An electromagnetic calorimeter cluster classifier as a beam background suppressant in $B^0 \to D^{*+}\ell^-\nu_\ell$ decays at the Belle II experiment.

Kyle Amirie

Department of Physics
McGill University
Montreal, Canada

November, 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Abstract

The presence of beam background-related photons in the extra energy distributions of events reconstructed using Belle II data can make it difficult to accurately identify energy missing from the detector that can be attributable to neutrinos or other hypothesized weakly interacting particles. By isolating samples of detector information relating to these beam background photons, and samples relating to photons from the intended collision event (signal photons), it is possible to perform a multivariate analysis using a boosted decision tree approach to train a classifier variable for distinguishing between the two types of photons. This classifier can be used to suppress the inclusion of beam background-related photons in extra energy distributions.

Six classifier tools were trained on electromagnetic calorimeter information relating to beam background and signal photons from $e^+e^- \to \gamma \to \mu^+\mu^-$ and $B^0 \to D^{*+}\ell^-\nu_\ell$ events. These tools were validated as a beam background suppressant when used in reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events to evaluate their performance. Described within is a full description of all six trained classifiers, an examination of the impact training sample characteristics have on their performance, and a number of possible future optimizations to this approach to beam background photon suppression.

Abrégé

Il y a des photons qui viennent du bruit dans le faisceau présents dans la distribution d'énergie supplémentaire des évènements qui ont été reconstruits en utilisant les données du détecteur Belle II. Par conséquent c'est difficile d'identifier d'une manière précise l'énergie manquante du détecteur qui peut être attribuée aux neutrinos ou aux autres particules hypothétiques interagissant faiblement. On peut isoler des échantillons d'information du détecteur associés aux photons de bruit du faisceau et ceux qui viennent des évènements de collisions attendues (les photons du signal). Avec ces informations c'est possible de faire une analyse multivariée en utilisant un arbre de décisions renforcé pour entrainer une variable classificatrice afin de faire la distinction entre les deux types de photons. Ce classificateur peut être utilisé pour supprimer les photon de bruit du faisceau, présent dans les distributions d'énergie supplémentaire.

Six outils classificateurs ont été entraînés avec des informations d'un calorimètre électromagnétique associées aux photons de bruit et du signal provenant des évènements $e^+e^- \to \gamma \to \mu^+\mu^-$ et $B^0 \to D^{*+}\ell^-\nu_\ell$. Ces outils ont démontré leur capacité de supprimer le bruit du faisceau quand on a évalué leur performance lors de la reconstruction des évènements $B^0 \to D^{*+}\ell^-\nu_\ell$. Cette thèse décrit complètement les six classificateurs, examine l'impact que les caractéristiques des échantillons de l'entrainement ont sur leur performance et décrit plusieurs optimisations futures possibles.

Acknowledgements

This thesis, and the work described herein, was made possible by the diligence, ingenuity, and hard work of the members of the Belle II collaboration over many years.

I would like to thank my supervisor Steven Robertson, not only for providing an opportunity for me to join the Belle II collaboration, but for the support and guidance he has provided over the past two years. I would also like to thank Racha Cheaib, whose experience, knowledge, and patience played an essential role in my completion of this thesis, and whose guidance was invaluable.

In no particular order, I would also like to thank Andrea Fodor, Hannah Wakeling, Trevor Shillington, Ryan MacGibbon, and Andreas Warburton for their willingness to provide advice and assistance over the past two years. I would also like to thank Michelle Sullivan for translating the abstract, and Fritz Buchinger for allowing me an opportunity to attend McGill University.

Preface

Unless otherwise stated, the work presented in this thesis has been completed independently by the author using the analysis framework provided by the Belle II collaboration. This framework includes the machine learning tools, analysis software, and data sets created by other contributors to the Belle II collaboration. As there are no co-authors of this thesis, all six chapters represent original content contributed by the author, with any references to the work of others accompanied by a citation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The study of matter and the fundamental forces that make up our universe is a key driver behind many areas of technological progress and the development of a better understanding of the world around us. The field of particle physics focuses on the examination of elementary particles and the interactions that occur between them, which often requires the use of particle accelerators and detectors like those in the Belle II experiment. This thesis presents work undertaken to optimize the analysis of information collected by the Belle II detector.

The current framework used to explain most modern particle physics observations is known as the Standard Model. The Standard Model describes nearly every observed particle physics phenomenon with a high degree of success, from the behaviour of elementary particles to the electromagnetic, weak, and strong force interactions they participate in. Characteristics of this model are often verified with increasing precision by the observation of unstable or short-lived particles at particle colliders.

When two relatively stable particles collide with sufficient energy it is possible for the collision to result in the production of such short-lived particles (within the constraints of momentum and energy conservation). Depending on the forces involved these particles can quickly decay into a number of lower mass particles that can in turn decay further. Many particle accelerators have been constructed over the past sixty years to collide particles at a carefully controlled center-of-mass energy in order to produce a large quantity of these short-lived particles for observation [1]. Progressively more sophisticated detectors have been used over time to identify the decay products of these particles in order to, with understandings provided by the Standard Model, reconstruct a chain of particle decays leading back to the original collision. Analysis of these particle decay chains provides information that can be used to measure characteristics of the observed particles, such as their lifetimes and decay branching ratios, which can in turn be used to further refine our understanding of particle behaviour.

The Belle II detector is one such apparatus and is currently collecting data at the SuperKEKB accelerator in the High Energy Accelerator Research Organization (KEK) particle physics laboratory in Tsukuba, Japan. The SuperKEKB accelerator collides electrons and positrons at 10.58 GeV in the center-of-mass frame, corresponding to the mass of the Υ(4S) resonance. The Υ(4S) from each collision event quickly decays into two B mesons, composed of a heavy bottom quark (or antiquark) and a lighter up or down antiquark (or quark), which can subsequently decay into a number of different particles. The Belle II detector is comprised of multiple sub-detector components that track charged particle trajectories and measure photon and electron energy in order to identify the particles passing through the detector and reconstruct decay chains leading back to the Υ(4S).

A key feature in the analysis of such decay chains is the handling of missing and extra energy. In an ideal decay reconstruction, the sum of the energies of detected decay products will equal the center-of-mass energy of the initial collision event. However, it is difficult to directly detect weakly-interacting particles like neutrinos that don't interact with the detector as they pass through it, so it is often necessary to infer their presence in a particle decay. This is typically accomplished by looking for a deficit of energy in the decay products, or missing energy, relative to the total energy of the initial collision, that can be attributed to such a particle. Conversely, background photons originating from particle interactions within the detector other than the Υ(4S) decay can be detected at the same time as Υ(4S) decay products. This can lead to a total energy attributed to the decay products of a given $\Upsilon(4S)$ that is greater than the center-of-mass energy of the initial collision. It is often imperative that such background-related extra energy can be distinguished from decay-related energy in order to correctly identify the amount of missing energy in a decay chain and determine if undetected particles were present. The work presented in this thesis has been conducted in an attempt to provide a method for generating such a distinguishing power for particles reconstructed using data collected by the Belle II detector. A multivariate analysis tool was trained on information collected by the detector's electromagnetic calorimeter to create a probabilistic classifier variable that indicates the likelihood that each observed photon is associated with beam-related backgrounds.

Chapter 2 provides an overview of the physics relevant to the Belle 2 experiment and the particle interactions that are reconstructed with data collected by it. Chapter 3 provides a structural overview of the SuperKEKB accelerator and the Belle 2 detector. The analysis tools and framework used to train the classifier tool are detailed in Chapter 4, while the results of the training and application of the classifier to reconstructed B meson decays are presented in Chapter 5. A brief summary concluding the thesis follows in Chapter 6.

Chapter 2

Background

The work presented in this thesis was undertaken to create a tool that can be used to efficiently identify if energy collected by the Belle II detector that is unused in particle reconstruction is attributable to beam-related backgrounds. The understanding of particle behaviour within the detector and particle decay chain reconstruction that is necessary for the development of this tool is currently provided by the Standard Model of particle physics, described briefly in this section. The full motivation behind this study and the machine learning principles used to develop this tool are also discussed.

2.1 The Standard Model of particle physics

The Standard Model of particle physics is a theoretical framework that has been developed over many years to provide an explanation for observed physics phenomena that relate to elementary particles. The theory is defined by an internal SU(3)×SU(2)×U(1) gauge symmetry, providing a description of fundamental particle interactions and allowing for a meaningful categorization of elementary particles. The SU(3) symmetry group governs quantum chromodynamics and strong force interactions, while the SU(2)×U(1) symmetry groups together govern electromagnetic and weak force interactions. In this model, the particles that mediate these interactions and act as force carriers have an integer intrinsic angular momentum (also known as spin) and are categorized as gauge bosons. Elementary particles with half-integer spin are categorized as fermions. Hadrons, composite particles comprised of multiple quarks or antiquarks (which are fermions), are similarly categorized; hadrons consisting of an even total number of quarks and antiquarks have an integer spin and are called mesons, while those consisting of an odd total number of quarks and antiquarks have a half-integer spin and are called baryons.

The Standard Model does not explain all physics phenomena. It does not reconcile the gravitational force with other fundamental forces and it does not provide an explanation for a number of cosmological observations, like the universe's accelerating expansion. It also provides no basis for predicting the existence of dark matter, a type of matter that has not yet been directly observed despite mounting evidence that it fills the universe [2]. It does, however, act as a rigorous and nearly complete framework for the study of observed particle physics phenomena.

2.1.1 Fundamental interactions and gauge bosons

Although the Standard Model does not provide a description of gravitational interactions, it is able to provide a description of electromagnetic, weak, and strong interactions, as well as the gauge bosons that mediate them. A summary of these forces can be found at the end of the section, in Table 2.1.

The electromagnetic force and photons

In classical theory, any particle with an electric charge is the source of an electric field, and any moving charged particle will generate a magnetic field. These two fields are known together as an electromagnetic field and serve to carry electromagnetic forces. The electromagnetic force is not only responsible for binding electrons to baryonic nuclei to form atoms, it is also responsible for the attractive force that allows chemical bonds to form between oppositely charged ions and is necessary for the formation of the molecules that constitute most stable matter. As a result, electromagnetic fields are a widely studied and exploited physics phenomena that play an integral role in many aspects of modern society.

In quantum field theory, forces are viewed to arise from the exchange of gauge bosons. In the case of electromagnetism, as described by quantum electrodynamics (QED), photons serve as the gauge boson force carrier for all electromagnetic interactions. For example, when an electron and positron interact they can deflect, or scatter, off each other in a process known as Bhabha scattering. Rather than simply interacting through field lines as in classical theory, the electromagnetic force felt between the two leptons is said to be communicated by a virtual photon (Fig 2.1).

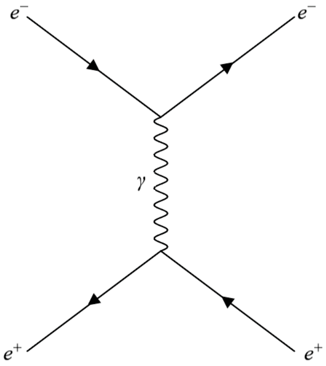


Figure 2.1: A Feynman diagram of the electron and positron interaction known as Bhabha scattering. The virtual photon is inherently un-observable and acts as a force carrier between the two particles.

The photon, γ, is a massless, electrically neutral elementary particle with a spin of 1. It exhibits a type of wave-particle duality, having behaviour that can be described mathematically with wave equations but interacting with matter in the discretized manner most often associated with particles.

For any of the fundamental forces the strength of the interaction relative to the strongest force can be described numerically by a unitless coupling constant, α. The coupling constant for electromagnetic interactions (also known as the fine-structure constant) has many physical interpretations, but perhaps can be best interpreted as a measure of the strength of interactions between charged particles and photons. The fine-structure constant is not numerically determined by the standard model, but rather has been measured experimentally [3]. For electromagnetic interactions around the 1 GeV energy level α_{EM} = $\frac{1}{137}$, a value that places the electromagnetic force above the weak force and below the strong force in regards to field strength. Although electromagnetic forces can theoretically act at an infinite range, their magnitude decreases with the square of the separation distance between the interacting particles.

One of the electromagnetic interactions that is highly relevant to the Belle II experiment and many other particle detectors is the emission of bremsstrahlung radiation. When a moving charged particle is slowed down due to interactions with the electromagnetic field of a nearby particle, the lost kinetic energy is released in the form of a radiated photon. Experiments make use of detector mediums designed to emit large quantities of such bremsstrahlung radiation in order to generate a signal from collected photons and identify when a charged particle has passed through the detector. At high energies, the amount of energy a charged particle loses in the form of bremsstrahlung radiation per unit of distance travelled within a medium, $\frac{-dE}{dx}$, can be approximated as [4]:

$$\frac{-dE}{dx} = 4\alpha \left(\frac{e^2}{4\pi\epsilon_0 M c^2} \right)^2 \left(\frac{N_A Z^2}{A} \right) z^2 E \ln \left(\frac{183}{Z^{1/3}} \right), \tag{2.1}$$

where α is the electromagnetic coupling constant, *e* is the elementary charge, ϵ_0 is the vacuum permittivity constant, *M*, *z*, and *E* are the mass, charge, and energy of the incident particle, *c* is the speed of light, N_A is the Avogadro constant, , and Z and A are the atomic number and mass of the nearby particle. As the energy lost is inversely proportional to the square of the charged particle mass, bremsstrahlung radiation is often the dominant form of energy loss for relatively light particles like electrons and positrons passing through a material.

The weak force, W bosons, and Z bosons

Of the three fundamental forces described by the standard model, the weak force is the most subtle. It effectively only acts on sub-atomic distance scales and has a coupling constant of around $\alpha_W = 10^{-6}$, making it considerably weaker than strong or electromagnetic forces over a similar distance. Weak interactions are responsible for the forces that mediate nuclear beta decay, the process by which a neutrino and an electron or positron are radiated from an atomic nucleus during the conversion between a proton and neutron (Fig 2.2). These types of interactions facilitate many hadronic decays, playing an important role in processes like nuclear fission, and are required for any particle interaction that includes changes in quark flavour. Although electromagnetic interactions are governed by conservation laws that preserve quantum numbers across each interaction, weak interactions have been observed to violate parity, isospin, and strangeness conservation.

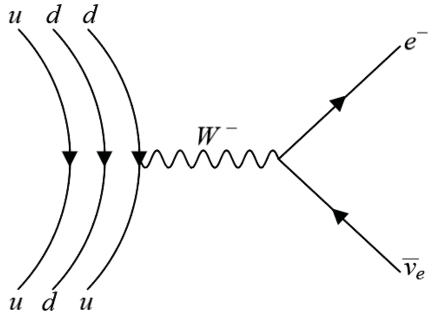


Figure 2.2: A Feynman diagram of beta decay, a common weak interaction that plays an important role in nuclear fission. One of the down quarks in the neutron transitions to an up quark by the emission of a W^- boson, creating a proton. The emitted W^- boson quickly decays into an electron and electron antineutrino.

There are three gauge bosons that act as force carriers for weak interactions, all with a lifetime on the order of 10^{-25} s. The electrically charged W^+ and its antiparticle W^- mediate interactions involving changes in lepton or quark flavour. The neutral Z^0 boson (which is its own antiparticle) is a mediator for neutrino scattering or other flavour-conserving lepton interactions [5]. Although gauge bosons are predicted by gauge theory to be massless [6], all three weak mediators have a relatively large mass (80.379 GeV/c^2 for the W bosons, 91.188 GeV/c^2 for the Z boson) which is responsible for the short range of weak force interactions.

The generation of this mass has been attributed to the Higgs mechanism, a component of the Standard Model that explains the separation of the electroweak force into distinct electromagnetic and weak forces, and provides an explanation for the presence of mass in many particles without violating the fundamental tenets of gauge theory [7]. It is postulated that in high temperature environments, like the early universe, all elementary particles are massless and electromagnetic and weak interactions are indistinguishable. As temperatures cool, a quantum field known as the Higgs field develops a non-zero vacuum expectation value and spontaneous symmetry breaking occurs, allowing interactions between elementary particles and the Higgs field to generate mass. Of the four degrees of freedom in the Higgs field, three mix with components of electroweak bosonic fields to generate mass in the three types of weak force gauge bosons. The electroweak bosonic field components that do not interact with the field remain massless and become photons in this lower energy environment, allowing electromagnetic and weak interactions to became distinct from one another. The fourth degree of freedom in the Higgs field results in a massive elementary scalar boson. The existence of this scalar boson, termed the Higgs boson, was experimentally verified in 2012 [8].

The strong nuclear force and gluons

The strong nuclear force, as the name implies, is the strongest of the fundamental forces at atomic distances. At a 1 GeV energy level, the strong force has a range of about 10^{-15}m and a coupling constant $\alpha_S = 1$, making it approximately 137 times stronger than the electromagnetic force and 10^6 times stronger than the weak force [9]. The strong force is responsible for binding quarks together to form hadrons, as well as the nuclear binding force that holds protons and neutrons together to form atomic nuclei.

The gauge boson that acts as the primary force carrier of these interactions is the gluon, a massless particle with a spin of 1. Although gluons are electrically neutral, they carry a combination of 'colour' charges, a conserved strong force analog to electric charge. There are six possible colour charges, each given a label by convention that does not reflect any particular physical characteristic; red, anti-red, blue, anti-blue, green, and anti-green. Gluons always carry a combination of one colour and one anti-colour and can exist in one of eight different linearly independent colour combination states. They interact with quarks, antiquarks, and other gluons through the exchange of this colour charge.

The field theory that governs these colour interactions is quantum chromodynamics (QCD), which provides a basis for asymptotic freedom and colour confinement. Asymptotic freedom is a gauge theory mechanism that suggests that the strength of a field will drop off asymptotically at a certain energy level or distance, allowing particles beyond that point to effectively become free of the field's influence. The presence of asymptotic freedom can be determined by examining the sign of the beta-function, β, that describes the variation in the field's coupling constant:

$$\beta = \frac{\alpha}{\pi}(\frac{-11N}{6} + \frac{n_f}{3}), \tag{2.2}$$

where α is the field coupling constant, *N* is the gauge group degree, and n_f is the number of flavours within the gauge group. As there are six known quark flavours and *N* = 3 for SU(3), this beta function is negative and indicates that particles interacting with a strong field will experience a sharp, decreased influence from the field as they move further away or energy decreases.

Although an analytical proof has not yet been fully formulated, colour confinement dictates that no particle can exist in isolation with a net colour charge, suggesting that quarks and gluons cannot be observed in isolation under normal conditions. As a quark and antiquark pair move apart from each other, the force communicated between them by a gluon does not diminish as it would in electromagnetic interactions and it eventually becomes energetically favourable for a second quark-antiquark pair to be produced to split the body into two quark pairs. In experiments or environments where quarks would be produced, this effect leads to the creation of hadrons rather than individual quarks (Fig. 2.3).

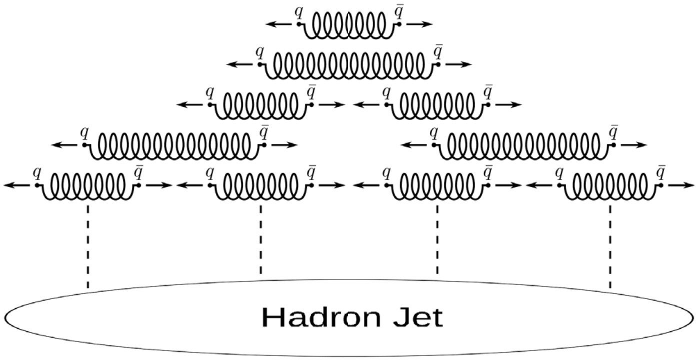


Figure 2.3: A diagram of the formation of hadron jets, one consequence of colour confinement. It becomes energetically favourable for the production of quark-antiquark pairs, which reduces the overall distance between quarks, before the quarks can be naturally isolated [10].

2.1.2 Fermions

There are two known categories of elementary particles with half-integer spin: leptons and quarks.

Leptons

Although both groups of elementary fermions contain six particles and six antiparticles separated into three generations, leptons are distinguished by a lack of interaction via the strong force and, consequently, a capacity to exist independently. Each lepton generation can be further divided into groups of charged and neutral particles. The first generation contains the electron, e^-, positron, e^+, the electron neutrino, ν_e, and the electron antineutrino, $\bar{\nu}_e$. The second generation contains the muon, μ^-, muon neutrino, ν_μ, and their associated antiparticles, while the third generation contains the tau, τ^-, tau neutrino, ν_τ, and their related antiparticles. All charged leptons have a charge of ± 1 and are capable of participating in electromagnetic and weak interactions. Neutrinos are electromagnetically neutral and have only been observed to interact via the weak force. See Table 2.2 at the end of the section for a summary of the leptons and their characteristics.

Compared to the other charged leptons, electrons and positrons are light, long-lived particles. Their stability allows them to play an essential role in the formation of matter and to participate in many other electromagnetic phenomena. Electrons emitted from matter via the photoelectric effect or any other ionization process are used in various types of particle detectors, and both electrons and positrons play a fundamental role in beta decay. A common source of these leptons in nature is the pair production that occurs when a photon interacts with a nearby atomic nucleus and splits into an electron and positron.

The negatively charged muon and positively charged antimuon that make up the second generation of charged leptons are over 200 times more massive than electrons and positrons and have an average lifetime of 2.2 x 10^{-6} s. For an unstable particle, this is a relatively long lifetime. Muons are able to exhibit this lifetime because there are few lower-mass particles for them to decay into while still obeying conservation laws, and they only decay through weak interactions (which take longer than decays via electromagnetic or strong interactions). Given that the amount of energy that particles lose to bremsstrahlung radiation is inversely proportional to their mass, the large mass of muons (relative to electrons) and the long lifetime they exhibit allow them to penetrate into materials much further than electrons and positrons. Muons are produced in nature most commonly by the decay of charged pions, which decay into a muon and muon neutrino 99.99% of the time [11]. However, the lifetime of the pion is such that pions produced in an accelerator experiment can often exit a detector before decaying into muons, preventing the muons from being detected. As a result, most muons detected in such experiments are those that arose from electromagnetic interactions or weak decays.

The tau and antitau particles exhibit a lifetime that is many orders of magnitude shorter than first or second generation charged leptons. They have a mass that is approximately 3500 times larger than an electron or positron, and decay with an average lifetime of 2.9 x 10^{-13} s. Despite losing little energy in the form of bremsstrahlung radiation due to their large mass, their short lifetime prevents them from travelling far in any given medium and makes them unlikely to be tracked over distances greater than a few centimetres. However, the large mass of the tau also permits a large number of different decay products, many of which have longer lifetimes and are easier to detect. This often allows for an analytical reconstruction of the tau, provided that the missing energy in the reconstruction associated to tau neutrinos is determined accurately.

The neutrinos that comprise the neutral sector of each lepton generation are particles that have been at the focus of many modern research endeavours. As they have only been observed to interact via the weak force and rarely interact with matter that they pass through, they eluded the type of incidental direct detection that historically provided evidence for the existence of many other particles. Although neutrinos were thought to be massless for many years, recent experiments have provided conclusive evidence that they are not. Early solar neutrino detectors were only sensitive to electron neutrinos and recorded a flux of solar neutrinos incident on earth that was only around one-third of the predicted flux. It was hypothesized that if neutrinos had mass they may be able to oscillate between electron, muon, and tau neutrino flavour states, explaining the deficit in observed solar neutrinos. Observations of muon neutrinos at the Super-Kamiokande detector in the late 1990's provided some evidence of neutrino oscillation [12], and data collected by the Sudbury Neutrino Observatory in the early 2000's provided convincing evidence that the predicted flux was correct and solar neutrinos detected in prior experiments only represented around one-third of the solar neutrinos incident on earth [13].

As a neutrino and antineutrino of the same flavour cannot be differentiated by their charge, it remains to be determined if they are the same particle (a Majorana particle) or two distinct particles (a Dirac particle). If they are Majorana particles, they would only be distinguishable by how they behave under a parity transformation, their chirality. Another consequence of being a Majorana particle is that neutrinos and antineutrinos could annihilate with themselves. This would permit lepton number violating particle interactions like neutrinoless double beta decay to occur (Fig 2.4). If this decay is possible it is expected to be quite rare, making observation of it difficult. If it were observed it would provide strong evidence that neutrinos are Majorana in nature [14].

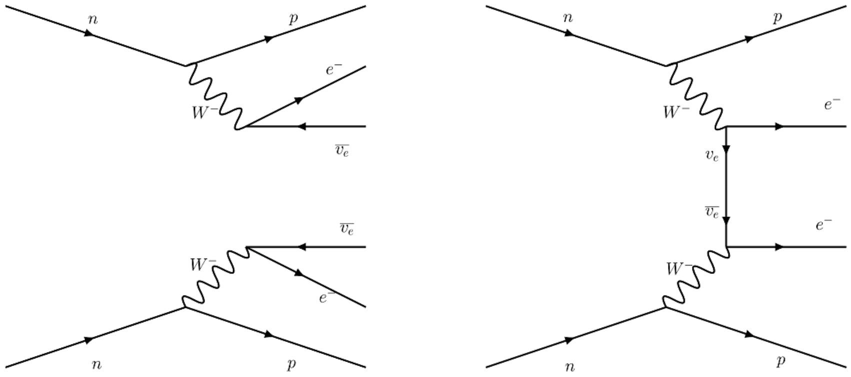


Figure 2.4: Feynman diagrams of double beta decay with the antineutrinos as non-annihilating Dirac particles (left) and neutrinoless double beta decay where the produced neutrino and antineutrino are Majorana particles that are able to annihilate with each other (right).

Quarks

Quarks are the only elementary particle in the Standard Model that interact via the strong force in addition to the other fundamental forces. Like leptons, there are six quark flavours, each with a corresponding antiparticle, organized into three generations. The first generation contains the lightest quarks, comprised of the up, u, antiup, $\bar{u}$, down, d, and antidown, $\bar{d}$, quarks. The charm, c, and strange, s, quarks together with their antiparticles make up the second generation, and the heaviest quarks, the top, t, bottom, b, and their associated antiparticles, make up the third generation. The up, charm, and top quarks have an electric charge of +$\frac{2}{3}$ and the down, strange, and bottom quarks have an electric charge of -$\frac{1}{3}$. Unlike leptons, individual quarks have not been observed to exist in isolation and are found bound to other quarks to form composite particles called hadrons (described in the next section). See Table 2.3 at the end of the section for a summary of the quarks and some of their key characteristics.

As quarks interact via the strong force, they carry a colour charge in addition to electric charge. Each quark can carry a red, blue, or green colour charge, and each antiquark can carry an anti-red, anti-blue, or anti-green colour charge. It is the presence of this colour charge that allows quarks to interact with gluons. As colour confinement dictates that no particle in isolation can exhibit a net colour charge at non-extreme temperatures, all quarks must be bound together in such a way as to have no net colour charge, or be ’colourless’. This necessitates that the quarks in all two-quark bound states have a colour and anti-colour of the same type, and that the quarks in three-quark bound states each have a different colour, or anti colour.

Although many interactions that quarks participate in are governed by the strong force, they are able to change flavour only through weak interactions. For example, in beta decay, neutrons (comprised of an up quark and two down quarks) become protons (comprised of two up quarks and a down quark) while radiating an electron and electron antineutrino. The mechanism by which this occurs is the emission of a W^- boson from one of the neutron’s down quarks. In order for a down quark to emit a particle with a charge of -1, conservation of charge dictates that the charge of the down quark must change by +1, in this case becoming an up quark. The W^- boson then decays into the electron and neutrino pair that one expects from beta decay.

The unitary matrix that parametrizes weak quark interactions and describes the probability that flavour changes will occur is known as the Cabibbo–Kobayashi–Maskawa (CKM) matrix. The matrix, along with the average of recent experimentally determined values, is as follows [15]:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.3}$$

$$V_{CKM} = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2221 \pm 0.0013 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{pmatrix}, \tag{2.4}$$

where V_{ij} represents the probability of a quark changing from flavour *i* to flavour *j*. The diagonal terms represent changes in flavour within the same generation and, like in beta decay, are mediated by a charged W boson and can be represented by a first order tree-level Feynman diagram. It is possible for quarks to change flavour across quark generations, although, as indicated by the off-diagonal terms, such interactions are suppressed and less likely to occur.

It is also possible that flavour changing interactions can occur where the generation of the quark changes but not the charge (for example, an up quark becoming a charm quark). The probabilities of these flavour-changing neutral currents are not represented in the CKM matrix as they are forbidden at the tree level in the Standard Model. These types of interactions require a higher-order loop-level Feynman diagram and are quite rare [16]. As decays involving flavour-changing neutral currents often include CP-violation, they can serve as a sensitive probe to new physics and have been studied extensively at experiments like Belle II [17].

2.1.3 Hadrons

When multiple quarks or antiquarks are bound together by strong force interactions, the resulting composite particle is known as a hadron. There are two types of hadrons; mesons, typically consisting of a quark and antiquark, and baryons, typically consisting of an odd total number of quarks or antiquarks. The lifetime of a hadron is generally inversely proportional to the mass of the constituent quarks, resulting in many of the hadrons that are comprised of top, bottom, or charm quarks quickly undergoing flavour changing weak decays shortly after being created. Consequently, the most common hadrons in nature are the more stable protons and neutrons, comprised of up and down quarks, that make up most of the matter that we interact with.

The rest mass of a hadron often dwarfs the combined rest mass of its constituent quarks. For example, neutrons (comprised of an up quark and two down quarks) have a rest mass of around 939.6 MeV/c^2, despite the rest mass of the three constituent quarks summing only to around 11.9 MeV/c^2. The remainder of the rest mass is largely attributable to the strong force binding energy described by QCD that holds the quarks together.

The possibility of mesons existing that are comprised of four or more quarks is being investigated [18][19], however all well-established mesons consist of only one quark and one antiquark. As quarks are fermions with a half-integer spin, mesons have an integer spin and obey bosonic statistics. They can exist with a net flavour (comprised of a quark and antiquark of different flavours) or in a flavourless state (with the constituent quark and antiquark being the same flavour). Whichever flavours are present in the meson, the net electric charge of the constituent quarks must sum to an integer value.

Mesons can be further categorized by their associated quantum numbers. The spin of the two constituent quarks can be aligned, resulting in the meson having an overall spin of 1 and existing in a spin triplet state. If the two quarks have oppositely aligned spins, the meson has an overall spin of 0 and exists in a spin singlet state. Additionally, mesons can interact via the weak force and, since weak interactions do not conserve parity, it is possible to have mesons that behave differently under parity. Mesons with an odd parity are known as vector mesons if they have an overall spin of 1, or pseudoscalar mesons if they have a spin of 0. Similarly, even parity mesons are known as pseudovector mesons if they have a spin of 1, and scalar mesons if they have a spin of 0.

Mesons that are of particular relevance to the Belle II experiment are B and $\overline{B}$ mesons. They consist of two quarks; a bottom or antibottom quark and a quark or antiquark from the first two quark generations. These mesons exhibit a number of characteristics that make them a useful tool for studying the Standard Model. Neutral B mesons have been observed to exhibit particle-antiparticle oscillation [20]. This phenomenon requires a spontaneous change in internal quantum number, in this case charge-parity, and allows the meson to transition to its antiparticle and back. B mesons also exhibit a number of charge-parity violating rare decays, the branching fractions of which provide insight into CKM matrix elements [21][22].

Although some pentaquarks have been reported in recent experiments and it has been speculated that other baryons with five or more quarks exist [23], well established baryons are comprised of three quarks. Given that they are comprised of an odd total number of quarks or antiquarks, baryons have a half-integer spin and obey fermionic statistics. Like mesons, only quark combinations that provide the baryon with an integer overall electric charge are possible. Similarly, the alignment of quark spins allows the baryons to have one of two different overall spin values, in this case either $\frac{1}{2}$ or $\frac{3}{2}$. Baryons are generally much heavier than mesons and, with the exception of protons and neutrons, decay quickly.

Table 2.1: A summary of the coupling constants and quantum number conservation violations permitted by the interactions of the three fundamental forces described by the Standard Model at the 1 GeV energy level. The associated gauge boson for each interaction type and their mass is also listed.

Interaction type	Coupling constant	Conservation violation	Gauge boson	Mass (GeV/c^2)
Electromagnetic	$\alpha_{em} = \frac{1}{137}$	None	Photon	Massless
Weak	$\alpha_w = 10^{-6}$	Parity, isospin, strangeness	$W^{\pm}$, Z^0	80.379, 91.188
Strong	$\alpha_s = 1$	Isospin	Gluon	Massless

Table 2.2: A summary of key characteristics for the leptons described by the Standard Model. The electron neutrino mass upper limit listed reflects 2019 findings by the KATRIN experiment and the muon and tau neutrino mass limits reflect aggregate values determined in 2016 [24][25].

Lepton	Symbol	Generation	Electric charge	Mass (MeV/c^2)	Lifetime (s)
Electron, positron	e^-, e^+	1	-1, +1	0.511	Stable
Electron (anti)neutrino	ν_e, ($\bar{\nu}_e$)	1	0	$<$ 0.0000011	TBD
Muon	μ^-, μ^+	2	-1, +1	105.658	2.2x10^{-6}
Muon (anti)neutrino	ν_μ, ($\bar{\nu}_\mu$)	2	0	$<$ 0.17	TBD
Tau	τ^-, τ^+	3	-1, +1	1776.84	2.9x10^{-13}
Tau (anti)neutrino	ν_τ, ($\bar{\nu}_\tau$)	3	0	$<$ 18.2	TBD

Table 2.3: A summary of key characteristics for the quarks described by the Standard Model. The third generation quarks are by far the heaviest and, as a result, hadrons comprised of them exhibit much shorter lifetimes than those comprised of up, down, or strange quarks.

Quark	Symbol	Generation	Electric charge	Mass (MeV/c^2) ± (stat) ± (sys)
Up, antiup	u, $\bar{u}$	1	+$\frac{2}{3}$, -$\frac{2}{3}$	2.3 ± 0.7 ± 0.5
Down, antidown	d, $\bar{d}$	1	-$\frac{1}{3}$, +$\frac{1}{3}$	4.8 ± 0.5 ± 0.3
Charm, anticharm	c, $\bar{c}$	2	+$\frac{2}{3}$, -$\frac{2}{3}$	1275 ± 25
Strange, antistrange	s, $\bar{s}$	2	-$\frac{1}{3}$, +$\frac{1}{3}$	95 ± 5
Top, antitop	t, $\bar{t}$	3	+$\frac{2}{3}$, -$\frac{2}{3}$	173210 ± 510 ± 710
Bottom, antibottom	b, $\bar{b}$	3	-$\frac{1}{3}$, +$\frac{1}{3}$	4180 ± 30

2.2 Analysis motivation and approach

2.2.1 Motivation

When electrons and positrons collide within the Belle II detector at a 10.58 GeV centre-of-mass energy, they annihilate to create a virtual photon which subsequently decays to a bottom and antibottom quark pair via the Υ(4S) meson resonance. The Υ(4S) quickly decays via the strong force into a $B\ \overline{B}$ meson pair and the further decay of these mesons is recorded by the detector and studied. They are of particular interest to physicists because they may exhibit many decay paths that require the violation of symmetries like charge-parity. This makes a thorough study of rare B meson decays fertile ground for observing new physics or further constraining the Standard Model.

B mesons primarily decay through weak interactions and thus produce neutrinos or antineutrinos in a large number of their decays. As neutrinos are nearly massless and don't participate in electromagnetic or strong interactions, they can be difficult to detect directly. Most analyses that examine decays involving neutrinos, like the B meson decays studied at Belle II, are required to sum the energy and momentum of directly detected decay products and compare it to the energy and momentum of the original particle in order to identify 'missing energy' that can be attributed to the presence of neutrinos. The sensitivity to neutrinos that is exhibited by studies that calculate missing energy in decays incidentally also provides sensitivity to the presence of other hypothesized weakly interacting particles, like dark matter.

Given the aforementioned potential for the observation of rare B meson decays to provide valuable insight into multiple areas of new physics, it is critically important that analysis methods used to identify missing energy in such studies are optimized. When particle decay chains are reconstructed in most Belle II analyses, any energy that was collected by the detector but unused in the reconstruction is categorized as 'extra energy'. The most common sources of extra energy are photons created in the detector by beam-related background events (described in detail in section 3.3). In order to correctly identify the amount of missing energy in a particle decay, it is necessary to simultaneously determine if energy collected by the detector is associated with these beam backgrounds or with the studied decay. The purpose of the work described by this thesis is to create an efficient method of distinguishing between the sources of this extra energy and to provide a new tool for beam background suppression. This should serve to improve the sensitivity of Belle II studies to new physics by optimizing neutrino identification and making it easier to identify when B meson decays have produced particles that are not accounted for in the reconstructed decay.

2.2.2 Machine learning

In order to develop a tool capable of efficiently distinguishing between sources of extra energy in reconstructed particle decays, it was necessary to use machine learning. Machine learning is a type of artificial intelligence that allows computer programs to develop a method to complete complex tasks without being given explicit instructions for each step beforehand. There are many different types of machine learning algorithms, each with their own subdivisions and complexities, but perhaps the two most distinct are unsupervised learning and supervised learning.

As the name suggests, unsupervised learning involves providing a program with only input data and allowing it to analyze the data without strict guidance. It is up to the program's algorithms to identify the structure of the data and the relationships that exist within it. This is approach is often used to identify new patterns or relationships in a set of data that could not have been originally programmed. It is especially useful in probability density modelling and the cluster analysis that is used in fields like climatology and medical imaging [26][27].

Supervised learning requires that the program is provided with an output sample, or end result, in addition to the input data. The provided data serves as a training sample for the program's algorithms to examine in order to identify patterns and relationships between the input data and output data. If successful, this allows the program to make accurate predictions about future outcomes when provided only with similar sets of input data. It is common in image and speech recognition software and is the type of machine learning used in the analysis described by this thesis [28].

Within each approach to machine learning there are a variety of algorithm types and models that are used. The machine learning model used in the work described by this thesis is the decision tree. Information is organized by relation into a series of paths or branches, with final states associated with particular chains of decisions, or a specific location on a given branch. It is generally used in supervised learning, where training data is supplied and the decisions or characteristics associated with each outcome are identified to later be used in a predictive fashion. Although this type of model is capable of handling very complex problems, it is especially valuable for its relatively clear organization of decision making and overall simplicity.

There are two main types of decision tree; classification trees where outcomes are organized into discrete classes or values, and regression trees where outcomes fall somewhere within a continuous range. For example, a classification tree may provide a binary output indicating that a sample either meets certain criteria or it does not, while a regression tree may provide a range of values that indicate the probability that a sample falls within a particular category. Another benefit of this machine learning model is that multiple poorly performing decision trees can be processed in stages and aggregated to produce a single, stronger performing decision tree in a process known as 'boosting'. Boosting allows the algorithm to consider the preceding decision trees when processing each subsequent one, amplifying the decision-making capability of the final output while reducing bias and variance. The extra energy discriminating tool developed in this thesis was produced by a multivariate analysis performed on Belle II data using a boosted regression decision tree.

Chapter 3

The Belle II experiment

The Belle II experiment operates out of the High Energy Accelerator Research Organization (KEK) particle physics laboratory in Tsukuba, Ibaraki Prefecture, Japan. The experiment studies aspects of the standard model of physics by examining the behaviour of B mesons that arise from e^+ e^- collisions induced by the SuperKEKB accelerator. A description of the SuperKEKB accelerator and the Belle II detector follow. Unless otherwise stated, the contents of this chapter are informed by the Belle II Technical Design Report [29].

3.1 The SuperKEKB accelerator

The KEKB accelerator was an asymmetric double-ring electron-positron collider that began operation in 1998 [30]. It accelerated positrons and electrons to 3.5 GeV and 8 GeV respectively, and was able to reach a peak luminosity of 2.11 x 10^{34} cm^{-2}s^{-1}. Operations ceased in 2010 when it was shut down to allow for the rings and injector linac to be upgraded in order to reduce beam size and increase luminosity. The revised accelerator, known as the SuperKEKB accelerator, began operating in 2016 with a target luminosity of 8 x 10^{35} cm^{-2}s^{-1}.

Like its predecessor, the SuperKEKB accelerator consists of an injector linac and two concentric ring-shaped chambers lined with magnets used to guide particles towards a collision. The rings have a circumference of 3016 m and extend over a large region of the KEK site, with the electrons and positrons colliding at an interaction point within the Belle II detector in the Tsukuba region (Fig 3.1). A major outcome of the accelerator upgrades is a reduction in the size of the electron and positron beams that collide at the interaction point.

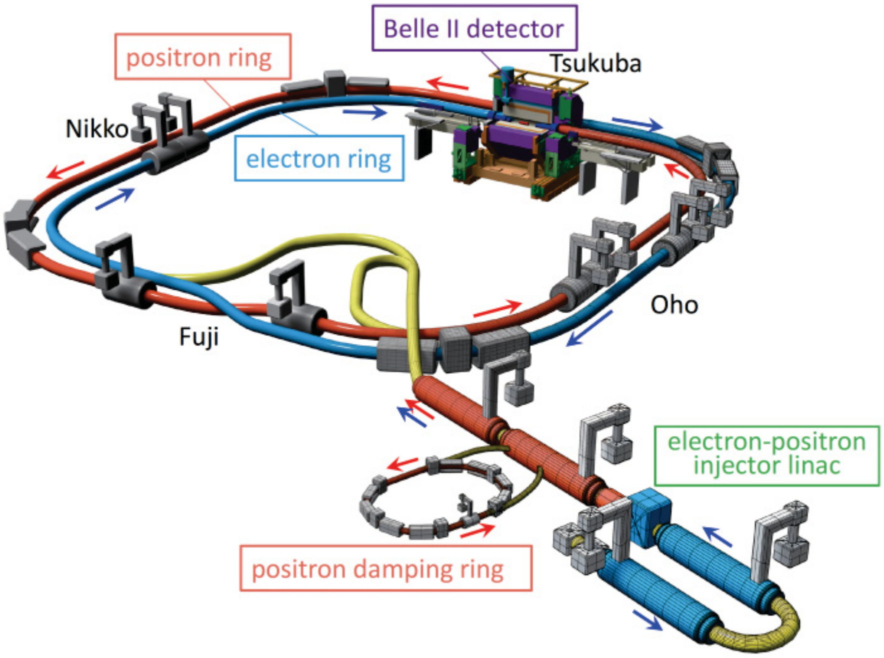


Figure 3.1: An overview of the electron-positron injector linac and storage rings that comprise the SuperKEKB accelerator. The location of the Belle II detector relative to the entire accelerator is also shown [29].

Although higher particle beam density allows for a higher luminosity to be obtained, it also increases the amount of particle scattering that occurs within the beams and increases the difficulty of maintaining low beam emittance. In order to keep electron emittance under 20 μm and electron bunch charge at 5 nC (two requirements necessary to meet the upgraded accelerator design objectives) a photocathode RF gun was developed to act as an electron source. The photocathode RF gun makes use of a high acceleration field gradient and small cathode size to produce low-emittance electrons that are accelerated to 7 GeV in the injector linac before being directed into the high-energy ring (HER). A separate electron pre-injector (a triode-type thermionic electron gun) produces 4 GeV electrons that irradiate a tungsten target to produce positrons which are subsequently accelerated to 4 GeV and directed into the low-energy ring (LER). The asymmetric energy of the colliding particles results in a non-zero net momentum in the collision products, displacing subsequent decay vertices from the initial interaction point and allowing for a more reliable determination of the distance between decaying B mesons. The decrease in electron beam energy from KEKB to SuperKEKB of 8 GeV to 7 GeV serves to further reduce beam emittance, while the increase in LER beam energy from 3.5 GeV to 4 GeV serves to reduce Touschek and intra-beam scattering effects. This increase in LER beam energy also contributes to an elevated positron beam emittance of 2.1 μm. To mitigate this effect, a 1 GeV damping ring is used partway through the injector linac that reduces positron beam emittance to around 13 nm. These changes allow SuperKEKB to collide the particles at the 10.58 GeV Υ(4S) resonance necessary for high B meson yield while maintaining significantly improved beam emittance and bunch charge levels relative to KEKB (Table 3.1). As of June 2020, the SuperKEKB accelerator has achieved a peak instantaneous luminosity of 2.4 x 10^{34} cm^{-2}s^{-1}, with a total integrated luminosity over 70 fb^{-1} [31].

Table 3.1: A summary of key design parameters for the electron and positron beams produced by the KEKB accelerator, and those hoped to be achieved by the upgraded SuperKEKB accelerator [29].

	KEKB		SuperKEKB	
	e^+	e^-	e^+	e^-
Beam energy (GeV)	3.5	8.0	4.0	7.0
Stored current (mA)	1600	1200	3600	2620
Beam lifetime (min)	150	200	10	10
Bunch charge (nC)	10.0/1.0	1.0	10.0/4.0	5.0
Beam emittance (μm)	2100	300	10	20
Energy spread σ_E/E (%)	0.125	0.05	0.07	0.08
Bunch length σ_z (mm)	2.6	1.3	0.5	1.3

The electrons and positrons are guided through the arcing beam pipe by dipole, quadrupole, and sextupole magnets arranged in a lattice. Approximately 30% more magnets are required in the HER than the LER due to the higher energy of the beam. In order to reduce the number of beam background events and ensure that the rate of beam particle loss from such events is less than the rate of particle loss by collision events, the beam chambers are held to a target vacuum pressure of 10^{-7} Pa. This is accomplished with strip-type non-evaporable getter pumps located in the antechamber of the beam pipe (Fig 3.2). Auxiliary sputter ion pumps are used to remove non-active gases in high-pressure areas of the chamber. As the high photon density in the HER and LER beams leads to high levels of synchrotron radiation and heat in the vacuum components of the accelerator, a cooling system circulates water on the outside of the beam antechambers to disperse heat.

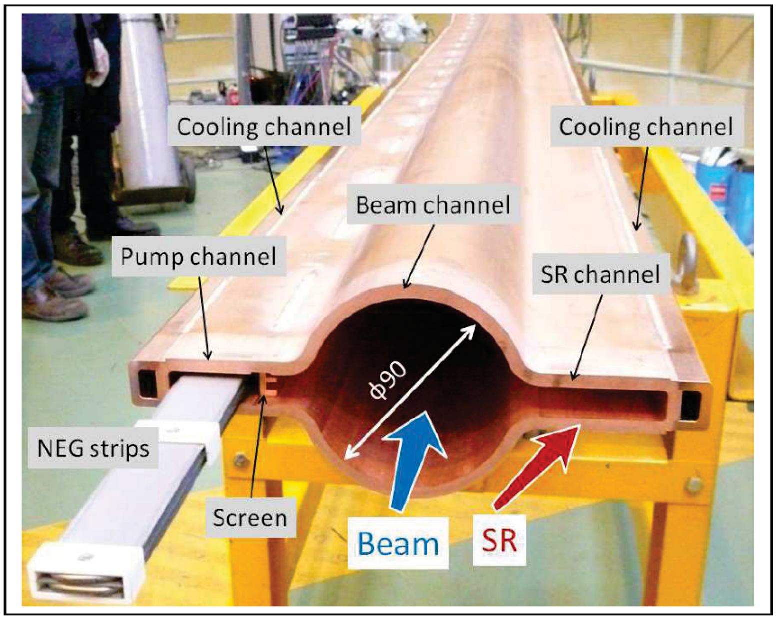


Figure 3.2: A cross-sectional view of a piece of beam pipe used in the SuperKEKB accelerator. Antechambers on either side hold vacuum pumps and are next to liquid cooling chambers [29].

A number of superconducting quadrupole magnets, permanent quadrupole magnets, and superconducting solenoids make up a focusing system that is used to cross the particle beams in the interaction point chamber with a crossing angle of 83 mrad. The magnetic flux produced by the superconducting magnets can also affect beam dynamics in the nearby HER, necessitating the inclusion of superconducting correction coils to cancel out the magnetic flux beyond the interaction point chamber. Like the beamline, the interaction point chamber is held in vacuum and uses a cooling system to maintain an effective operating temperature. However, as the central part of the chamber is made with beryllium that may react with water, purified liquid paraffin ($C_{10}H_{22}$) is circulated around the outside of the chamber as a coolant.

3.2 The Belle II detector

The Belle II detector, located at the Tsukuba region of the storage rings, is an upgraded version of the original Belle detector. It is comprised of seven primary sub-detector components arranged cylindrically around the crossing point of the electron and positron beams (Fig 3.3). The innermost chamber surrounds the beam pipe at the interaction point and consists of a silicon pixel detector (PXD) and silicon vertex detector (SVD) which together form the vertex detector (VXD) that tracks the movement of charged particles. The large central drift chamber (CDC) contains 56 layers of particle identifying sensors and encloses the VXD. Moving outward, the time of propagation counter (TOP) radially envelops the CDC and the aerogel ring-imaging Cherenkov detector (ARICH) covers a portion of the forward end-cap region of the CDC. These two components use photo-sensors to measure the Cherenkov radiation of particles and improve the effectiveness of particle identification. Further out lies the electromagnetic calorimeter (ECL) that uses thallium-doped caesium iodide crystals to determine the trajectories and energies of photons and electrons passing through the detector. Beyond the ECL lies a superconducting coil and the outermost component, the K_L^0-muon detector (KLM), that identifies the trajectories of muons and K_L^0 mesons in the detector.

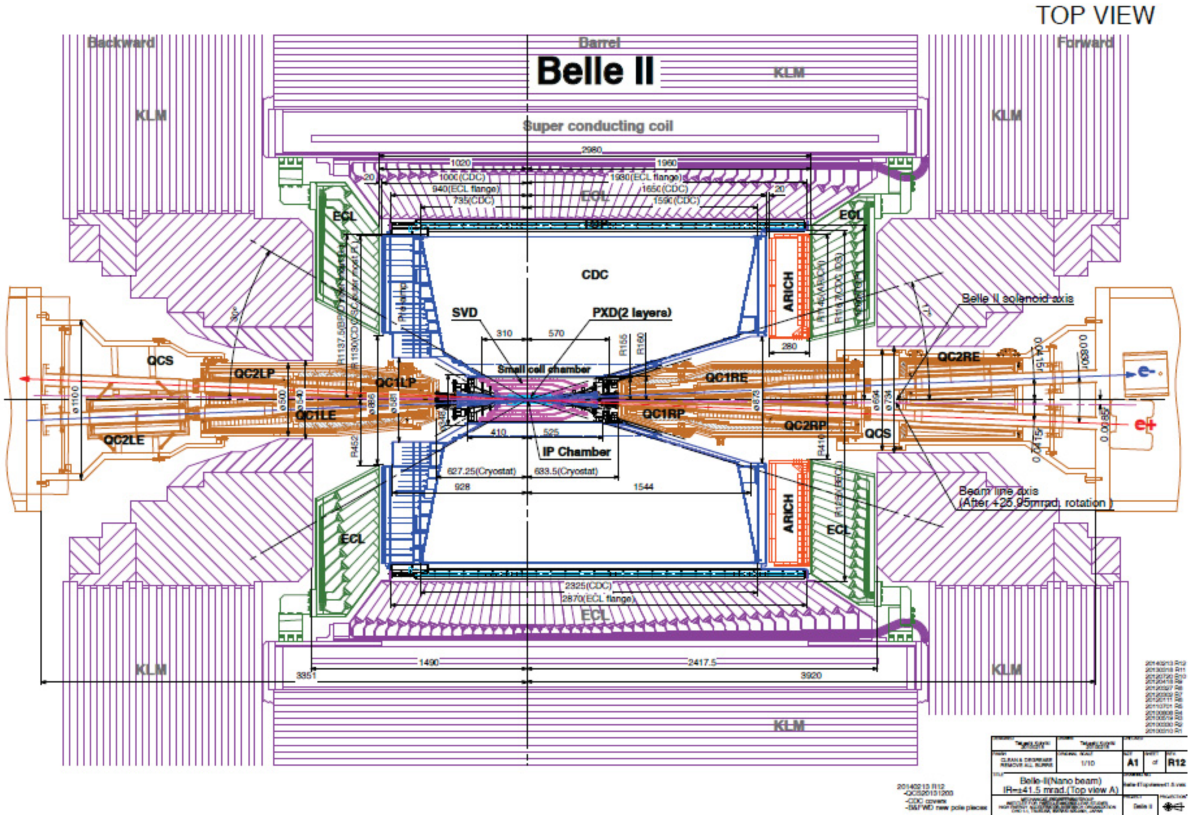


Figure 3.3: A complete schematic of the Belle II detector. Moving outward, most components nearly completely envelop all of the inner components before it, providing a high degree of coverage [29].

Locations within the Belle II detector are often described in one of two coordinate systems (Fig 3.4). The Cartesian system, with the positive z-axis directed along the electron beamline, the positive x-axis pointing away from the center of the containment rings, and the positive y-axis directed towards the top of the detector, is useful when describing certain detector components. However, the radial symmetry of many parts of the detector lends itself to a description in polar coordinates. The radial distance, r, is measured from the center of the detector, the azimuthal angle, ϕ, is the angle relative to the x-axis such that +y corresponds to 90°, and the polar (or zenith) angle, θ, is the angle along the beamline relative to the positive z-axis.

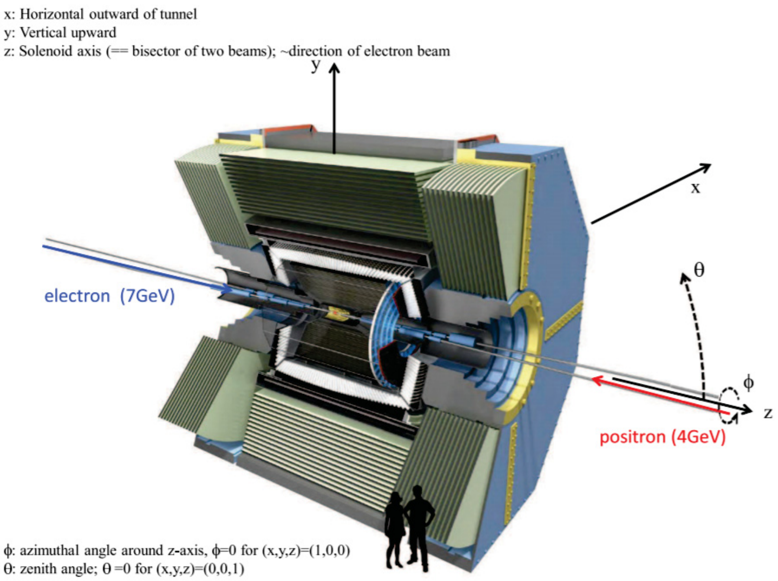


Figure 3.4: An illustration of the coordinate system that is used to describe the Belle II detector. θ=0 corresponds to (x,y,z) = (0,0,1), and ϕ=0 corresponds to (x,y,z) = (1,0,0) [32].

3.2.1 Pixel detector (PXD)

The vertex detector (VXD) is comprised of a pixel detector (PXD) and a silicon vertex detector (SVD) that collect the spatial information of charged particles as they pass through, allowing for a reconstruction of their trajectories and the vertex of the decay they originated from.

The PXD is a cylindrical silicon self-supporting structure that wraps around the 10 mm radius beam pipe, with two sensor layers at a radius of 14 mm and 22 mm (Fig 3.5). There are 8 different 15 x 90 mm^2 sensor regions on the inner layer and 12 15 x 123 mm^2 regions on the outer layer, each containing a matrix of 50 x 50 μm^2 pixels. The arrangement of the sensors provides complete coverage of the interaction region in the azimuthal plane. The close proximity of the sensor layers to the beam pipe allows for considerable coverage in the polar plane as well, from 17° to 150°. The asymmetry in polar coverage is intended to address the center-of-mass boost present in collision products arising from the asymmetry in beam energies.

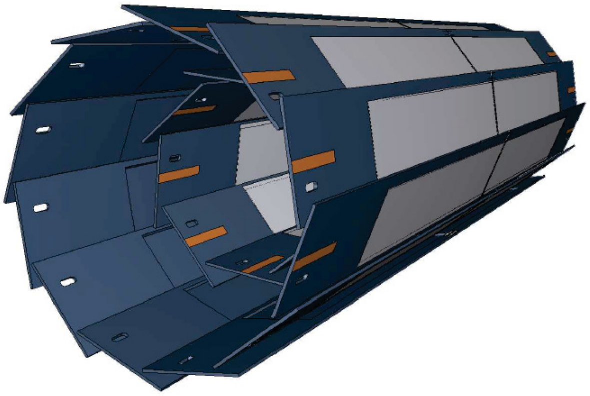


Figure 3.5: A schematic of the inner and outer layer of the PXD. Grey areas represent the 20 regions of pixels that allow the PXD to obtain complete azimuthal coverage and nearly complete polar coverage of the interaction point [29].

Ionizing particles passing through the pixel sensor regions produce electron-hole pairs that propagate through the sensor and an accumulation of charge generates a signal. The sensors are comprised of silicon semiconductor diodes based on depleted field effect transistor (DEPFET) technology that uses internal gates to allow for electron accumulation and signal charge amplification near the source of electron generation, preventing pre-amplification charge transfer loss. The small capacitance of the internal gates also allows for very low noise performance.

The near-source amplification provided by DEPFET technology also reduces the required thickness of the sensor region, which can reduce the likelihood of multiple scattering events occurring within the sensor. As a result, the Belle II PXD sensors are 75 μm thick, approximately 0.2$\%$ of a radiation length [33]. Given that the pixels only generate heat when activated, dry-air cooling is sufficient to maintain the sensors at the desired operating temperature. The readout electronics generate considerably more heat but they are located at the end of the PXD away from the acceptance region, allowing for an active liquid cooling system to be used.

The location and timing of activated pixels is used in track-finding algorithms to reconstruct the trajectories of particles passing through the PXD. The large number of small pixels in the sensors provides highly granular coverage of the acceptance region, allowing for a spatial resolution as small as 10 μm. The small spatial resolution so close to the interaction point provides strong vertex resolution, which is especially important for accurately tracking any low momentum decay products that do not reach areas of the detector outside of the PXD.

3.2.2 Silicon vertex detector (SVD)

Like the PXD, the SVD consists of layers of silicon sensor regions aligned cylindrically around the interaction point chamber that are used to identify particle trajectories and reconstruct decay vertices. The four layers of SVD sensors are further away from the interaction point than the PXD sensors, at radii 38 mm, 80 mm, 115 mm, and 140 mm. As radiation intensity drops off exponentially with distance the SVD sensors experience a lower degree of radiation damage than the PXD sensors, allowing for the use of double-sided silicon microstrip detectors instead of the more radiation-hard pixel sensors. The strips are 12.3 cm long and approximately 300 μm thick and are again arranged to provide polar-angle coverage from 17° to 150° and full azimuthal-angle coverage. Sensors in the forward region of the detector are slanted towards the beamline to reduce the number of sensors necessary to obtain this level of coverage and are trapezoidal in shape rather than rectangular.

Near the outer layer of the Belle II detector is a solenoid that generates a 1.5 T magnetic field that is aligned with the beam trajectory (along the z-axis). This magnetic field has the adverse affect of applying a Lorentz force on the electron-hole pairs produced in the strip sensors as charged particles pass through them. As electrons are able to move through the strip sensors more freely than holes this leads to a greater spread of electrons within the sensor (Fig 3.6). To address this and minimize the overall charge spread in the sensors, the strips are tilted and arranged in a slightly overlapping ’windmill’ pattern around the beamline (Fig 3.7).

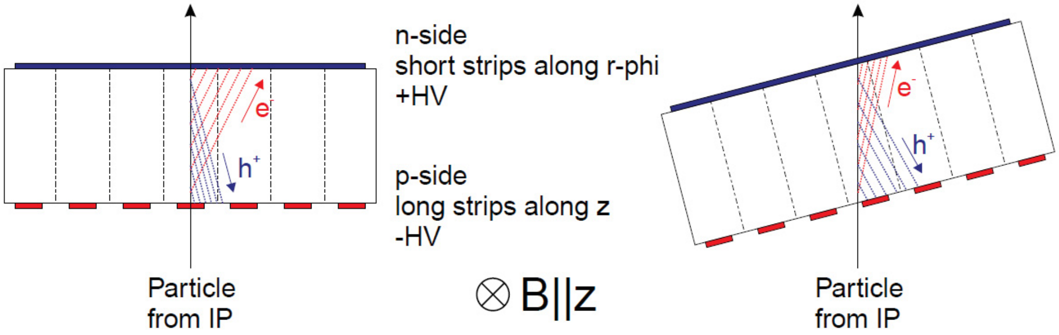


Figure 3.6: An illustration of charge spread in the strip sensors arising from the presence of an orthogonal magnetic field. Adjusting the angle of the strip sensor serves to minimize the difference in spread between the electrons and holes [29].

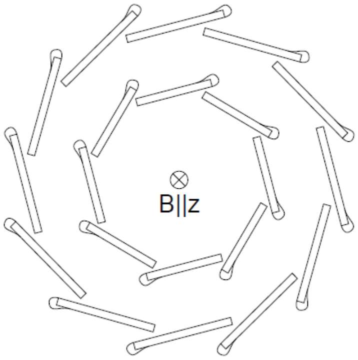


Figure 3.7: The overlapping windmill arrangement of strip sensors in the SVD used to minimize charge spread within the sensors. The overlap accounts for 8 to 10% of the sensor area [29].

Given the close proximity of the PXD to the interaction point and the relatively large radius of the outer SVD layers, the two components combined allow for trajectory reconstruction over a large area and excellent vertex resolution. The large radius of the SVD is also important for accurately tracking the decay products of long-lifetime particles. Some B meson decay channels do not contain any charged particles other than the pion daughters of K_S^0, which often do not decay until they are outside of the PXD.

3.2.3 Central drift chamber (CDC)

The central drift chamber (CDC) is a cylindrical gas chamber enveloping the VXD that is used to reconstruct charged particle trajectories, generate trigger signals, measure particle momentum, and identify particles by measuring their energy loss as they pass through the detector. It is 232.5 cm long, extending from an inner radius of 160 mm to an outer radius of 1130 mm with a slight polar-angle asymmetry to compensate for the center-of-mass boost arising from the asymmetric beam energies.

The chamber contains 14336 sensor wires arranged in 56 layers, grouped into nine super-layers. The super-layers are either axial (A) in orientation to align with the magnetic field produced by the outer solenoid or skewed in one of two 'stereo' (U and V) orientations (Fig 3.8). The different alignments are necessary to achieve accurate three-dimensional reconstruction of particle tracks.

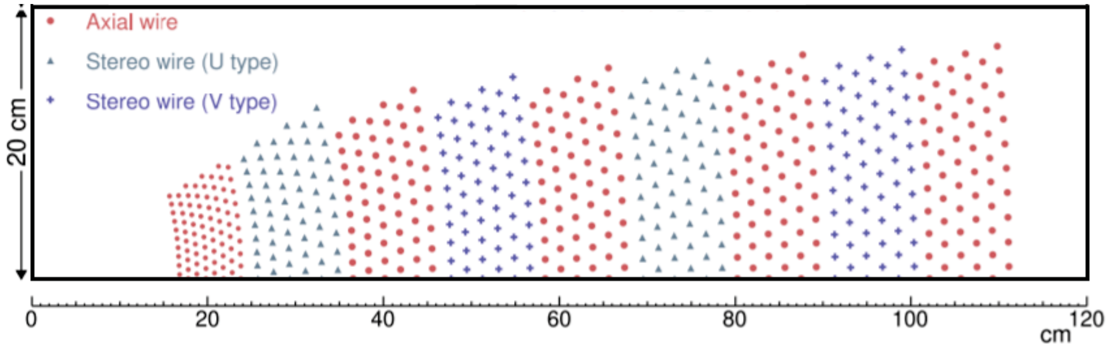


Figure 3.8: The configuration of sensor wires in the CDC. The eight outermost super-layers contain six layers of wires and the final innermost super-layer contains eight. The layers alternate between axial and stereo orientation in an 'AUAVAUAVA' pattern [34].

Each sensor wire is surrounded by eight field wires to form a cell. Cells in the eight outermost super-layers extend 18.2 mm radially, while the innermost super-layer is more compact with cells extending 10 mm radially. The density of these cells provides a spatial resolution between 50-120 μm, depending on the layer and incident angle of the particle [34].

A 50:50 mixture of helium and ethane (C_2H_6) gas is circulated through the chamber by a pressure-controlled pump system that slowly feeds fresh gas into the detector. This gas mixture was chosen for its low radiation length and fast drift velocity. As charged particles pass through the gas, electrons are displaced and generate a signal in nearby sensor wires. The magnetic field provided by the outer solenoid curves the trajectory of charged particles in the CDC and the extent of the curvature is used to determine the particle's momentum. The charge deposited on each sensor wire is also used to determine the particle's specific ionization (dE/dx), which is used in particle identification (Fig 3.9). The large size of the CDC allows for better momentum resolution, as the curvature of long tracks can be better fit. It is expected that the CDC will achieve a momentum resolution, $\frac{\sigma_p}{p_T}$, of 0.2% [35]. Because the position resolution and precision of trajectory reconstruction varies with the incident angle of the charged particle, the dE/dx resolution varies as well, ranging between 8.5% and 12%.

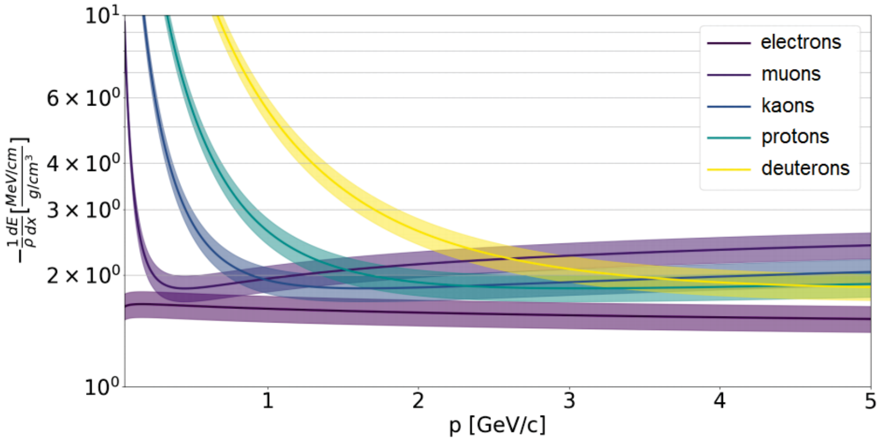


Figure 3.9: The specific ionization of electrons, muons, kaons, protons, and deuterons as a function of momentum. A comparison of these distributions to data collected by the CDC allow for particle identification [36].

3.2.4 Time of propagation counter (TOP)

Moving outward from the CDC, the Belle II detector is comprised of three cylindrical regions; a forward end-cap, a central barrel, and a backward end-cap. The time of propagation counter (TOP) is a particle identification apparatus that wraps around the outside of the CDC in the barrel region and collects Cherenkov light produced by the passage of particles through the detector modules. Rather than an upgrade of an already existing Belle detector component, the TOP is a new device made for Belle II. It consists of sixteen identical modules, each composed of two fused synthetic silica (quartz) bars 125 cm in length, a mirror located at the forward end, and a 10 cm long prism attached to a photomultiplier tube (PMT) array at the back end (Fig 3.10). The quartz bars act as Cherenkov light radiators and, due to their high refractive index (n = 1.47 for 410 nm light [37]), are able to capture and transmit some of that light to the PMTs. The prism is used to expand the light rings reaching the PMTs, slightly increasing the effective number of received photons and improving wavelength discrimination.

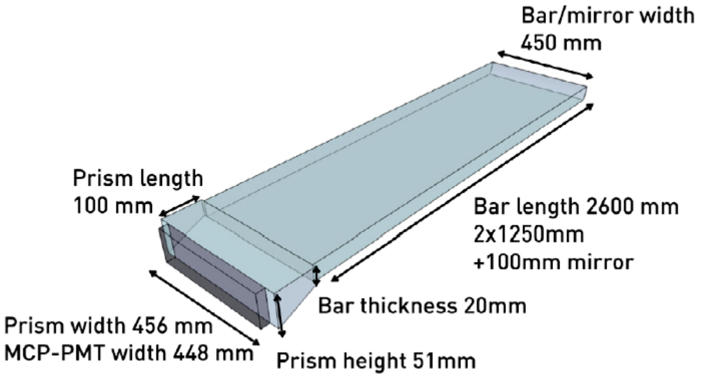


Figure 3.10: An illustration of one of the TOP modules. The two fused synthetic silica bars are connected longitudinally for a total radiation medium length of 2600 mm, and a mirror on the forward end ensures that captured light will eventually reach the PMT array opposite it [38].

The modules are aligned around the CDC barrel in equal 22.5° azimuthal segments at a radius of 1200 mm. This arrangement contains small gaps between each module, leading to a 7% loss in azimuthal coverage. The modules provide polar angle coverage from 31° to 128°. The PMT array attached to each module contains 32 PMTs divided into two rows of 16 and each PMT contains 16 5 x 5 mm pixels. In addition to providing an approximate measurement of the location of incident photons, the PMTs have a transit time spread under 50 ps which allows the detection time of incident photons to be determined precisely [38].

The purpose of the apparatus is to measure the time that has elapsed between the initial $e^+ \ e^-$ collision and the moment the Cherenkov light reaches a PMT by total internal reflection through one of the modules. As light incident on a PMT from different angles will have taken different amounts of time to reflect through the module, and given that the angle of Cherenkov light emission depends on the momentum of the transiting particle, it is possible to use the collected timing information to estimate the mass and velocity of the particle. A distribution of arrival times across the PMT array is compared to expected probability density functions for electrons, muons, pions, kaons, rho mesons, and down quarks. This comparison can be used in conjunction with information from other detector components to determine particle identity probabilities (Fig 3.11).

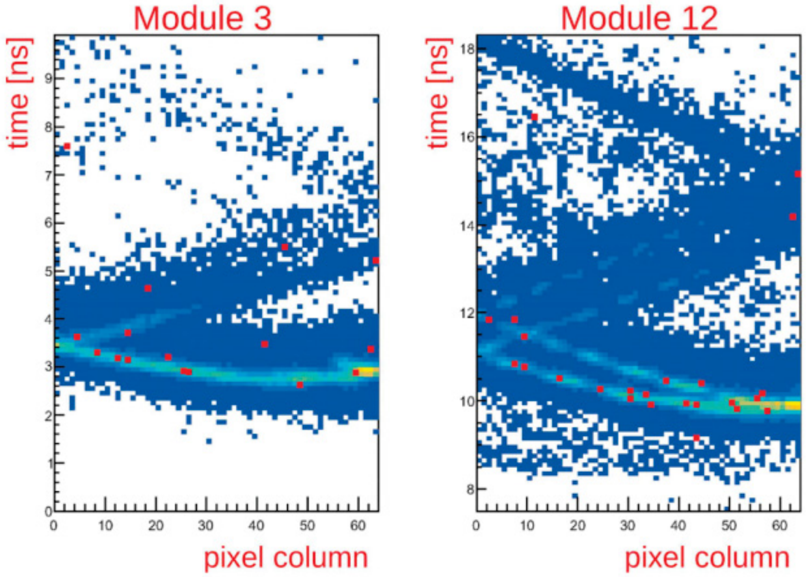


Figure 3.11: The time of detection and pixel column within two TOP modules for photons produced by a particle suspected to be a cosmic muon that was observed by the Belle II detector in 2018. Coloured bands represent simulated timing and position information for muons, and red data points represent collected data. Each pixel column corresponds to a position along the TOP module [39].

3.2.5 Aerogel ring-imaging Cherenkov detector (ARICH)

The aerogel ring-imaging Cherenkov detector (ARICH) is located just outside the CDC in the forward end-cap region of the detector and, like the TOP, collects Cherenkov light to identify transiting particles. It is comprised of two silica aerogel radiator layers, an expansion volume, and an array of hybrid avalanche photodiodes (HAPDs). The two aerogel layers are 2 cm thick and have different refractive indices (n = 1.045 upstream and n = 1.055 downstream). Having two radiation mediums with different refractive indices allows for a longer overall radiation medium (so more light is produced) while still focusing the emitted light rings from both mediums so they overlap when reaching the HAPDs. The aerogel layers are constructed in tiles and 248 of them are arranged in a 'donut' shape to cover the forward face of the CDC (Fig 3.12). The 16 cm long expansion volume allows Cherenkov rings to expand before they reach the array of 420 HAPDs, and planar mirrors are placed around the edges of the ARICH cylinder to prevent the loss of light incident on the sides.

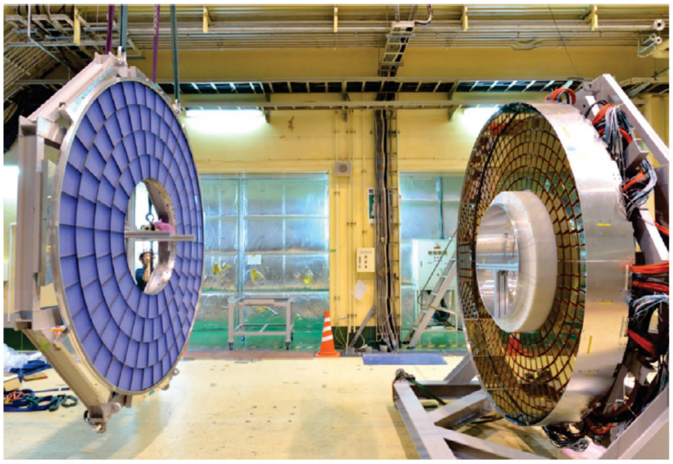


Figure 3.12: The 248 silica aerogel radiator tiles (left) and the array of 420 HAPDs (right) during construction of the ARICH. The cylindrical design provides a high degree of coverage of the forward end-cap region [40].

Incident photoelectrons entering a HAPD are accelerated over a 10kV potential difference within a vacuum tube in order to produce an avalanche of electron-hole pairs within the diode and generate a signal. Due to the need for large area coverage, the constant presence of a 1.5 T magnetic field in the detector during data collection, and the likelihood of radiation damage over a ten-year operating lifetime, the HAPDs used in the ARICH had to be developed specifically for this detector component. Each sensor is 73 x 73 mm^2 with 144 readout channels and a 5 mm spatial resolution. They are designed to have a high sensitivity to single photons and are capable of reaching a quantum efficiency greater than 30% while maintaining immunity to the magnetic field [41].

The primary function of the ARICH is to allow for discrimination between pions and kaons in the 0.5 GeV/c to 4.0 GeV/c energy range and pions, electrons, and muons below 1 GeV/c. As the angle of emitted Cherenkov light depends on the mass and velocity of the transiting particle, two particles with the same velocity and different mass should produce rings of Cherenkov light that will reach different sizes by the time they arrive at the photodiodes (Fig 3.13). As with the TOP detector, an analysis of these rings in conjunction with information provided by other detector components allows for a determination of particle identity probabilities. The design objective of reaching 4σ separation between pions and kaons was achieved in early evaluations of ARICH performance based on Belle II colliding-beam data [40].

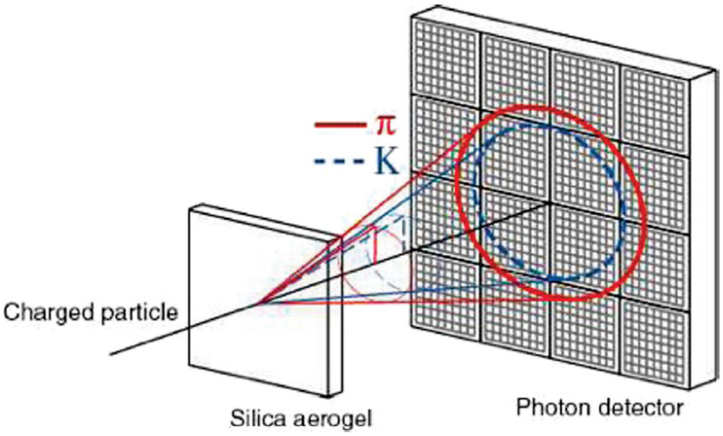


Figure 3.13: An illustration of the design principle of the ARICH. The angle of Cherenkov photon emission is a function of incident particle momentum, so a pion and kaon with similar velocity would produce Cherenkov rings of different radii [40].

3.2.6 Electromagnetic calorimeter (ECL)

The electromagnetic calorimeter (ECL) is situated outside of the TOP and ARICH and is segmented into a barrel, forward end-cap, and backward end-cap region that encloses the inner detector components. It is filled with thallium-doped cesium iodide CsI(Tl) scintillation crystals that absorb energy from incident electrons and photons and produce a proportional amount of light. The cylindrical barrel region contains 6624 crystals and is 3 m long, with an inner radius of 125 cm and an outer radius of 162 cm. The forward and backward end-cap contain 2112 crystals and are approximately 42 cm wide, beginning at z = 1.96 m and z = -1.02 m, respectively (Fig 3.14). Aside from 1° gaps between the barrel and end-caps, the ECL provides polar angle coverage from 12.1° to 157.1° and nearly complete azimuthal angle coverage.

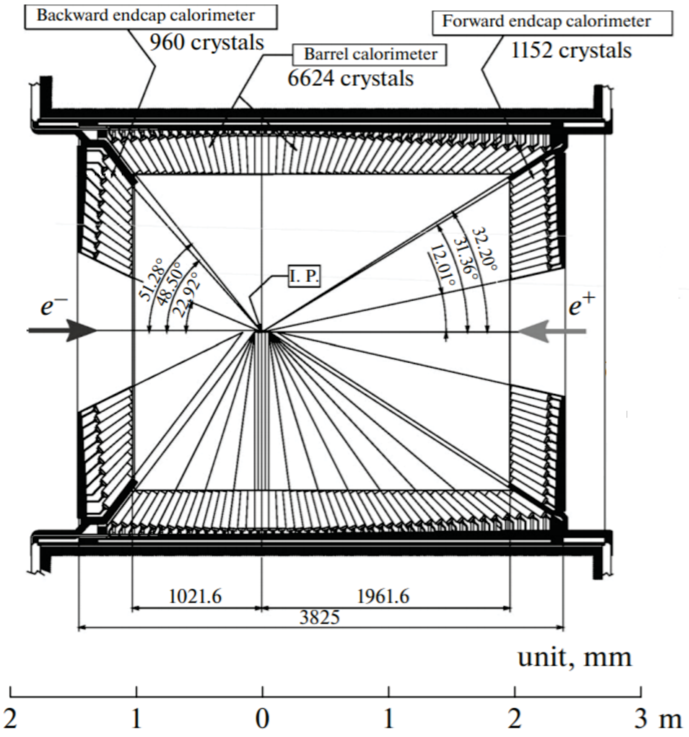


Figure 3.14: A schematic of the arrangement of CsI(Tl) crystals around the interaction point in the barrel, forward end-cap, and backward end-cap of the electromagnetic calorimeter [42].

Although the crystals vary in shape, they are all around 30 cm long (which represents 16.2 radiation lengths) and their average face is approximately 6 x 6 cm^2. Every crystal is wrapped in a 200 μm layer of porous Teflon to minimize light loss within the crystals, and then wrapped in a 60μm layer of aluminized LAVSAN to shield them from electrical background sources [42]. Two 10 x 20 mm^2 PIN photodiodes connected to preamplifiers are attached to the back of each crystal to collect light and produce an electrical signal.

When an incident electron passes through one of the crystals it can emit bremsstrahlung photons as it interacts with the Coulomb field of the CsI(Tl) nuclei. These photons, or any other incident photons, can similarly interact with the nuclei to undergo pair production, producing an electron and positron. These electrons and positrons can then emit further bremsstrahlung radiation, creating a self-feeding system of photon production until there is insufficient energy available to undergo pair production or emit photons (Fig 3.15). These photon showers can be absorbed by the crystals, causing the emission of scintillation light that is collected by the photodiodes and converted into an electrical signal. Often, photon showers from a single passing particle are produced in a group of crystals and the ECL operates by measuring the timing, position, and magnitude of these clusters of signals. Given that bremsstrahlung radiation for charged particles is dependent on particle mass, hadrons like pions that are much heavier than electrons will experience a different amount of energy loss due to bremsstrahlung radiation. As well, particles like pions can ionize CsI(Tl) nuclei which leads to the emission of photons when the ejected electrons eventually recombine with the ions. This emitted light is similarly collected by the photodiodes and has different characteristics than the light resulting from bremsstrahlung radiation. These varying mechanisms for energy deposition allow information collected by the ECL to facilitate discrimination between the types of particles that pass through.

Once the CsI(Tl) crystals have begun emitting light it takes around 1 μs for the light emission to cease and the crystals to return to their initial state. As the Belle II experiment operates in a relatively high luminosity range, it is possible that the associated high levels

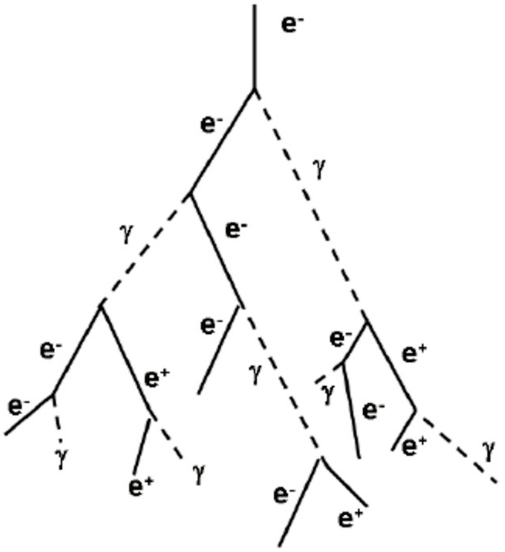


Figure 3.15: A diagram of an electromagnetic shower showing an electron emitting bremsstrahlung photons and causing a cascade of photon, electron, and positron production. These showers lead to the emission of scintillation light that is collected by the photodiodes at the back of each crystal [43].

of beam backgrounds in the detector could lead to a loss in signal efficiency when electrons and photons from multiple events reach a crystal before it has stopped scintillating from a prior event. To address this potential signal pileup, wave-form-sampling electronics are used to fit the signal pulse shape and event time to shorten the signal-shaping time to 0.5 μs to reduce the likelihood that non-background events are lost.

One of the most common decay products of B mesons is the neutral pion which has a branching fraction of approximately 98.8% for the decay $\pi^0 \to 2\gamma$ [11]. Around one third of all decay products of B mesons are neutral particles that similarly decay into photons [21]. As many other components in the Belle II detector are only able to detect charged particles, the ECL is the primary component for detecting photons and discriminating between electrons and hadrons. It also assists in K_L^0 detection and the generation of signals for triggering. These factors emphasize the need for good timing and energy resolution in the ECL. Recent evaluations of ECL performance indicate a timing resolution of 12 ns in the barrel and 22 ns in the end-caps, as well as an energy resolution of approximately 2 MeV in the barrel and 5 MeV in the end-caps [44].

3.2.7 K_L^0 muon detector (KLM)

The K_L^0 muon detector (KLM) is the largest sub-detector in the Belle II experiment apparatus and is segmented into three octagonal sections that enclose the rest of the detector; a barrel region, a forward end-cap region, and backward end-cap region (Fig 3.16). Between the KLM and ECL barrels is a solenoid used to generate a 1.5 T beam-aligned magnetic field within the detector. The KLM measures the position, timing, and intensity of light produced by the transit of particles in order to identify muons and K_L^0 mesons passing through the detector.

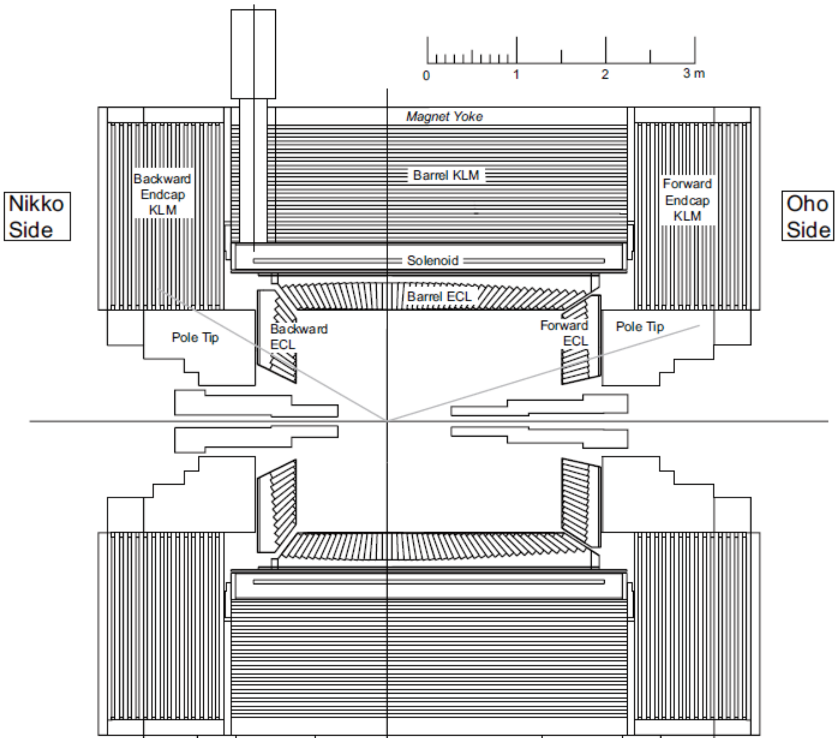


Figure 3.16: A schematic of the KLM. The barrel, forward end-cap, and backward end-cap comprised of alternating layers of iron and detector medium enclose the solenoid and other sub-detector components that make up the Belle II detector [29].

The three regions are comprised of alternating layers of iron plates and detector medium. The 4.7 cm thick iron plates act as the flux return for the magnetic field and represent approximately 3.9 interaction lengths, providing ample opportunity for transiting particles to interact with the iron and eject charged particles or photons. Each region contains 15 layers of iron plates and 14 layers of detector medium, the outermost 12 of which in the barrel are glass-electrode resistive plate chambers (RPCs). The RPCs consist of two rectangular parallel float glass electrodes held at a voltage difference and filled with a mixture of argon, butane-silver, and HFC-134a gas. Charged particles passing through the gas can ionize the gas molecules, creating free electrons that are accelerated by the electric field between the electrodes. These electrons can further ionize gas molecules, instigating an 'avalanche' of electrons that are directed to the electrodes to produce a signal.

The RPCs experience a dead time, a period of insensitivity while the detector medium resets after receiving a signal. The anticipated luminosity of the Belle II experiment is associated with high levels of beam background events that may result in an event hit rate that exceeds this dead time, significantly reducing detection efficiency. This effect would be especially prominent in the edges of the inner barrel layers and the entire forward and backward end-cap regions, as they are closer in angular alignment to the beam trajectories. To address this, scintillator strips are used as the detector medium instead of RPCs in the inner two barrel layers and all 14 end-cap layers.

The scintillator strips provide a similar granularity to the RPCs and vary in size depending on physical constraints imposed by the detector, having a cross section varying from 7 x 40 mm to 10 x 40 mm and a length up to 2.8 m (Fig 3.17). The strips are composed of polystyrene doped with scintillating phosphate and benzene compounds and are grooved to hold a wavelength-shifting optical fiber. Particles and photons passing through the scintillating material can cause the emission of light that subsequently enters the fiber and is transported to a silicon photomultiplier (SiPM). Each SiPM consists of a matrix of silicon photodiode pixels that collect the light and generate an electrical signal. The wavelength-shifting properties of the fiber allow light to reach SiPMs at a wavelength associated with higher quantum efficiency. The use of SiPMs was necessary as limited detector space and the presence of a magnetic field prevented the use of conventional photomultiplier tubes [45].

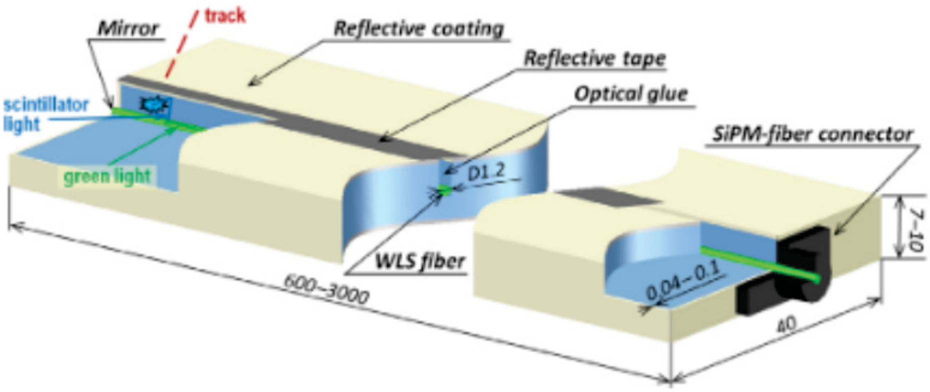


Figure 3.17: An illustration of the scintillator strip and optical fiber used in the end-cap and lower-barrel layers of the KLM. A mirror is used to prevent light from escaping from one end of the optical fiber, and a silicon photomultiplier is used to collect light at the other end [45].

The combination of RPCs and silicon photomultipliers in the three KLM regions provides a polar angle coverage from 20° to 155°. The RPCs in the barrel region have a spatial resolution between 1.1 cm and 1.7 cm, depending on how many cathode strips generate a signal at once. Due to difficulties in identifying where along the fiber light entered before reflecting to a photodiode, the spatial resolution of the SiPMs length-wise is only around 12 cm. However, evaluations of electronics performance indicate they can reach a time resolution of approximately 0.7 ns [45].

3.3 Background sources

In addition to typical environmental background sources, like cosmic particles and decay products from nearby materials that are incident on the detector, there are many background sources that arise due to the presence of the beams of accelerated electrons and positrons. The five primary sources are Touschek scattering, beam-gas scattering, synchrotron radiation, Bhabha scattering, and two photon processes [46].

In a beam of electrons or positrons within a storage ring or accelerator, the particles will often be quite close to each other. If the beam is dense enough, it is possible for individual particles within a bunch to feel a strong enough repulsive electromagnetic force from other particles in the bunch to undergo Coulomb scattering. This intra-beam scattering, known as Touschek scattering, can cause a number of particles to deviate from the intended beam trajectory and, in some cases, travel towards sensitive Belle II detector regions rather than the interaction point. Unfortunately, the narrow beam size needed to achieve the high luminosity provided by the SuperKEKB accelerator also serves to amplify the amount of Touschek scattering that occurs. To address this, collimators are used throughout the beam pipe to remove scattered beam particles before they reach the interaction region and metal shields are used on the outside of the detector to protect sensitive regions from these particles. However, it is expected that the detector may still collect energy deposited by some of these particles.

Despite efforts to hold the beam pipe in complete vacuum, a small number of gas molecules may remain. Beam-gas scattering occurs when one of these stray molecules within the beam pipe drifts close enough to the beamline to interact with the accelerated particles. If the particles interact via Coulomb scattering, the electrons and positrons can be deflected away from the beamline, similar to those that experience Touschek scattering. The electrons and positrons can also emit bremsstrahlung radiation if they pass close enough by a charged gas molecule, losing energy and drifting away from the beam line. The collimators and metal shielding used to mitigate Touschek scattering also serve to reduce the number of beam-gas scattered particles that are incident upon sensitive detector regions.

Synchrotron radiation is a type of polarized electromagnetic radiation that arises any time a charged particle is accelerated radially. It can be thought of as a magnetic analog to bremsstrahlung radiation, in that it originates from the presence of magnetic fields that slow down a charged particle along a given axis, rather than originating from the presence of another nearby charged particle. Strong focusing magnets are used in the Belle II detector near the interaction point to guide the beams together, creating synchrotron radiation close in proximity to sensitive regions of the detector. As the energy of synchrotron radiation is proportional to beam energy squared and the magnetic field strength squared, the high energy of the two beams leads to the emission of radiation capable of damaging the detector. To reduce the impact of this effect, a layer of gold that covers the inner surface of the beam pipe is used to absorb synchrotron radiation emitted, and the beam pipe is structured in a way that minimizes the number of photons that are scattered towards the detector.

When the electrons and positrons from the two beams collide, they will often scatter off each other in the process $e^+e^- \rightarrow e^+e^-$, rather than form an Υ(4S). This Bhabha scattering, discussed briefly in section 2.1.1, can cause the trajectory of at least one of the originally scattered particles to deviate significantly from the beamline and produce electromagnetic showers if it collides with detector elements. Bhabha scattering can also lead to the emission of a photon that will often propagate along the beam axis and interact with the iron magnets along the beam pipe. These interactions result in the production of low energy neutrons which occasionally enter the KLM region of the detector. The amount of background particles created by this process is proportional to the beam luminosity. Despite the inclusion of focusing magnets meant to reduce the impact of Bhabha scattering, the high luminosity of the beams makes such events a non-negligible source of background particles.

In addition to Bhabha scattering and the formation of an Υ(4S), the interaction of an electron and positron can also lead to the two photon process $e^+e^- \rightarrow (e^+e^-)f^+f^-$. The initial scattering of the electron and positron results in an emission of two virtual photons which then interact to produce a pair of leptons or hadrons (Fig 3.18). The produced photons are typically low energy, allowing only for their interaction to produce a low energy electron-positron, muon-antimuon, or positively and negatively charged pion pair (though other pairs are possible). If of sufficiently low momentum, this pair may be caught by the magnetic field of the detector's solenoid and spiral through inner detector regions, like the PXD, multiple times. The original scattered electron and positron in this interaction are often not deflected very much, remaining nearly colinear with the beamline and avoiding most detector components.

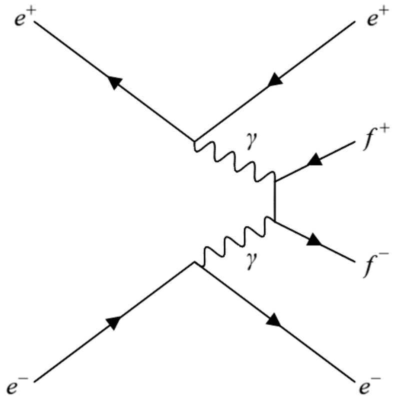


Figure 3.18: A Feynman diagram of the two photon process $e^+e^- \rightarrow (e^+e^-)f^+f^-$. Most commonly, the photons produce an electron-positron, muon-antimuon or positively and negatively charged pion pair. The original electron and positron will often not be scattered enough to enter the detector acceptance region and continue down the beam pipe undetected.

Chapter 4

Analysis

The goal of this analysis is to use a supervised machine learning technique to create a tool for suppressing beam background-related extra energy when analyzing data from the Belle II detector. The nature of supervised learning algorithms requires the use of samples from two states of data as training information for the program. Consequently, in order to train an algorithm to distinguish between extra energy arising from beam background sources and extra energy arising from photons relating to the reconstructed event (signal photons), it is necessary to provide it with pure samples of both. To do so, $e^+e^- \to \gamma \to \mu^+\mu^-$ and $B^0 \to D^{*+}\ell^-\nu_\ell$ events are reconstructed from Belle II data and the extra detector information not used in the reconstruction is aggregated for each event. Specific selections are made, discussed herein, that allow for an isolation of this leftover detector information that relates only to the desired photons for each sample. The created samples are then merged and used for training.

In this case, the mechanism by which the algorithm learns from the provided samples is an analysis of the information collected by ECL crystal clusters in each sample. When photons activate crystals in the ECL, the information recorded by the detector electronics is processed and organized into a number of variables that describe each activated crystal cluster. These variables are studied to determine which exhibit a distribution shape that differs between event and background photons while also showing adequate agreement between simulated events and experimental data. A set of variable distributions are chosen based on these characteristics, as described in this chapter, and used as training data in a boosted decision tree program. The output of this training is a classifier object that can be used to assign a probabilistic variable to any activated ECL cluster that indicates the likelihood that it was activated by a photon originating from a signal event. If successful, selections on this variable can be used to suppress the inclusion of ECL clusters activated by background photons in the extra energy distributions of reconstructed events.

4.1 Event reconstruction

4.1.1 Di-muon event reconstruction

In addition to the background sources described in Chapter 3, the electron-positron collisions induced by the accelerator can also cause di-muon production, an interaction that may also produce a photon by radiation from one of the involved leptons (Fig 4.1).

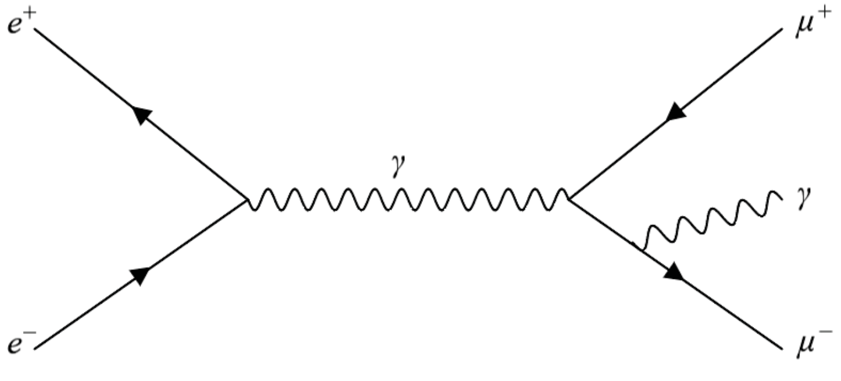


Figure 4.1: A Feynman diagram of an electron and positron colliding to produce a positive and negative muon. In addition to being mediated by a virtual photon, it is possible for this interaction to produce a photon by means of initial or final-state radiation.

With respect to the number of involved particles this is a relatively simple interaction, making event reconstruction straight-forward. This simplicity allows direct cuts on certain detector parameters to be very effective in isolating specific components of the interaction, which in turn makes it possible to examine these events in a relatively ’clean’ environment that contains only the desired particles. These features make this event type an excellent reconstruction candidate for the purposes of creating pure samples of background or signal photons for training.

Although the number of electron-positron collisions that result in the production of two muons is small relative to other particles at the SuperKEKB collision energy, identifying these interactions is possible with minimal cuts. As only two charged particles are expected as products from this interaction, reconstructing events that result in only two charged tracks being recorded in the CDC serves to exclude a number of other events from the reconstruction process. It is also required that two oppositely charged particles are detected in the KLM detector with a timing that corresponds to the charged tracks. A further selection is made on a single particle identification variable, muonID, requiring that the probability that the detected particles are muons is 90% or greater. MuonID is part of a set of particle identification (PID) variables used in the Belle II collaboration that assign particle likelihood values based on an assessment of a variety of information collected by the different components of the detector [47].

Because it is unclear at this time if the forward and backward end-caps of the KLM detector have a comparable performance to the barrel, only muon pairs detected in the barrel region are used in event reconstruction. This was accomplished by applying an angular selection on the location of the detected muons which removes any that weren’t detected between 45° and 125° in the detector, the angular limits of the KLM barrel region.

Once two oppositely charged muons are identified and correlated to an electron-positron collision event by the timing of their detection, a di-muon candidate is reconstructed. Any information collected by the detector around the same time that isn’t used in the reconstruction is organized into a ’Rest of Event’ (ROE) grouping for that di-muon candidate. This reconstruction process allows an output file to be created containing a ROE with ECL cluster information organized into various cluster variables, which is then used as training information for the multivariate analysis.

4.1.2 Semileptonic event reconstruction

As previously mentioned, the primary purpose of colliding electrons and positrons within the Belle II detector at the Υ(4S) resonance is the high probability of a decay that produces B meson pairs. Although it is possible to perform the multivariate analysis training with background and signal ROE photons isolated from di-muon events (described in section 4.3), there are some concerns about the applicability of the resulting tool to other studies. The signal photons from semileptonic events that can be isolated in a ROE are on average higher energy than those that can be isolated from di-muon events. It is possible that a classifier tool trained on lower energy signal photons may perform poorly when applied to ROE ECL clusters for event reconstructions involving higher energy photons. The large number of studies performed using Belle II data that involve such semileptonic decays emphasizes the need for this possibility to be investigated. For this reason, in addition to training samples made from di-muon events, signal ROE photon training samples made from the semileptonic decay $B^0 \to D^{*+}\ell^-\nu_\ell$ are also used in this study.

Reconstructing these events is reasonably more complicated than reconstructing di-muon events; entire studies have been performed on developing and optimizing the procedures necessary to do so. Rather than independently repeating these endeavours, this study leverages the work already completed by one such recent study that was completed in parallel to the work presented here. Although a brief description of the selections and tools used in the reconstruction process follow, a thorough explanation can be found in the original publication [48].

As the B^0 decay produces a neutrino that does not interact with the Belle II detector, a successful reconstruction requires the accurate identification of all the other Υ(4S) decay products in order to determine the missing energy, relative to the initial 10.58 GeV collision energy, that can be attributed to a neutrino. This limits the reconstruction to events where only one of the B mesons produced by the decay of the Υ(4S) has a neutrino as a product. The other B meson must therefore decay completely hadronically and be reconstructed first.

This hadronically decaying B meson, termed B_{tag}, is reconstructed using a Full Event Interpretation (FEI) algorithm [49]. This algorithm uses a series of boosted decision trees to determine the most probable particle candidates for each decay step, starting with the final decay products and using the output of each prior decision tree to work back towards the original B meson (Fig 4.2). Tracks, vertex information, and neutral energy deposits in the detector are first analyzed in order to identify the probability that they correspond to final state particles like photons, electrons, protons, and muons. These particle candidates and their associated probabilities are then combined, using known parent particle branching fractions, to determine the probability that certain parent particles, like Λ baryons or J/ψ mesons, existed in the detector. A further step occurs to similarly determine the probability of the presence of D mesons, and the combined output of each step is again processed to identify the presence of B meson candidates.

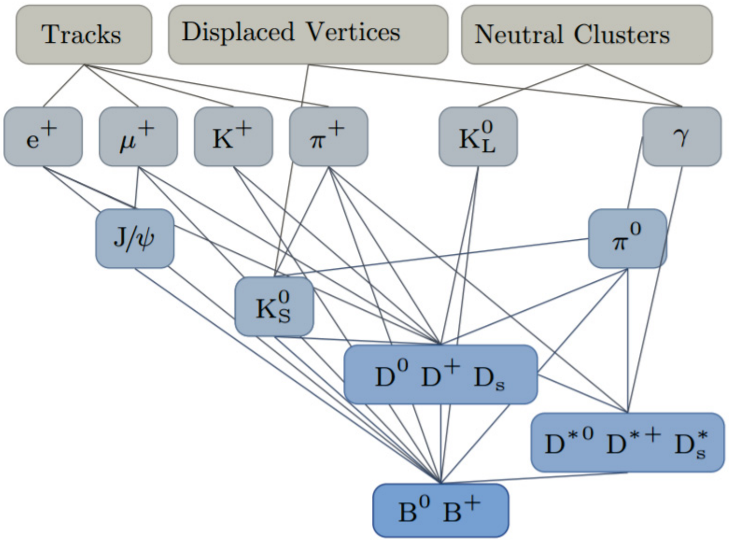


Figure 4.2: An overview of the FEI reconstruction process of hadronically decaying B mesons. Track, vertex, and neutral cluster information is analyzed to determine particle candidates in stages, eventually leading back to a reconstruction of the original B meson [49].

Once the B_{tag} meson candidate has been identified, the other B meson, termed the B_{sig}, is reconstructed from the remaining detector information. In this case, B_{sig} corresponds to either a B^0 meson decaying as $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$, with D^{*+} decaying as $D^{*+} \rightarrow D^0\pi^+$ and D^0 decaying as $D^0 \rightarrow K^-\pi^+$, or the charge conjugate equivalent. A number of selections are used to improve the reconstruction process. For example, the reconstructed D^0 meson must have an invariant mass that falls within around 10 MeV of the expected mass, and the difference in mass between the D^* meson and its daughter D^0 meson must lie within a range of 143 - 148 MeV/c^2. The leptons are required to have a center-of-mass momentum of at least 1 GeV/c and exhibit the characteristics necessary to meet the criteria for a 90% particle identification probability or greater. In the case where this process provides multiple reconstruction paths that meet the stated criteria, the path with a D^* mass that more closely aligns with the world average is chosen.

At this stage, the four-momenta of the B_{tag} and reconstructed B_{sig} components are subtracted from that of the initial electron-positron collision to determine the square of the missing mass that is attributable to the single neutrino expected from the B_{sig} decay, with a value near zero expected. With this reconstruction complete, a ROE is built from the remaining detector data. When this process is applied to events simulated without backgrounds, it is possible to obtain a ROE containing primarily signal-type photons that can be used for multivariate analysis training.

4.2 Data samples

This reconstruction of events that is necessary to build a ROE and create the samples needed for training is conducted on both simulated and experimental data collected by the detector.

The Belle II detector collects data under various conditions, both at the 10.58 GeV Υ(4S) resonance and off-resonance. Although the detector has achieved a total integrated luminosity of over 70 fb^{-1} thus far, the collection of data from these events occurs in discrete data collecting runs. The experimental data used in this analysis was collected on-resonance and processed during 2019. It is referred to as 'proc10' data and corresponds to an integrated luminosity of 5023.4±0.9 pb^{-1}.

The simulated background and signal event data is created using a Monte Carlo (MC) simulation approach, often referred to as background MC and signal MC, respectively. In this case, the generation of MC simulation relies on iterative random sampling within specified parameters and supplied probabilities to simulate the frequency of specific particle interactions within the detector and the varying kinematic characteristics they can exhibit. The simulated Belle II data is created in campaigns, with each one aiming to resolve inaccuracies or errors found within the prior set. The early stages of this analysis were performed using MC from the 15th and 16th campaigns (termed MC11 and MC12, respectively), but all of the final work presented in this thesis was completed using MC13, from the 17th campaign. MC simulation is a broad field of study with applications in numerous areas of science and mathematics. Although it falls outside the scope of this analysis, in-depth reviews of the simulation techniques and generators used in the Belle II experiment exist elsewhere [46].

There are different types of MC available for use in Belle II studies. Signal MC models a large number of a chosen type of event, and can exist with beam backgrounds present or without. For example, although the previously described di-muon events are relatively rare in the detector, simulated data sets containing a large quantity of them can be generated. The beam backgrounds in these samples can be either run-dependent (referred to as MCb), reflecting the characteristics of beam backgrounds observed in specific runs of data collection, or run-independent (referred to as MCa), exhibiting generalized characteristics of beam backgrounds.

Generic MC also exists, containing particle interactions in the proportions that they are expected to be observed in the detector. Generic MC is most commonly broken down into seven categories of event type; B^+B^- and $B^0\overline{B}^0$, representing charged and neutral B meson production from the decay of the Υ(4S) resonance, $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, and $c\bar{c}$ for quark pair production from the electron-positron collisions (referred to as continuum events), and $\tau^-\tau^+$ for the production of tau pairs from the collisions. Most comprehensive analyses are required to perform particle reconstruction over all of these types of generic MC to obtain an accurate simulation of how the same reconstruction procedures and selections will perform when applied to actual detector data.

4.3 Training sample creation and training process

The multivariate analysis performed to create the beam background suppression tool makes use of a boosted decision tree program, as described in section 2.2.2. The program is provided with two different input data samples and is 'trained' to recognize differences between the two. The output of this training is a classifier object that can be applied to new data samples to quickly identify the probability that its contents resemble one of the input samples or the other.

In this case, the tool is being developed to recognize differences in ECL cluster variable distributions between signal event-related and beam background-related ROE photons. The creation of two samples containing these photons requires reconstructing di-muon and semileptonic events, as described in section 4.1, and building a ROE with the selections necessary to remove unwanted photons.

As beam backgrounds are generally not dependent on the event reconstructed, a sample of beam background ROE photons created by reconstructing di-muon events will sufficiently represent the beam backgrounds present for other events. When di-muon events radiate a photon, conservation of momentum requires that the trajectories of the two muons are offset by the component of momentum attributed to the photon. This means that the two muons will rarely be detected in the KLM detector with a 180° separation, or 'back-to-back', if a photon is radiated. Applying a selection that requires the angular separation of the two muons to be 179.62° or greater removes any that are not back-to-back from the reconstruction process and prevents events with a radiated photon from being reconstructed. The inclusion of this selection ensures that the created ROE will not contain detector information from any of these radiated photons and results in the creation of a data sample with a ROE containing ECL cluster information associated only with beam background-related photons.

As described in section 4.1.2, variations in photon energy necessitate an examination of ROE signal photons from both di-muon and semileptonic reconstructions. It is possible to isolate signal photons in the ROE of di-muon events by performing a similar reconstruction process without the back-to-back selection on MC13a events that have beam backgrounds removed. This results in the creation of a data sample with a ROE containing only ECL cluster information associated with signal photons that have been radiated from the di-muon interactions. To instead isolate signal photons from semileptonic events, the same $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events described in section 4.1.2 are reconstructed. When these events are reconstructed using beam background-free signal MC13a simulation, the constructed ROE contains primarily signal-type photons.

Although the choice of data sets used in creating each training sample and the other selections made in the reconstruction process adequately remove most unwanted activated clusters from the ROE, it does not remove them all. In B meson decays, charged hadrons can interact with detector components to produce pions, neutrons, or other hadrons. These pions can subsequently decay into two photons which can then be collected by ECL crystals, and the other hadrons can interact hadronically to deposit energy in ECL crystals that can resemble the deposits of photons. These sources of non-beam background-related ECL cluster activation, referred to as 'hadronic splitoffs', do not have a charged track associated with them and will often result in additional energy being included in the ROE of reconstructed semileptonic events. Although this type of background contribution to a reconstructed ROE could potentially be suppressed by developing a classifier similar to the one described by this thesis, such a tool falls outside the scope of this study and a further selection is required to remove hadronic splitoff clusters from the training samples used here.

The boosted decision tree program used in the multivariate analysis is structured to distinguish between input data distributions by a designated target variable, which in this case was chosen to be ‘isSignal’. This variable is a binary classifier, returning a 1 if the ROE photon meets the criteria of being equivalent to a generated signal event-level photon, and 0 if it does not. In addition to the selections used to reconstruct di-muon and semileptonic events, the construction of the ROE also has an isSignal requirement. The signal ROE photon samples have the requirement that ROE photons have an isSignal value of 1, which effectively removes clusters activated by hadronic splitoffs from semileptonic signal ROE photon samples. The beam background ROE photon samples are also required to have an isSignal value of 0. This variable is only meaningful when examining simulated particles, so the isSignal selection on training samples created from experimental data has no impact. Fortunately, particles from experimental data have an isSignal value of 0 by default and are only used for creating beam background training samples in this study, so the use of isSignal as a target variable remains functional.

Once pure beam background and signal samples are created containing a similar number of ROE photons, they are merged into a single file to be used in training. The boosted decision tree program analyzes the ECL cluster variable distributions contained in the file to identify differences between ROE photons with an isSignal value of 0 and those with a value of 1. The classifier output of this process can then be used to determine a probability that a given ROE photon is associated with either beam background or signal events. The final step of the training process is the application of this classifier to a second merged file containing a new mix of beam background and signal ROE photons that serves as a testing sample, allowing for an initial evaluation of the performance of the created classifier object.

4.4 Cluster variables relevant to analysis

When ECL crystal clusters are activated by a photon, the detector electronics system records a variety of information. The location of the cluster within the detector, the timing of the cluster's activation, and the overall magnitude of the signal produced are among the types of information recorded that can later be organized into distinct variables for each activated crystal cluster. The distribution of these variables over many events can then be used in the multivariate analysis performed by the supervised learning program.

The suitability of a cluster variable as a tool for suppressing background photons from the ROE is determined by two factors. Primarily, the variable distributions must exhibit some divergence in shape between background photon distributions and signal photon distributions for the machine learning program to identify as a distinguishing characteristic. As well, the variables must show good agreement between simulated distributions and those from experimentally collected data. This is to ensure that training performed on simulated data will result in a tool that is applicable to experimental data without any significant loss in efficacy. These factors are determined largely by visual observation of the variable distributions.

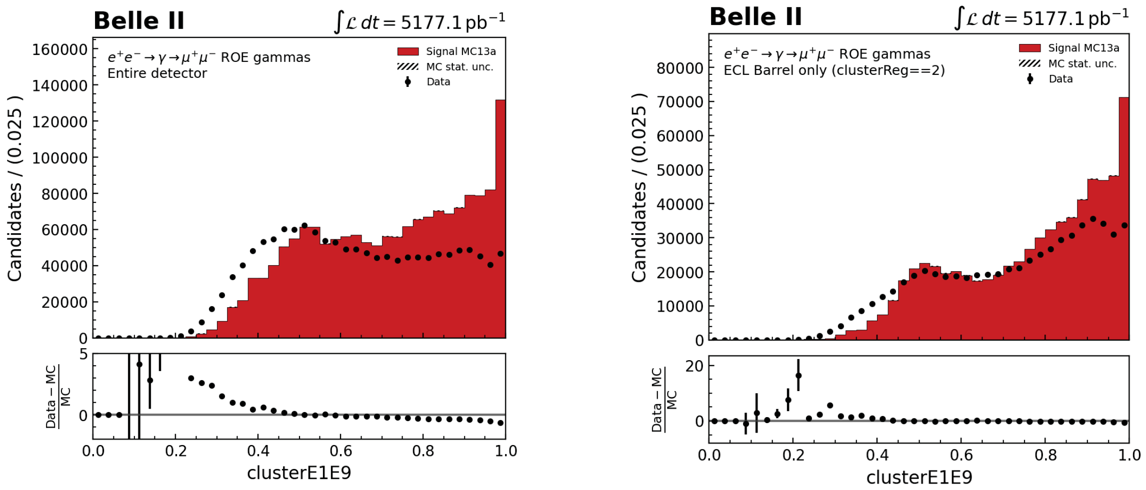
To look for divergence in variable distributions between beam background and signal photons, $e^+e^- \to \gamma \to \mu^+\mu^-$ events are reconstructed on run-independent MC13 and the resulting ROE distributions are compared to those of photons generated with a selection on energy to simulate signal photons. These generated 'tight' photons are required to have an energy over 50 MeV if located in the barrel or forward end-cap regions of the ECL, or over 75 MeV if located in the backward end-cap. To examine the agreement between simulated and experimental data, the same MC13 ROE distributions are compared to those of di-muon events reconstructed on data. Although a large number of variables were considered for this analysis, the Belle II experiment only recently began collecting data and it has not yet been possible to address every deficiency that exists in the performance and quality of some variables. As a result, only seven are deemed optimal for use in training at this time; clusterE1E9, clusterLAT, clusterSecondMoment, clusterZernikeMVA, clusterPhi, clusterTheta, and clusterE. These variables, as well as some that are not used but may be of value in any comparable future analysis, are described in this section.

4.4.1 clusterE1E9

The variable clusterE1E9 is a ratio of the energy of the central crystal in a cluster of activated ECL crystals to the summed energy of the 9x9 crystal grid that it centers. As the central crystal is included in the sum, values can range between 0 to 1.

This variable is reasonably well modelled in the simulated data. When considering the entire detector, clusterE1E9 agreement between data and MC is reasonable but not perfect (Fig 4.3a). However, if only the barrel region of the ECL is isolated, agreement appears to improve considerably (Fig 4.3b). A higher performing beam background suppression tool could possibly be created by training only on these barrel clusters instead of the entire detector, however the tool would only be optimized for the same region of the detector in future studies. It was deemed more worthwhile to develop a procedure for making a beam background suppression tool that would not preclude a large portion of collected ECL information (which could limit future studies it may be used in), and instead allow the performance of future iterations of the tool to be optimized by ongoing improvements to simulation agreement.

When MC13 ROE clusterE1E9 distributions are compared to those of photons generated with a tight selection on energy, shape differences are apparent (Fig 4.4). A peak in the tight photon distribution exists around clusterE1E9 = 0.85 that is not present in the MC13 ROE photon distribution. This feature should allow for some amount of discrimination between signal and beam background photons and make clusterE1E9 a reasonable candidate variable for use in the training samples.



(a) **(b)**

Figure 4.3: A comparison of clusterE1E9 distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. Values are presented from (a) from the entire detector and (b) the barrel region only.

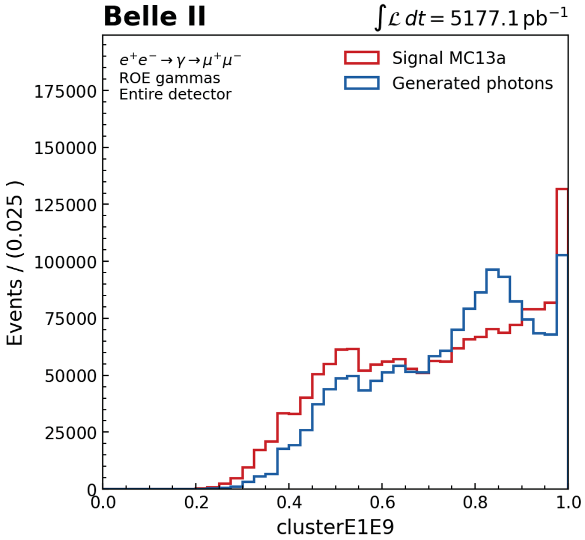


Figure 4.4: A comparison of clusterE1E9 distributions between ROE photons from reconstructed MC13a $e^+e^- \to \gamma \to \mu^+\mu^-$ events and generated tight photons. A clear shape difference exists around the value 0.85.

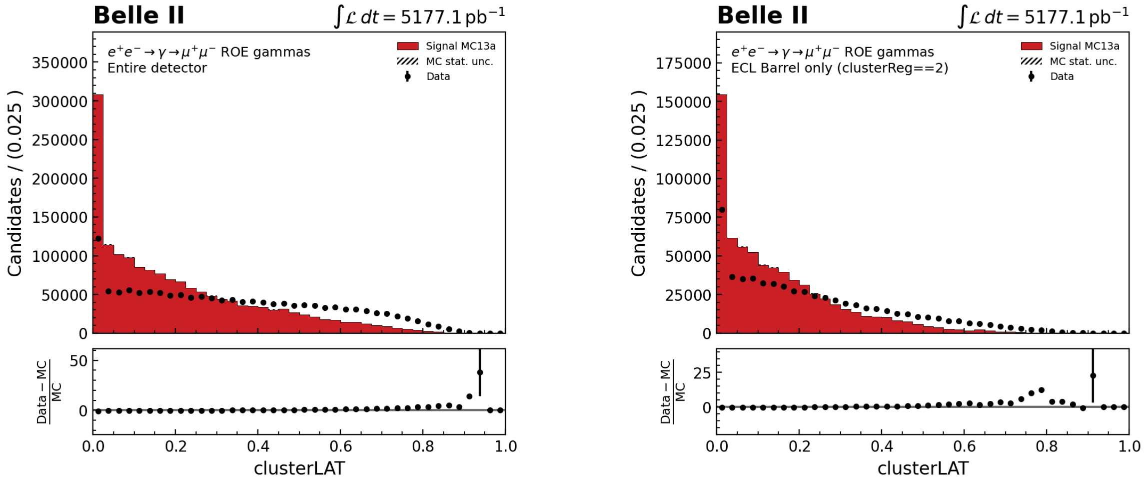
4.4.2 clusterLAT

The variable clusterLAT represents the lateral energy distribution of a crystal cluster and, like clusterE1E9, ranges from 0 to 1. The value corresponds to a ratio of the weighted sum of the product of the crystal energy and the distance from the cluster center squared for all but the two most energetic crystals, compared to the same weighted sum for all the crystals in the cluster. The clusterLAT value, S, is explicitly defined as:

$$S = \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^{n} w_i E_i r_i^2}, \tag{4.1}$$

where E_i is the energy of the ith crystal in a cluster containing n+1 crystals, listed in order of descending energy, and r_i is the distance of the ith crystal from the center of the crystal cluster. The minimum distance between two crystals, r_0, is 5 cm. The weighting factor for each crystal, w_i, is determined as a ratio of the crystal energy to the total energy of the crystal cluster.

This variable describes the spread of energy on the axis perpendicular to the axis of the electromagnetic shower trajectory. If the central crystals, E_0 and E_1, contain the majority of the cluster energy and the remaining crystals are evenly distributed a short distance from the center, the value will be closer to 0. Such radially symmetric showers, like those associated with isolated photon events, generally result in a clusterLAT value around 0.3. This is evidenced by the pronounced peak in the distribution for generated tight photons that does not exist in the distribution for MC13 ROE photons (Fig 4.6). Hadronic events or lepton events with nearby radiated photons often result in a less symmetric lateral energy distribution and are associated with larger values. Despite sub-optimal agreement between experimental and simulated data, primarily in the ECL end-caps (Fig 4.5), these characteristics makes this variable a sufficient candidate for use in training. It is hoped that the revisions that come with future simulation campaigns will further improve agreement.



(a) **(b)**

Figure 4.5: A comparison of clusterLAT distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. Values are presented from (a) from the entire detector and (b) the barrel region only. Although both figures do not show excellent agreement, it is evident that removing end-cap data improves agreement.

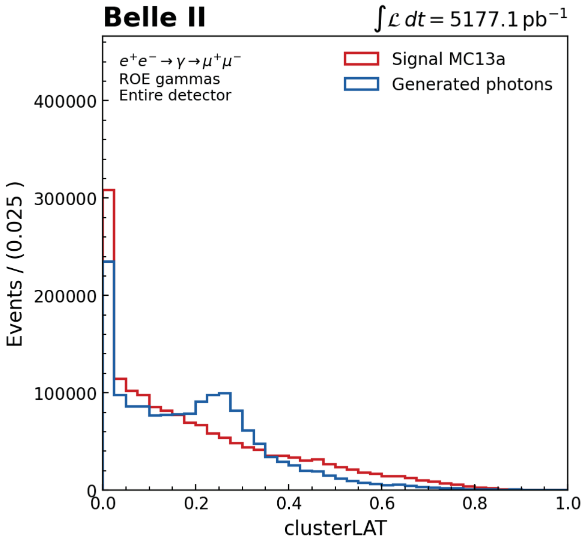


Figure 4.6: A comparison of clusterLAT distributions between ROE photons from reconstructed MC13a $e^+e^- \to \gamma \to \mu^+\mu^-$ events and generated tight photons. A clear shape difference exists around the value 0.3.

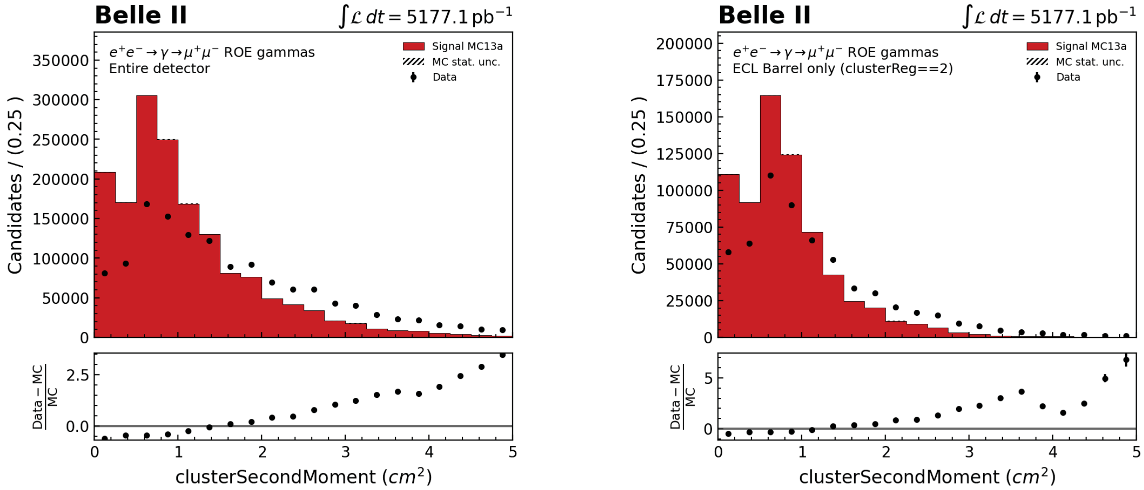
4.4.3 clusterSecondMoment

The variable clusterSecondMoment is similar to clusterLAT in that it describes a distribution of energy within a cluster with respect to a plane arbitrarily defined to be perpendicular to the axis of the photon shower trajectory. However, unlike the dimensionless clusterLAT, it involves the ratio of a weighted sum of the product of the crystal energy and the distance from the cluster center squared for all of the crystals in the cluster, to a weighted sum of the crystal energies, resulting in values with dimensions of distance-squared that typically range from 0 to 40. It is defined explicitly as:

$$S = \frac{\sum_{i=0}^{n} w_i E_i r_i^2}{\sum_{i=0}^{n} w_i E_i}, \tag{4.2}$$

where again, E_i is the energy of the *i*th crystal in a cluster containing *n*+1 crystals, r_i is the distance of the *i*th crystal from the center of the crystal cluster, and w_i is an energy-based weight value. While clusterLAT may be viewed as a numeric representation of the amount of energy in a cluster displaced from the center, clusterSecondMoment is more of a description of the overall layout of the cluster's energy.

Like clusterE1E9 and clusterLAT, clusterSecondMoment shows sufficient enough agreement between experimental and simulated data to be used in training, with mild improvement when only considering the barrel region (Fig 4.7). Although clusterSecondMoment distribution shape differences between reconstructed ROE photons and generated photons are not quite as distinct as in the aforementioned variables (Fig 4.8), it was determined that the minor differences were enough to warrant the use of this variable in training.



(a) **(b)**

Figure 4.7: A comparison of clusterSecondMoment distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. Figure (a) presents values from the entire detector, while figure (b) presents values from only the barrel region. As in prior variable distributions, agreement is evident and improves when considering only the barrel region of the ECL.

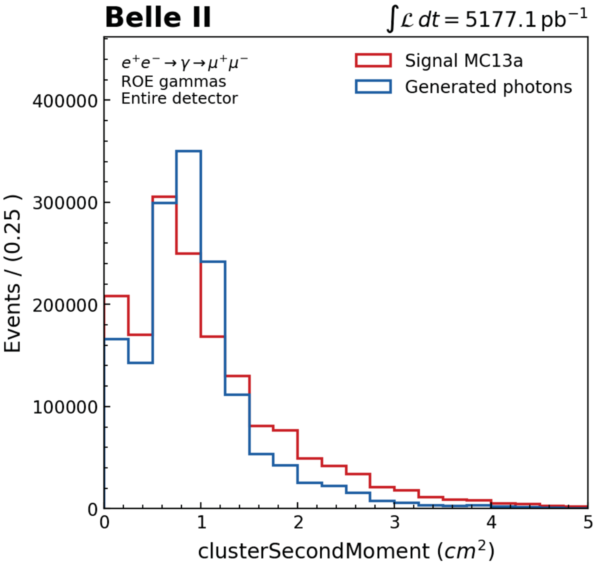


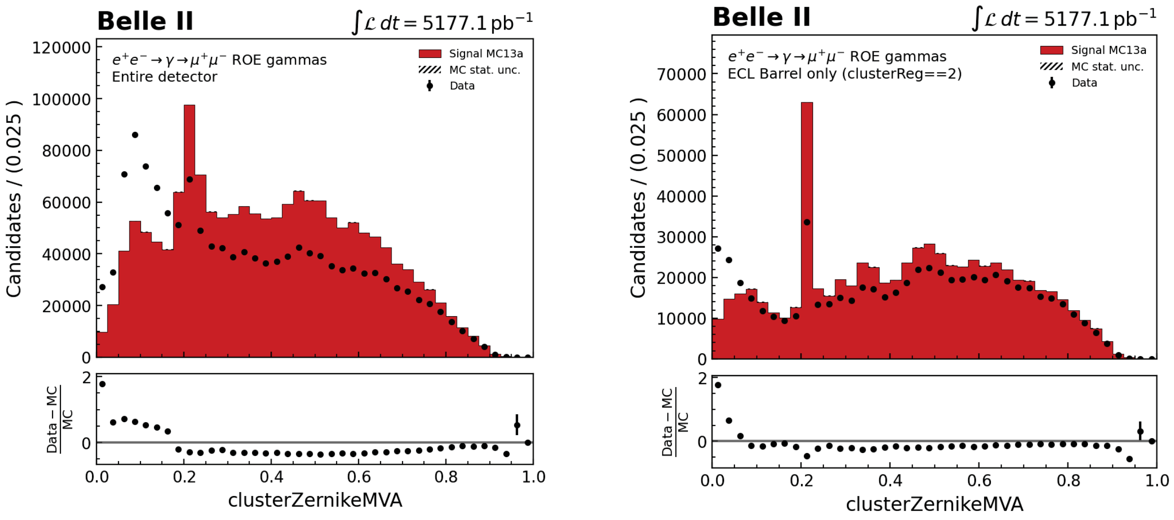
Figure 4.8: A comparison of clusterSecondMoment distributions between ROE photons from reconstructed MC13a $e^+e^- \to \gamma \to \mu^+\mu^-$ events and generated tight photons.

4.4.4 clusterZernikeMVA

Zernike polynomials are complex two-variable polynomials that play an important role in the mathematics used in image recognition, optics analysis, and in the case of Belle II, particle identification [50]. There are a number of different Zernike polynomials, all of which are orthogonal on the unit disk and invariant under rotation, with values ranging from -1 to 1. These polynomials can be used to generate Zernike moment values that characterize the shape of detected electromagnetic showers.

Although a large number of Zernike polynomials and their associated moments exist, only eleven are used in the current Belle II framework. A multivariate analysis is performed, much like the one described in this thesis, in order to aggregate these eleven values into a single 'clusterZernikeMVA' variable that is primarily used to distinguish between photon and K_L^0 events in the ECL. Values range between 0 and 1, with a value closer to 1 indicating a higher likelihood that the crystal cluster was activated by a photon event. A thorough description of Zernike moments and clusterZernikeMVA can be found in internal Belle II documents [51].

In the samples examined, agreement between experimental and simulated data is reasonably strong in the barrel region, but somewhat weaker when considering the entire detector (Fig 4.9). Although a peak close to the origin exists in experimental data that is not present in simulated events, agreement in other regions of the distribution is adequate and it is hoped that future iterations of the simulated data will resolve the difference. A strong divergence exists between simulated ROE photons and generated signal photons at large clusterZernikeMVA values, suggesting that this variable may provide strong distinguishing power when used in training (Fig 4.10).



(a) **(b)**

Figure 4.9: A comparison of clusterZernikeMVA distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. Figure (a) presents values from the entire detector, while figure (b) presents values from only the barrel region.

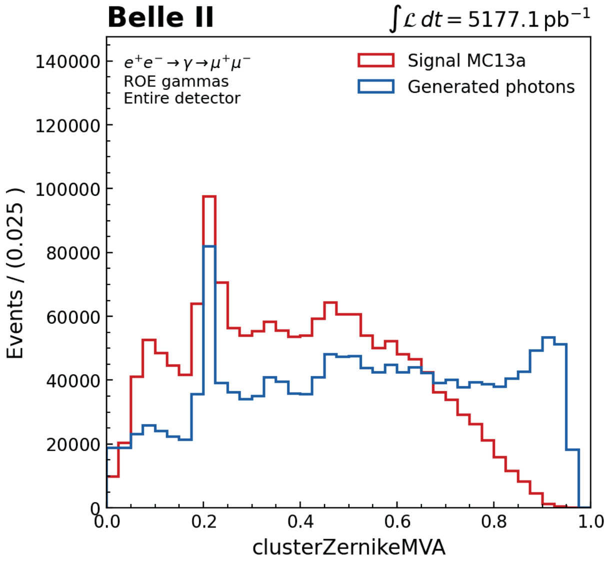


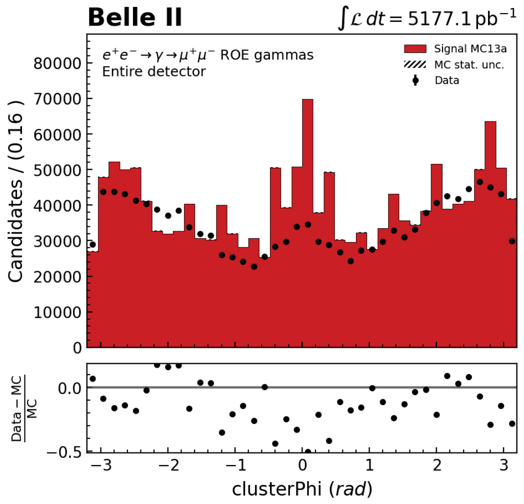
Figure 4.10: A comparison of clusterZernikeMVA distributions between ROE photons from reconstructed MC13a $e^+e^- \to \gamma \to \mu^+\mu^-$ events and generated tight photons. Considerable shape differences exist between the two samples, suggesting that clusterZernikeMVA may perform well as a background distinguishing variable.

4.4.5 clusterPhi and clusterTheta

The variables clusterPhi and clusterTheta represent the angular position of a crystal cluster in the ECL with respect to the center of the detector. ClusterTheta corresponds to the polar angle of the cluster, with $\theta = 0^\circ$ aligning with the beam axis at the forward end of the detector and $\theta = 180^\circ$ aligning with the beam axis at the backward end. ClusterPhi is the azimuthal angle of the cluster, with $\phi = 0^\circ$ corresponding to the horizontal axis directed outside of the accelerator ring. The central position of a cluster is determined as a weighted average position of activated crystals, with a bias towards higher energy crystals.

In contrast to the other variables describes thus far, neither variable shows particularly improved agreement between simulated and experimental data when considering only the ECL barrel. As is demonstrated by Fig 4.11, the distributions for clusterPhi show a generally similar shape, with some differences around clusterPhi = 0. However, overall agreement appears a bit stronger when considering the entire detector. Similarly, clusterTheta shows relatively strong agreement throughout the detector, with a few notable disagreements in the region clusterTheta < 0.9, the forward end-cap and forward region of the barrel (Fig 4.13).

Although the clusterPhi distribution for generated photons shows minor deviations from simulated ROE photons (Fig 4.12), the clusterTheta deviation shows a significant deviation (Fig 4.14). For clusterTheta > 2.3 (the backward end-cap region), a considerably larger number of activated clusters are present in the simulated ROE photon sample compared to the generated photons. This is likely due to the presence of beam background-related photons in the simulated ROE photon sample and suggests that clusterTheta will make an effective discriminating tool in the training samples. Given the reasonably strong agreement between experimental and simulated data for both variables, and the tendency for certain beam background-related photons to exhibit a positional bias, it was deemed worthwhile use both clusterPhi and clusterTheta in the multivariate analysis.



(a) **(b)**

Figure 4.11: A comparison of clusterPhi distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. Figure (a) presents values from the entire detector, while figure (b) presents values from only the barrel region.

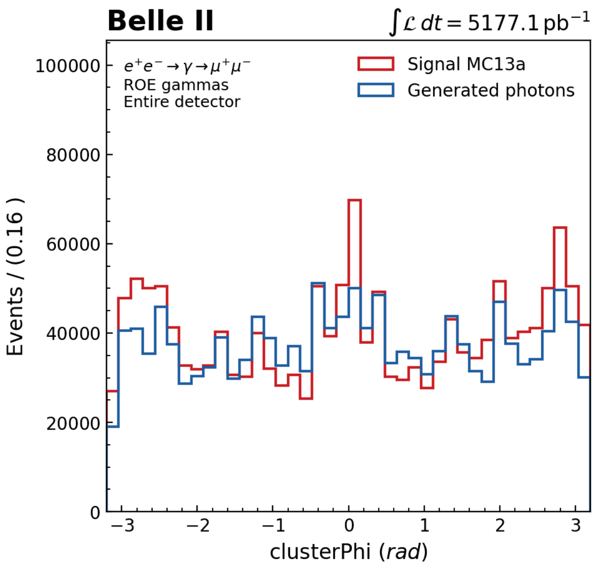
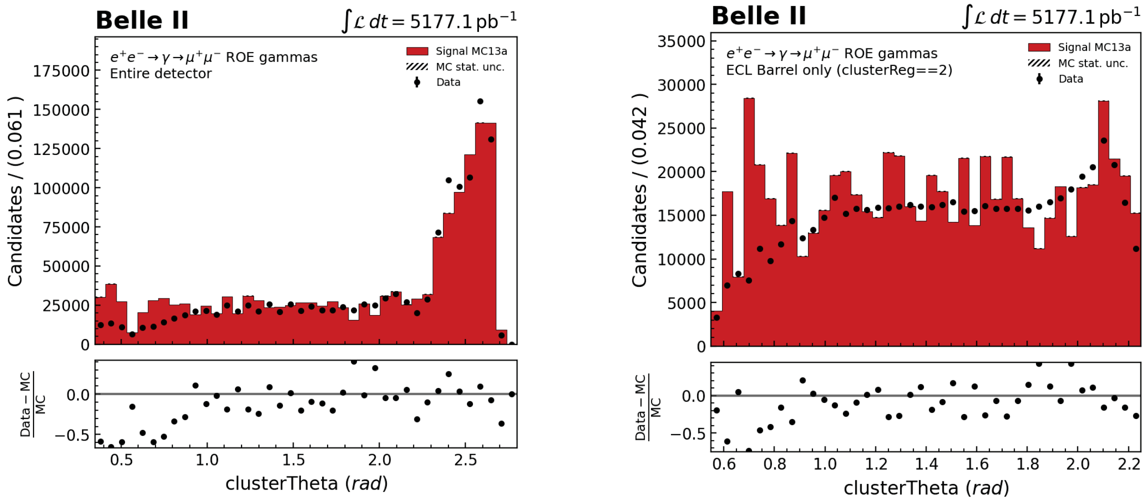


Figure 4.12: A comparison of clusterPhi distributions between ROE photons from reconstructed MC13a $e^+e^- \to \gamma \to \mu^+\mu^-$ events and generated tight photons.



(a)

(b)

Figure 4.13: A comparison of clusterTheta distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. Figure (a) presents values from the entire detector, while figure (b) presents values from only the barrel region.

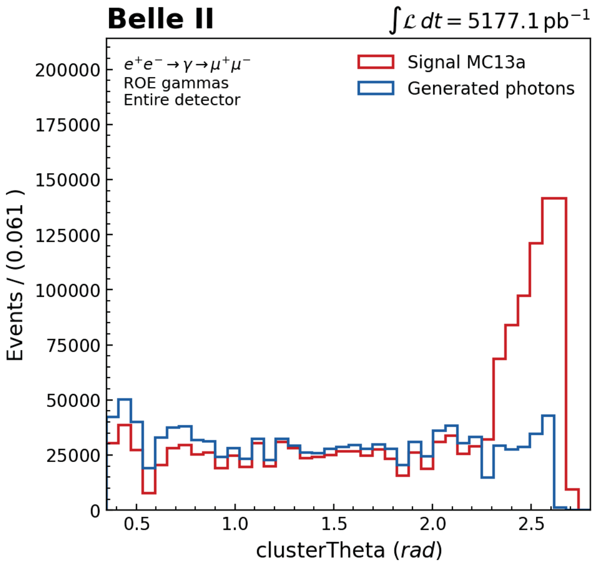
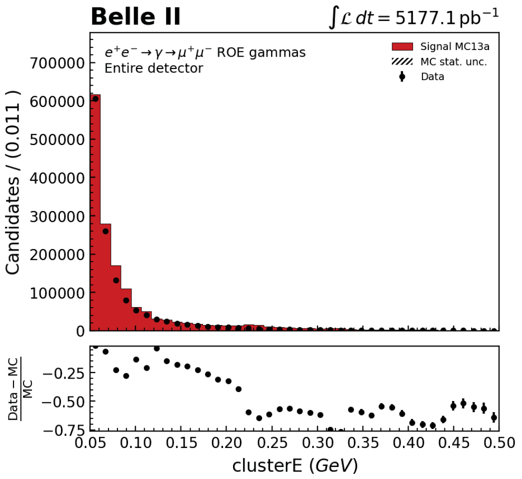


Figure 4.14: A comparison of clusterTheta distributions between ROE photons from reconstructed MC13a $e^+e^- \to \gamma \to \mu^+\mu^-$ events and generated tight photons.

4.4.6 clusterE

The variable clusterE returns the total energy of a cluster of activated crystals. The value represents a weighted sum of the energy of each crystal in the cluster, with weights used to separate crystal energy associated with any overlapping clusters. Transverse and longitudinal energy leakage in the ECL manifests as a low energy tail in photon energy distributions, so a correction factor is applied. The correction factor is calculated as the ratio of reconstructed photon energy for simulated ECL photons to the original designated energy of the photons. For experimental photon data collected in the ECL, a calibration is performed based on disagreement with simulated data in order to achieve an energy resolution precision of 1.8% or less. Only clusters with a total energy of 20 MeV or greater are recorded. For each reconstructed particle, the sum of clusterE for all of the ECL clusters in the ROE is what comprises the extra energy distribution.

Agreement between simulated and experimental clusterE values is strong (Fig 4.15). Unlike many of the other variables examined here, agreement within the barrel region appears similar to the detector as a whole, suggesting the variable is modelled well in all regions of the detector. A comparison of simulated ROE photons to generated tight photons suggests a higher proportion of non-signal photons are lower energy, as might be expected of beam background photons (Fig 4.16).



(a) **(b)**

Figure 4.15: A comparison of clusterE distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \rightarrow \gamma \rightarrow \mu^+\mu^-$ events. Figure (a) presents values from the entire detector, while figure (b) presents values from only the barrel region.

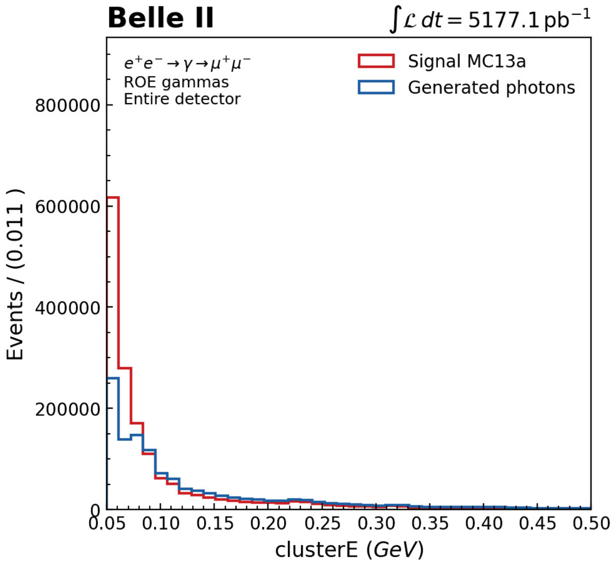


Figure 4.16: A comparison of clusterE distributions between ROE photons from reconstructed MC13a $e^+e^- \rightarrow \gamma \rightarrow \mu^+\mu^-$ events and generated tight photons.

4.4.7 Potential future variables

Although the aforementioned seven variables are used in the multivariate analysis presented here, other variables were considered. Variable distributions do not necessarily need to show a drastic difference between signal and beam background-related photons for the variable to be useful in a multivariate analysis, however the scope of this study called for the use of a limited number of variables and therefore compelled the inclusion of only the highest-impact variables available. The variables expected to show strong discriminating power between signal and beam background-related photons that were not used in this multivariate analysis are described below, along with the reason for their exclusion.

clusterTiming

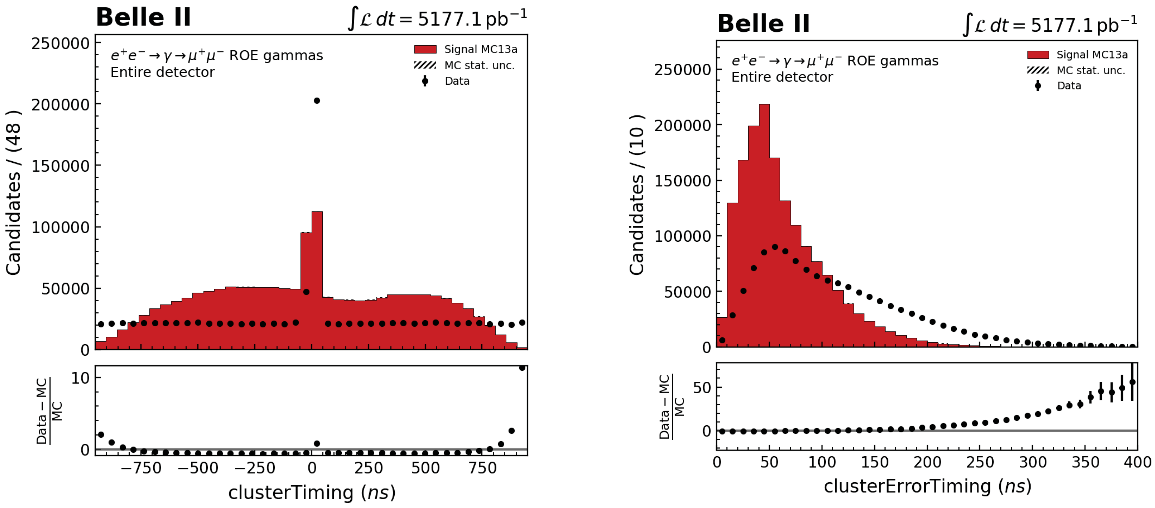
The variable clusterTiming corresponds to the ‘photon timing’ of a cluster relative to the overall event time determined from other detector subsystems. The photon timing is determined as the fitted time for the waveform of the highest energy crystal within a cluster, and the event time is the recorded time of the electron-positron collision at the interaction point. The photon timing relies extensively on calibration and is adjusted for the estimated Time-Of-Flight. Under ideal conditions, the photon timing of a photon produced at the interaction point would equal the event time, resulting in a clusterTiming value of zero. For photons produced outside of the interaction point, like those related to beam backgrounds, the net time described by clusterTiming would generally be non-zero, allowing for straight-forward identification of them. However, the waveform fitting, calibration, and correction processes necessary for this variable to be meaningful were not yet functioning correctly in simulated events at the time this thesis was written [52]. The lack of agreement between experimental and simulated data for clusterTiming (Fig 4.17a) prevents its use in this analysis, though it is expected that this variable will provide a high degree of discriminating power between signal and beam background photons once agreement improves.

clusterErrorTiming

A derivative of clusterTiming is clusterErrorTiming. This variable corresponds to the uncertainty in cluster activation timing that would be necessary for the timing range to contain 99% of signal-related photons for the given collision. For any cluster activation in the ECL, the uncertainty in its activation timing is determined using simulated data and is contingent on the energy of the highest energy crystal in the cluster and the level of beam background present at the time of activation. A distribution of these uncertainties for a given cluster is typically non-gaussian and is processed to determine the uncertainty value that 99% of the signal-related photons fall within. As the evaluation of the timing necessary to determine this variable is dependent on the timing of simulated events that are not yet accurately modelled, this variable, like clusterTiming, was not deemed usable for this study (Fig 4.17b).

clusterPulseShapeDiscriminationMVA

Like clusterZernikeMVA and the variable developed in this study, clusterPulseShapeDiscriminationMVA is the output of a multivariate analysis that is used to classify clusters within the ECL [53]. In this case, the variable was trained on pulse shapes to distinguish between electromagnetic and hadronic showers. It ranges between 0 and 1, indicating the probability that the cluster was activated by an electromagnetic shower. Being able to distinguish between the shower type that activated a crystal cluster within the ECL should assist in discriminating between some types of background and event photons. However, many of the clusters used in the training samples of this study did not have a clusterPulseShapeDiscriminationMVA value associated with them, preventing its inclusion as a training variable.



(a) **(b)**

Figure 4.17: A comparison of (a) clusterTiming and (b) clusterErrorTiming distributions between proc10 data and signal MC13a for ROE photons from reconstructed $e^+e^- \to \gamma \to \mu^+\mu^-$ events. The lack of agreement prevents the use of these variables in the multivariate analysis.

4.5 Training performed

The output of the aforementioned training procedure can be optimized by adjusting the combination of variables used in the training, as well as by using different types of simulated and experimental data as a source for the background and signal ROE photons. The sources and variable combinations chosen are described in this section.

As many of the variables described in section 4.4 exhibit some level of dependence on the energy of the clusters, it is expected that a classifier trained on these variables will be at least somewhat sensitive to cluster energy. It is not entirely clear if the inclusion of the clusterE variable in the training process, in addition to the other variables, will overemphasize the energy of a cluster as a discriminating factor. Additionally, future studies that make use of such a classifier may have their own selections on clusterE that impact the effectiveness of using it in training. For this reason, many of the trainings have been performed twice; once with all seven of the variables described in section 4.4

(clusterE1E9, clusterLAT, clusterSecondMoment, clusterZernikeMVA, clusterTheta, clusterPhi, clusterE), and once with all but clusterE.

The first classifier was trained primarily as an exercise to assess the feasibility of the approach. The training samples were created using both signal and beam background ROE photons from di-muon event reconstructions performed on run-independent signal MC13 simulation. As the final classifier made in this study is intended to eventually be applied to B meson reconstructions that have a lower limit on ROE cluster energy of 50 MeV, this same lower limit was applied to the photons used in these training samples. The performance of the classifier resulting from this training, using all seven cluster variables, is shown in Fig 4.18. The classifier values assigned to the signal clusters used in training are shown in blue, and the values assigned to the beam background-related clusters are shown in orange. The comparable result from applying this classifier to a new group of clusters from the same data set is shown, noted as test values. This output demonstrates a clear separation of signal and beam background-related ROE photons by the trained classifier.

After this training was performed, additional samples were made using the same signal ROE photons, but with the di-muon beam background ROE photons coming instead from events reconstructed with run-dependent signal MC13 simulation. This was done to evaluate the impact of variations to simulated beam background photon characteristics on the performance of the classifier training. The overtraining plot of this classifier, trained using the same variables as the prior classifier, is shown in Fig 4.19. The performance of the classifier trained on run-dependent MC13 is extremely similar to the prior classifier trained on run-independent MC13, suggesting that run-related variations in simulated beam background photons will not significantly impact the performance of the resulting classifier.

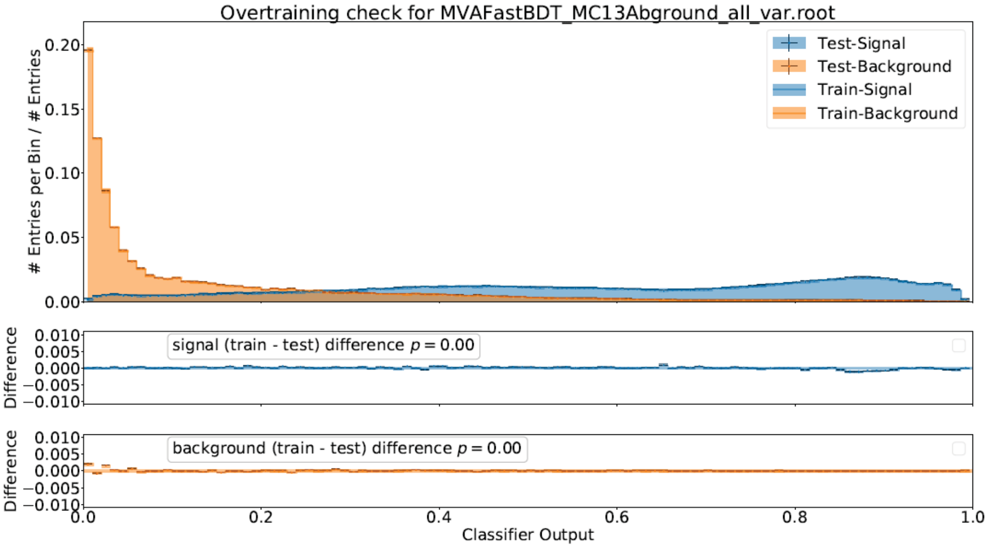


Figure 4.18: An overtraining plot for the first classifier trained using signal and beam background ROE photons from run-independent MC13 $e^+e^- \to \gamma \to \mu^+\mu^-$ events.

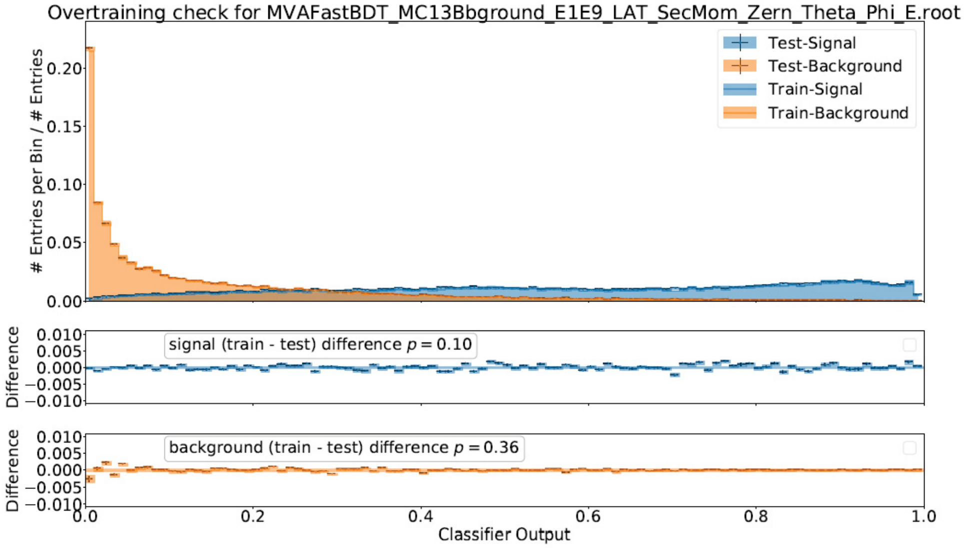


Figure 4.19: An overtraining plot for the classifier trained using beam background ROE photons from run-dependent MC13 $e^+e^- \to \gamma \to \mu^+\mu^-$ events. The distribution is similar to the analogous figure for the classifier made using run-independent MC13.

The next training samples used were again made with signal ROE photons from run-independent MC13 simulation, but with the beam background ROE photons collected from di-muon reconstructions using proc10 detector data rather than simulated data. The classifiers trained on these samples shows comparable performance to the prior classifiers trained using beam background ROE photons from simulated data. As the classifier is intended to eventually be applied as a selection tool to events reconstructed on experimental data, it was determined going forward that using experimental data as the source of beam background ROE photons is preferable. Fig 4.20 shows the overtraining plot when clusterE is not present in the training, and Fig 4.21 shows the overtraining plot for when it is included alongside the other six cluster variables. Separation between signal and beam background photons in the training and testing samples in these figures suggest that distinguishing power is slightly improved when including clusterE in the analysis. This is further evidenced by the receiver operating characteristic (ROC) curve for the same two classifiers, comparing background rejection to signal efficiency (Fig 4.26). For a cut on a given value of the classifier, the number of ROE signal photons in the test samples that pass the selection is larger when clusterE is present in the training. This allows for a higher level of background rejection to be achieved while maintaining an acceptable level of signal efficiency. Similarly, a plot of the efficiency and purity of the test samples indicates that a higher level of purity and efficiency can be achieved for a given classifier value while using the classifier trained on all seven of the cluster variables (Fig 4.27).

To further explore the impact of the source of ROE photons on the performance of the resulting classifier, two other groups of training samples were used. The source of beam background ROE photons for both groups continued to be di-muon events reconstructed from experimental data, while the signal ROE photons were collected from $B^0 \to D^{*+}\ell^-\nu_\ell$ events reconstructed from simulated data. To increase the number of events and ROE photons available for use in the samples, a combined dataset of signal MC13 and generic MC13 was used. The first group of these samples was created with the same selections as the prior samples, with a lower limit on cluster energy of 50 MeV. In order to examine the influence of low energy photons on classifier performance, the second group of samples was created with a selection allowing the inclusion of only clusters with an energy between the minimum recorded ECL energy of 20 MeV and an upper limit of 200 MeV.

The overtraining plots corresponding to the first pair of these samples are shown in Figures 4.22 and 4.23. As the photons used in the training samples were collected from the ROE of reconstructed semileptonic B meson decays that are relatively rare, the samples contain fewer photons than comparable samples made with photons from the ROE of di-muon events. As a result of these smaller sample sizes, the output for the training and testing distributions vary slightly from each other. Despite the low statistics, the distributions still show reasonable separation between background and signal photons, with the training samples containing clusterE providing a more effective classifier. The ROC curves (Fig 4.28) and efficiency-purity plots (Fig 4.29) for the classifier trained on these samples indicate this as well. Although the performance is slightly weaker when the classifier is not trained directly on cluster energy, it is still able to achieve over 75% background rejection while maintaining 75% signal efficiency. This is a very similar level of performance to the classifier trained on signal ROE photons from di-muon events.

When the second group of samples, made of signal and beam background ROE photons with an energy range from 20 MeV to 200 MeV, are used in training, the overtraining plots for the resulting classifier show a strong divergence in classifier values assigned to background and signal photons (Figs 4.24, 4.25). The divergence is extremely pronounced for the classifier trained with clusterE, suggesting cluster energy is a dominating factor when discriminating between low energy beam background and signal photons. The ROC curves show that this classifier can achieve 80% background rejection while maintaining as high as 90% signal efficiency (Fig 4.30). At this 90% efficiency level, a purity over 40% can be achieved (Fig 4.31), a value much higher than what is expected from the prior trained classifiers. These results indicate that the boosted decision tree training process provides strong discriminating power when performed with photons below 50 MeV included in the training samples, especially when clusterE is used as one of the training variables.

These two groups of training samples made with signal ROE photons from reconstructed B meson events, and the aforementioned two training samples made using signal and beam background ROE photons from reconstructed di-muon events, comprise the group of samples used to train six classifiers that are then applied to the ROE of reconstructed events. A summary of these final six sample mixes can be found in Table 4.1, and the application process is described in the following section.

Table 4.1: A summary of the beam background and signal ROE photon sources, variables, and cluster energy limits used in the creation of training samples. The six training sample mixes described below were used to train the final classifiers created in this analysis.

Sample number	Background photon source (data)	Signal photon source (MC13)	clusterE used	Training photon energy range (MeV)
1	Di-muon events	Di-muon events	No	$50 <$ clusterE
2	Di-muon events	Di-muon events	Yes	$50 <$ clusterE
3	Di-muon events	Semileptonic B meson decays	No	$50 <$ clusterE
4	Di-muon events	Semileptonic B meson decays	Yes	$50 <$ clusterE
5	Di-muon events	Semileptonic B meson decays	No	$20 <$ clusterE < 200
6	Di-muon events	Semileptonic B meson decays	Yes	$20 <$ clusterE < 200

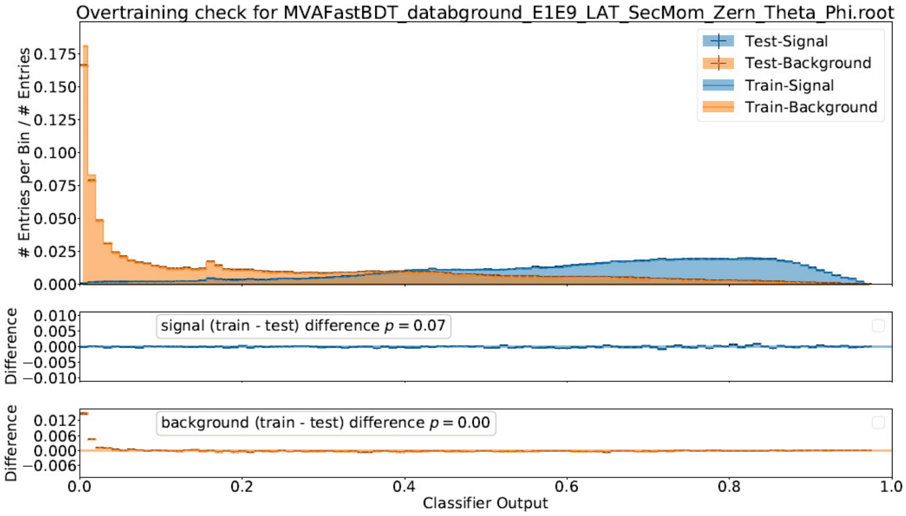


Figure 4.20: The overtraining plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $e^+e^- \to \gamma \to \mu^+\mu^-$ events, corresponding to sample 1.

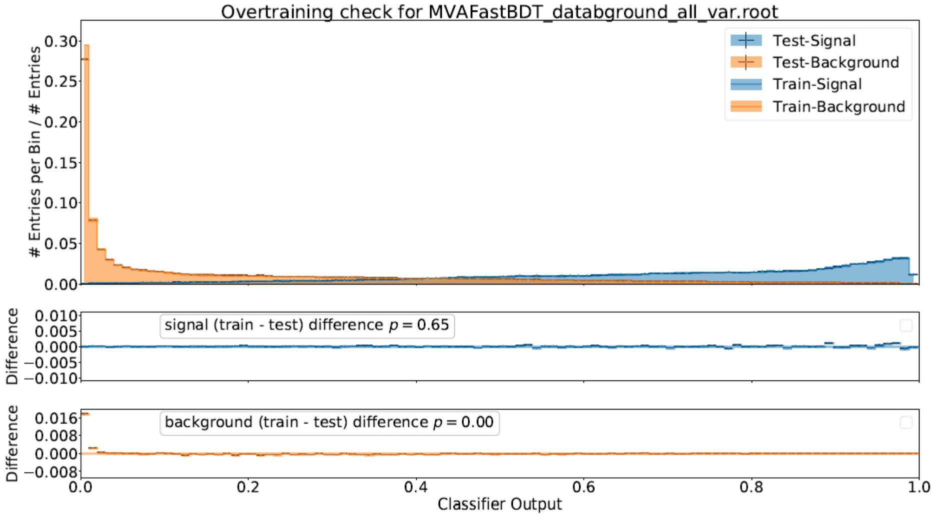


Figure 4.21: The overtraining plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $e^+e^- \to \gamma \to \mu^+\mu^-$ events, corresponding to sample 2.

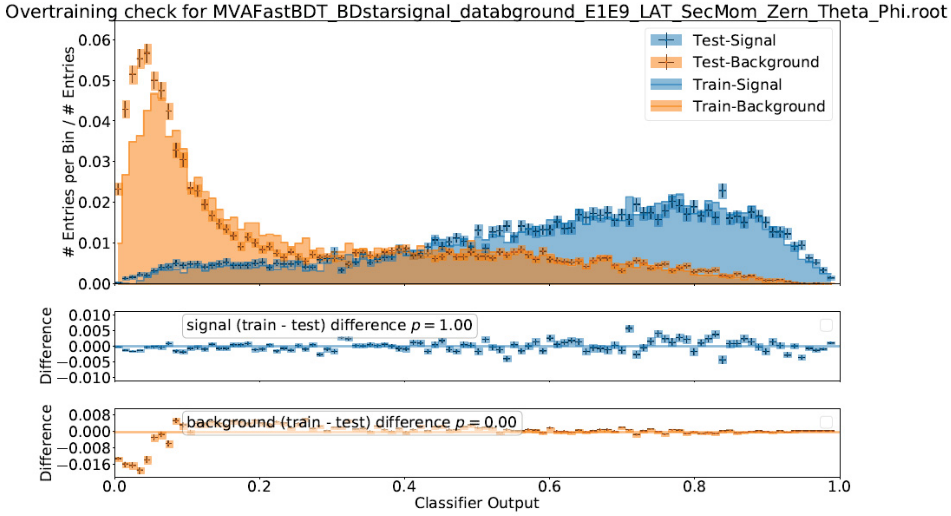


Figure 4.22: The overtraining plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events, corresponding to sample 3.

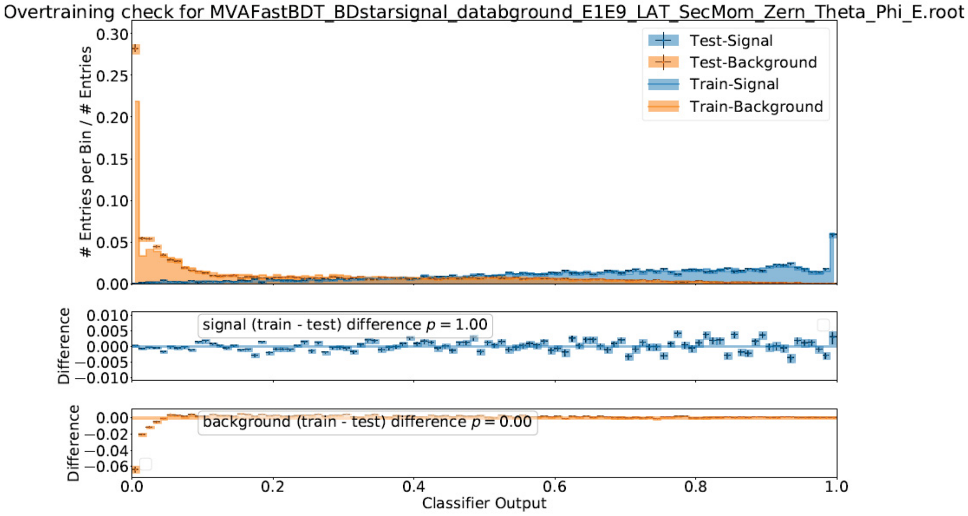


Figure 4.23: The overtraining plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events, corresponding to sample 4.

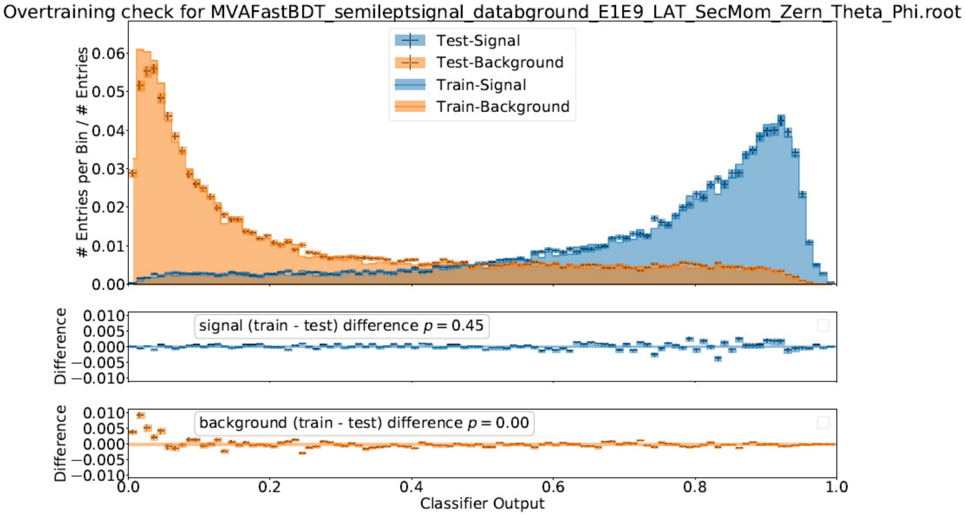


Figure 4.24: An overtraining plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events, within the photon energy range 20 - 200 MeV, corresponding to sample 5.

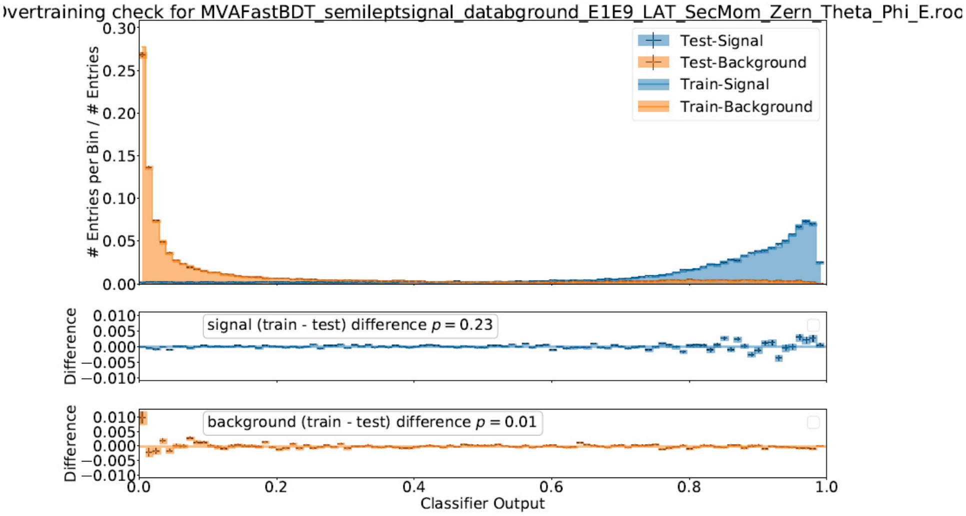
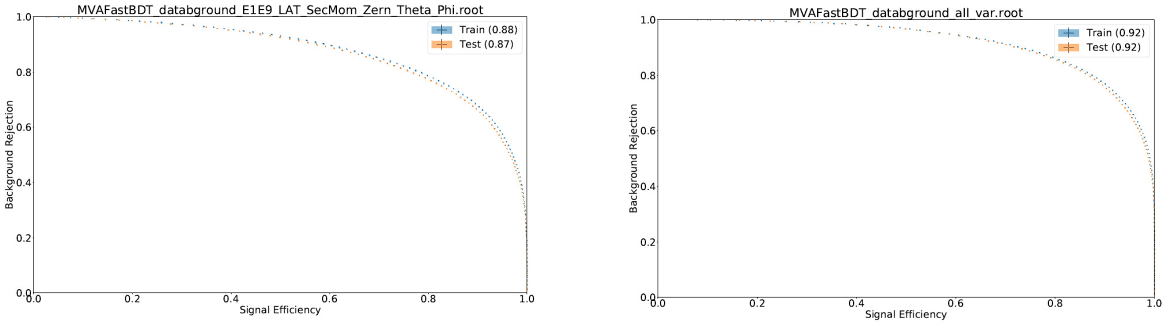
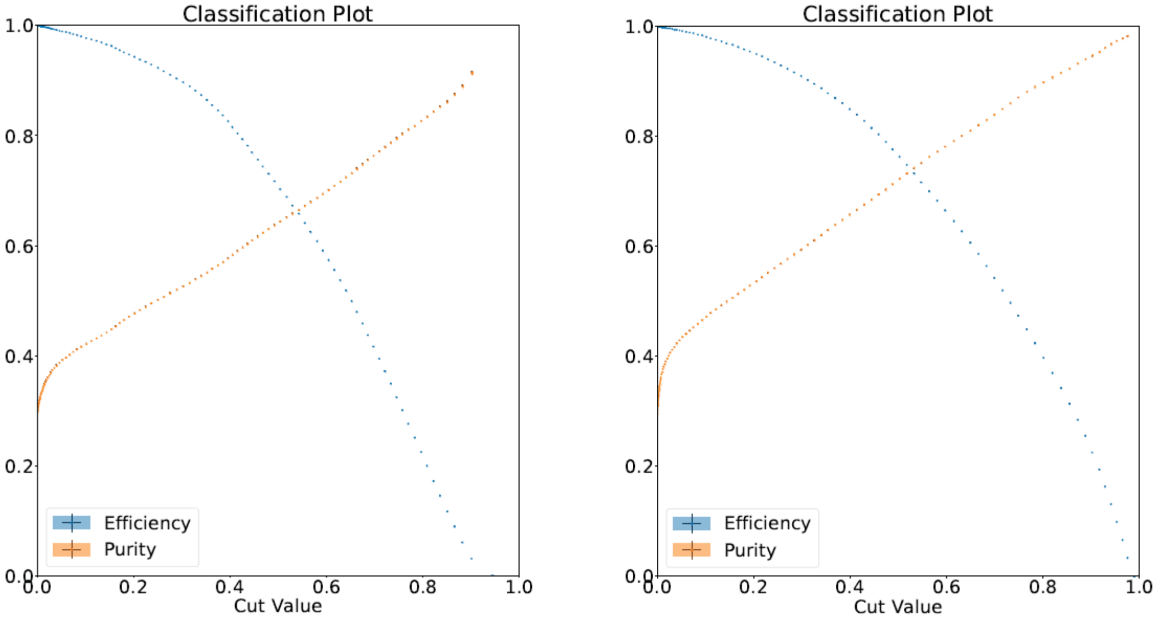


Figure 4.25: An overtraining plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events, within the photon energy range 20 - 200 MeV, corresponding to sample 6.



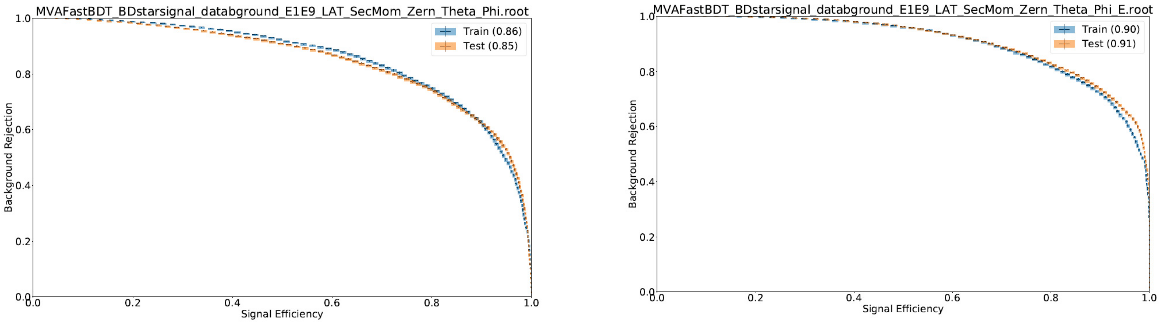
(a) **(b)**

Figure 4.26: The background rejection to signal efficiency curve for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $e^+e^- \to \gamma \to \mu^+\mu^-$ events. The trainings were performed (a) without clusterE (sample 1) and (b) with clusterE (sample 2).



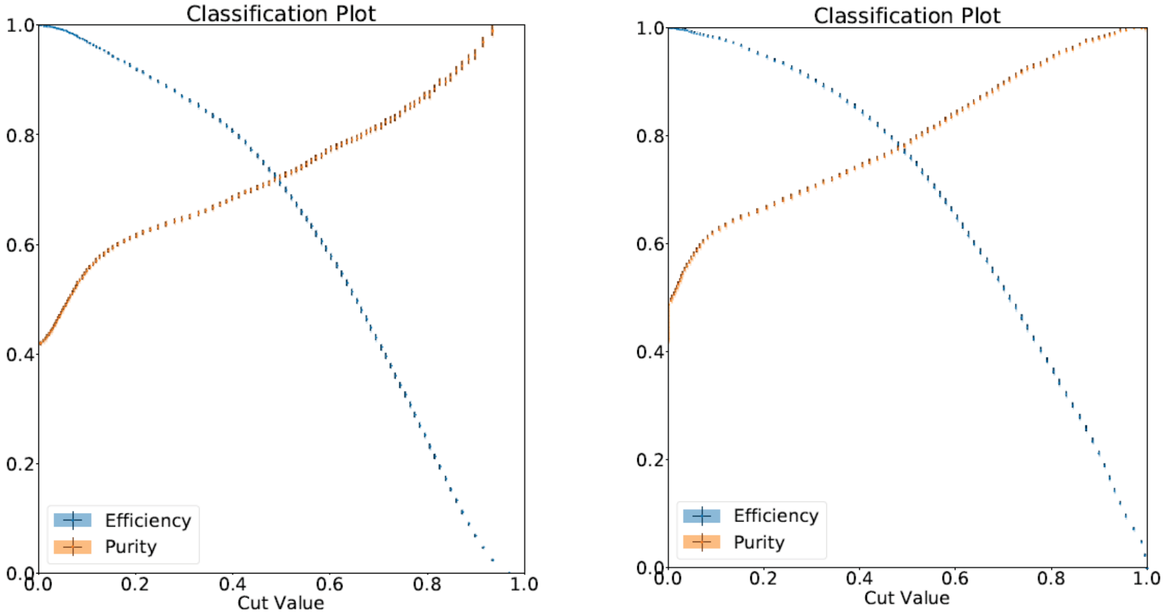
(a) **(b)**

Figure 4.27: The efficiency to purity plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $e^+e^- \to \gamma \to \mu^+\mu^-$ events. The trainings were performed (a) without clusterE (sample 1) and (b) with clusterE (sample 2).



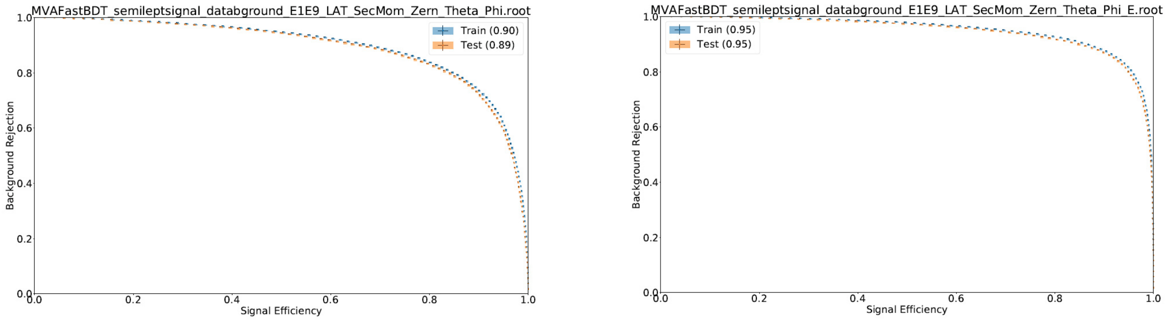
(a) **(b)**

Figure 4.28: The background rejection to signal efficiency curve for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The trainings were performed (a) without clusterE (sample 3) and (b) with clusterE (sample 4).



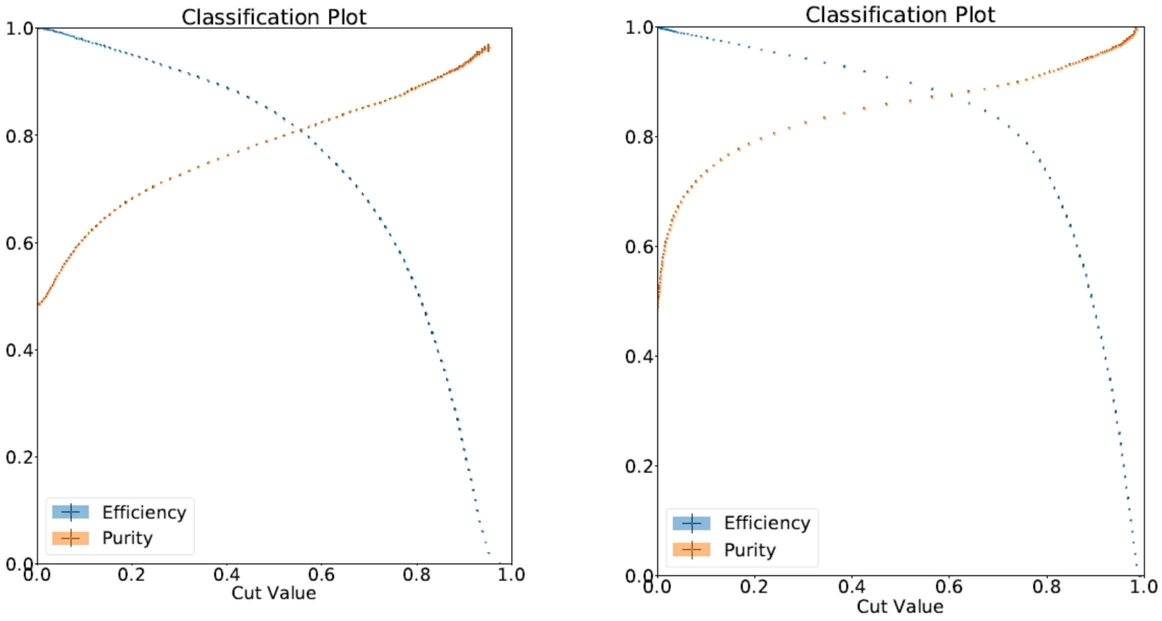
(a) **(b)**

Figure 4.29: The efficiency to purity plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The trainings were performed (a) without clusterE (sample 3) and (b) with clusterE (sample 4).



(a) **(b)**

Figure 4.30: The background rejection to signal efficiency curve for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events, within the photon energy range 20 - 200 MeV. The trainings were performed (a) without clusterE (sample 5) and (b) with clusterE (sample 6).



(a) **(b)**

Figure 4.31: The efficiency to purity plot for the classifier trained using beam background ROE photons from proc10 data and signal ROE photons from run-independent MC13 $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events, within the photon energy range 20 - 200 MeV. The trainings were performed (a) without clusterE (sample 5) and (b) with clusterE (sample 6).

4.6 Application of BDTgamma classifier

The efficacy of the aforementioned training process and resulting classifier objects can be further evaluated by applying the classifiers to clusters in the ROE of other reconstructed events. When the classifier object is used in a reconstruction, each cluster in the ROE is given a value between 0 and 1 that indicates the probability that the cluster was activated by a signal photon, referred to as its 'BDTgamma' value. Examining the resulting BDTgamma distribution for a group of reconstructed events allows for a choice of selection on the variable, removing any clusters from the ROE that do not meet the desired probability. The sum of the energies of all of the clusters associated with a neutral particle in a reconstructed event's ROE is referred to as its neutral extra energy (ROE_neextra). A comparison of this extra energy distribution for a group of reconstructed events before and after such a selection is applied can provide insight into the level of background photon suppression that has been achieved.

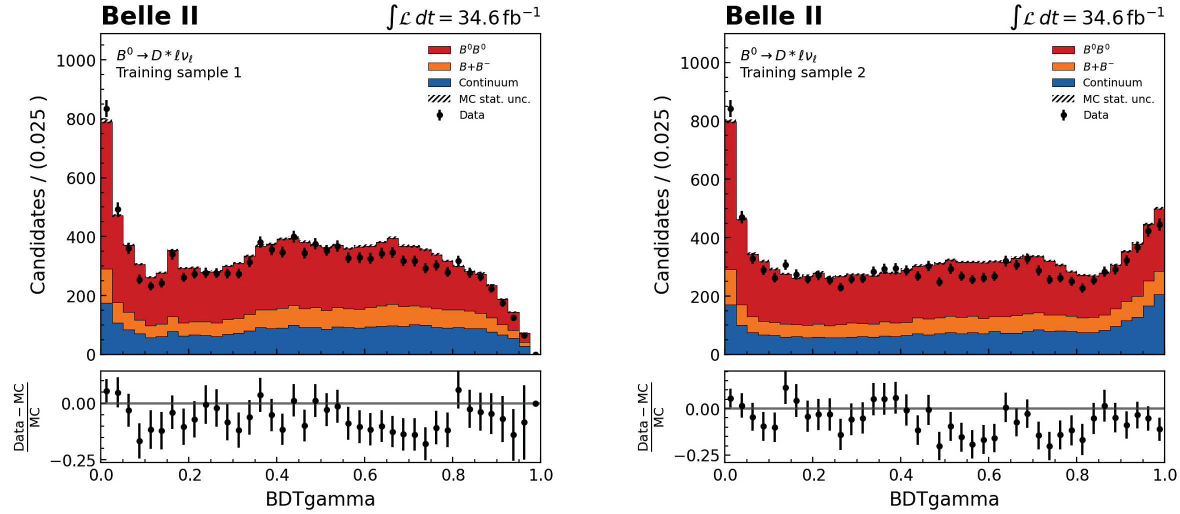
To evaluate the performance of the classifiers trained using the final six training samples described in the previous section, they were applied to $B^0 \to D^{*+}\ell^-\nu_\ell$ events reconstructed using both generic MC13 and proc10 data. The reconstruction process follows the same procedure described in section 4.1.2, with the ROE clusters then each assigned a BDTgamma value using the criteria defined by each classifier object. A selection of BDTgamma > 0.3 is applied for all six reconstructions, with any clusters not meeting this criterion removed from the ROE. A comparison of the neutral extra energy distributions of these events before and after this selection is applied, as well as the BDTgamma distributions themselves, are detailed in the next chapter.

Chapter 5

Results

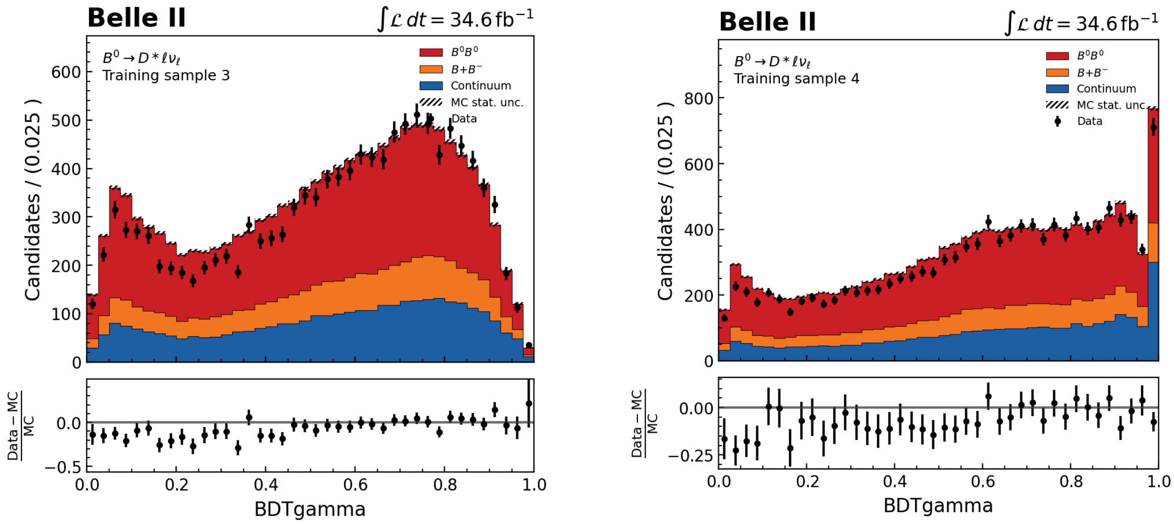
5.1 BDTgamma distributions

The application of the six classifiers to $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events reconstructed using both generic MC13 and proc10 data provides the following BDTgamma distributions:



(a) **(b)**

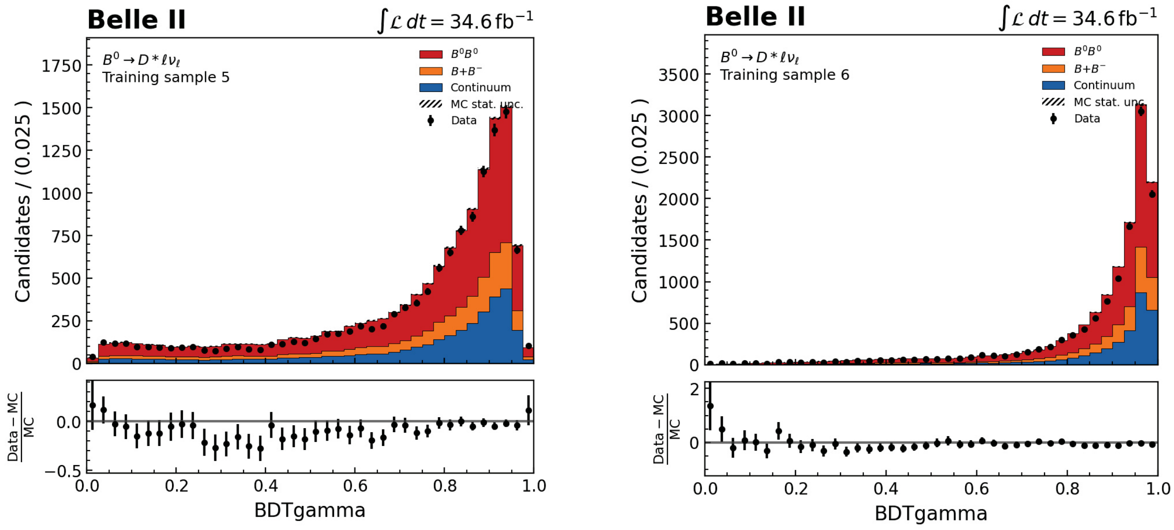
Figure 5.1: The BDTgamma distribution that results when applying the classifier trained using the background and signal ROE photon samples described in Table 4.1 as (a) sample 1 and (b) sample 2, to the ROE clusters of reconstructed $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events.



(a)

(b)

Figure 5.2: The BDTgamma distribution that results when applying the classifier trained using the background and signal ROE photon samples described in Table 4.1 as (a) sample 3 and (b) sample 4, to the ROE clusters of reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events.



(a)

(b)

Figure 5.3: The BDTgamma distribution that results when applying the classifier trained using the background and signal ROE photon samples described in Table 4.1 as (a) sample 5 and (b) sample 6, to the ROE clusters of reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events.

The BDTgamma distribution for the two classifiers trained on signal ROE photons over 50 MeV from di-muon events, referred to as training samples 1 and 2, are shown in Fig 5.1. A clear peak exists near zero, representing clusters identified as highly unlikely to have been activated by signal photons. Additionally, a more prominent bias towards 1 exists for some clusters evaluated with the classifier trained on sample 2, due to the inclusion of clusterE in the training. These features align with the expectations provided by the training output described in the previous chapter, and provide evidence for the possibility that including clusterE in the training can improve separating power.

The BDTgamma distribution for the two classifiers trained on samples 3 and 4, containing signal ROE photons from B meson decays in the same energy range as training samples 1 and 2, exhibit similar features (Fig 5.2). The distributions are faintly bimodal, with a peak near zero and a peak near 1, as predicted by the overtraining plots for these classifiers. Again, the classifier trained on clusterE, in addition to the other variables, shows a more distinct peak at 1. The extent of this difference suggests that a direct examination of a cluster's energy provides far stronger distinguishing power between beam background-related and signal photons than the other six variables.

The BDTgamma distribution for the two classifiers trained on samples 5 and 6, containing signal ROE photons from B meson decays ranging from 20 MeV to 200 MeV, provides useful insight into classifier training tendencies (Fig 5.3). In contrast to the BDTgamma distributions associated with the prior classifiers trained on photons above 50 MeV, most clusters in the ROE of these reconstructed events are deemed by the classifier to have a high probability of being signal event-related. This suggests that the presence of many photons below 50 MeV in the training process severely impacts the performance of the classifier when it is applied to clusters above 50 MeV. The clusterE distributions of the signal and beam background ROE photons used in training samples 5 and 6 are shown in Fig 5.4. The majority of background clusters in this distribution have an energy below 50 MeV, with a strong bias towards the 20 MeV minimum value. It is clear that the classifier training process is identifying this extreme bias and weighting it in a way that dominates the influence of other cluster variables when evaluating further clusters. This suggests that an approach using this type of boosted decision tree may be limited to using training samples containing ROE photons strictly in the same energy range as the ROE photons the eventual classifier is expected to be applied to.

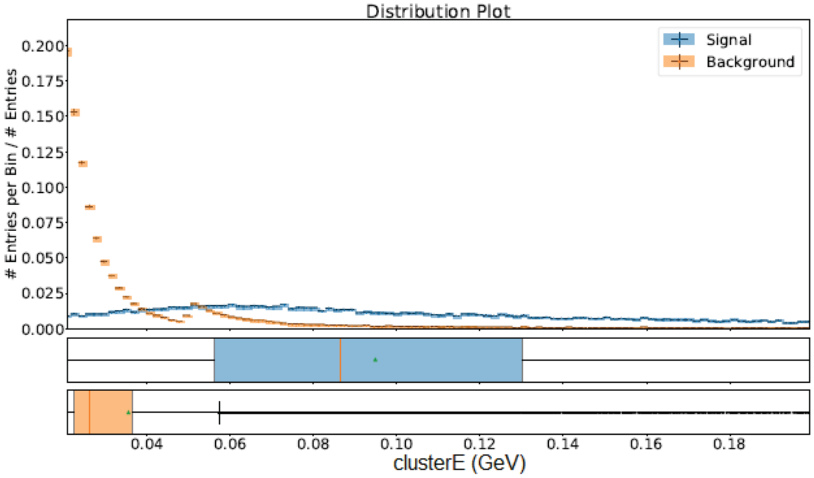


Figure 5.4: The clusterE distribution of signal and beam background ROE photons used in training samples 5 and 6. The secondary peak in the background distribution at 50 MeV is due to an innate detector clusterTiming selection that is only applied to activated clusters within the ECL below 50 MeV.

Although minor disagreements between data and MC in these six BDTgamma distributions are present, very few significant deviations exist. This suggests that the influence of disagreements between data and MC in the distributions of cluster variables used in training was averaged out or mitigated through the training process. It is possible that future improvements to the agreement of simulated cluster variables with data, specifically in the end-cap regions of the ECL, may even further improve the extent of agreement observed when applying the classifier to reconstructed events.

5.2 Neutral ROE extra energy

The neutral ROE extra energy distribution for the reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events, before any selections on BDTgamma are applied, is shown in Fig 5.5. A distinct non-zero peak exists in the distribution, suggesting the presence of beam background-related photons in the ROE. The mean of this distribution is 1.356±0.005 GeV for MC13 and 1.486±0.025 GeV for data.

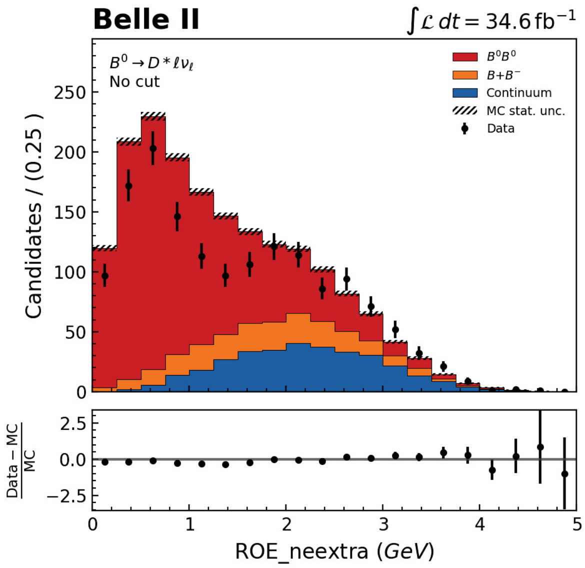
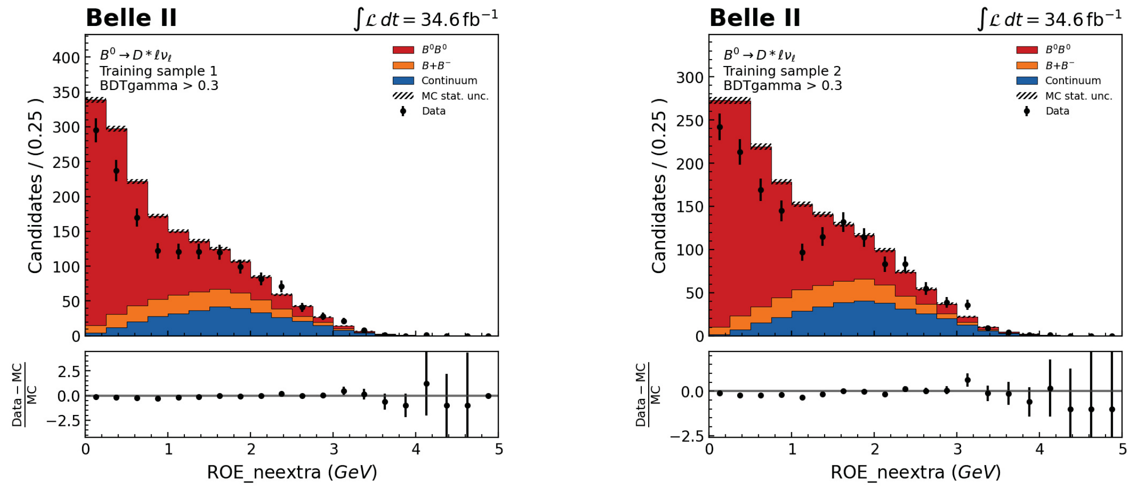


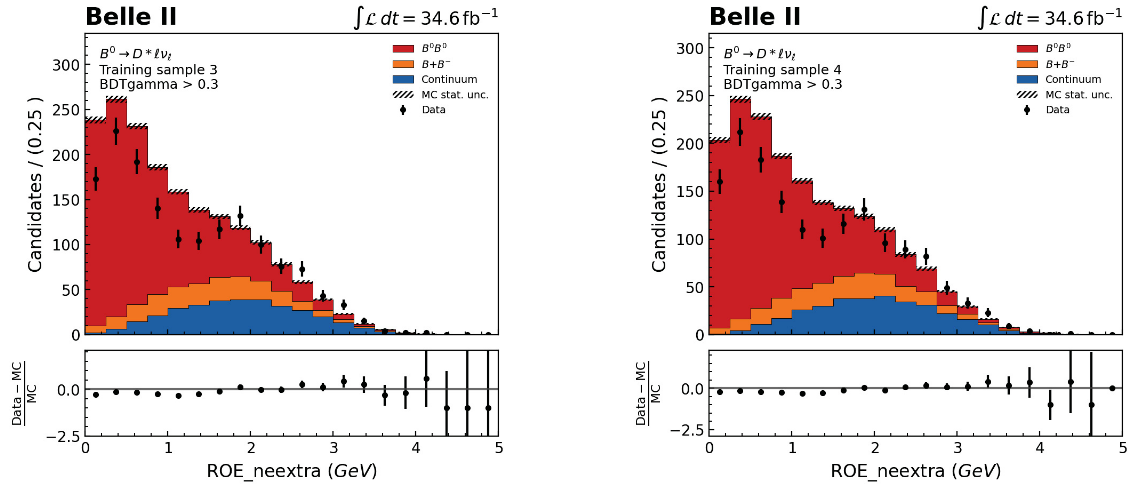
Figure 5.5: The neutral ROE extra energy distribution for reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events before any selection on a BDTgamma classifier is applied.

Separately applying the selection BDTgamma > 0.3 on the clusters in the ROE using each of the six aforementioned BDTgamma distributions results in six neutral ROE extra energy distributions that have been optimized by the suppression of beam-related background photons, to varying extents. The mean ROE neutral extra energy values exhibited by these distributions, for both MC13 and data, are summarized in Table 5.1. The distributions themselves are as follows:



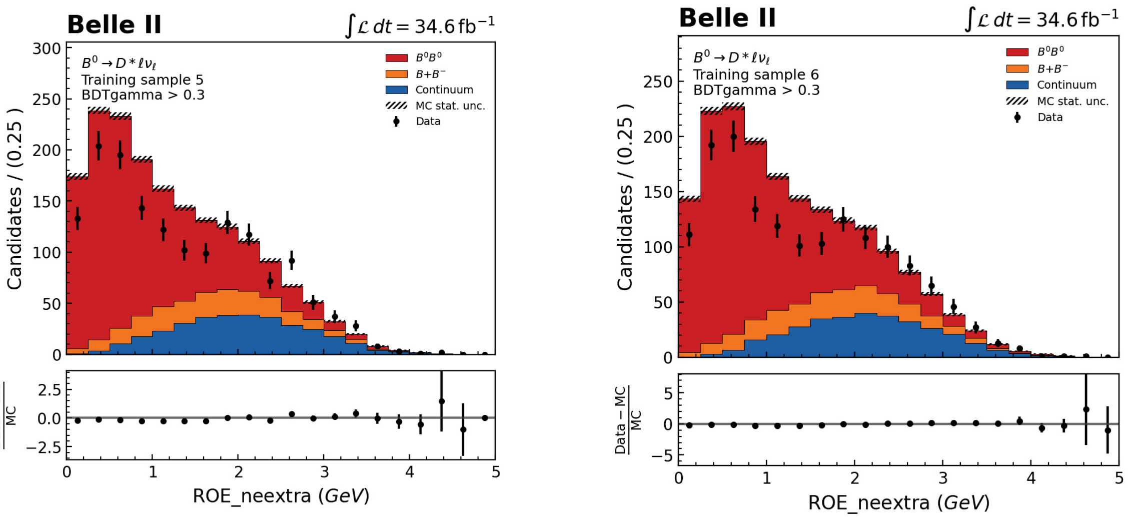
(a) **(b)**

Figure 5.6: The neutral ROE extra energy distribution for reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events with the selection BDTgamma $>$ 0.3 applied using the classifier trained using (a) sample 1 and (b) sample 2.



(a) **(b)**

Figure 5.7: The neutral ROE extra energy distribution for reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events with the selection BDTgamma $>$ 0.3 applied using the classifier trained using (a) sample 3 and (b) sample 4.



(a) **(b)**

Figure 5.8: The neutral ROE extra energy distribution for reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ events with the selection BDTgamma $>$ 0.3 applied using the classifier trained using (a) sample 5 and (b) sample 6.

A comparison of Fig 5.6 to the original ROE_neextra distribution in Fig 5.5 shows a clear change in the distribution shape. The non-zero peak is shifted towards the origin, suggesting a large number of beam background-related photons have been excluded by the chosen selection on BDTgamma. Explicitly, the mean of the sample 1 MC13 distribution is 0.977±0.005 GeV and the mean in data is 1.067±0.021 GeV, representing a reduction of 27.95±0.53% and 28.20±2.25%, respectively. Similarly, sample 2 shows a reduction in mean of 19.84±0.53% and 20.32±2.31% in MC13 and data. This reconciles with expectations, as the sample 1 and 2 BDTgamma distributions shown in Fig 5.1 indicate a large number of clusters with a nearly 0% probability of being signal event-related which have been removed by the selection.

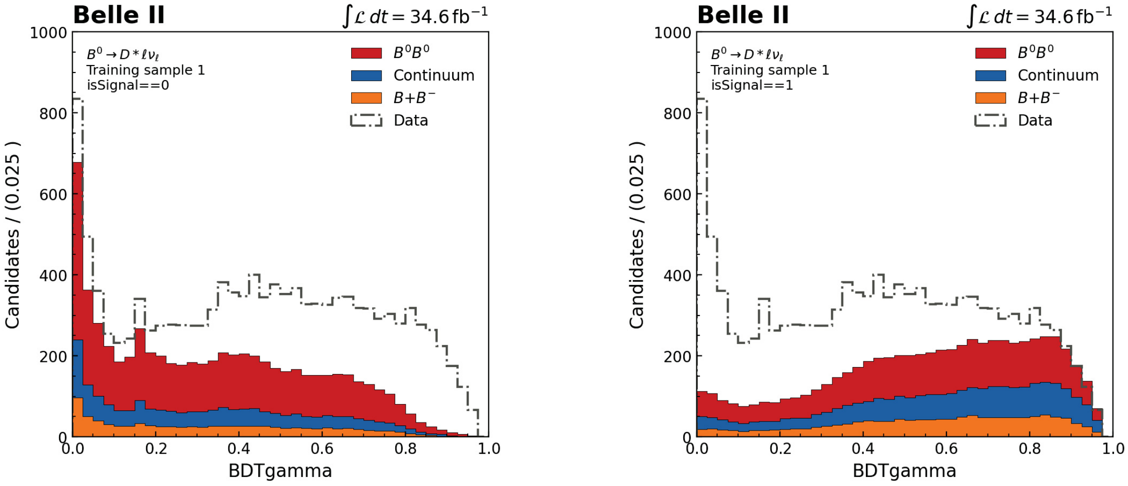
A comparison of Fig 5.7 to the original ROE_neextra distribution shows a similar peak shift, though the change is less pronounced. The sample 3 distribution shows a reduction in mean of 17.04±0.53% in MC13 and 15.34±2.30% in data, while sample 4 shows a reduction of 11.87±0.52% in MC13 and 11.10±2.29% in data. Again, the BDTgamma distributions used for this selection show a number of clusters that fall below a value of 0.3 that are highly likely to be beam background-related. It appears that the presence of clusterE in the classifier training for samples 2 and 4 reduces the number of clusters below this threshold in the BDTgamma distribution by a non-negligible amount, reducing the impact of the selection on the ROE_neextra distribution. Despite this, these first four distributions clearly demonstrate a successful application of the trained classifier objects. Adjustments to the chosen BDTgamma selection and refinements to the training process and variables used should serve only to further improve the optimization of the ROE_neextra distribution that results.

Table 5.1: The changes in ROE neutral extra energy means when the six different classifiers are applied to the ROE of reconstructed $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ events and a selection of BDTgamma $>$ 0.3 is made. The uncertainties presented are statistical uncertainties. Smaller uncertainties are attributed to MC13 values because the distributions originally contained far more events before being scaled down to match the luminosity of the data.

ROE Extra Energy mean (GeV)				
Sample number	MC13	Change	Proc10 Data	Change
No cut	1.356±0.005	N/A	1.486±0.025	N/A
1	0.977±0.005	27.95±0.53%	1.067±0.021	28.20±2.25%
2	1.087±0.005	19.84±0.53%	1.184±0.023	20.32±2.31%
3	1.125±0.005	17.04±0.53%	1.258±0.023	15.34±2.30%
4	1.195±0.005	11.87±0.52%	1.321±0.023	11.10±2.29%
5	1.237±0.005	8.78±0.52%	1.358±0.024	8.61±2.34%
6	1.303±0.005	3.91±0.52%	1.428±0.024	3.90±2.33%

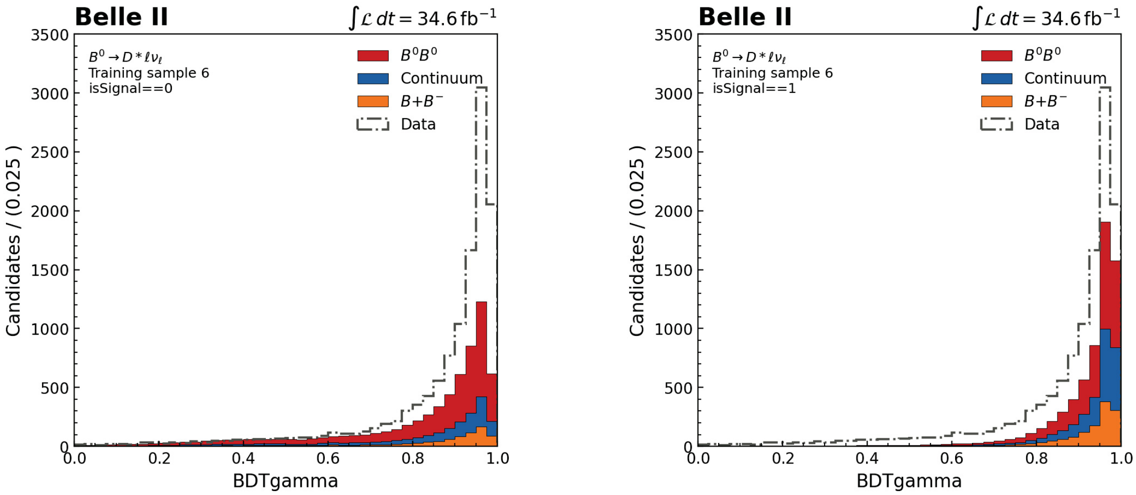
The selection of BDTgamma $>$ 0.3 was used for all six samples in order to remain consistent, however it will clearly not remove many clusters from samples 5 and 6. Consequently, the resulting ROE_neextra distribution, shown in Fig 5.8, varies little from the original distribution. The mean of the sample 5 distribution is decreased by 8.78±0.52% in MC13 and 8.61±2.34% in data, while the mean of the sample 6 distribution is decreased by 3.91±0.52% in MC13 and 3.90±2.33% in data. Although a cut on a higher BDTgamma value would remove a larger number of clusters and reduce the average value further, the distribution is so heavily skewed towards large values it is not clear that any impactful cut would remove more beam background-related photons than signal-related photons.

It may be possible to determine a more optimal BDTgamma selection by considering the location of background and signal photons within the BDTgamma distributions. This can be accomplished by re-examining the BDTgamma distributions for the six samples after they have been separated into one distribution with isSignal==0, and one with isSignal==1. As ROE photons from events reconstructed from experimental data cannot have a non-zero isSignal value, this approach is only possible for examining the distributions for simulated events. For the figures shown, the data distribution remains plotted to provide a visual approximation of the portion of clusters that did not pass the applied isSignal selection.



(a) **(b)**

Figure 5.9: The BDTgamma distributions that result from applying the classifier trained on sample 1 to the ROE clusters of reconstructed MC13 $B^0 \rightarrow D^{*+} \ell^- \nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.

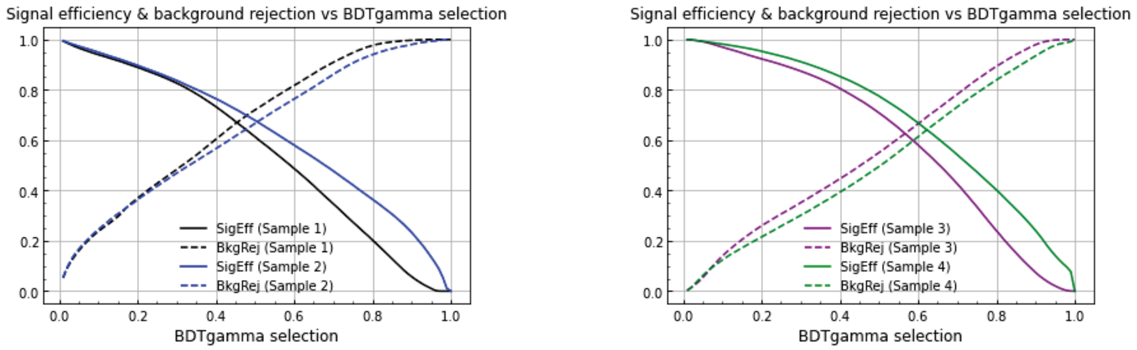


(a) **(b)**

Figure 5.10: The BDTgamma distributions that result from applying the classifier trained on sample 6 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^- \nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.

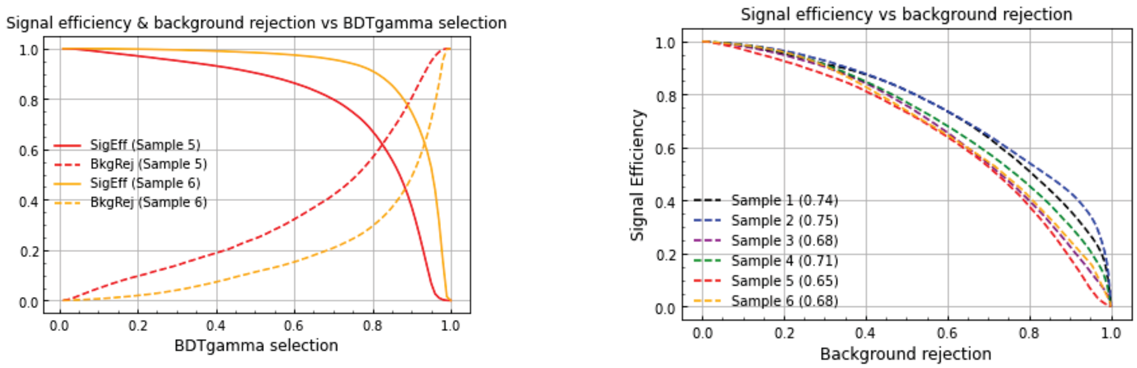
As shown by Fig 5.9, a BDTgamma selection of 0.3 using the classifier trained on sample 1 will remove a much larger portion of the isSignal==0 ROE photons than isSignal==1 ROE photons. This allows the classifier to function appropriately as a beam background-related photon suppressant while not significantly diminishing signal efficiency. Conversely, the same distributions for sample 6, shown in Fig 5.10, display a drastically different organization of ROE photons. Very few ROE clusters were assigned a BDTgamma value less than 0.8 by the classifier, and the distinction between isSignal==0 and isSignal==1 clusters is not as clear. This makes it difficult to determine a BDTgamma selection that will allow the classifier to function effectively as a background photon suppressant. BDTgamma distributions with both isSignal selections for the classifiers trained on all six samples are available in Appendix A.

The extent of separation between beam background and signal ROE photons in a BDTgamma distribution can be further evaluated by a direct examination of the signal efficiency and background rejection achieved as a function of BDTgamma selection (Figs 5.11, 5.12a). Although it is clear that the amount of signal retention that can be achieved for a given level of background rejection varies slightly depending on the training samples used to create the classifier, the classifiers all exhibit a relatively similar level of performance. Samples 1 and 2 perform slightly better than the others, and samples 5 and 6 slightly worse. Fig 5.12b shows signal efficiency versus background rejection for all six samples, with the area under each curve providing a numerical indication of the efficacy of the classifier in separating beam background and signal ROE photons. In all cases, including clusterE slightly improves the performance of the classifier.



(a) **(b)**

Figure 5.11: Signal efficiency and background rejection as a function of BDTgamma selection for the classifiers trained on (a) samples 1 & 2 and (b) samples 3 & 4.



(a) **(b)**

Figure 5.12: (a) Signal efficiency and background rejection as a function of BDTgamma selection for the classifiers trained on samples 5 & 6 and (b) The ROC curve for all six classifiers. The number corresponding to each sample is the area under the curve.

In all of the ROE_neextra distributions observed, the agreement between data and MC is imperfect. A distinct second peak in the distributions exist in data at the location where clusters associated with continuum events are at a maximum, and this peak is less pronounced in simulated events due to an excess of lower ROE_neextra values. Fortunately, the selection on BDTgamma appears to diminish this difference. Although the reason for this is not explicitly clear, it is plausible that the run-independent MC13 used to simulate these reconstructed events does not correctly model the amount of neutral extra energy in the detector relating to beam backgrounds, and the classifier is attributing a low BDTgamma value to a disproportionately large number of these inaccurately modelled clusters.

Overall, these results suggest that using a boosted decision tree machine learning program to train a classifier for beam background photon suppression in events reconstructed from Belle II data is a viable approach. At this stage, the success of the classifier appears dependent on an alignment in cluster energy between the training samples and applied samples. However, the inclusion of variables like clusterTiming in the training process, when available, may significantly mitigate this dependence and allow the classifier to perform at an acceptable level when trained on a less strict range of cluster energies. It may also provide the classifier with even better distinguishing power and allow for a classifier variable selection that achieves high levels of background rejection without significantly compromising signal efficiency.

Chapter 6

Conclusion

Presented here is the formulation of a suppression tool capable of discriminating between beam background and signal photons detected in the ECL of the Belle II detector. Initial evaluations of the performance of the tool provide evidence for the feasibility of this approach. The classifier object trained on signal ROE photons above 50 MeV from the ROE of reconstructed di-muon events provides clear distinguishing power when applied to clusters in the ROE of reconstructed $B^0 \to D^{*+}\ell^-\nu_\ell$ decays. A selection requiring the classifier-determined probability of a cluster being related to a signal photon to be greater than 30% provides a reduction in the mean of the ROE_neextra distribution for simulated events of 19.84±0.53% when clusterE is used in training, and 27.95±0.53% when it is not. For data, the reduction is 20.32±2.31% with clusterE in the training and 28.20±2.25% without it. The classifier trained on signal ROE photons over 50 MeV from the ROE of the same reconstructed B meson decays provides similar results, with a reduction to the ROE_neextra distribution mean for simulated events of 11.87±0.52% when clusterE is used in training, and 17.04±0.53% when it is not. The reduction in data is 11.10±2.29% when clusterE is used in training, and 15.34±2.30% when it is not. These changes to the ROE_neextra distribution are achieved without large losses to signal efficiency.

The performance of the fifth and sixth classifiers using the same 30% classifier probability selection, trained on photons between 20 MeV and 200 MeV from reconstructed B

meson decays and di-muon events, is worse. The change in the ROE_neextra distribution mean for simulated events is only 3.91±0.52% when clusterE is used in the training, and 8.78±0.52% when it is not. The analogous changes in mean for data are 3.90±2.33% and 8.61±2.34%. This decrease in performance relative to the other classifiers is attributable to the presence of a large number of clusters in the background training samples with an energy near 20 MeV. The training was found to be significantly influenced by these clusters and this causes the resulting classifier to deem that most clusters with an energy above 50 MeV are highly likely to be related to signal photons, and thus few are removed by a 30% probability selection. Although a higher probability selection will counter this effect, these two classifiers are shown to have a lower level of background rejection for a given amount of signal retention, compared to the prior four classifiers. This suggests that any study using this type of boosted decision tree approach may not be as effective when including very low energy ROE photons, and may be limited to using ROE photons in the training samples that closely match the energy of the ROE photons the eventual classifier is to be applied to.

Although the results presented in this thesis largely demonstrate the viability of the methods used, this work represents only a first step. In addition to the need for further consideration regarding the energy of photons used in the training samples, additional variables may be useful in the training process. Cluster timing and pulse shape variables that are expected to be useful discriminating tools for training purposes were not available at the time of this study. With these variables included in future iterations of this analysis, along with the continually improving agreement between experimental and simulated Belle II data, the efficacy of the resulting classifier tool and the level of extra energy optimization that is achievable is likely to increase. With these improvements made, the use of such a tool may substantially improve the ability of studies using Belle II data to suppress beam background-related photons from extra energy distributions and more accurately identify missing energy.

Bibliography

[1] Bryant, P. J., “A brief history and review of accelerators.”, (1994).

[2] Bertone, G., Hooper, D., & Silk, J., “Particle dark matter: evidence, candidates and constraints.”, Physics reports **405**, 279–390 (2005).

[3] Aoyama, T., *et al*, “Tenth-order QED contribution to the electron g-2 and an improved value of the fine structure constant.”, Physical Review Letters **109**, 111807 (2012).

[4] Grupen, C., Shwartz, B. A., *Particle detectors*, 2nd ed., 19-20 (Cambridge university press, 2008).

[5] Hasert, F. J., *et al*, “Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment.”, Nuclear Physics B **73**, 1–22 (1974).

[6] Dobrescu, B. A., “Massless gauge bosons other than the photon.”, Physical review letters **94**, 151802 (2005).

[7] Higgs, P. W., “Broken symmetries and the masses of gauge bosons.”, Physical Review Letters **13**, 508 (1964).

[8] Aad, G., *et al*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.”, Physics Letters B **716**, 1–29 (2012).

[9] Gaillard, M.K., Grannis, P.D., Sciulli, F.J., “Particle dark matter: evidence, candidates and constraints.”, Reviews of Modern Physics **71**, S96 (1999).

[10] Lokal_Profil, *Quark confinement*. `commons.wikimedia.org/w/index.php?curid%20=8235410`, 2009, (accessed: 06.10.2020).

[11] Zyla, P.A., *et al*, “Light unflavored mesons”, Particle Data Group **Prog. Theor. Exp. Phys. 2020**, 083C01 (2020).

[12] Kajita, T., “Atmospheric neutrino results from Super-Kamiokande and Kamiokande—evidence for $\nu\mu$ oscillations.”, Nuclear Physics B-Proceedings Supplements **77**, 123–132 (1999).

[13] Ahmad, Q. R., *et al*, "Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory.", Physical review letters **89**, 011301 (2002).

[14] Rodejohann, W., "Neutrino-less double beta decay and particle physics.", International Journal of Modern Physics E **20**, 1833–1930 (2011).

[15] Zyla, P.A., *et al*, "CKM Quark-Mixing Matrix.", Particle Data Group **Prog. Theor. Exp. Phys. 2020**, 083C01 (2020).

[16] Battaglia, M., *et al*, "The CKM matrix and the unitarity triangle.", arXiv preprint hep-ph/0304132 (2003).

[17] Cunliffe, S., "Prospects for rare B decays at Belle II.", arXiv preprint arXiv:1708.09423 (2017).

[18] Liu, Y. R., *et al*, "Pentaquark and Tetraquark states.", Progress in Particle and Nuclear Physics **107**, 237–320 (2019).

[19] Chen, R., *et al*, "Strong LHCb evidence supporting the existence of the hidden-charm molecular pentaquarks.", Physical Review D **100**, 011502 (2019).

[20] Opal Collaboration, "A study of B meson oscillations using hadronic Z 0 decays containing leptons.", Zeitschrift für Physik C Particles and Fields **76**, 401–415 (1997).

[21] Zyla, P.A., *et al*, "Bottom mesons.", Particle Data Group **Prog. Theor. Exp. Phys. 2020**, 083C01 (2020).

[22] Blake, T., Lanfranchi, G., & Straub, D. M., "Rare B decays as tests of the standard model.", International Journal of Modern Physics E **92**, 50–91 (2017).

[23] Liu, X. H., Wang, Q., & Zhao, Q., "Understanding the newly observed heavy pentaquark candidates.", Physics Letters B **757**, 231–236 (2016).

[24] Aker, M., *et al*, "Improved upper limit on the neutrino mass from a direct kinematic method by KATRIN.", Physical review letters **123**, 221802 (2019).

[25] Patrignani, C., *et al*, "Neutrino properties.", Particle Data Group Chin. Phys. C **40**, 100001 (2016).

[26] Sonnewald, M., Wunsch, C., & Heimbach, P., "Unsupervised learning reveals geography of global ocean dynamical regions.", in, Vol. 6, ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing 5 (IEEE, 2019), 784–794.

[27] Balakrishnan, G., *et al*, "An unsupervised learning model for deformable medical image registration.", in, Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE, 2018), 9252–9260.

[28] Ravanelli, M., *et al*, "Multi-task self-supervised learning for Robust Speech Recognition.", in, ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE, 2020), 6989–6993.

[29] Abe, T., *et al*, "Belle II technical design report", arXiv preprint arXiv:1011.0352 (2010).

[30] Abe, T., *et al*, "SuperKEKB collider.", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **907**, 188–199 (2010).

[31] CERNcourrier, *KEK reclaims luminosity record*, `cerncourier.com/a/kek-reclaims-luminosity-record/`, Accessed: 2020-10-02, 2020.

[32] Hara, T., Kuhr, T., Ushiroda, Y., "Belle II Coordinate System and Guideline of Belle II Numbering Scheme.", 2011.

[33] Paschen, B., *et al*, "Belle II pixel detector: Performance of final DEPFET modules.", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **958**, 162222 (2020).

[34] Dong, T. V., *et al*, "Calibration and alignment of the Belle II central drift chamber.", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **930**, 132–141 (2019).

[35]Svidras, H., “The Central Drift Chamber of Belle 2”, in, Vol. 6, Belle II starter kit workshop (Belle II, 2020), p. 23.

[36]Tamponi, U., & Strube, J., “Charged Particle ID in Belle II”, in, Vol. 5, Belle II starter kit workshop (Belle II, 2019), p. 5.

[37]Bessner, M., “Performance of the Belle II imaging Time-Of-Propagation (iTOP) detector in first collisions.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **958**, 162318 (2020).

[38]Tamponi, U., & Belle II collaboration, “The TOP counter of Belle II: status and first results.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **952**, 162208 (2020).

[39]Gaz, A., & Belle II collaboration, “Calibration and commissioning of the Time Of Propagation PID detector at the Belle II Experiment.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **936**, 511–514 (2019).

[40]Kindo, H., *et al*, “Initial performance of the Aerogel RICH detector of the Belle II experiment.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **952**, 162252 (2020).

[41]Shiizuka, S., *et al*, “Study of 144-channel hybrid avalanche photo-detector for Belle II RICH counter.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **628**, 315–318 (2011).

[42]Shebalin, V. E., “Electromagnetic Calorimeter of the Belle-II Detector.”, Physics of Particles and Nuclei **49**, 793–798 (2018).

[43]La Rocca, P. & Riggi, F., “The Use of Avalanche Photodiodes in High Energy Electromagnetic Calorimetry.”, Advances in Photodiodes, 251 (2011).

[44]Aulchenko, V., *et al*, “Time and energy reconstruction at the electromagnetic calorimeter of the Belle-II detector.”, Journal of Instrumentation **12**, C08001 (2017).

[45] Aushev, T., *et al*, “A scintillator based endcap KL and muon detector for the Belle II experiment.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **789**, 134–142 (2015).

[46] Kou, E., *et al*, “The Belle II physics book.”, Progress of Theoretical and Experimental Physics **2019**, 47–55 (2019).

[47] The Belle II collaboration, “Muon and electron identification efficiencies and hadron-lepton mis-identification probabilities”, BELLE2-NOTE-PL-2020-XXX (2020).

[48] Abudinen, F., *et al*, “Measurement of the semileptonic $\bar{B}^0 \to D^{*+}\ell^-\nu_\ell$ branching fraction with fully reconstructed B meson decays and 34.6 fb^{-1} of Belle II data.”, arXiv preprint arXiv:2008.10299 (2020).

[49] Keck, T., *et al*, “The Full Event Interpretation.”, Computing and Software for Big Science **3** (2019).

[50] Lakshminarayanan, V. & Fleck, A., “Zernike polynomials: a guide.”, Journal of Modern Optics **58**, 545–561 (2011).

[51] Hershenhorn, A., Ferber, T., & Hearty, C., “ECL shower shape variables based on Zernike moments.”, BELLE2-NOTE-TE-2017-001 (2017).

[52] Hill, E., Hearty, C., “Time calibration of the Belle II ECL crystals and crates”, BELLE2-NOTE-TE-2019-002 (2019).

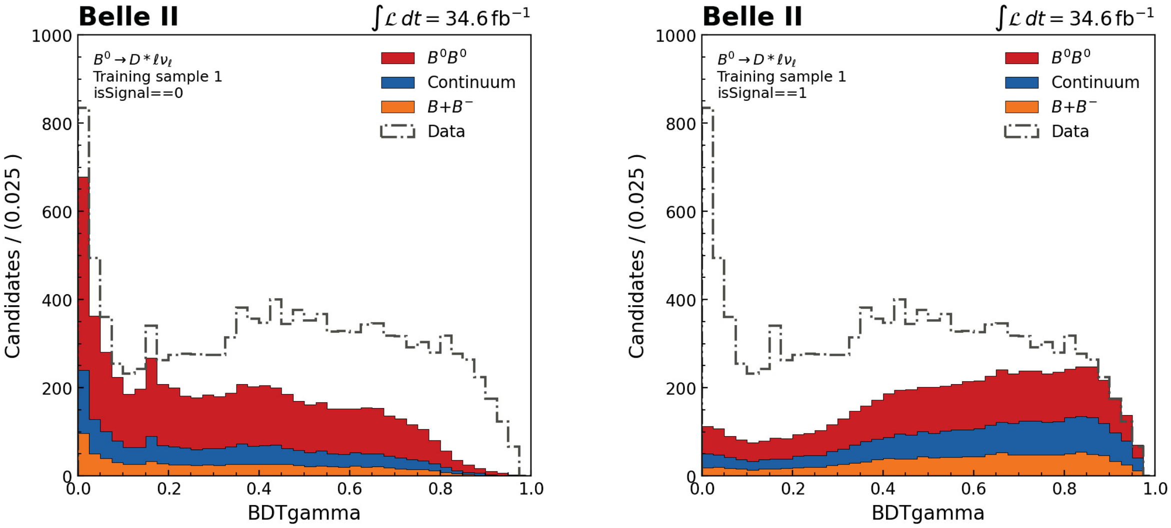
[53] Longo, S., *et al*, “CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders.”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **982**, 164562 (2020).

Appendices

Appendix A

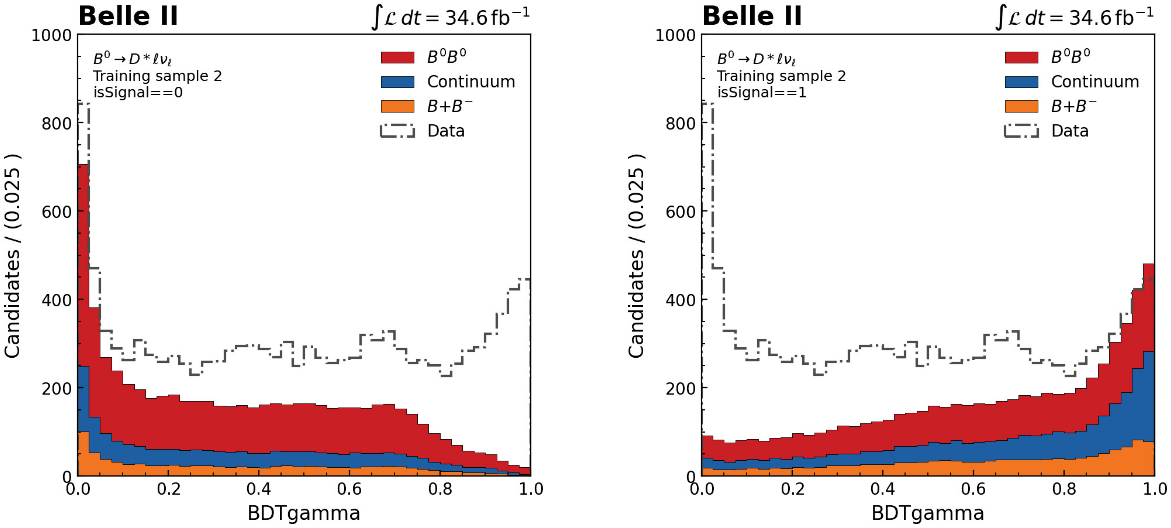
BDTgamma isSignal selections

As the trained classifiers are applied to the ROE ECL clusters of simulated $B^0 \to D^{*+}\ell^-\nu_\ell$ events, these clusters also have an associated isSignal value. It is therefore possible to directly examine the BDTgamma variable assigned to background (isSignal==0) and signal (isSignal==1) clusters. The plotted events reconstructed from data do not have an associated isSignal value and have been left only to provide a visual approximation of the clusters that did not pass the applied isSignal selection.



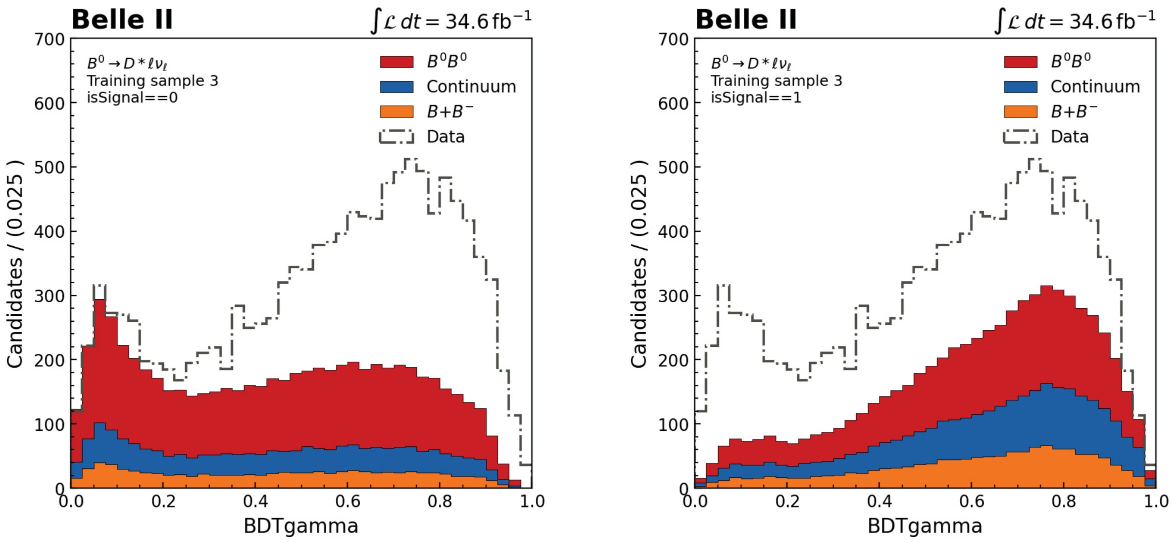
(a) **(b)**

Figure A.1: The BDTgamma distributions resulting from applying the classifier trained on sample 1 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.



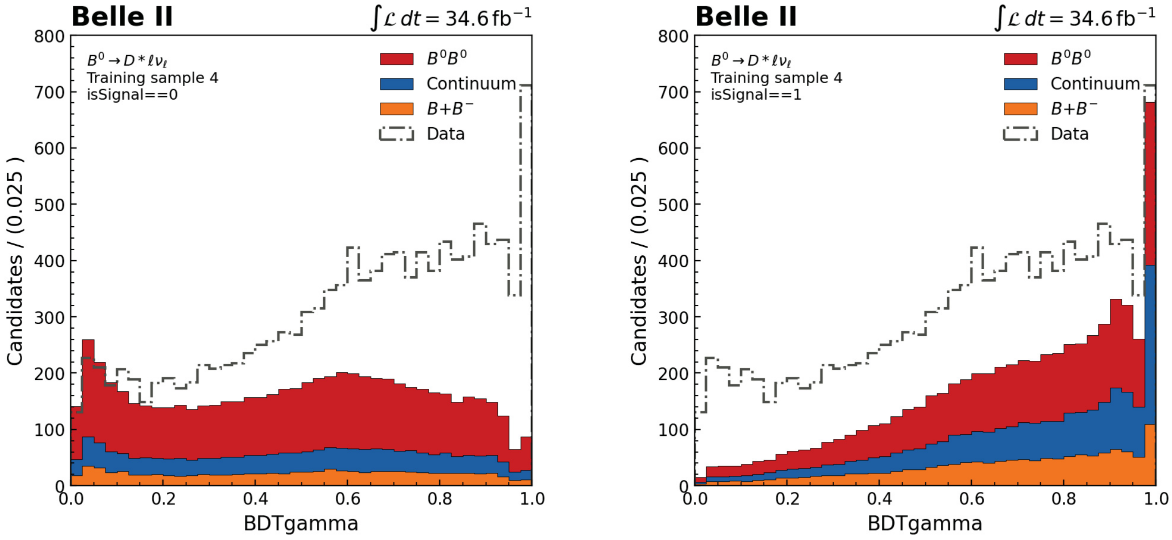
(a) **(b)**

Figure A.2: The BDTgamma distributions resulting from applying the classifier trained on sample 2 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.



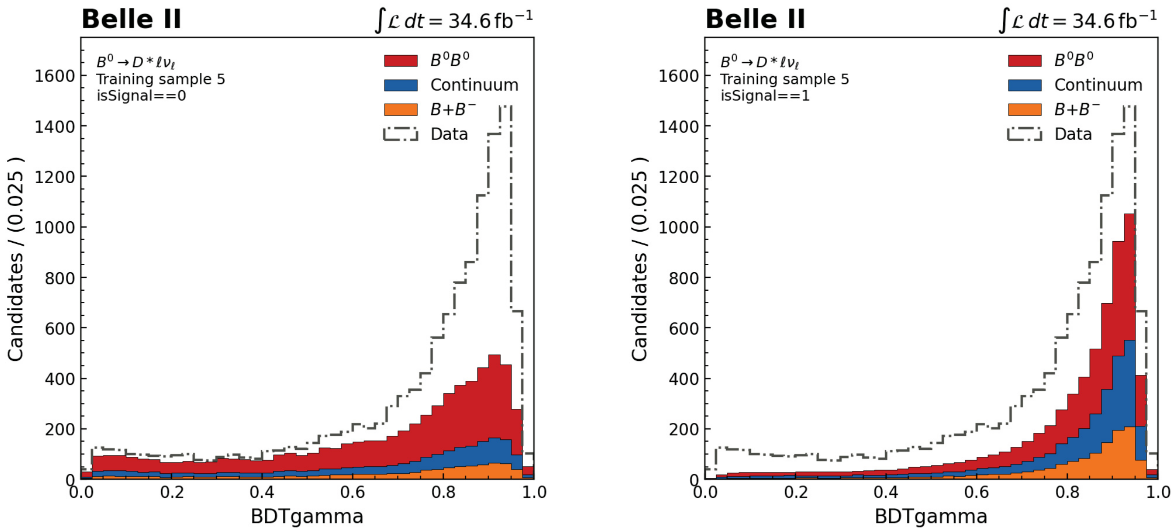
(a) **(b)**

Figure A.3: The BDTgamma distributions resulting from applying the classifier trained on sample 3 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.



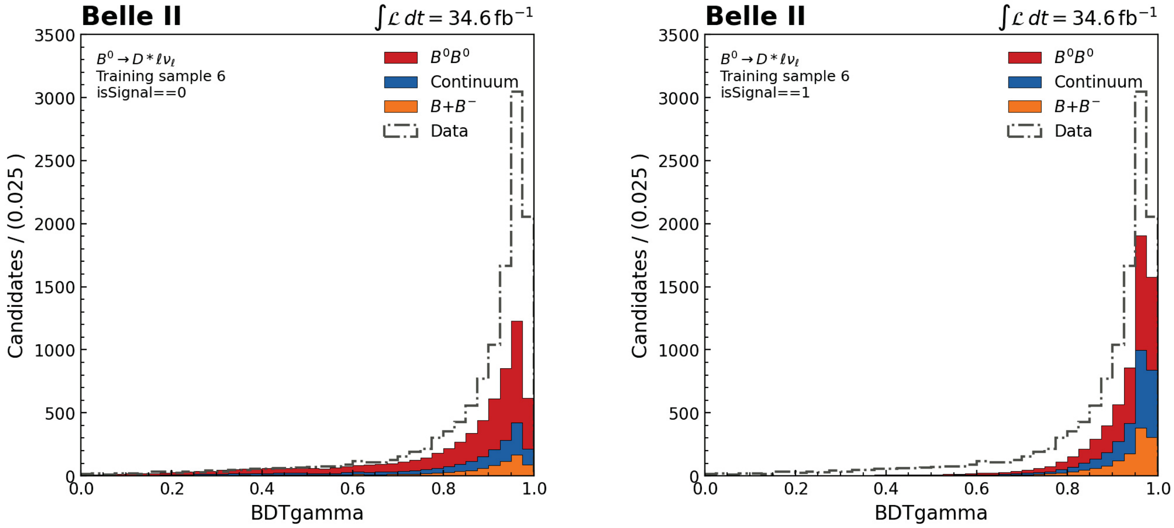
(a) **(b)**

Figure A.4: The BDTgamma distributions resulting from applying the classifier trained on sample 4 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.



(a) **(b)**

Figure A.5: The BDTgamma distributions resulting from applying the classifier trained on sample 5 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.



(a) **(b)**

Figure A.6: The BDTgamma distributions resulting from applying the classifier trained on sample 6 to the ROE clusters of reconstructed MC13 $B^0 \to D^{*+}\ell^-\nu_\ell$ events. The clusters shown in each distribution correspond to those with an isSignal value of (a) 0 and (b) 1.