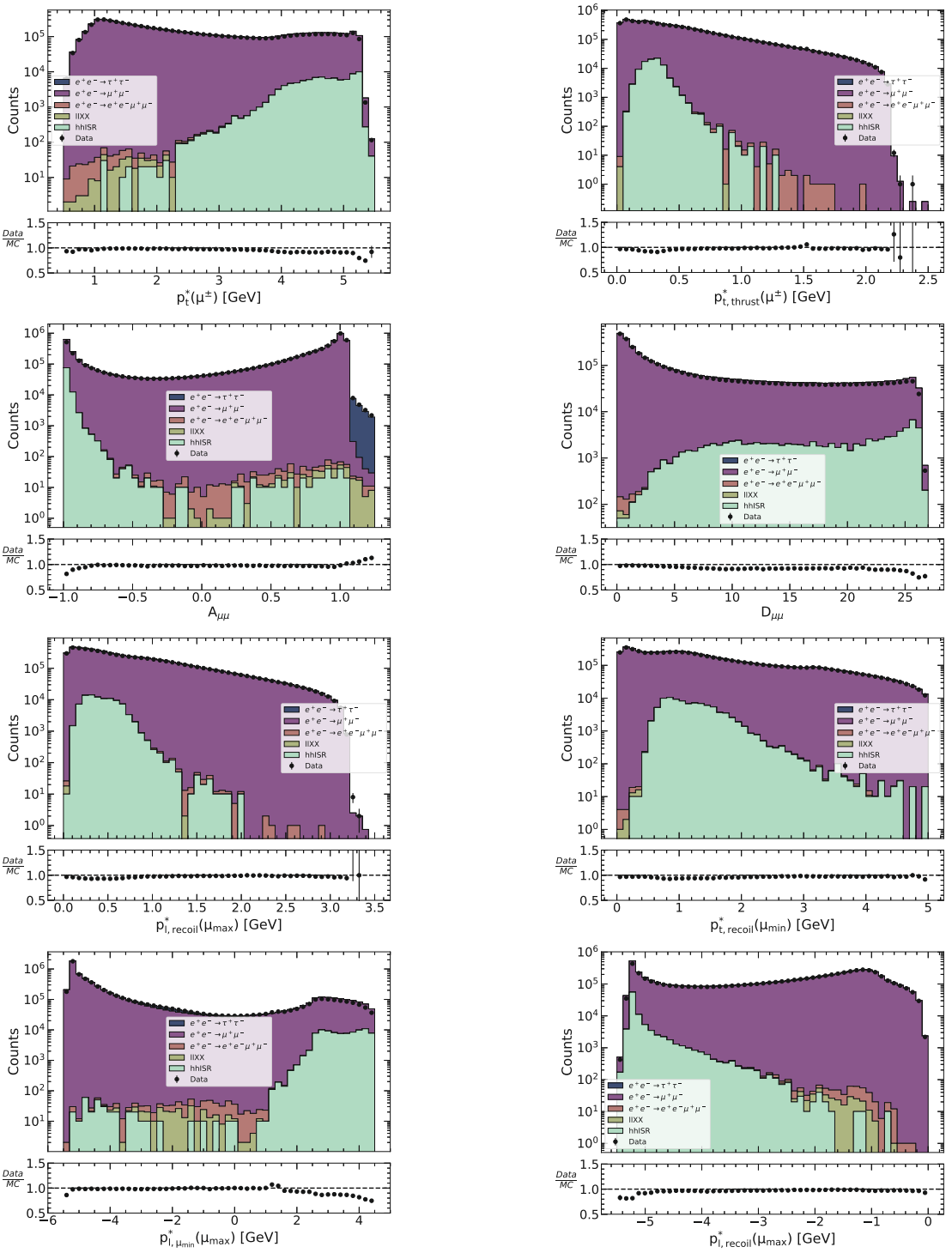

Figure 3.3: Data and MC distributions of Punzi-Net input variables with the $\mu^+\mu^-\gamma$ control sample

3.1.2 $e^+e^- \to e^\pm\mu^\mp$

The distribution of M^2_{recoil} observed in data and MC is shown in figure 3.4 below, both before the application of the Punzi-net (left), and after (right). Here, the low recoil mass corrections (table 3.1) defined using the $\mu^+\mu^-\gamma$ control sample are applied showing good agreement at low mass. A closer look at the effect of this reweighting is shown in figure 3.5, with the unweighted distribution on the left and the reweighted on the right. One can see that these weights defined using the $\mu^+\mu^-\gamma$ control sample appear to transfer well to this independent sample and accurately flatten the data-to-MC ratio around 0 GeV^2/c^4.

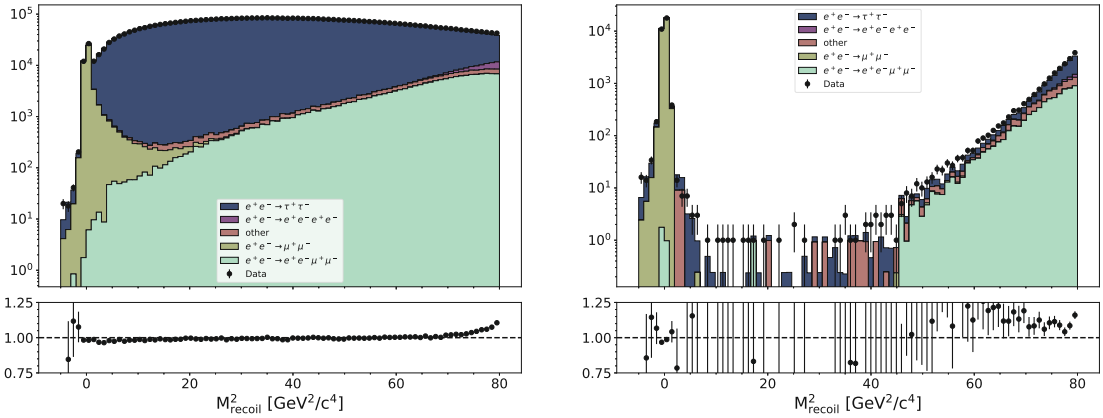

Figure 3.4: M^2_{recoil} distribution of data and MC events in the $e^+e^- \to e^\pm\mu^\mp$ control sample before (left) and after (right) the application of the Punzi-Net.

One can see in the left of fig. 3.4 that the $e^+e^- \to e^\pm\mu^\mp$ control sample is almost entirely dominated by $e^+e^- \to \tau^+\tau^-$ events, where one of the tau leptons decays to an electron and the other to a muon. There is a small peaking contribution of $\mu^+\mu^-$ events where one of the muons has been misidentified as an electron, around 0 GeV^2/c^4, and there is a notable contribution of $e^+e^-\mu^+\mu^-$ events in the high recoil mass region. As observed also in the nominal $\mu^+\mu^-$ selected MC, the $e^+e^- \to \tau^+\tau^-$ events are largely removed across the full range of M^2_{recoil} by the application of the Punzi-Net. This leaves just a peak of $\mu^+\mu^-$ events around 0 GeV^2/c^4, and a small contribution of $e^+e^-\mu^+\mu^-$ events in higher recoil. Data-to-MC agreement appears to remain consistent with the application of the neural network, without affecting data or MC differently.

Figure 3.6 shows the data and MC distributions of the 8 input variables of the Punzi-Net, before its application. In general the agreement between data and MC for each of these looks good. The agreements show flat distributions with a few regions of slight disagreement, typically in regions with higher statistical uncertainty.

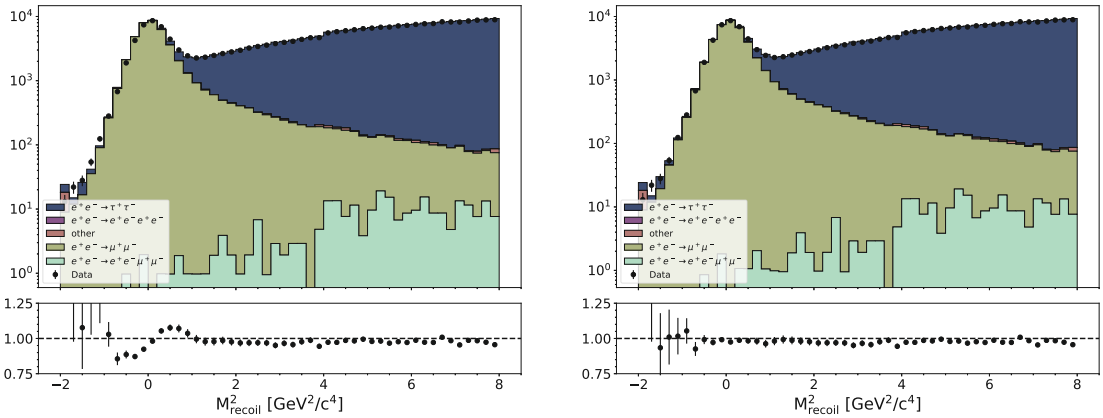

Figure 3.5: M^2_{recoil} distribution of data and MC events in the low recoil mass region of the $e^+e^- \to e^\pm\mu^\mp$ control sample. Left: before reweighting of MC distribution to reflect data. Right: after applying the weights defined in table 3.1 to MC events.

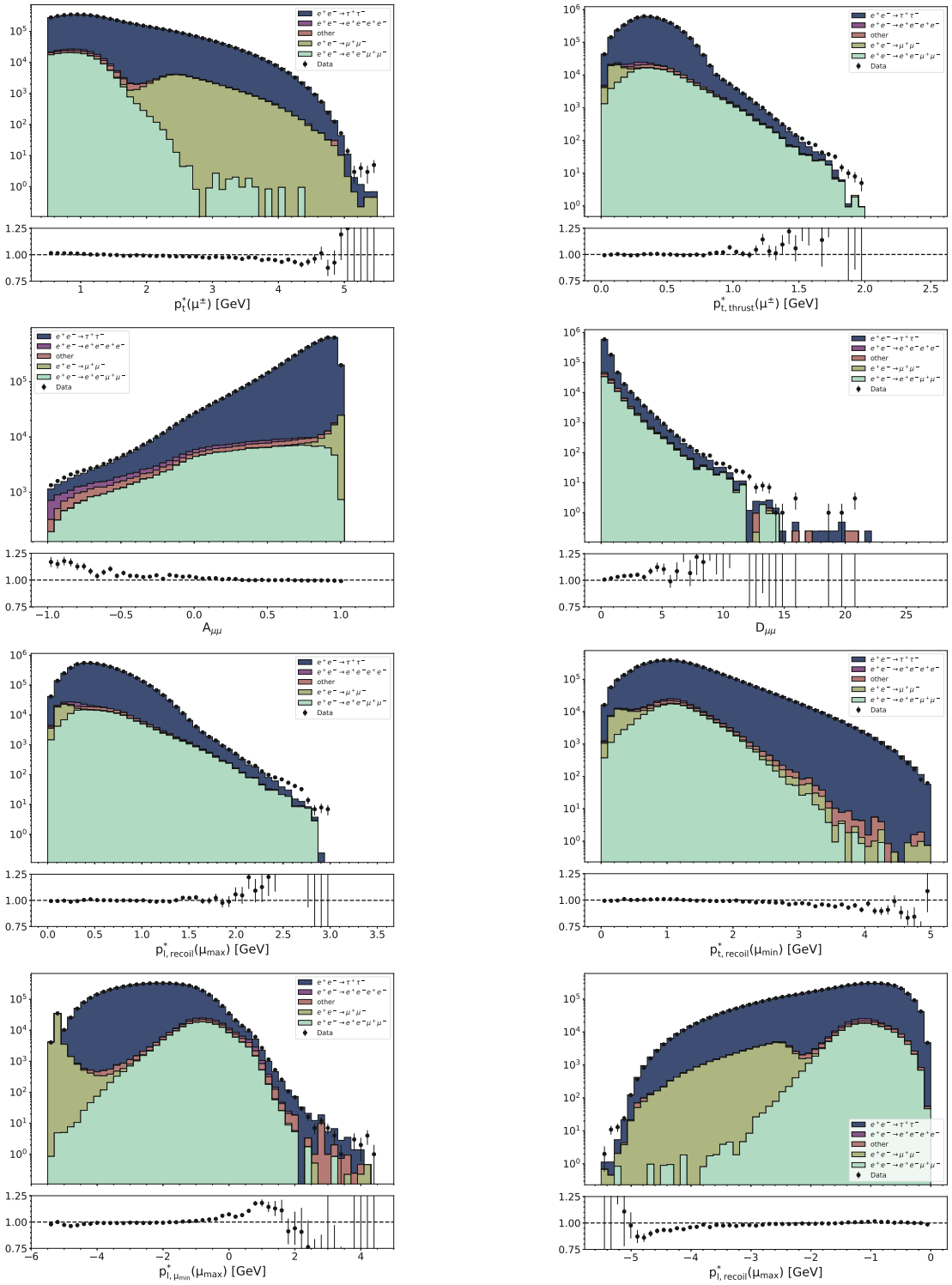

Figure 3.6: Data and MC distributions of Punzi-Net input variables with the $e^+e^- \to e^\pm\mu^\mp$ control sample

3.1.3 Validation of muonphilic dark scalar selection

Due to the use of a separate BDT for background rejection in the search for a muonphilic scalar particle, the data-to-MC agreement found in both the $e^+e^- \to e^\pm\mu^\mp$ and $e^+e^- \to \mu^+\mu^-$ after its application are investigated. The data and MC distributions of the squared recoil mass and recoil polar angle are shown for each of these in figures 3.7 and 3.8, where the BDT has been applied. The low recoil mass corrections (table 3.1) defined using the $\mu^+\mu^-\gamma$ control sample are again applied. Both control channels show good agreement, with consistent data-to-MC ratios across the respective distributions. There is a mild excess of data present in the $e^+e^- \to \mu^+\mu^-\gamma$. This is not deemed to be problematic, as the agreement appears flat across both variables shown and so they are not being shaped by the application of the BDT.

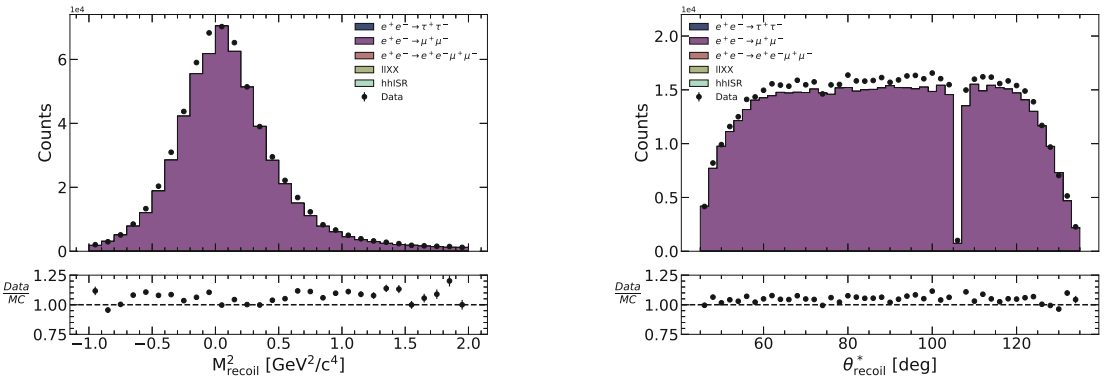

Figure 3.7: The data and MC distributions of the $e^+e^- \to \mu^+\mu^-\gamma$ control channel after the application of the BDT. The left hand plot shows the squared recoil mass and the right shows the polar angle of the recoil vector.

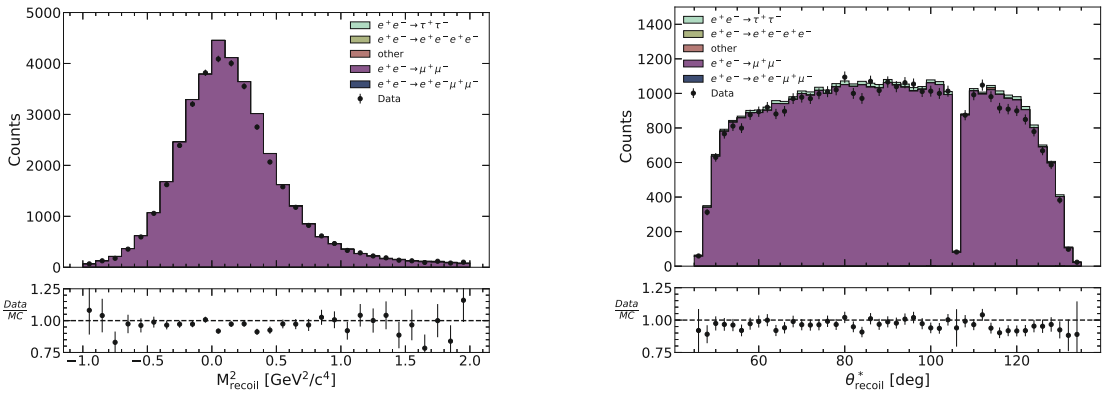

Figure 3.8: The data and MC distributions of the $e^+e^- \to e^\pm\mu^\mp$ control channel after the application of the BDT. The left hand plot shows the squared recoil mass and the right shows the polar angle of the recoil vector.

3.1.4 Photon veto inefficiency

The photon veto selection criteria presents a unique risk in which mismodelling could possibly produce a signal-like distribution in data. The previous iteration of the search for an invisibly decaying Z′ was one of the first to implement such a selection during analysis of a large data set at Belle II. Upon unblinding, it was found that the efficiency of the veto was substantially overestimated in the region $M^2_{\text{recoil}} < 2\ \text{GeV}^2/\text{c}^4$. This then produced a large peak in data, with clear structures appearing in the distribution of the recoil polar angle that indicated it was background and not signal. Investigation of this then led to the implementation of corrections and systematics associated with the selection.

In this current iteration of the study, the photon veto is now expanded to then further include a secondary KLM-based selection on top of the initial ECL-based photon veto. To study the data and MC efficiency of these, and define associated systmatics, neither the $e^{\pm}\mu^{\mp}$ nor $\mu^+\mu^-\gamma$ control channel will suffice. The latter of these of course being due to the fact that a photon is explicitly required, without which one would simply unblind the whole study. Therefore a third $e^+e^-\gamma$ control channel is included, which avoids any risk of unblinding as under the L_μ-L_τ framework the Z′ is not expected to couple to electrons.

The control sample is defined using all the same preselection cuts as outlined in section 2.2, of course with replacement of the muonID criteria with electronID > 0.5, for both tracks in the event. In contrast to the $\mu^+\mu^-$ final state, electrons are expected to commonly produce FSR photons, typically by bremsstrahlung. It was found that the modelling of such events were not well modelled in MC and so are suppressed with extra selection criteria so as to maximise likeness to the nominal $\mu^+\mu^-$ events. The following cuts are applied:

- For events with $p_{\text{recoil}} < 4$ GeV/c, the opening angle between the recoil vector and either electron in the event is required to be greater than 60°.

- For events with $p_{\text{recoil}} \geq 4$ GeV/c, the opening angle between the recoil vector and either electron in the event is required to be greater than 10°.

The lml1 L1 trigger bit is used, as all triggers used in the nominal selection criteria include a bhabha veto which explicitly kills the events needed here. The lml1 line requires at least one single ECL cluster of energy 2 GeV in the event. This trigger line does become prescaled (i.e. sampled) 50%. The resulting distribution of events in MC is shown in fig. 3.9, with the recoil momentum in the left hand plot and the recoil polar angle in the right hand plot.

One can see clear peaks in the in polar angle distribution, centred around the far ends of the full detector, the joins between the forward/backward end caps and the barrel region ($\approx 30°/130°$) and at the 90-degree gap. These are all regions in which gaps exist in the ECL instrumentation and so these events are those in which the photon has evaded the ECL-based veto.

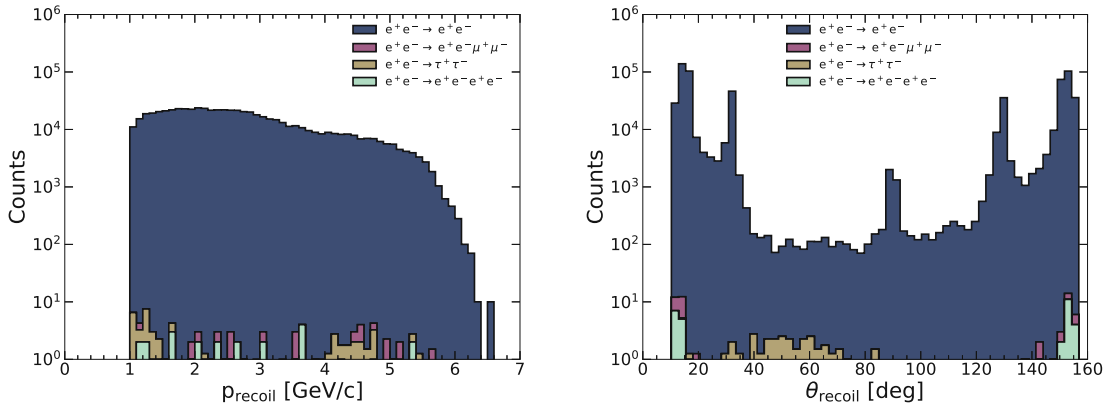

Figure 3.9: Recoil momentum (left) and polar angle (right) of bhabha MC events after the anti-FSR selection criteria is applied.

Figure 3.10 shows a 2D distribution, the y-axis of which is defined as the difference between the reconstructed recoil energy and the MC generated energy of the unreconstructed photon (referred to as the *MC truth value*). The x-axis instead shows the difference between the recoil polar angle and the photon polar angle truth value. One can see that these events are very much constrained to the lower left of the distribution, indicating that the selected sample of MC events contains predominantly radiative bhabha events with a single photon that has not been reconstructed. These events can be studied to define the inefficiency of the photon veto, both the initial ECL-based and then with the additional KLM-based photon veto. Comparisons can be made with the a data sample with the same selection applied and systematic uncertainty on the full photon veto selection is defined.

The photon veto *inefficiency* can be defined by the ratio of events surviving the application of the photon veto to those present before. This is calculated first for the ECL-based photon veto alone, and then for the additional KLM-based photon veto. Furthermore in each case the calculations are conducted in three recoil momentum regimes: low ($p_{recoil} < 2$GeV/c), mid ($4 \leq p_{recoil} < 4$GeV/c) and high ($p_{recoil} \geq 2$GeV/c). The data-to-MC ratios found can then be used to define correction factors to be applied to the nominal $\mu^+\mu^-$ sample. Data and MC are binned in 10 irregular bins that cover the recoil polar angle distribution and are selected to properly capture the regions of the ECL that lack instrumentation.

Figure 3.11 shows the final data-to-MC ratio of the full measured photon veto inefficiency after both the ECL and KLM veto criteria are applied. The three distributions show the calculated inefficiency ratios in the three momentum regimes. The polar angle regions of particular interest where ECL coverage is lacking are highlighted with dashed lines. One can see that a clear problematic region exists at the joining of the forward endcap and barrel region of the detector. It should be noted that in the previous iteration of this study, the agreement between data and MC was considerably worse however

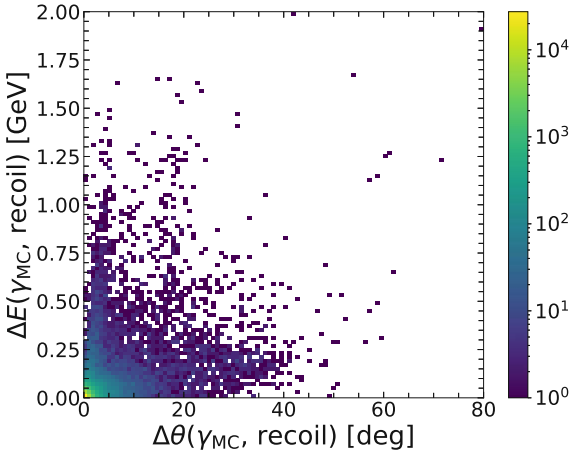

Figure 3.10: 2D distribution showing the difference between MC generated truth value of the leading photon and the reconstructed recoil. The y-axis shows the difference in energy, the x-axis shows the difference in polar angle.

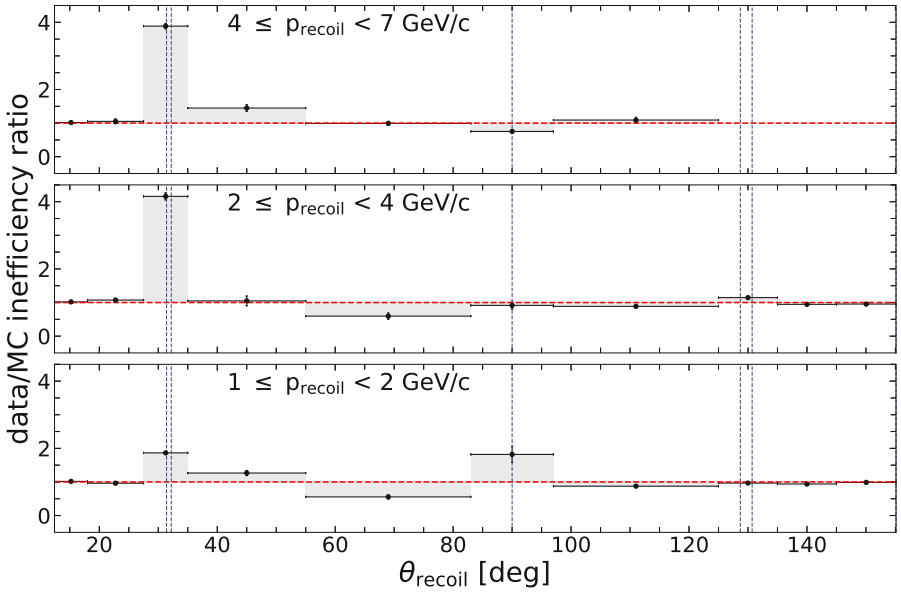

Figure 3.11: The ratio of measured photon veto inefficiency in data and MC in bins across the polar recoil angle. Distributions are shown with values calculated in three recoil momentum ranges ranging across 1 to 7 GeV/c. Regions in which ECL instrumentation is lacking are highlighted with dashed lines.

the newer simulated MC that has been produced appears to much better reflect the real response of the ECL observed in data.

3.2 Systematic effects

Having now discussed the overall agreement observed in each of the control channels, one can investigate the accuracy of the simulation with regards to observed data in various aspects, and thereby quantify uncertainty on the expected background distributions. Due to the nature of this study, where one searches for a peaking signal over this background, underestimation of uncertainty could lead to some unexpected structure in data being interpreted as a signal.

The statistical analysis is carried out on the two dimensional M^2_{recoil} vs θ^*_{recoil} distribution, and so modelling of each variable is carried out separately and percentage uncertainty of data-to-MC agreement is defined. This is a more general method that includes any expected systematic effects due to the trigger, particle identification and track reconstruction. The following systematic sources are studied here,

- *Detector resolution modelling*: The resolution modelling will effect the width of any signal peak in M^2_{recoil}. Agreement between data and MC is studied using the $e^+e^- \rightarrow \mu^+\mu^-\gamma$ control channel and a percentage uncertainty on this width is defined.

- *Punzi-Net efficiency*: The Punzi-Net efficiency refers to the rate of event rejection. This is studied in both data and MC, and the data-to-MC ratio of these efficiencies is calculated.

- *Signal efficiency*: The signal efficiency uncertainty is investigated with consideration of multiple sources. The final value considers the Punzi-Net efficiency uncertainty and the initial data-to-MC agreement of the control channels. The agreement in signal efficiency between run dependent and run independent signal samples is also included.

- θ^*_{recoil} *modelling*: The data-to-MC agreement in the recoil polar angle variable is studied with the $e^+e^- \rightarrow \mu^+\mu^-\gamma$ and $e^+e^- \rightarrow e^\pm\mu^\mp$ control channels. The data-to-MC ratios in each is combined to provide a percentage uncertainty on the accuracy of the θ^*_{recoil} axis in the 2D templates that are fitted to data.

- M^2_{recoil} *modelling*: Similar to the method described for the θ^*_{recoil} uncertainty. This time data-to-MC ratios in the control channels are combined to provide a percentage uncertainty on the M^2_{recoil} axis of the 2D templates fitted to data.

- *Photon veto*: The uncertainty associated with the photon veto inefficiency corrections are defined using the $e^+e^- \rightarrow e^+e^-\gamma$ control channel. Values are derived separately for the θ^*_{recoil} and M^2_{recoil} axes, with consideration of how the corrections might effect template shape uncertainty in the respective axis.

The uncertainties are then propagated to the binned maximum-likelihood fit in the form of nuisance parameters. This is discussed in section 4.2.

Detector Resolution

Due to the peaking structure one would expect any signal decay to have in the squared recoil mass distribution (be it an L_μ-L_τ Z′ or a muonphilic scalar), it is important to validate the modelling of the detector's resolution in this variable. The $e^+e^- \to \mu^+\mu^-\gamma$ control sample provides an excellent opportunity for this due to the similar event kinematics. Specifically in this case, it is important to use only events where these kinematics are well constrained. This is done by including an additional requirement that the sum of the energies of the two muons and the most energetic photon in the event lie near to the centre-of-mass energy ($10.38 < E(\mu^+\mu^-\gamma) < 10.78$ GeV), thereby ensuring that only events with a single hard radiated photon contribute.

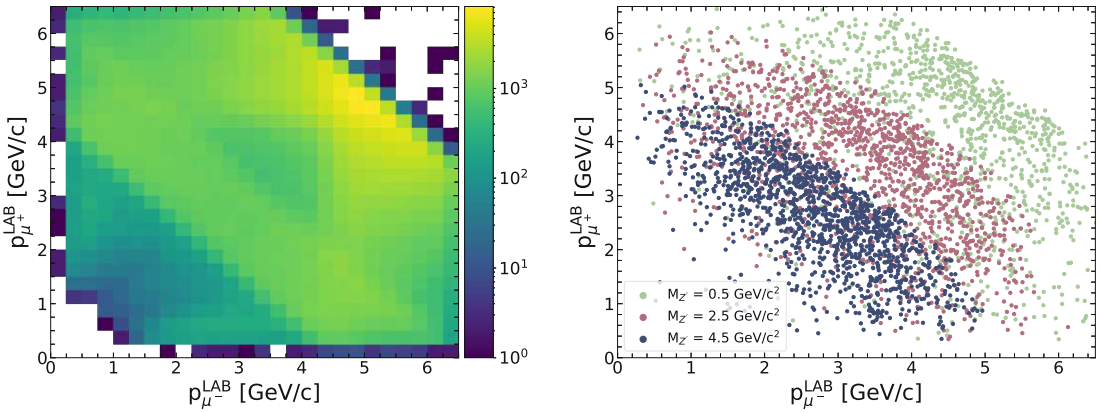

Figure 3.12: 2D distribution of the muon momenta. The left plot shows the distribution of the $\mu^+\mu^-\gamma$ events in MC, the right hand plot instead the same distribution for a set of three Z′ signal masses ($M_{Z'}$ = 0.5, 2.5 and 4.5 GeV/c^2)

One can then seek to approximate the kinematics of a signal event with some specific mass of Z′ by re-weighting the muon momenta of $e^+e^- \to \mu^+\mu^-\gamma$ events based on the distribution of the muon momenta in the signal events. Figure 3.12 shows the 2D distribution of the muon momenta in the LAB frame for both the $\mu^+\mu^-\gamma$ data sample (left) and a set of MC-simulated Z′ signals at three mass points, 0.5, 2.5 and 4.5 GeV/c^2. So for each mass point studied, the left hand distribution is weighted to reflect the right hand distribution.

As one might expect, it can be seen that as the mass increases the two muons tend to have lower respective momenta. This study is limited to only masses below approximately 6 GeV/c^2. With masses higher than this, the associated muon momenta distribution occupies the far bottom left region where $\mu^+\mu^-\gamma$ event statistics are limited.

For each mass hypothesis, the M^2_{recoil} distribution of the re-weighted $\mu^+\mu^-\gamma$ data and MC samples are then fit with a Crystall Ball function, the widths of which are then compared by taking the ratio as a means of quantifying the resolution modelling. This process is done both before and after the application of the Punzi-Net, and the respective results of each are shown in tables 3.2 and 3.3, and in figure 3.13. The results

show relatively flat agreement across the mass range, with the agreement actually being improved slightly by the Punzi-Net. This is due to the slight shaping of muon momenta variables that is induced by the application of the network. A flat uncertainty of 4% on the resolution modelling is taken from this.

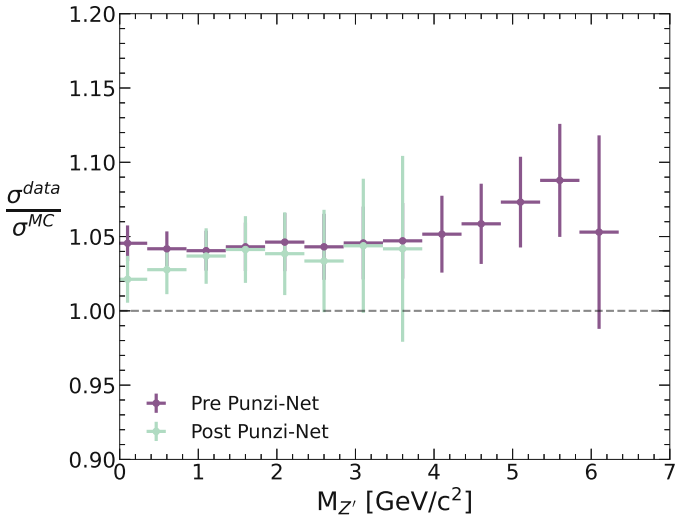

Figure 3.13: The ratio of measured widths in data and MC, both before (purple) and after (green) the application of the Punzi-Net. The number of data points in the later case is reduced due to the lack of statistics at these masses.

Table 3.2: : The data and MC widths corresponding to various Z′ signal masses, along with the ratio.

$M_{Z'}$ [GeV/c^2]	σ^{data} [GeV/c^2]	σ^{MC} [GeV/c^2]	$\sigma^{data}/\sigma^{MC}$
0.1	0.190 ± 0.002	0.181 ± 0.001	1.05 ± 0.01
0.5	0.203 ± 0.002	0.194 ± 0.002	1.04 ± 0.01
1.0	0.225 ± 0.002	0.216 ± 0.002	1.04 ± 0.02
1.5	0.269 ± 0.003	0.258 ± 0.003	1.04 ± 0.02
2.0	0.305 ± 0.004	0.292 ± 0.004	1.05 ± 0.02
2.5	0.329 ± 0.005	0.314 ± 0.005	1.05 ± 0.02
3.0	0.342 ± 0.006	0.328 ± 0.005	1.04 ± 0.03
3.5	0.347 ± 0.006	0.332 ± 0.006	1.05 ± 0.03
4.0	0.349 ± 0.006	0.332 ± 0.006	1.05 ± 0.03
4.5	0.352 ± 0.006	0.334 ± 0.006	1.06 ± 0.03
5.0	0.357 ± 0.007	0.334 ± 0.006	1.07 ± 0.04
5.5	0.358 ± 0.008	0.330 ± 0.008	1.07 ± 0.06
6.0	0.338 ± 0.013	0.317 ± 0.011	1.02 ± 0.21

Table 3.3: : The data and MC weighted widths corresponding to various Z′ signal masses, along with the ratio, after Punzi-net selection.

$M_{Z'}$ [GeV/c^2]	σ^{data} [GeV/c^2]	σ^{MC} [GeV/c^2]	$\sigma^{data}/\sigma^{MC}$
0.1	0.235 ± 0.003	0.230 ± 0.002	1.02 ± 0.02
0.5	0.238 ± 0.003	0.232 ± 0.003	1.03 ± 0.02
1.0	0.266 ± 0.003	0.257 ± 0.003	1.04 ± 0.02
1.5	0.294 ± 0.004	0.283 ± 0.004	1.04 ± 0.02
2.0	0.327 ± 0.006	0.315 ± 0.006	1.04 ± 0.03
2.5	0.354 ± 0.008	0.341 ± 0.008	1.04 ± 0.03
3.0	0.371 ± 0.011	0.356 ± 0.010	1.04 ± 0.04
3.5	0.376 ± 0.014	0.362 ± 0.014	1.04 ± 0.06
4.0	0.386 ± 0.023	0.375 ± 0.023	1.03 ± 0.09
4.5	0.399 ± 0.162	0.398 ± 0.052	1.01 ± 0.62
5.0	0.399 ± 0.197	0.392 ± 0.164	1.02 ± 0.58
5.5	-	-	-
6.0	-	-	-

Punzi-Net efficiency

The back rejection power of the Punzi-Net is studied in both the $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples, and the discrepancy between data and MC is measured. This is quantified as:

$$\sigma_{NN} = \frac{\epsilon_{NN}^{data}}{\epsilon_{NN}^{MC}} \tag{3.1}$$

Where $\epsilon_{NN}^{data(MC)}$ is defined as the ratio of events after and before the application of the Punzi-Net in data (MC). The ratio of this value calculated in data and MC then gives a measure of the expected difference in background rejection power between the simulated MC and real data. This is calculated across M^2_{recoil} in bins of width 1 GeV^2/c^4, the results of which are shown in figure 3.14. Note that the $e^+e^- \to \mu^+\mu^-\gamma$ control channel is only used up to 40 GeV^2/c^4 as the sample is mis-modelled in MC above this due to the already discussed lack of ISR in some MC samples. The $e^+e^- \to e^\pm\mu^\mp$ sample on provides results above 60 GeV^2/c^4 due to the complete removal of nearly all events below this by the Punzi-Net.

The results from the two samples do compliment each other's coverage and, as there is reasonable stability across the squared recoil mass range, a 0^{th} order polynomial is fit to the points. This fit finds a 7% higher efficiency in MC than in data, meaning the Punzi-net is expected to reject 7% less background in data than in MC. This effect appears quite stable across M^2_{recoil} aside for in the region below approximately 5 GeV^2/c^4 where the Punzi-Net performs similarly on data and MC.

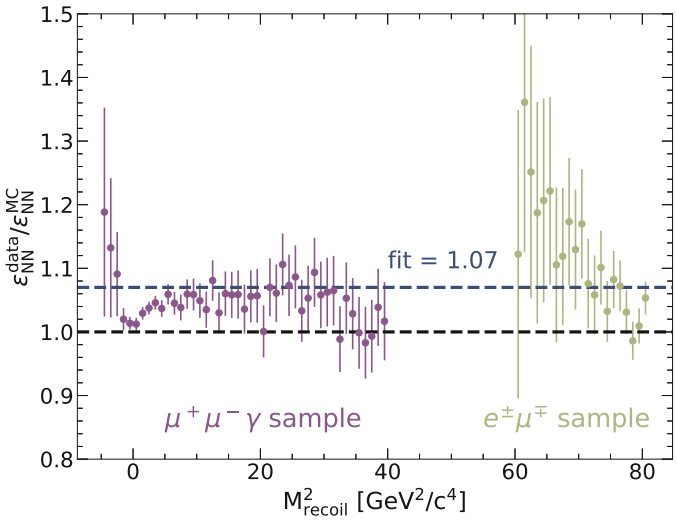

Figure 3.14: The ratio of Punzi-Net efficiency found in data to that found in MC. The values for each control sample on cover regions in which the data is modelled well both before and after the application of the network. A line is fitted at 1.07.

Signal efficiency

As discussed in section 2.4.2, run-dependent MC signal samples are known to have variation in the final signal efficiency depending on the experiment and run period. A set of signal samples spanning some example experiments/runs (shown in fig. 2.20) across the full data-taking period are studied and the total signal efficiency from these, weighted by the luminosity of the samples is shown in vlue in figure 3.15. The run-independent MC (which is used in the final signal extraction fit) is shown in light brown. The agreement between the two appears to be good, nevertheless the root mean square error is calculated with an average of 0.3% found. Based on this, a conservative value of 0.5% systematic uncertainty associated with the agreement between run dependent and run independent signal MC is propagated.

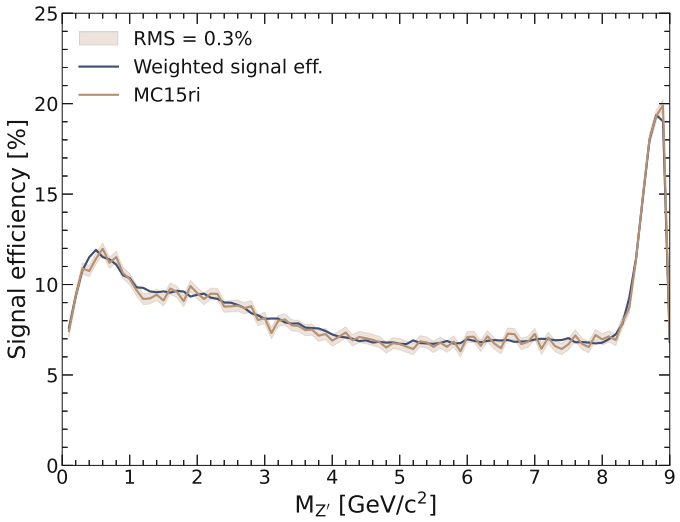

Figure 3.15: Signal efficiency across mass hypotheses after all selection criteria are applied. The amalgamated luminosity weighted run dependent MC sample is shown by a blue line, while the run independent sample is shown in brown. The calculated RMS error between the two is represented by the shaded region.

This is then combined with data-to-MC agreement ratios from the control channels and the Punzi-Net efficiency agreement to define a normalisation uncertainty on the signal templates. The data-to-MC ratio found before the application of the Punzi-Net in both the $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples is shown in figure 3.16. The agreement is in general good, with trends upwards at either end of the range. From this a conservative uncertainty of 4% is taken across the full range, which when combined in quadrature with the Punzi-Net efficiency uncertainty and the run dependent uncertainty discussed previously gives a total signal efficiency uncertainty of 8%. This is applied as a normalisation uncertainty on the signal template.

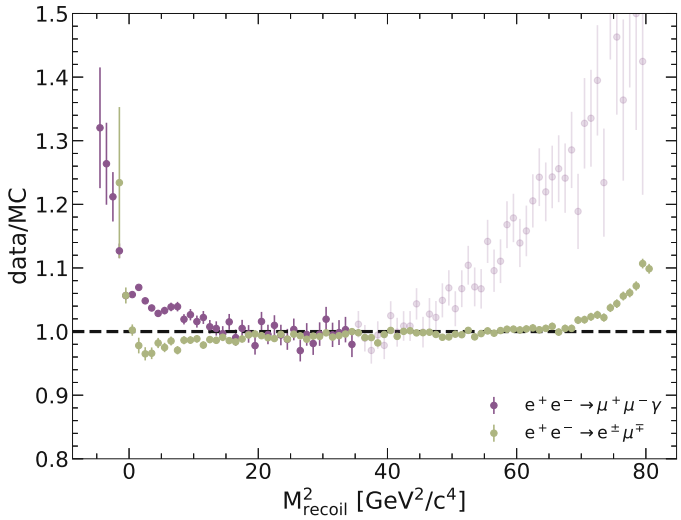

Figure 3.16: The data-to-MC ratio found before the application of the Punzi-Net in both the $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples. The former of these is faded out in the region above $36\text{GeV}^2/c^4$ where the agreement is worse due to the lack of ISR in MC samples.

θ^*_{recoil} agreement

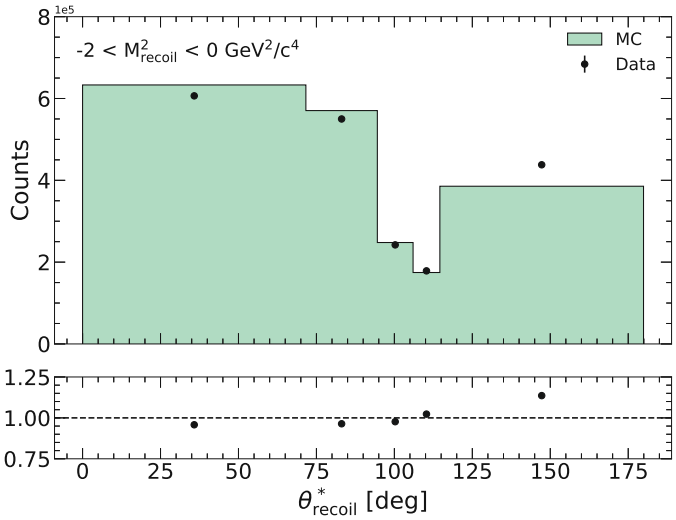

Figure 3.17: Data and MC distributions of θ^*_{recoil} for $e^+e^- \to \mu^+\mu^-\gamma$ events, where only events in the range $M^2_{\text{recoil}} < 0\ \text{GeV}^2/c^4$ are shown.

The agreement in modelling of the θ^*_{recoil} distribution between data and MC-generated

events is checked. This variable is strongly influenced by the simulation of the Belle II detector itself, as discussed previously (section 2.4.1), the detector construction produces specific structures in the MC-generated background distributions that could mimic signals if large discrepancy between data and MC is observed. This aspect is studied in both the $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples, and a systematic uncertainty is derived with consideration of both results.

'Events were split into bins of M^2_{recoil}, based on regions in which consistent data-to-MC agreement was observed. In each of these, the data and MC agreement in the θ^*_{recoil} variable was studied, with events binned as described in section 3.1. The difference between the largest and smallest data-to-MC ratio found across these bins is calculated. This value is then halved. For example when using the $e^+e^- \to \mu^+\mu^-\gamma$, in the first region ($M^2_{recoil} < 0$ GeV^2/c^4) the minimum and maximum data-to-MC ratio values found across the θ^*_{recoil} bins are 0.96 ($\theta^*_{recoil} \lesssim 75°$) and 1.14 ($115° \lesssim \theta^*_{recoil} \lesssim 180°$) respectively, so the calculated shape uncertainty is (1.14 - 0.96)/2 = 0.09. The data and MC θ^*_{recoil} distribution for this example is shown in figure 3.17.

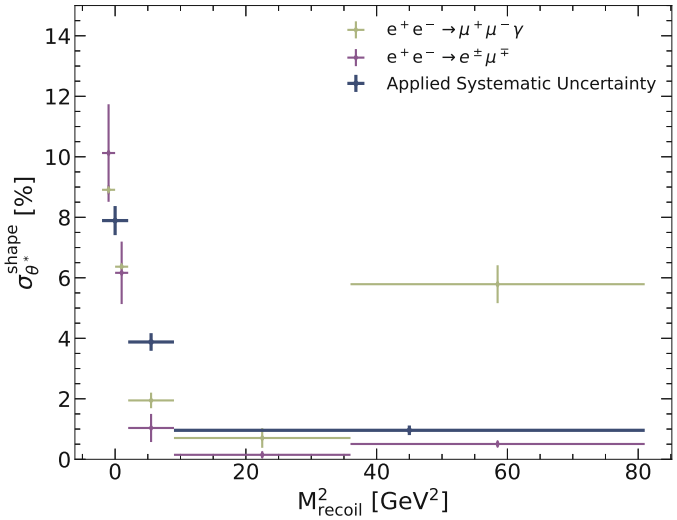

Figure 3.18: The calculated θ^*_{recoil} shape uncertainty in regions of the squared recoil mass. The $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples are shown in green and purple, and the total value that is taken as a systematic uncertainty is shown in blue.

Only bins which contain at least 500 events in both data and MC are included in calculation, so as to avoid effects from statistical uncertainty. The resulting values for both control channels are defined in table 3.4, along with the selected value of systematic uncertainty that is propagated to the maximum likelihood fit. The results are shown graphically in figure 3.18. The systematic uncertainty that is applied is shown in blue, where the value is chosen to reflect both control channels. Conservative values are chosen, for instance it was decided to take 4% uncertainty in the region $2 \leq M^2_{recoil} < 9$ GeV^2/c^4 which is above the calculated value in both control channels. This is done to provide

more smooth transition of values from the low mass region to higher. While a higher uncertainty is observed in the negative region, the value around 0 is taken as the average of the positive and negative sides.

Table 3.4: Template θ^*_{recoil} shape uncertainties applied in the maximum likelihood fit, defined in regions of M^2_{recoil}.

$M^2_{\text{rec.}}$ [GeV^2/c^4]	$\mu^+\mu^-\gamma$			$e^\pm\mu^\mp$			Sys.
	min	max	$\sigma^{\text{shape}}_{\theta^*}$	min	max	$\sigma^{\text{shape}}_{\theta^*}$	
$-2 < M^2_{\text{recoil}} < 0$	0.96	1.14	0.09	0.99	1.19	0.10	0.08
$0 \leq M^2_{\text{recoil}} < 2$	0.91	1.04	0.06	0.92	1.04	0.06	0.08
$2 \leq M^2_{\text{recoil}} < 9$	0.96	1.01	0.02	0.99	1.01	0.01	0.04
$9 \leq M^2_{\text{recoil}} < 36$	0.96	0.98	0.01	1.02	1.03	0.00	0.01
$36 \leq M^2_{\text{recoil}} < 81$	1.00	1.12	0.06	1.05	1.06	0.01	0.01

M^2_{recoil} agreement

A similar check is then conducted on the squared recoil mass distribution, this time the data-to-MC agreement is checked for individual Z′ signal mass hypotheses. For each of these, events in a $\pm 10\sigma$ M^2_{recoil} window around the given mass hypothesis are studied, where σ is the measured width of the signal peak[1]. Within these windows again the difference between the greatest and smallest data-to-MC ratio is found and the value is then halved to provide an associated uncertainty for the mass hypothesis. These are then averaged across three mass regions in which agreement is consistent. The resulting averaged uncertainties from with each control channel individually, and also the combined value, are shown in figure 3.19, and the values are summarised in table 3.5.

Table 3.5: The calculated M^2_{recoil} shape uncertainties in both $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples, with the final value that is taken as a systematic uncertainty.

	$\mu^+\mu^-\gamma$	$e^\pm\mu^\mp$	Sys. uncert.
$M_{Z'} < 1.4$ GeV/c^2	0.062 ± 0.004	0.065 ± 0.006	0.064 ± 0.004
$1.4 \leq M_{Z'} < 3$ GeV/c^2	0.043 ± 0.005	0.028 ± 0.004	0.036 ± 0.003
$3 \leq M_{Z'} < 9$ GeV/c^2	0.008 ± 0.002	0.022 ± 0.001	0.015 ± 0.001

[1] This definition of the squared recoil mass window is discussed fully in section 4.1.1

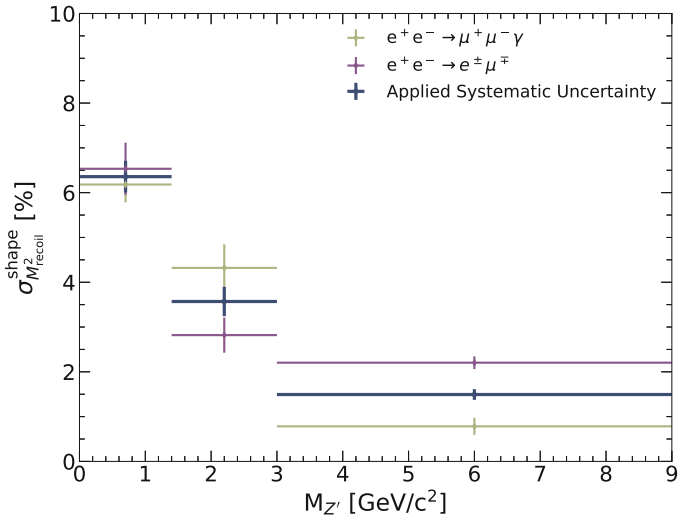

Figure 3.19: The calculated M^2_{recoil} shape uncertainty calculated in regions of $M_{Z'}$ in which the behaviour is consistent. The results are shown for the $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control samples in green and purple respectively, with the final value that is propagated as a systematic uncertainty then shown in blue.

Photon Veto

Systematic uncertainties associated with the photon veto inefficiency corrections, discussed in section 3.1.4, are defined with study of various possible factors that could play a role in any incorrect measurement of the corrections. These include:

- Contamination from physics processes other than radiative bhabha decays, i.e. $e^+e^- \to \tau^+\tau^-$ or $e^+e^- \to e^+e^-e^+e^-$, which can be seen to be present, albeit in small quantities in fig. 3.9. These could bias the calculation of corrections and so a systematic uncertainty is defined by subtracting their presence from MC, and the equal number of events from data.
- Punzi-Net selection can mildly effect the distributions of events in recoil mass and polar angle, and so calculations are conducted both before and after its application.
- The effect of the anti-FSR selection cuts (described in section 3.1.4) are investigated by mildly varying the values and calculating the corrections in these cases.

In each of these cases the systematic uncertainty is defined as the resulting variation of the background template with the given correction values. Similar to the previously discussed uncertainties found in the data-to-MC agreement of θ^*_{recoil} and M^2_{recoil} distributions, these are defined individually for each of the two respective axes. The resulting values are defined in table 3.6.

Table 3.6: The % shape uncertainty applied in the recoil theta and squared mass axes respectively.

M^2_{recoil} range	θ^*_{recoil} unc. [%]	M^2_{recoil} unc. [%]
$M^2_{recoil} < 2$ GeV^2	3	6
$2 \leq M^2_{recoil} < 9$ GeV^2	2	2
$9 \leq M^2_{recoil} < 81$ GeV^2	1	1

4. Statistical Analysis

4.1 Binned Maximum likelihood fit

Statistical interpretation of the observed data, after all selection criteria are applied, is done by way of a *binned maximum likelihood fit*. In this study the `pyhf` [68] python library is used for the task. This provides a method of checking compatibility between observed data and the set of Z' boson (and muonphilic scalar) mass hypotheses investigated.

For each of these one can define a statistical model, $f(x|\phi)$, which describes the probability of some observation, x, given the model parameters, ϕ, that constrain the model. The number of events, n, observed in the given bins, is defined in the model by the event rate of the measured samples (in this case these are just signal and background). The rate of events is subject to parameterised variations, which come in the form of free parameters η or constrained parameters, χ. The latter of these particularly pertains to the defined systematic uncertainties. These constraint terms can be described by *auxiliary measurements*, a, which together with the event counts define the full observation, $x = (n, a)$. The model parameters can be similarly deconstructed into the *parameter of interest*, ψ, and *nuisance parameters*, θ. The model can thus be defined as:

$$f(n, a|\eta, \chi) = \prod_{b}^{N} \text{Pois}(n_{cb}|\nu_{cb}(\eta, \chi)) \prod_{\chi \in \chi} c_{\chi}(a_{\chi}|\chi) \tag{4.1}$$

The first product term runs over all N bins, b, of the channel[1], c, and contains a Poisson distribution associated with the observation of n_{cb} events due to the sample rates ν_{cb} with the given unconstrained and constrained parameters η and χ. The latter product term pertains to observed auxiliary measurements, a_{χ} given the constrained parameters, χ. Furthermore, the event rates are defined as:

$$\nu_{cb}(\phi) = \sum_{s \in samples} \nu_{scb}(\eta, \chi) = \sum_{s \in samples} \left(\prod_{\kappa \in \kappa} \kappa_{scb}(\eta, \chi) \right) \left(\nu^{0}_{scb}(\eta, \chi) + \sum_{\Delta \in \mathbf{\Delta}} \Delta_{scb}(\eta, \chi) \right) \tag{4.2}$$

[1]Some implementations can utilise multiple channels, i.e. variable distributions, to constrain the parameter of interest. This is done in section 3, where the different q^2 distributions comprise the channels of the fit.

This describes how the nominal event rate, ν^0_{scb}, is altered by both multiplicative and additive *rate modifiers*, $\kappa_{scb}(\eta, \chi)$ and $\Delta_{scb}(\eta, \chi)$ respectively. These modifiers are thus dependent on the unconstrained and constrained parameters η and χ, and with the exception of the parameter of interest (which here is the signal strength) they are parameterised by nuisance parameters. A summary of the symbols used to define the statistical model is given in table 4.1.

Symbol	Description
n	observed event count
a	auxiliary data
$\nu(\phi)$	calculated event rate
η	free parameters
χ	constrained parameters
ψ	parameter of interest
θ	nuisance parameters
$\phi = \{\eta, \chi\} = \{\psi, \theta\}$	All parameters
$\kappa(\phi)$	multiplicative rate modifiers
$\Delta(\phi)$	additive rate modifiers
$c_\chi(a_\chi\|\chi)$	Constraint term for χ
σ_χ	Relative uncertainty on χ

Table 4.1: Definition of symbols used in definition of statistical model.

4.1.1 Hypothesis Testing

With this statistical model defined, must then seek to find the model parameters that best fit the observed data, both the event counts and the associated auxiliary measurements. This study makes use of the $\tilde{q}_\mu$ test statistic, which is modified for the specific purposes of testing models in which the signal strength μ cannot take a negative value. This is defined as [80]:

$$\tilde{q}_\mu = \begin{cases} -2ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{4.3}$$

where $\tilde{\lambda}(\mu)$ is the constrained *likelihood ratio*. In the limit of large datasets this will asymptotically tend towards a χ^2 distribution, and is defined as:

$$\tilde{\lambda}(\mu) = \begin{cases} -2ln\frac{L(\mu, \hat{\hat{\phi}}(\mu))}{L(\hat{\mu}, \hat{\hat{\phi}}(0))} & \hat{\mu} \leq \mu \\ -2ln\frac{L(\mu, \hat{\hat{\phi}}(\mu))}{L(\hat{\mu}, \hat{\phi})} & \hat{\mu} > \mu \end{cases} \tag{4.4}$$

The numerator value in this likelihood ratio, $L(\mu, \hat{\hat{\phi}}(\mu))$, represents the likelihood of the signal strength, μ, given all other model parameters are at the maximum likelihood

estimate, $\hat{\hat{\phi}}$ (i.e. at the best values that best fit the observations). The denominator on the other hand gives the likelihood when the signal strength is also at its maximum likelihood estimate.

Typically the *p-value* is used as a measure of the incompatibility between the signal hypothesis and the background-only (null) hypothesis. When testing a given signal hypothesis, a low p-value therefore means that there is a low probability of observing the given data under the background-only hypothesis thus implying confidence in some signal hypothesis. Commonly in particle physics studies, the significance of an observation is quoted in standard deviations of a unit Gaussian with the one-sided tail probability of the observed p-value. This is defined as:

$$Z = \Phi^{-1}(1 - p) \tag{4.5}$$

where Φ is a cumulative distribution function of the Gaussian distribution. One then can decide to reject the background-only hypothesis based on some p-value lower threshold level, α_0. In this study a threshold of α=0.1 is defined as the threshold to reject the given hypothesis. Furthermore, for each of these a range of signal strengths can be tested and thus a 90% confidence level upper limit can be defined.

Template Binning

The maximum likelihood fit is conducted using templates comprised of a 2D binned distribution of the squared recoil mass, M^2_{recoil}, against the CMS-frame recoil polar angle, θ^*_{recoil}. The widths of the M^2_{recoil} bins are defined by the signal width at the given squared recoil mass, which is found by way of fitting with a Crystal ball function as discussed in section 2.4. This results in 702 bins that span the range [-2, 81]GeV^2/c^4. A mass hypothesis is tested corresponding to each bin (excluding of course negative values), giving a total of 692 mass hypotheses that are tested. The masses tested are equal to the central value of each bin.

For each mass hypothesis that is studied, a fit window comprising a sub-region of the full M^2_{recoil} spectrum is defined. This is done to improve the fitting speed, as for any given mass hypothesis, signal events are naturally only expected to exist in the region near to the corresponding squared mass in M^2_{recoil}. It was found that a fit window with a width of $\pm 10 \times 1\sigma$ bins around the given squared mass was wide enough to allow for accurate fitting of the signal and background templates. Smaller windows limit the ability to fit the background template as all bins would contain a non-trivial number of events from the signal template, and no bins with purely background contributions.

As discussed previously, it was found that the performance of the ECL-based photon veto suffered due to photons being missed in gaps of the calorimeter instrumentation. This resulted in specific band like structures appearing in the 2D M^2_{recoil} vs. θ^*_{recoil} distribution. One can see this clearly in figure 4.1, which shows $\mu^+\mu^-(\gamma)$ MC events after the application of the Punzi-Net. The 2D distribution on the left is shown with a red contour defined using the kernel density estimate method. This highlights particularly the large number of low recoil mass $\mu^+\mu^-(\gamma)$ events surviving in the $1.75 \lesssim \theta^*_{\text{recoil}} \lesssim 2.05$

region, where a there is a gap in ECL instrumentation, commonly referred to as the *90 degree gap*, leading to larger numbers of photons going undetected.

The right hand subplot of fig. 4.1 shows, in red, the chosen bin edges in the θ^*_{recoil} axis. There are 4 distinct regions in which binning defined such as to best capture the expected structures in the $\mu^+\mu^-(\gamma)$ events:

- $\mathbf{M^2_{recoil}} \leq 2.0\ \mathbf{GeV^2/c^4}$: Bin edges in this low recoil mass region are selected to isolate the previously discussed region around the 90 degree gap. This allows for better constraint of the background and signal models during fitting, due to the expectation that background events will predominantly appear in those bins around the gap, whereas signal events are expected to describe a flat distribution in θ^*_{recoil}. The bin edges are chosen as [0, 72, 95, 106, 115, 180] degrees.

- $\mathbf{2.0 < M^2_{recoil}} \leq 30.0\ \mathbf{GeV^2/c^4}$: This region has little-to-no expected backgrounds at all, with only a very small contribution of $\mu^+\mu^-(\gamma)$ events. These reside mainly in a diagonal band that originates from the 90° gap and extends towards the forward detector direction. The region is split into 3 θ^*_{recoil} bins, with varied edges defined to bound the observed structure in the central bin, thereby defining it as the bin in which any background events are expected to reside.

- $\mathbf{30.0 < M^2_{recoil}} \leq 60.0\ \mathbf{GeV^2/c^4}$: This region contains 2 notable bands that originate from the forward and backward directions of the detector polar angle at approximately 30 GeV^2/c^4, and curve to the centre where they join at around 50 GeV^2/c^4. In this region just 3 bins are deemed necessary to capture the shape of thist structure, with edges at [0, 80, 103, 180] degrees. In the lower end of this squared recoil mass region, this provides 2 edge bins in which the background $\mu^+\mu^-(\gamma)$ events are expected to predominantly appear. Conversely in the higher squared recoil mass side, one sees that the middle bin is expected to contain the bulk of the backgrounds, with lesser numbers expected in the two side bins.

- $\mathbf{M^2_{recoil}} > 60.0\ \mathbf{GeV^2/c^4}$: In this region there appears to be no difference in expected distributions between signal and background events, with both describing flat distributions across θ^*_{recoil}. However it was decided to continue to use the three bins defined in the region below this. This allows for robustness against unexpected variation observed in data without much impact on the fitting speed.

One should note that this binning schema is defined based on the observed distribution of $\mu^+\mu^-(\gamma)$ backgrounds before the application of the additional KLM-based photon veto (described fully in section 2.4). This provides further coverage in regions with limited ECL instrumentation, and so is expected to reduce much of the band-like structures that the binning is designed to capture. One can see this in figure 2.19, which shows the 2D $\text{M}^2_{\text{recoil}}$ vs θ^*_{recoil} distribution of $\mu^+\mu^-(\gamma)$ events before and after the application. The band structure between 2 and 30 GeV^2/c^4 is largely removed and the lower portion of the bands between 30 and 60 GeV^2/c^4 is heavily reduced. Nevertheless the binning described here is kept unchanged as it provides robustness against possible mis-modelling of these

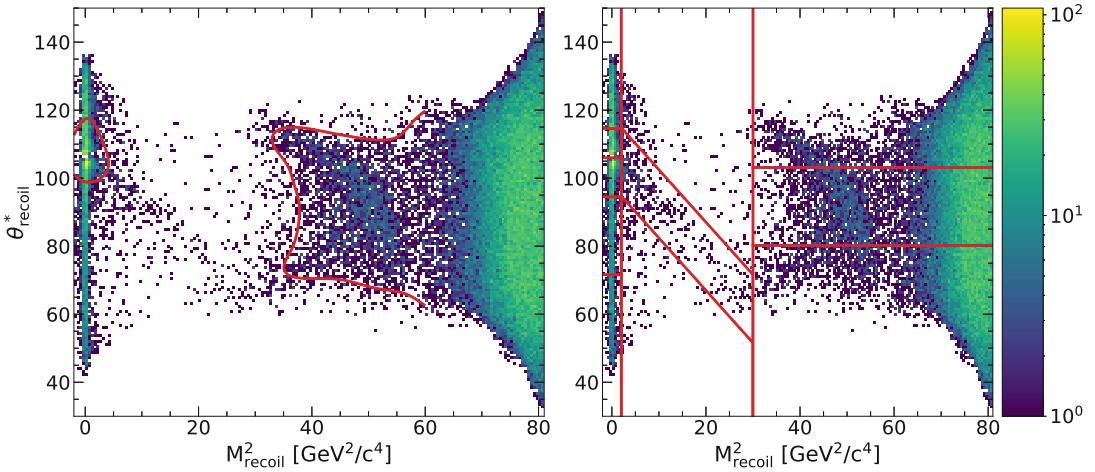

Figure 4.1: The 2D θ^*_{recoil} vs. M^2_{recoil} distribution for $\mu^+\mu^-(\gamma)$ MC events surviving all selection criteria. The left hand plot shows a contour defined by kernel density estimation which highlights the large peaking structure of events around $\approx 105^{\circ}$. The right hand plot then shows the chosen bin edges in the recoil polar angle, with 3 regions with particular binning schemes.

radiative background processes in simulated events. This same binning scheme is applied in the case of the muonphilic scalar search, which is of course just limited to the lowest squared recoil mass region outlined here.

Template Binning for fitting with varied α_D

The results of the analysis are interpreted under multiple theoretical model assumptions, the basic case being that of the vanilla L_μ - L_τ Z' in which the invisible decay is purely to SM particles and there is no dark matter contribution. When extending this to study models with some non-zero coupling to DM particles, one must consider the effect that will be incurred on the expected distribution of signal events, and thus template shape used in the binned maximum-likelihood fit.

As discussed in section 1.3, the expected width of the Z' is dependent on this coupling α_D, with larger coupling producing a larger width. This effect is shown as a function of the recoil mass in figure 1.10. However, signal extraction is conducted instead in M^2_{recoil} and thus one must convert the decay width, $\Gamma_{(Z'\to\chi\bar{\chi})}$, for a given value of α_D to the corresponding value in the squared recoil mass. A new quantity, Γ', is therefore defined:

$$\Gamma' = 2\Gamma_{(Z'\to\chi\bar{\chi})} M_{Z'} \tag{4.6}$$

Figure 4.2 shows Γ' as a function of M^2_{recoil} for a set of dark matter coupling strengths. The detector resolution is shown with a black dashed line. In the standard L_μ - L_τ model,

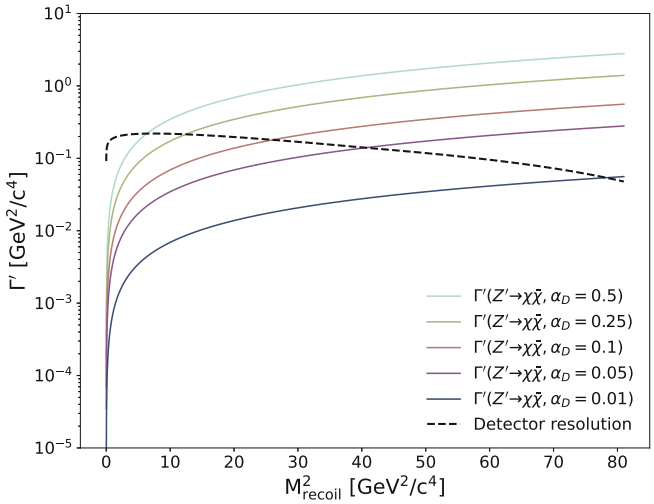

Figure 4.2: The Z′ width in M^2_{recoil}, defined as Γ' in eqn. 4.6, shown as a function of M^2_{recoil} for a set of dark matter coupling strengths.

the physical decay width is negligible and thus limited by this. That is to say, the physical decay is thinner but the detector is unable to resolve below a certain threshold. One can see however that with non-zero coupling to DM this is no longer the case and in fact the width starts to surpass the detector resolution, particularly at higher squared recoil mass.

To study such cases with non-negligible width due to DM coupling, additional sets of simulated signals are generated across the full mass range with $\alpha_D = 0.05$, 0.1, 0.25 and 0.5. For each of these a new binning schema is defined such that the range is covered by bins of width $\approx 1\sigma_{\alpha_D}$, where σ_{α_D} is the width of a signal peak with some DM coupling α_D at the given mass (the mass being defined as the square root of the central value of the M^2_{recoil} bin). For larger values of α_D, this can produce notably larger bin widths than in the vanilla case. At lower mass hypotheses where the width is still negligible, the bins have a width equal to that defined by the detector resolution just as in the case of the vanilla model. Signal MC samples are generated at mass hypotheses such that each defined bin in these schema has a corresponding signal hypothesis to test, just as is done in the vanilla search.

Figure 4.3 shows the distribution of the resulting mass hypotheses and their respective bin widths (in the squared recoil mass distribution). One can see that for different values of α_D the widths become non-negligible at different points at which they deviate from the vanilla binning schema. Furthermore, due to the wider bins, there are fewer mass hypotheses tested for larger values of α_D. Figure 4.4 shows the comparison of a set of templates with increasing values of α_D with vanilla model templates of the same mass. These highlight just how much the signal peaks begin to spread out at higher mass.

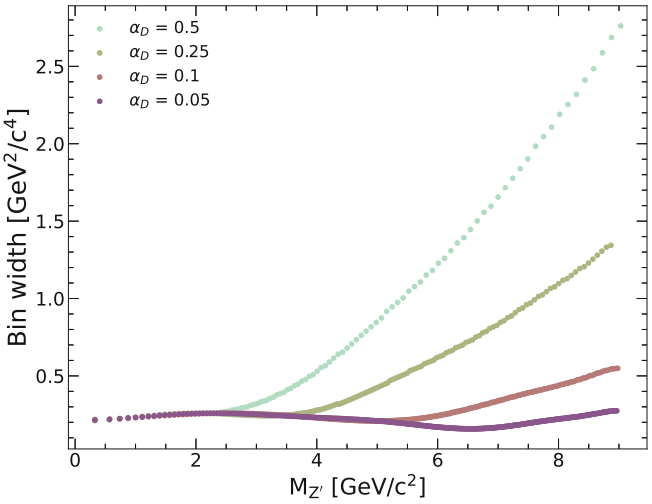

Figure 4.3: The generated MC signal mass hypotheses and their respective bin width. Entries are shown for dark matter couplings α_D=[0.05, 0.1, 0.25, 0.5]. Each distribution begins to trend upwards at the point where the width is no longer negligible with respect to the detector resolution and so the bin widths must increase to accommodate this.

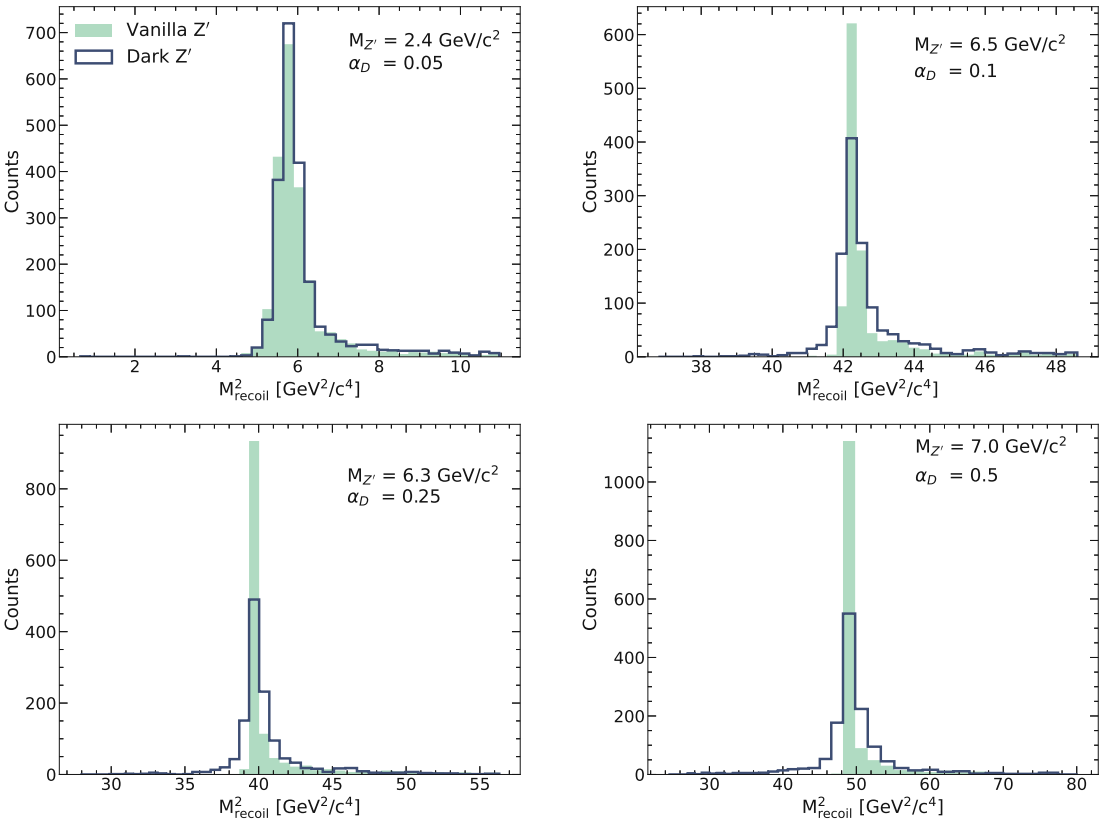

Figure 4.4: Comparison of vanilla Z′ templates with negligible decay widths (green) to those width varied couplings to dark matter (α_D=0.05 (top left), 0.1 (top right), 0.25 (bottom left) and 0.5 (bottom right))

4.2 Nuisance Parameters

The systematic uncertainties associated with the measurement are captured by nuisance parameters in the binned maximum likelihood fit, each being constrained by a Gaussian term in the likelihood itself. These can come in the form of single normalisation values or shape uncertainties which may be comprised of multiple individual fitted nuisance parameters.

Correlated shape uncertainties in both the M^2_{recoil} and θ^*_{recoil} axes of the templates are used. These provide stronger constraint on the uncertainty that they parameterise due to the correlation across the opposite axis of the 2D template. For instance, take the case of the correlated shape uncertainties that are applied in the θ^*_{recoil} axis. By doing this, each of the polar angle bins in the template are assigned some percentage uncertainty. A given bin can then fluctuate during fitting by this given amount. However, the effect is correlated across the M^2_{recoil} bins within that polar angle bin. If there was some observed excess in one of the θ^*_{recoil} bins, it would also require a presence in the corresponding bins across the M^2_{recoil} bins of the template to notably impact the fit. This is designed to improve robustness against any mismodelling in the polar angle distribution, as was observed in the previous iteration of the analysis where a peaking structure was present in M^2_{recoil} which then corresponded to a specific region of the θ^*_{recoil} - there was a correlation. It was the result of the previously discussed overestimation of the ECL photon veto efficiency in data.

Table 4.2: The applied systematic uncertainties.

Source	$N_{nuis.\ par.}$	σ		
		< 2 GeV^2	2 - 9GeV^2	>9 GeV^2
Signal efficiency	1 Global		8%	
Signal resolution	$N_{M^2_{rec}bins}$ Shape		4%	
θ^*_{recoil} shape	$N_{\theta^*_{rec}bins}$ Shape	8%	4%	1%
M^2_{recoil} shape	$N_{M^2_{rec}bins}$ Shape	6.5%	3.5%	2%
γ veto θ^*_{recoil}	$N_{\theta^*_{rec}bins}$ Shape	3%	2%	1%
γ veto M^2_{recoil}	$N_{M^2_{rec}bins}$ Shape	6%	2%	1%
bgd. norm.	1 Global		Free floating	
Lumi	1 Global		0.45%	

Each of systematic uncertainties included the nuisance parameters in the binned maximum-likelihood fit are detailed below, and summarised in table 4.2. As the studies are in general quite similar, systematic uncertainties outlined here are applied also in the case of the muonphilic scalar search.

- *Detector resolution modelling*: The resolution modelling will effect the width of any signal peak in M^2_{recoil}. Agreement between data and MC is studied using the $e^+e^- \to \mu^+\mu^-\gamma$ control channel and a percentage uncertainty on this width is defined. This is implemented in the signal template as a shape uncertainty in the

M^2_{recoil} axis, with correlation across the θ^*_{recoil} axis. This is done because any effect from resolution modelling should only impact shape agreement in the M^2_{recoil} axis and not the polar angle. A value of 4% is used.

- *Signal efficiency*: The signal efficiency uncertainty is investigated with consideration of multiple sources. The final value considers the Punzi-Net efficiency uncertainty and the initial data-to-MC agreement of the control channels. The agreement in signal efficiency between run dependent and run independent signal samples is also included. This is applied as a normalisation uncertainty on the signal template, with a value of 8% used.

- θ^*_{recoil} *modelling*: The data-to-MC agreement in the recoil polar angle variable is studied with the $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to e^\pm\mu^\mp$ control channels. The data-to-MC ratios in each is combined to provide a percentage uncertainty on the accuracy of the θ^*_{recoil} axis in the 2D templates that are fitted to data. This is applied as a correlated shape uncertainty, with each polar angle bin then having an associated nuisance parameter. Different values are used in regions of squared recoil mass, defined in table 4.2.

- M^2_{recoil} *modelling*: Similar to the method described for the θ^*_{recoil} uncertainty. This time data-to-MC ratios in the control channels are combined to provide a percentage uncertainty on the M^2_{recoil} axis of the 2D templates fitted to data. This is applied as a correlated shape uncertainty, with each M^2_{recoil} bin then having an associated nuisance parameter. Different values are used in regions of squared recoil mass, defined in table 4.2.

- *Photon veto*: The uncertainty associated with the photon veto inefficiency corrections are defined using the $e^+e^- \to e^+e^-\gamma$ control channel. Values are derived separately for the θ^*_{recoil} and M^2_{recoil} axes, with consideration of how the corrections might effect template shape uncertainty in the respective axis. Correlated shape uncertainties are then defined for each of the two axes. Different values are used in regions of squared recoil mass, defined in table 4.2.

- *Background template normalisation*: The background template is left free floating, meaning it can fit to any total normalisation rather than within some range defined by an uncertainty. This is a more conservative approach.

4.3 Projected limits

This chapter shows the expected limits that will be set on the cross section and coupling constants in the different models that are studied as part of this thesis work. As discussed previously, the iteration of this analysis presented here is an update to a previous work that was completed and published as part of the PhD study. The results of this are discussed in section 1.5. The updated analysis outlined in this work has not yet completed the data unblinding process within the Belle II collaboration and so only the projected limits from simulated MC data can be shown here.

The vanilla Z′ model

The 90% confidence level upper limits on the cross section of the $e^+e^- \to \mu^+\mu^- Z'[\to inv]$ process are shown as a function of the studied mass hypotheses in figure 4.5. This shows the cross section found in the *vanilla* L_μ-L_τ model, where the Z′ is assumed to decay only to SM neutrinos with a decay width that is negligible with respect to the detector resolution.

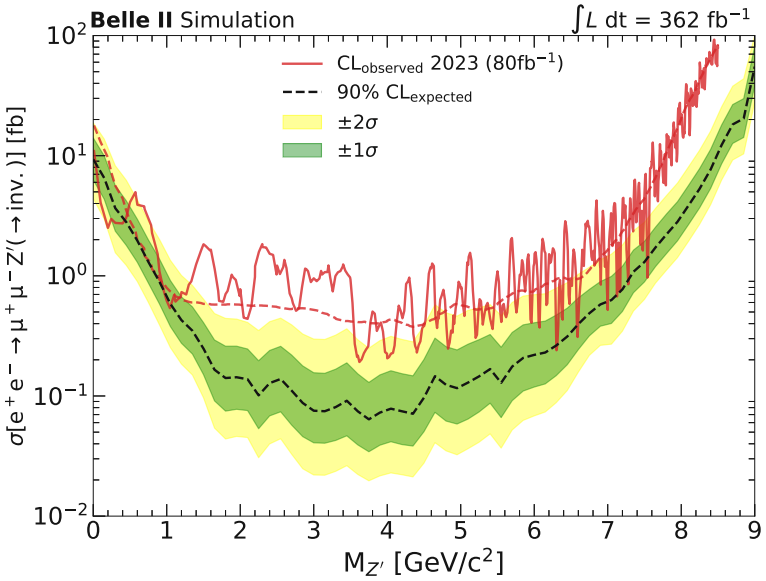

Figure 4.5: The expected 90% CL upper limits on the cross-section of the Z′ boson in the *vanilla* L_μ-L_τ model, with the $\pm 1, 2\sigma$ uncertainty bands shown in green and yellow. The result obtained in the previous iteration of the analysis that used $80 fb^{-1}$ is shown in red, where the dashed line is the expected upper limit and the solid is the final observed limit.

The dashed and solid red lines show the expected and observed (after unblinding) result obtained in the previous iteration of the analysis with $80 fb^{-1}$ of data. For masses above $1 GeV/c^2$, the updated result shows improvements beyond that which one might

expect from the increase in statistics alone (which scale with the square root of the luminosity, therefore the limit would be expected to improve by a factor of ≈0.5).

The improvement is notably lesser below this, which is understood to be due to changes in the MC samples being used during training of the Punzi-Net. As mentioned, in the previous iteration there was a notable mismodelling of the photon veto in MC which led to an unexpected excess of data in this region. This was largely fixed in the more recently produce MC samples and so the Punzi-Net was able to account for this irreducible background during training. The loss function optimises the average sensitivity across the full mass range. With this new, more accurate MC, the network learns it is easier to achieve this by optimising selection for higher masses. This therefore also accounts for the large improvement in the higher masses at the same time.

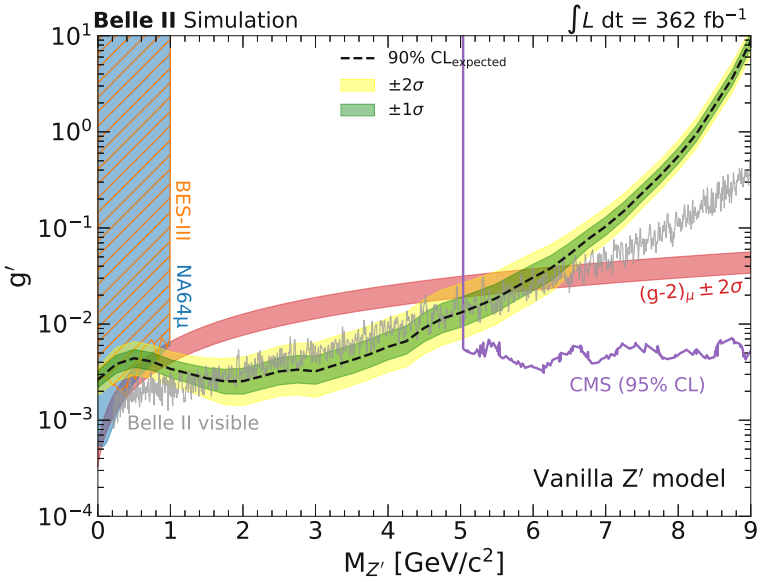

Figure 4.6: The expected 90% CL upper limits on the g′ coupling constant under the *vanilla* L_μ-L_τ model. The $\pm 1, 2\sigma$ uncertainty bands are shown in green and yellow.

The cross section limits are then translated into 90% CL upper limits on the g′ coupling constant, which is shown in figure 4.6. This is therefore done considering the BF of the Z′ invisible decay mode as shown by the blue line in figure 1.9. The upper limits are shown in comparison to the region of g′ that could then explain the muon g-2 anomaly (discussed in section 1.2.2), which is depicted by the red band.

A few notable experimental results are shown in addition. The low mass region ($M_{Z'} < 1 GeV/c^2$) is dominated by a recent study published by the NA64μ experiment (blue) [44], along with a recent result from the BES-III collaboration [47] (orange). The mass range above 5 GeV/c^2 is heavily constrained by the CMS collaboration [41] (purple). The Belle II study of the visibly decaying Z′ boson is shown in grey [36]. A previous analysis by the Belle collaboration [40] set very similar limits as this Belle II study. These

have been omitted from the plot here for clarity. One may note that the study described within this thesis is expected to set highly competitive and perhaps world-leading limits in the Z′ mass range between approximately 2 and 5 GeV/c^2.

The dark Z′ model

The 90% CL upper limits on the Z′ cross section under the dark Z′ model are shown in figure 4.7. The results are shown for fits conducted with signal templates that are constructed with $\alpha_D = 0.01$, 0.05, 0.1, 0.25 and 0.5. As expected these converge well at the low mass where the widths are expected to be negligible with respect to the detector resolution. There are slight differences around 2 GeV/c^2, which are likely statistical fluctuations. As expected the limits associated with larger values of α_D diverge first as the width of the signal template increases, thus making it harder to distinguish from backgrounds (events spread and the peak becomes less clear). The limits for $\alpha_D = 0.25$ and 0.5 are only shown to 8.5 and 8.0 GeV/c^2 respectively. It was found that above these points the associated signal templates grew too wide reducing the effectiveness of the fitting procedure.

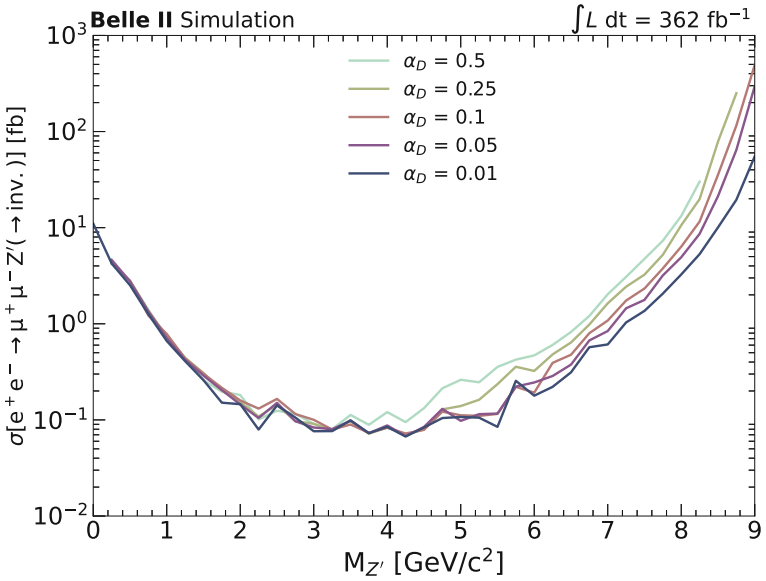

Figure 4.7: The expected 90% CL upper limits on the cross-section of the Z′ boson in the dark Z′ model, with varied coupling to dark matter ($\alpha_D = 0.01$, 0.05, 0.1, 0.25, 0.5). The limits for $\alpha_D = 0.25$ and 0.5 are only shown to 8.5 and 8.0 GeV/c^2 respectively. It was found that above this the fitting procedure became ineffective.

Again the cross section limits are then translated into limits on the g'_D coupling constant. This is shown in figure 4.8 again in comparison with the region of g'_D that could explain the muon g-2 anomaly. Only two other experimental results constrain this dark Z′ model. The result from BES-III is shown in orange [47], and the result from the

NA64 is shown in blue [44]. These both constrain the low mass region heavily, however, the study outlined in this thesis is able to constrain a large region of the mass range, ruling out an invisibly decaying Z′ with coupling to dark matter and a non-negligible decay width as an explanation for the muon g-2 anomaly for masses up to approximately 6 GeV/c^2. It it is important to note that this plot shows only results from direct searches, however, there are also stringent limits set by reinterpretation of data from neutrino experiments [31, 81, 82]. Even so, the results shown here provide world leading limits on the g'_D coupling constant in the Z′ approximate mass range of 1 to 5 GeV/C^2.

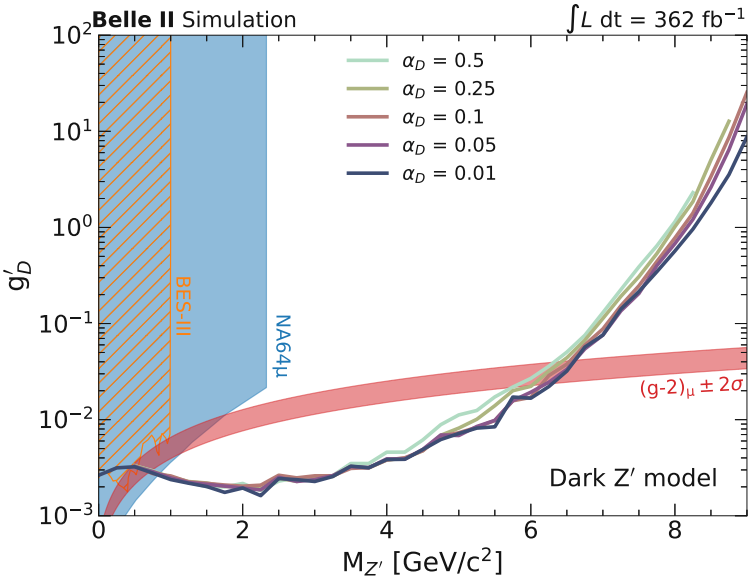

Figure 4.8: The 90% CL upper limits on the g'_D coupling constant, with result shown for varied coupling to dark matter (α_D = 0.01, 0.05, 0.1, 0.25, 0.5). The limits for α_D = 0.25 and 0.5 are only shown to 8.5 and 8.0 GeV/c^2 respectively. It was found that above this the fitting procedure became ineffective.

The muonphilic scalar model

Due to the fact that the search for a muonphilic scalar is constrained to the region below the muon mass limit ($< 2M_\mu$), a full scan of mass values is not conducted. Instead the final result will be given with a single fit at a single mass value. However for now three mass values spanning the range are studied as a check for stability. The 90% CL upper limits on the cross section for the mass hypotheses 10, 100 and 200 MeV/c^2 are shown in figure 4.9, along with the $\pm 1, 2\sigma$ uncertainty bands shown in green and yellow.

Finally the resulting 90% CL upper limits on the g_S coupling constant associated with the scalar are shown in figure 4.10. Again three mass hypotheses (M_S = 10, 100 and 200 MeV/c^2) are shown, along with the region associated with the muon g-2 anomaly. The

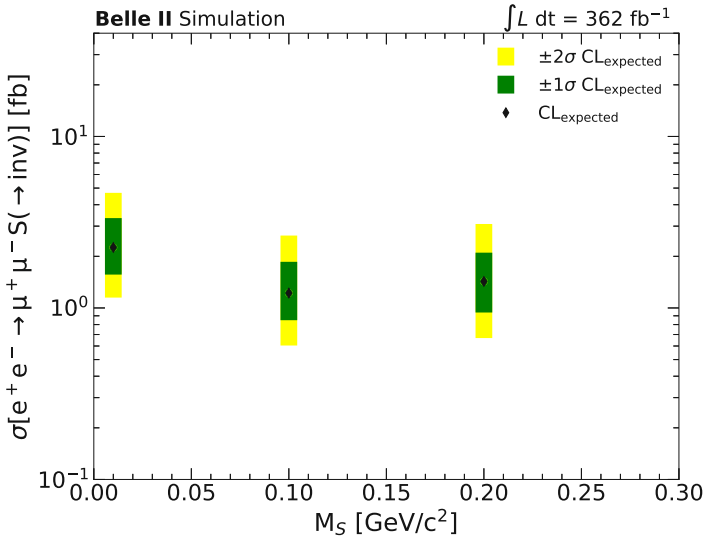

Figure 4.9: The 90% CL upper limits on the cross section for the mass hypotheses 10, 100 and 200 MeV/c^2, along with the $\pm 1, 2\sigma$ uncertainty bands shown in green and yellow.

limits set by the BES-III collaboration [47] are shown in orange. The limits on g_S set by this study are therefore competitive with those published by the BES-III collaboration.

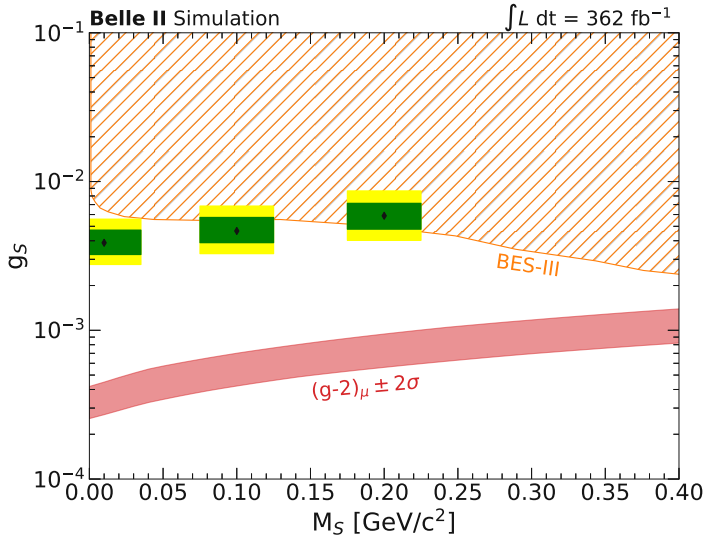

Figure 4.10: The 90% CL upper limits on the g_S coupling constant for three mass hypotheses at 10, 100 and 200 MeV/c^2. The region associated with a possible explanation of the muon g-2 anomaly is shown in red, and the limits published by the BES-III collaboration are shown in orange [47].

4.4 Fit Tests

4.4.1 Signal Injection Test

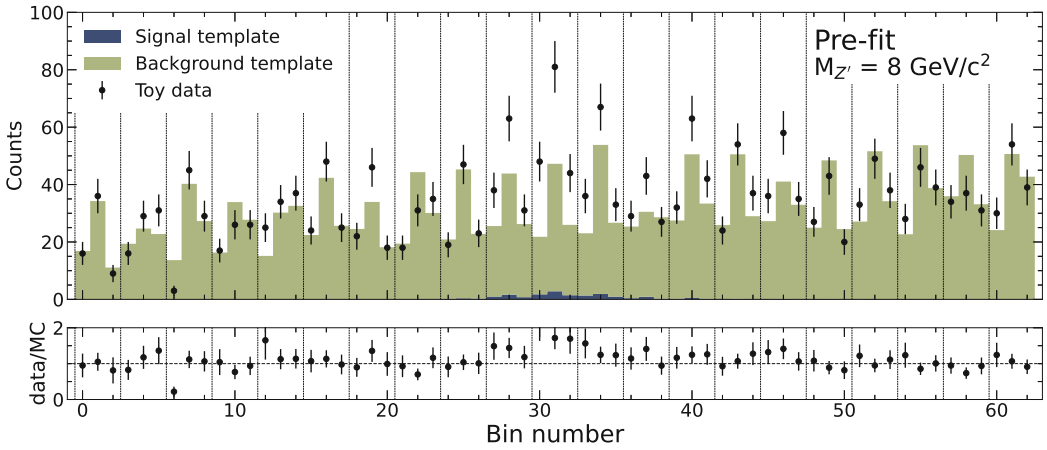

Figure 4.11: Toy data sampled from the background template with an 8 GeV/c^2 vanilla Z′ signal injected at a cross section of σ = 15fb. The background and signal templates are shown at their pre-fit levels. The full window contains 63 bins in total, comprised of the 21 M^2_{recoil} bins and 3 θ^*_{recoil}. The former of these is marked by the vertical dashed lines, each then containing the 3 θ^*_{recoil} bins.

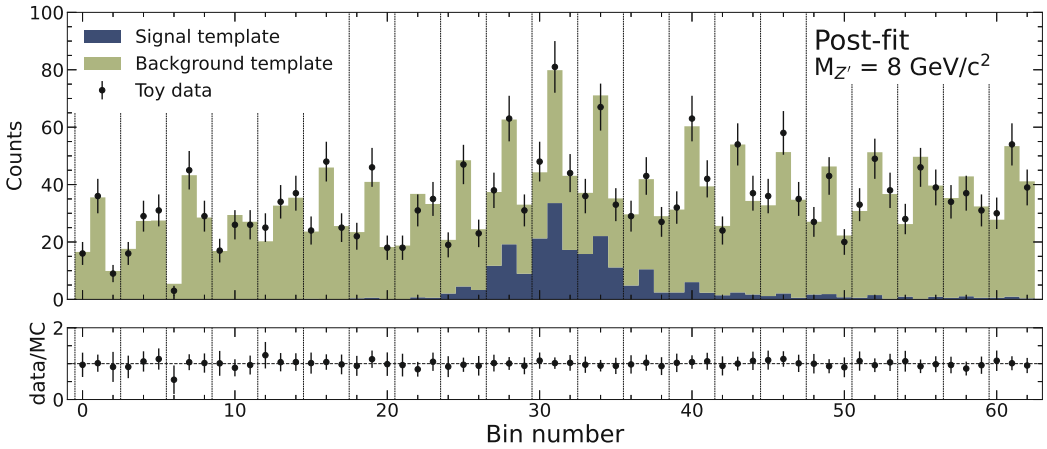

Figure 4.12: Toy data sampled from the background template with an 8 GeV/c^2 vanilla Z′ signal injected at a cross section of σ = 15fb. The background and signal templates are shown at their post-fit levels.

It is of course important to validate the fitting procedure and ensure that some observed signal would in fact be measured accurately. This can be done by way of fitting

toy data samples in which a signal hypothesis with some chosen cross section is added on top of the background.

An example of this procedure with an injected 8 GeV/c^2 Z′ signal at a cross section of σ = 15fb is shown in figures 4.11 and 4.12. These contain the sampled to data along with the background and signal templates before and after the fitting procedure respectively. The figures each show 63 bins in total, which result from the chosen fit window of ±10 bins around the bin centred on the mass hypothesis itself (resulting in a total of 21 bins spanning the M^2_{recoil} axis) and the additional three bins defined in the recoil polar angle axis, θ^*_{recoil}. The figures include dashed lines that define the individual M^2_{recoil} bins, within which one then finds the three θ^*_{recoil} bins.

Of course while doing an individual toy signal injection fit can be informative, it is better to conduct many such fits with randomly sampled toy data. This can reveal any bias factor that may be present in the fitting strategy. One region in which such a check is indeed important is in the low mass region, considering events with M^2_{recoil} near to 0 GeV^2/c^4. Due to the peaking structure of the residual $\mu^+\mu^-(\gamma)$ background events here (as shown in figures 2.14 and 2.15), it is important to check that there is not a tendency for the fitting procedure to attribute signal events as background. To study this, 2500 toy fits are conducted with an injected signal at a mass of $M_{Z'}$=0.515GeV/c^2 and cross section of σ=20fb. Figure 4.13 shows the *pull plots* of the fitted cross section (left) and background template normalisation (right). The histograms show the distribution of the fitted value ($\hat{\psi}$) minus the truth value (ψ_0), divided by the uncertainty on the fitted value (σ_ψ). If the fits are not in any way biased, one can expect these distributions to follow a Gaussian of mean 0 and standard deviation 1. One can see that indeed the behaviour in fitting both the signal and background templates is thus as expected, with no clear shift, the presence of which might indicate that signal events are being fitted as background, or vice versa.

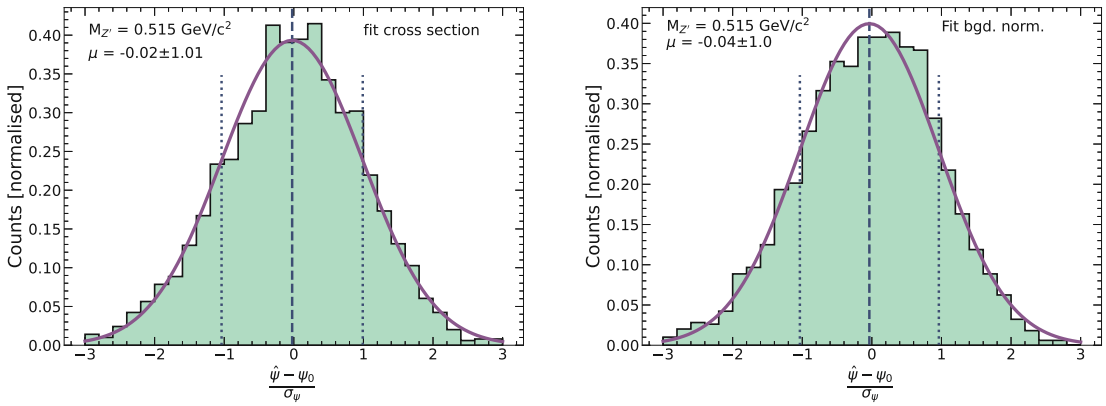

Figure 4.13: The pull plots of the fitted cross section (left) and background template normalisation (right) for 2500 toy data fits. An injected signal of mass M'_Z0.515 GeV/c^2 and cross-section σ =20fb^{-1} is present in the toy data.

Finally, one can also study the effect the injection of some signal may have on the limits that are set on cross section. Figure 4.14 shows the $\pm 1, 2\sigma$ uncertainty bands (green and yellow respectively) around the expected 90% CL upper limit on the cross section of the invisible Z′ decay in the vanilla model. The resulting observed limits from individual mass hypothesis fits to the toy data are shown by black points. The injected signal cross sections and masses ($M_{Z'}$ = 6.5 and 8.0 GeV/c^2, σ=0.7, 15fb).

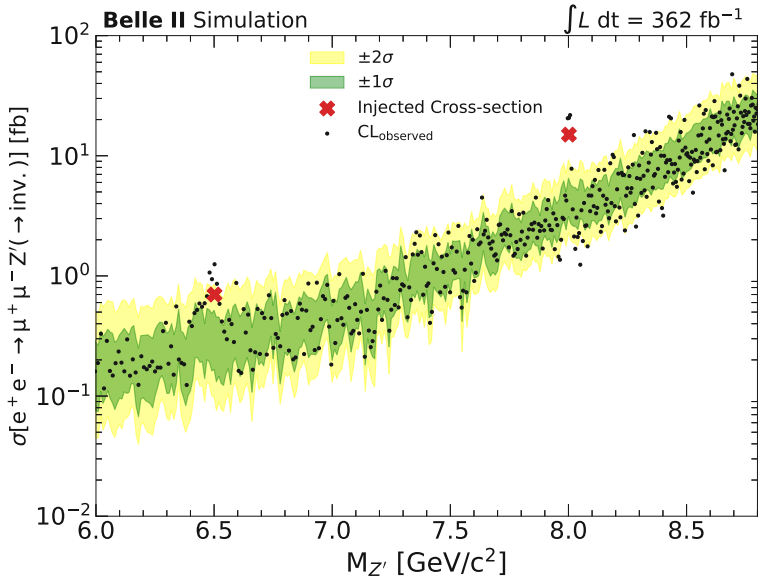

Figure 4.14: The 90% CL upper limits set in the mass range range between 6 and 8.8GeV/c^2. The $\pm 1, 2\sigma$ uncertainty bands shown in green and yellow, and the result from individual mass hypothesis fits are represented with black dots. The injected signals are represented by red crosses at $M_{Z'}$ = 6.5 and 8.0 GeV/c^2.

In the regions far from these injected signals the distribution of observed limit appear as one might expect, generally distributed around the central value with statistical fluctuation bringing some to the yellow $\pm 2\sigma$ uncertainty band and slightly beyond[2]. In the areas immediately around the injected signals however, one can see that the observed limits then diverge from this expected distribution and indeed set cross section limits above the injected signal regions.

4.4.2 The Look Elsewhere Effect

When testing a large set of hypotheses spanning some parameter space one must consider the possibility that an individual measurement may present a seemingly significant result that is in fact just the result of expected statistical fluctuations in the search region. Such

[2]This is to be expected when conducting a study with a large number of hypotheses being tested and is accounted for in the next section 4.4.2

effects can be seen in figure 4.14, where in the regions far from injected signals, there is a natural fluctuation of the observed limits around the expected value, with some in fact lying around the edges of the yellow $\pm 2\sigma$ uncertainty band. The *local significance* of such individual tests must therefore be translated into a *global significance* that takes into consideration statistical fluctuations across the full parameter space.

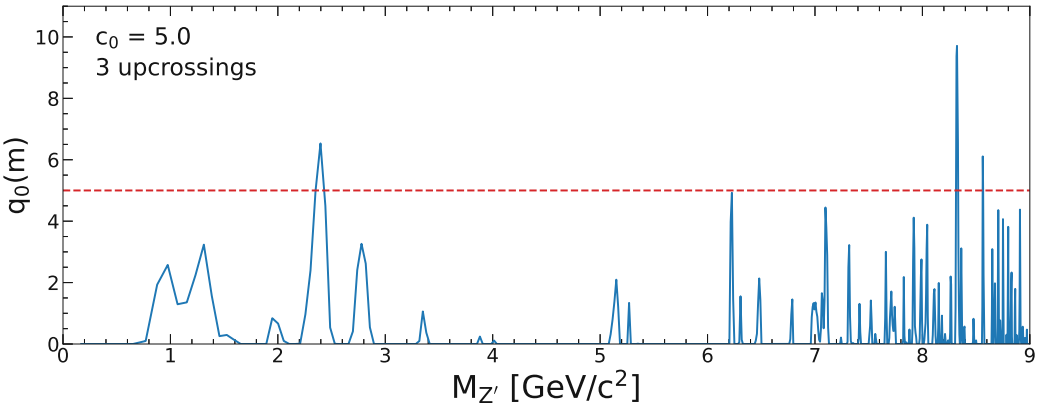

Figure 4.15: The value of $q_0(m)$ for all mass hypotheses when fit to a single toy data set. An example threshold of c_0=5.0 is shown which givens 3 up-crossings. The density of tested mass hypotheses increases as the mass increases due to the tighter signal peaks in M^2_{recoil} expected here.

To study this effect, the $q_0(m)$ test statistic is used and a similar strategy to that outlined in [83] is employed. 10000 toy data sets are sampled from the background MC distribution after all selections are applied and for each of these the test statistic is calculated across the mass hypotheses. The test statistic is expected to follow a χ^2_s distribution with s degrees of freedom (where here $s = 1$ as there is just one parameter of interest). The maximum value of this test statistic, $q(\hat{m})$, over the mass hypotheses is then bounded by the function:

$$P(q(\hat{m}) > c) \leq P(\chi^2_s > c) + \langle N(c) \rangle \tag{4.7}$$

where $\langle N(c) \rangle$ is the average number of *upcrossings* past the threshold value of c of the test statistic. For large c ($c >> s$) one can approximate to the bound and give the probability of observing a value of $q_0(m)$ greater than the threshold:

$$P(q(\hat{m}) > c) \approx P(\chi^2_s > c) + \langle N(c) \rangle \tag{4.8}$$

With just a single degree of freedom $s = 1$, the average upcrossings is given by:

$$\langle N(c) \rangle = \mathcal{N} e^{-c/2} \tag{4.9}$$

Where $\mathcal{N}$ will, in this large c limit, be a constant value that is independent of the value of c. With this one can therefore define the estimated number of up-crossings above

the threshold c, by measuring the number at some lower threshold c_0 using the toy data sets:

$$\langle N(c)\rangle = \langle N(c_0)\rangle e^{-(c-c_0)/2} \tag{4.10}$$

Finally the trial factor can then be defined as the ratio of the probability of observing the test statistic value at a fixed mass point, to that of observing it anywhere in the given range:

$$trial\# = \frac{P(q_0(\hat{m}) > c)}{P(q_0(m) > c)} \tag{4.11}$$

Thus providing a method of directly relating an observed local significance to a global significance that considers all mass points studied. Figure 4.16 shows on the left this trial factor, and on the right the global significance, both as a function of the local significance. Both plots show the distribution found analytically by way of the fits to to data, and that described by the asymptotic formula as discussed in [83, 80]. For these, a threshold value of $c_0 = 9$ is used. One can see that the analytical method appears to follow the asymptotic bound until around a local significance of approximately 3, at which point statistical uncertainty due to the sample size of toys takes effect. Above this one can then use the asymptotic bound as a means of translating the local to global significance. For instance a local significance of 5σ (often deemed as the level at which one can claim a discover in particle physics), the real global value with proper consideration of the look elsewhere effect is just 3.9σ.

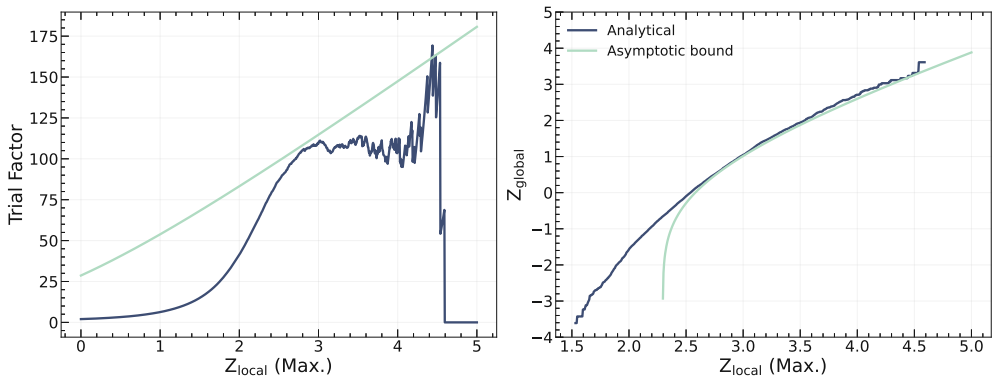

Figure 4.16: The trial factor (left) and global significance (right) against local significance. Both figures show the result from analytical calculation using to data sets and the asymptotic bound.

Conclusion

Overview

This thesis work has presented a comprehensive overview of the searches for invisibly decaying Z′ and muonphilic scalar bosons at the Belle II collider. The motivation behind these searches is introduced with discussion of the current landscape in particles physics, and the questions that are as yet unanswered in the Standard Model. One such gap that exists in the understanding of the universe is the nature of dark matter. The Z′ boson can provide a way to describe the observed relic density of dark matter, and thus models in which it may couple to dark matter are investigated. The *vanilla* L_μ-L_τ model in which the Z′ may only couple to SM particles is introduced and, while it may not couple to DM, its ability to describe the muon g-2 anomaly is discussed.

The study is then expanded to include also a search for the muonphilic scalar boson in the mass region below the muon mass threshold ($M_S < 2M_\mu$). Previous experimental searches for these particles are presented, with particular attention paid to the results of an earlier search for an invisibly decaying Z′ using 79.7fb^{-1} of data collected by the Belle II experiment. This study was completed as part of this thesis work and was published in *Physical Review Letters* in June 2023 [46]. The work then described within this thesis outlines the progress towards an updated study, now using 362fb^{-1}.

A basic overview of particle physics at collider experiments is provided, and the Belle II experiment itself, along with its various subsystems are discussed. A study is presented that details the reinterpretation of the measured branching fraction of the $B \to K \nu \bar{\nu}$ decay process. This focuses on a measurement by the Belle II collaboration that finds a BF at ~ 3 standard deviations above the SM prediction [63]. This is reinterpreted, along with measurements by the BaBar collaboration, under a light new physics model in which the observed excess is explained by the two-body $B^+ \to K^+ X$ decay process. This is then used to constrain parameters on a Z′ model. The results are published in Physical Review D [62].

In part II of this thesis, two studies that investigate the performance of the Belle II detector are presented. The first looks at measuring the efficiency of the cluster-track matching algorithm 1, for which a tag-and-probe method is utilised to study radiative bhabha events. The rate at which the probe side tracks are mis-reconstructed as photons is calculated in both data and MC, and a subsequent per-track uncertainty value is defined that can then be propagated to other physics studies. The second performance study details the measurement of the total efficiency of the trigger lines adopted for the

Z′ and muonphilic scalar searches 2. These are studied by way of comparing with an orthogonal trigger line, and the efficiencies in both data and MC are calculated across a few key variables. Required correction factors to be applied to MC events are then defined as a function of the minimum track transverse momentum.

In the third chapter of part II 3, the development of the Punzi-Loss function is discussed. This loss function is based on the Punzi figure-of-merit, which provides a method of measuring the sensitivity to possible new physics signals. A neural network that is then trained using this loss function (dubbed the Punzi-Net) introduces notable improvements in sensitivity to new physics signals when compared to more commonly used methods. The Punzi-Net provides a single classifier, with a single optimal cut value to the output for a range of new physics signal hypotheses spanning a wide parameter space. This study was published in the European Physics Journal [73].

The final part of this thesis details the full analysis procedure in the search for the Z′ and the muonphilic scalar S. The data sets utilised, both simulated and real, are defined and the basic initial pre-selection criteria is outlined, with the resulting distribution of background events then shown. The final selection process using the Punzi-Net (or BDT for the muonphilic scalar search) is outlined, with the input variables described and the expected final distribution of events given. A detailed analysis of the surviving radiative $\mu^+\mu^-$ events is provided. The MC modelling of this background source was found to be quite inaccurate in the 2023 iteration of the study and thus it is critically important to understand the mechanisms by which they are produced. The implementation of an additional KLM-based photon veto that provides further robustness to such backgrounds is described.

Chapter 3.1 then outlines the three control channels that are used to study various aspects of the data and MC agreement, and derive subsequent systematic uncertainties from sources such as the Punzi-Net, detector resolution modelling or the photon-veto selection. Finally the statistical analysis of the data is outlined in Chapter 4. The method of hypothesis testing by way of a binned maximum likelihood fit is introduced and the chosen approach to binning and template construction is defined, for both the vanilla L_μ-L_τ and dark Z′ models. The implementation of the defined systematic uncertainties as nuisance parameters in the binned maximum likelihood fit is described, where sources are attributed as shape or normalisation uncertainties on the fit templates.

Finally, the expected 90% CL upper limits on cross section of the invisibly decaying Z′ boson are shown, under the vanilla L_μ-L_τ model assumption (fig. 4.5). These are then translated to limits on the g′ coupling strength (fig. 4.6), which are display along with competitive measurements from other experimental efforts. The expected 90% CL upper limits on cross section and g'_D coupling strength are then presented under the dark Z′ model in figures 4.7 and 4.8, were results are shown for assumed DM couplings of α_D = 0.01, 0.05, 0.1, 0.25 and 0.5. Again these are shown with relevant limits from other experimental analyses included. Finally the cross section of the invisible decay of the muonphilic dark scalar S and subsequent coupling strength g_S (figures 4.9 and 4.10) are shown for a set of three mass hypotheses. Checks are conducted using toy simulated data with injected signal events, and considerations of the look elsewhere effect are detailed.

Discussion and outlook

The projected limits set on the g′ coupling constant under the vanilla L_μ-L_τ (fig 4.6) provide highly competitive limits in the mass range between 2 and 5 GeV/c^2, slightly exceeding those set by the Belle II study that searched for a visible decay to a muon pair. This analysis did utilise about half the data sample ($178 fb^{-1}$) and so the advantage is mainly statistical in nature.

Under the dark Z′ model however this study provides one of the only direct search results, providing new leading upper limits on the g'_D coupling constant in a direct search for an invisibly decaying Z′ that couples to dark matter with non-negligible decay width 4.8. Finally the limits set on the g_S coupling strength to a muonphilic scalar boson (fig. 4.10) are found to be competitive with those published by the BES-III collaboration in the mass range $M_S \lesssim 0.2$ GeV/c^2. As discussed the analysis is still undergoing internal review within the Belle II collaboration and is expected to go ahead with unblinding in the near future.

The Belle II experiment is expected to collect multiple inverse attobarns of integrated luminosity over its full runtime. With that, one could certainly hope to greatly expand on the limits presented here. While the door appears to be closing on the muon g-2 anomaly, with the newer theoretical predictions moving towards agreement with experimental results, the mystery of dark matter remains stubborn. Further study of the dark Z′ invisible decay model with this increased data set can at least serve to constrain our understanding of it, if not reveal to us its very nature.

Bibliography

[1] M. Thomson. *Modern Particle Physics*. University Printing House, Cambridge, United Kingdom: Cambridge University Press, 2019.

[2] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helvetica Physica Acta* 6 (Jan. 1933), pp. 110–127.

[3] K. Freese. "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars". In: *EAS Publications Series* 36 (2009), pp. 113–126. DOI: 10.1051/eas/0936016. URL: https://doi.org/10.1051%5C%2Feas%5C%2F0936016.

[4] A. Arbey and F. Mahmoudi. "Dark matter and the early Universe: A review". In: *Progress in Particle and Nuclear Physics* 119 (July 2021), p. 103865. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2021.103865. URL: http://dx.doi.org/10.1016/j.ppnp.2021.103865.

[5] N. Aghanim et al. "Planck2018 results: I. Overview and the cosmological legacy ofPlanck". In: *Astronomy amp; Astrophysics* 641 (Sept. 2020), A1. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833880. URL: http://dx.doi.org/10.1051/0004-6361/201833880.

[6] Mordehai Milgrom. *MOND–a pedagogical review*. 2001. arXiv: astro-ph/0112069 [astro-ph]. URL: https://arxiv.org/abs/astro-ph/0112069.

[7] Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". In: *The Astrophysical Journal* 648.2 (Aug. 2006), pp. L109–L113. ISSN: 1538-4357. DOI: 10.1086/508162. URL: http://dx.doi.org/10.1086/508162.

[8] L. Perivolaropoulos and F. Skara. "Challenges for ΛCDM: An update". In: *New Astronomy Reviews* 95 (Dec. 2022), p. 101659. ISSN: 1387-6473. DOI: 10.1016/j.newar.2022.101659. URL: http://dx.doi.org/10.1016/j.newar.2022.101659.

[9] Jonathan L. Feng. "Dark Matter Candidates from Particle Physics and Methods of Detection". In: *Annual Review of Astronomy and Astrophysics* 48.1 (Aug. 2010), pp. 495–545. ISSN: 1545-4282. DOI: 10.1146/annurev-astro-082708-101659. URL: http://dx.doi.org/10.1146/annurev-astro-082708-101659.

[10] P. S. Bhupal Dev, Anupam Mazumdar, and Saleh Qutub. "Constraining non-thermal and thermal properties of Dark Matter". In: *Frontiers in Physics* 2 (May 2014). ISSN: 2296-424X. DOI: 10.3389/fphy.2014.00026. URL: http://dx.doi.org/10.3389/fphy.2014.00026.

[11] Giorgio Arcadi et al. *The Waning of the WIMP: Endgame?* 2024. arXiv: 2403.15860 [hep-ph]. URL: https://arxiv.org/abs/2403.15860.

[12] Roberto Franceschini and Xiaoran Zhao. "Going all the way in the search for WIMP dark matter at the muon collider through precision measurements". In: *The European Physical Journal C* 83.6 (June 2023). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11724-3. URL: http://dx.doi.org/10.1140/epjc/s10052-023-11724-3.

[13] C. Bœhm and P. Fayet. "Scalar dark matter candidates". In: *Nuclear Physics B* 683.1–2 (Apr. 2004), pp. 219–263. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2004.01.015. URL: http://dx.doi.org/10.1016/j.nuclphysb.2004.01.015.

[14] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle dark matter: evidence, candidates and constraints". In: *Physics Reports* 405.5–6 (Jan. 2005), pp. 279–390. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2004.08.031. URL: http://dx.doi.org/10.1016/j.physrep.2004.08.031.

[15] Julien Billard et al. "Direct detection of dark matter—APPEC committee report*". In: *Reports on Progress in Physics* 85.5 (Apr. 2022), p. 056201. ISSN: 1361-6633. DOI: 10.1088/1361-6633/ac5754. URL: http://dx.doi.org/10.1088/1361-6633/ac5754.

[16] E. Aprile et al. "First Dark Matter Search with Nuclear Recoils from the XENONnT Experiment". In: *Phys. Rev. Lett.* 131 (4 July 2023), p. 041003. DOI: 10.1103/PhysRevLett.131.041003. URL: https://link.aps.org/doi/10.1103/PhysRevLett.131.041003.

[17] Carlos Pérez de los Heros. "Status, Challenges and Directions in Indirect Dark Matter Searches". In: *Symmetry* 12.10 (Oct. 2020), p. 1648. ISSN: 2073-8994. DOI: 10.3390/sym12101648. URL: http://dx.doi.org/10.3390/sym12101648.

[18] Antonio Boveia and Caterina Doglioni. "Dark Matter Searches at Colliders". In: *Annual Review of Nuclear and Particle Science* 68.1 (Oct. 2018), pp. 429–459. ISSN: 1545-4134. DOI: 10.1146/annurev-nucl-101917-021008. URL: http://dx.doi.org/10.1146/annurev-nucl-101917-021008.

[19] Alex Keshavarzi, Kim Siang Khaw, and Tamaki Yoshioka. "Muon g2: A review". In: *Nuclear Physics B* 975 (2022), p. 115675. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2022.115675. URL: https://www.sciencedirect.com/science/article/pii/S0550321322000268.

[20] T. Aoyama et al. "The anomalous magnetic moment of the muon in the Standard Model". In: *Physics Reports* 887 (2020). The anomalous magnetic moment of the muon in the Standard Model, pp. 1–166. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2020.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0370157320302556.

[21] D. P. Aguillard et al. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm". In: *Phys. Rev. Lett.* 131 (16 Oct. 2023), p. 161802. DOI: 10.1103/PhysRevLett.131.161802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.131.161802.

[22] 2 Collaboration et al. *Combination of KLOE* $\sigma\left(e^+e^- \to \pi^+\pi^-\gamma(\gamma)\right)$ *measurements and determination of* $a_\mu^{\pi^+\pi^-}$ *in the energy range* $0.10 < s < 0.95\ GeV^2$. 2018. arXiv: 1711.03085 [hep-ex]. URL: https://arxiv.org/abs/1711.03085.

[23] J. P. Lees et al. "Precise measurement of the $e^+e^- \to \pi^+\pi^-\gamma$ cross section with the initial-state radiation method at BABAR". In: *Physical Review D* 86.3 (Aug. 2012). ISSN: 1550-2368. DOI: 10.1103/physrevd.86.032013. URL: http://dx.doi.org/10.1103/PhysRevD.86.032013.

[24] F. V. Ignatov et al. "Measurement of the $e+e-ß\pi+\pi-$ cross section from threshold to 1.2 GeV with the CMD-3 detector". In: *Phys. Rev. D* 109.11 (2024), p. 112002. DOI: 10.1103/PhysRevD.109.112002. arXiv: 2302.08834 [hep-ex].

[25] Sz. Borsanyi et al. "Leading hadronic contribution to the muon magnetic moment from lattice QCD". In: *Nature* 593.7857 (May 2021), pp. 51–55. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03418-1. URL: https://doi.org/10.1038/s41586-021-03418-1.

[26] A. Boccaletti et al. *High precision calculation of the hadronic vacuum polarisation contribution to the muon anomaly*. 2024. arXiv: 2407.10913 [hep-lat]. URL: https://arxiv.org/abs/2407.10913.

[27] X. -G. He et al. "New-$Z^{'}$ phenomenology". In: *Phys. Rev. D* 43 (1991), R22. DOI: 10.1103/PhysRevD.43.R22.

[28] Brian Shuve and Itay Yavin. "Dark matter progenitor: Light vector boson decay into sterile neutrinos". In: *Physical Review D* 89.11 (June 2014). ISSN: 1550-2368. DOI: 10.1103/physrevd.89.113004. URL: http://dx.doi.org/10.1103/PhysRevD.89.113004.

[29] Wolfgang Altmannshofer et al. "Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model". In: *J. High Energy Phys.* 12 (2016), p. 106. DOI: 10.1007/JHEP12(2016)106.

[30] Ben Allanach et al. "Z models for the LHCb and g-2 muon anomalies". In: *Physical Review D* 93.5 (Mar. 2016). ISSN: 2470-0029. DOI: 10.1103/physrevd.93.055045. URL: http://dx.doi.org/10.1103/PhysRevD.93.055045.

[31] Wolfgang Altmannshofer et al. "Neutrino Trident Production: A Powerful Probe of New Physics with Neutrino Beams". In: *Physical Review Letters* 113.9 (Aug. 2014). ISSN: 1079-7114. DOI: 10.1103/physrevlett.113.091801. URL: http://dx.doi.org/10.1103/PhysRevLett.113.091801.

[32] R. Aaij et al. "Measurement of lepton universality parameters in $B^{+}\to l^{+}l^{-}$ and $B^{0*0}l^{+}l$ decays". In: *Physical Review D* 108.3 (Aug. 2023). ISSN: 2470-0029. DOI: 10.1103/physrevd.108.032002. URL: http://dx.doi.org/10.1103/PhysRevD.108.032002.

[33] J. P. Lees et al. "Measurement of an excess of $\bar{B}\to D^{(*)}\tau^{-}\bar{\nu}_{\tau}$ decays and implications for charged Higgs bosons". In: *Physical Review D* 88.7 (Oct. 2013). ISSN: 1550-2368. DOI: 10.1103/physrevd.88.072012. URL: http://dx.doi.org/10.1103/PhysRevD.88.072012.

[34] M. Huschle et al. "Measurement of the branching ratio of $\overline{B}\to D^{(*)}\tau^{-}\overline{\nu}_{\tau}$ relative to $\overline{B}\to D^{(*)}\ell^{-}\overline{\nu}_{\ell}$ decays with hadronic tagging at Belle". In: *Phys. Rev. D* 92 (7 Oct. 2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. URL: https://link.aps.org/doi/10.1103/PhysRevD.92.072014.

[35] Wolfgang Altmannshofer et al. "Explaining dark matter and B decay anomalies with an Lμ Lτ model". In: *Journal of High Energy Physics* 2016.12 (Dec. 2016). ISSN: 1029-8479. DOI: 10.1007/jhep12(2016)106. URL: http://dx.doi.org/10.1007/JHEP12(2016)106.

[36] Belle II Collaboration et al. *Search for a $\mu^{+}\mu^{-}$ resonance in four-muon final states at Belle II*. 2024. arXiv: 2403.02841 [hep-ex].

[37] Stefania Gori et al. *Dark Sector Physics at High-Intensity Experiments*. 2022. arXiv: 2209.04671 [hep-ph].

[38] Philip Harris, Philip Schuster, and Jure Zupan. *Snowmass White Paper: New flavors and rich structures in dark sectors*. 2022. arXiv: 2207.08990 [hep-ph].

[39] J. P. Lees et al. "Search for a muonic dark force at BaBar". In: *Phys. Rev. D* 94 (2016), p. 011102. DOI: 10.1103/PhysRevD.94.011102.

[40] T. Czank et al. "Search for $Z^{'}\to\mu^{+}\mu^{-}$ in the $L_{\mu}-L_{\tau}$ gauge-symmetric model at Belle". In: *Phys. Rev. D* 106.1 (2022), p. 012003. DOI: 10.1103/PhysRevD.106.012003.

[41] Albert M Sirunyan et al. "Search for an $L_{\mu}-L_{\tau}$ gauge boson using $Z\to 4\mu$ events in proton-proton collisions at $\sqrt{s}=13$ TeV". In: *Phys. Lett. B* 792 (2019), p. 345. DOI: 10.1016/j.physletb.2019.01.072.

[42] G. Aad et al. "Search for a new Z gauge boson in 4μ events with the ATLAS experiment". In: *Journal of High Energy Physics* 2023.7 (July 2023). ISSN: 1029-8479. DOI: 10.1007/jhep07(2023)090. URL: http://dx.doi.org/10.1007/JHEP07(2023)090.

[43] Yu. M. Andreev et al. "Search for a light Z' in the $L\mu - L\tau$ scenario with the NA64-e experiment at CERN". In: *Phys. Rev. D* 106.3 (2022), p. 032015. DOI: 10.1103/PhysRevD.106.032015.

[44] Yu. M. Andreev et al. *Exploration of the Muon $g-2$ and Light Dark Matter explanations in NA64 with the CERN SPS high energy muon beam.* 2024. arXiv: 2401.01708 [hep-ex]. URL: https://arxiv.org/abs/2401.01708.

[45] I. Adachi et al. "Search for an Invisibly Decaying Z' Boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ Plus Missing Energy Final States". In: *Phys. Rev. Lett.* 124 (14 2020), p. 141801. DOI: 10.1103/PhysRevLett.124.141801.

[46] I. Adachi et al. "Search for an Invisible Z′ in a Final State with Two Muons and Missing Energy at Belle II". In: *Physical Review Letters* 130.23 (June 2023). ISSN: 1079-7114. DOI: 10.1103/physrevlett.130.231801. URL: http://dx.doi.org/10.1103/PhysRevLett.130.231801.

[47] BESIII Collaboration et al. *Search for a muonphilic scalar X_0 or vector X_1 via $J/\psi \to \mu^+\mu^-$ + invisible decays at BESII.* 2024. arXiv: 2311.01076 [hep-ex]. URL: https://arxiv.org/abs/2311.01076.

[48] Simon Knapen and Steven Lowette. *A guide to hunting long-lived particles at the LHC.* 2022. arXiv: 2212.03883 [hep-ph]. URL: https://arxiv.org/abs/2212.03883.

[49] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. URL: https://doi.org/10.1016%2Fj.physletb.2012.08.020.

[50] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. URL: https://doi.org/10.1016%2Fj.physletb.2012.08.021.

[51] K. Abe et al. "Observation of Large CP Violation in the Neutral B Meson System". In: *Physical Review Letters* 87.9 (Aug. 2001). ISSN: 1079-7114. DOI: 10.1103/physrevlett.87.091802. URL: http://dx.doi.org/10.1103/PhysRevLett.87.091802.

[52] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. "SuperKEKB collider". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (Nov. 2018), pp. 188–199. ISSN: 0168-9002. DOI: 10.1016/j.nima.2018.08.017. URL: http://dx.doi.org/10.1016/j.nima.2018.08.017.

[53] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097. eprint: https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf. URL: https://doi.org/10.1093/ptep/ptac097.

[54] The Belle II Collaboration et al. *Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*. 2024. arXiv: 2407.00965 [hep-ex]. URL: https://arxiv.org/abs/2407.00965.

[55] E Kou et al. "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). DOI: 10.1093/ptep/ptz106. URL: https://doi.org/10.1093%2Fptep%2Fptz106.

[56] Katsuro Nakamura et al. "The Belle II SVD detector". In: *PoS* Vertex 2016 (2017), p. 012. DOI: 10.22323/1.287.0012.

[57] Marc Neu et al. "Real-Time Graph Building on FPGAs for Machine Learning Trigger Applications in Particle Physics". In: *Computing and Software for Big Science* 8.1 (Mar. 2024). ISSN: 2510-2044. DOI: 10.1007/s41781-024-00117-0. URL: http://dx.doi.org/10.1007/s41781-024-00117-0.

[58] Valerio Bertacchi et al. "Track finding at Belle II". In: *Computer Physics Communications* 259 (Feb. 2021), p. 107610. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2020.107610. URL: http://dx.doi.org/10.1016/j.cpc.2020.107610.

[59] Umberto Tamponi. *The TOP counter of Belle II: status and first results*. 2018. arXiv: 1811.04532 [hep-ex]. URL: https://arxiv.org/abs/1811.04532.

[60] Milesi, Marco, Tan, Justin, and Urquijo, Phillip. "Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers". In: *EPJ Web Conf.* 245 (2020), p. 06023. DOI: 10.1051/epjconf/202024506023. URL: https://doi.org/10.1051/epjconf/202024506023.

[61] T. Kuhr et al. "The Belle II Core Software". In: *Computing and Software for Big Science* 3.1 (Nov. 2018), p. 1. ISSN: 2510-2044. DOI: 10.1007/s41781-018-0017-9. URL: https://doi.org/10.1007/s41781-018-0017-9.

[62] Wolfgang Altmannshofer et al. "Light new physics in $B \to K^{(*)}\nu\bar{\nu}$?" In: *Phys. Rev. D* 109 (7 Apr. 2024), p. 075008. DOI: 10.1103/PhysRevD.109.075008. URL: https://link.aps.org/doi/10.1103/PhysRevD.109.075008.

[63] Belle II Collaboration et al. *Evidence for $B^+ \to K^+\nu\bar{\nu}$ Decays*. 2024. arXiv: 2311.14647 [hep-ex].

[64] Eldar Ganiev. (On behalf of the Belle II Collaboration). Talk given at EPS-HEP 2023 "On Radiative and Electroweak Penguin Decays", Hamburg, DESY. 2023.

[65] J. P. Lees et al. "Search for $B \to K^{(*)}\nu\overline{\nu}$ and invisible quarkonium decays". In: *Phys. Rev. D* 87 (11 June 2013), p. 112005. DOI: 10.1103/PhysRevD.87.112005. URL: https://link.aps.org/doi/10.1103/PhysRevD.87.112005.

[66] P. del Amo Sanchez et al. "Search for the rare decay $B \to K\nu\overline{\nu}$". In: *Phys. Rev. D* 82 (11 Dec. 2010), p. 112002. DOI: 10.1103/PhysRevD.82.112002. URL: https://link.aps.org/doi/10.1103/PhysRevD.82.112002.

[67] P. del Amo Sanchez et al. "Search for the rare decay $B \to K\nu\overline{\nu}$". In: *Phys. Rev. D* 82 (11 Dec. 2010), p. 112002. DOI: 10.1103/PhysRevD.82.112002. URL: https://link.aps.org/doi/10.1103/PhysRevD.82.112002.

[68] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *scikit-hep/pyhf*. Version v0.5.3. Oct. 2020. DOI: 10.5281/zenodo.4110938. URL: https://doi.org/10.5281/zenodo.4110938.

[69] Lukas Heinrich et al. "pyhf: pure-Python implementation of HistFactory statistical models". In: *Journal of Open Source Software* 6.58 (2021), p. 2823. DOI: 10.21105/joss.02823. URL: https://doi.org/10.21105/joss.02823.

[70] W. G. Parrott, C. Bouchard, and C. T. H. Davies. "Standard Model predictions for $B \to K^+$, $B \to K_{1\,2}^{-+}$ and $B \to \nu\nu$ using form factors from $N_f = 2+1+1$ lattice QCD". In: *Physical Review D* 107.1 (Jan. 2023). ISSN: 2470-0029. DOI: 10.1103/physrevd.107.014511. URL: http://dx.doi.org/10.1103/PhysRevD.107.014511.

[71] Wolfgang Altmannshofer et al. "Light resonances and the low-q^2 bin of R_{K^*}". In: *Journal of High Energy Physics* 2018.3 (Mar. 2018). ISSN: 1029-8479. DOI: 10.1007/jhep03(2018)188. URL: http://dx.doi.org/10.1007/JHEP03(2018)188.

[72] S. Bähr et al. *The Neural Network First-Level Hardware Track Trigger of the Belle II Experiment*. 2024. arXiv: 2402.14962 [hep-ex]. URL: https://arxiv.org/abs/2402.14962.

[73] F. Abudinén et al. "Punzi-loss: A non-differentiable metric approximation for sensitivity optimisation in the search for new particles". In: *The European Physical Journal C* 82.2 (Feb. 2022). DOI: 10.1140/epjc/s10052-022-10070-0. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-022-10070-0.

[74] Giovanni Punzi. *Sensitivity of searches for new signals and its optimization*. 2003. arXiv: physics/0308063 [physics.data-an].

[75] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035.

[76] Johan Alwall et al. "MadGraph 5: going beyond". In: *Journal of High Energy Physics* 2011.6 (June 2011). ISSN: 1029-8479. DOI: 10.1007/jhep06(2011)128. URL: http://dx.doi.org/10.1007/JHEP06(2011)128.

[77] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785. URL: http://dx.doi.org/10.1145/2939672.2939785.

[78] Tao Han, Ian-Woo Kim, and Jeonghyeon Song. "Kinematic cusps: Determining the missing particle mass at colliders". In: *Physics Letters B* 693.5 (2010), pp. 575–579. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2010.09.010. URL: https://www.sciencedirect.com/science/article/pii/S0370269310010610.

[79] Francesca Acanfora et al. *Fusing photons into nothing, a new search for invisible ALPs and Dark Matter at Belle II*. 2023. arXiv: 2307.06369 [hep-ph]. URL: https://arxiv.org/abs/2307.06369.

[80] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[81] Ayuki Kamada et al. "Self-interacting dark matter and muon (g 2) in a gauged $U(1)_{L_\mu - L_\tau}$ model". In: *Journal of High Energy Physics* 2018.6 (June 2018). ISSN: 1029-8479. DOI: 10.1007/jhep06(2018)117. URL: http://dx.doi.org/10.1007/JHEP06(2018)117.

[82] G. Bellini et al. "Precision Measurement of the ^{7}Be Solar Neutrino Interaction Rate in Borexino". In: *Physical Review Letters* 107.14 (Sept. 2011). ISSN: 1079-7114. DOI: 10.1103/physrevlett.107.141302. URL: http://dx.doi.org/10.1103/PhysRevLett.107.141302.

[83] Eilam Gross and Ofer Vitells. "Trial factors for the look elsewhere effect in high energy physics". In: *The European Physical Journal C* 70.1–2 (Oct. 2010), pp. 525–530. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1470-8. URL: http://dx.doi.org/10.1140/epjc/s10052-010-1470-8.