Development and testing of a neutron identification algorithm for the Belle II experiment at SuperKEKB

Georgios Alexandris

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

November 2025

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...
Date Signature

1st Supervisor: Prof. Dr. Slavomira Stefkova
2nd Supervisor: Prof. Dr. Florian Bernlochner

Contents

CHAPTER 1

Introduction

In contrast to the large particle experiments operating at the LHC [1], Belle II [2], as well as other experiments operating at $e^+ e^-$ colliders, have the advantage of knowing the initial conditions of each collision with great accuracy. This makes studying particle decays with invisible particles, such as neutrinos, much more feasible in this experimental environment [3]. One such result from the Belle II collaboration is the observation of evidence for the $B \to K\nu\bar{\nu}$ decay [4], which contains two invisible particles. When studying such decays, an important background component is commonly associated with imprecisely modelled processes involving neutral hadrons, and especially neutrons, which tend to not interact much with the detector. It is therefore important for these analyses to improve the modeling of their neutral hadron backgrounds.

The most popular database for researching the properties of particles and specific decays, the Particle Data Group Particle Listings, which are published at the annual Review of particle physics [5], contains a significant lack of measurements of B decays containing neutrons, which is disproportionate with respect to the measured number of B meson decays containing protons. From symmetry arguments, it is expected that the number of B meson decay channels containing (anti-)protons and (anti-)neutrons should be equal. Therefore, the reasons that the neutron decay channels are not measured as often as the corresponding proton channels has to be the immense difference in the reliable reconstruction and identification of these particles, with protons being the significantly easier to detect.

This thesis aims at contributing to a better understanding of neutral hadrons, and especially neutrons, in Belle II, by developing a neutron classifier. Following some introductory remarks about the interactions of particles with matter in Chapter 2, Chapter 3 describes the experimental context of the SuperKEKB collider and the Belle II experiment. In Chapter 4, the interaction rates of photons and neutral hadrons with the electromagnetic calorimeter of Belle II are investigated. In Chapter 5, the relative particle abundances of neutral particles as a function of particle momentum is calculated according to generic B meson pair decay simulations. Three neutron classifiers for three momentum ranges are developed and are discussed in Chapter 6, including evaluation of their performance on the training samples and projections to generic B meson decays in Belle II. Finally, in Chapter 7 the $e^+e^- \to \Lambda\bar{\Lambda}\gamma$ process is studied and a reconstruction attempt is made for the specific decays of $\Lambda \to n\pi^0$ and $\bar{\Lambda} \to \bar{p}\pi^+$.

CHAPTER 2

Interaction of particles with matter

The Sandard Model (SM) of Particle Physics [5], describes the current understanding of Physics at the most fundamental level, that of elementary particles and their interactions. Notably, this understanding lacks the inclusion of the gravitational force, but the particle masses and energies concerning the experiments currently being conducted worldwide can safely ignore the effects of this interaction. The SM includes a unified description of the electromagnetic (EM) and weak nuclear interaction, under the Glashow–Weinberg–Salam model [6], using the Brout–Englert–Higgs mechanism [7] and of the strong nuclear force, under the Quantum Chromodynamics (QCD) framework [8].

Based on the ways particles interact with one another, they are grouped into *fermions*, the 'building blocks' of matter, *gauge bosons*, which are mediators of the fundamental forces and the scalar Higgs boson, the excitation of the Higgs field which gives mass to the elementary particles. Fermions can be further subdivided in *quarks*, which can interact with the strong force, and *leptons*, which cannot. This makes it possible for leptons to freely propagate[1], which is not possible for quarks, except at very high energies, where they reach the asymptotic freedom limit of QCD. For lower energies, such as those encountered in everyday life or in experiments such as Belle II, the strong interaction forces quarks and gluons to combine into composite particles which are neutral under the QCD colour charge, called hadrons. This phenomenon is known as *quark confinement* and it happens through a process called *hadronisation*.

Hadrons can be split into *mesons* and *baryons*, depending on the number of constituent quarks. Mesons consist of a quark–anti-quark pair, while (anti-)baryons contain three (anti-) quarks with a total colour charge of zero. A thorough description of the way quarks can be combined to form hadrons is given in Ref. [9].

Detector design and development, as well as particle identification, needs thorough understanding of the interaction of each particle species with the materials commonly used in particle detectors. Although these interactions are fundamentally described by the SM, calculations for macroscopic matter are not possible due to the large amount of complexity involved. To tackle this limitation, several models have been developed for different types of interactions. The following sections highlight the most important interaction mechanisms of elementary particles with matter. Section 2.1 covers the interactions of charged particles and photons, while Section 2.2 focuses on neutral hadrons. Inteactions of neutrinos with matter are beyond the scope of this thesis.

[1] In the case of muons and tau leptons, they freely propagate until they decay

2.1 Interaction of charged particles and photons with matter

2.1.1 Interaction of charged particles with matter

The charged particles that are of particular interest for high energy phyiscs experiments are $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$ and $p/\overline{p}$, because they are stable with respect to the typical detector length scales and leave traceable signatures in the detector systems. As particles travel through a medium, they interact electromagnetically with its atoms and/or nuclei. During each of these interactions, small fractions of the initial energy are transfered to the medium. Averaging these interactions over the whole passage of a particle though a material yields the Bethe-Bloch equation [10] which gives a mathematical expression for the mean rate of energy loss per unit length, also referred to as *mass stoppping power*:

$$\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{2.1}$$

where

- $K = 4\pi N_A r_e^2 m_e c^2$, where N_A is Avogadro's number, r_e is the classical electron radius, m_e is the electron mass and c is the speed of light,
- z is the charge of the incident particle in units of the electric charge of the electron,
- Z is the atomic number of the material,
- A is the atomic mass number of the material,
- $\beta = v/c$ is the relativistic speed of the incident particle,
- $\gamma = 1/\sqrt{1-\beta^2}$,
- $W_{\max}$ is the maximum possible energy transfer to an electron for each collision, in MeV,
- I is the mean excitation energy, in eV, and
- $\delta(\beta\gamma)$ is a density effect correction.

This equation describes incident particles with $0.1 \lesssim \beta\gamma \lesssim 10^3$, which can be described as moderately relativistic. In the typical energies appearing in high energy physics experiments, all of the aforementioned particles, except the electrons can be described by this equation. A more complete picture of the mass stopping power at and beyond the range of validity are shown in Fig. 2.1 Since electrons are the lightest stable charged particle that can be produced in particle collisions, the typical values of $\beta\gamma$ are well over 10^3, meaning that electrons are in the radiative loss area of the plot. This means that they electrons are significantly decelerated, which causes them to emit Bremstrahlung radiation at a large quantity.

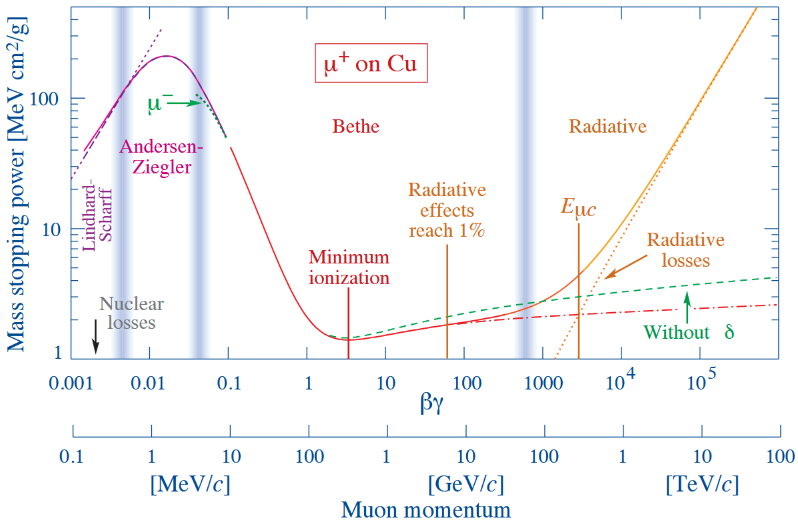


Figure 2.1: Mass stopping power for μ^{+} as a function of $\beta\gamma$, in copper. From Ref. [10]

2.1.2 Interaction of photons with matter

Photons interact with matter with three main mechanisms, the significance of which changes depending on the incident photon energy, as well as the material which is considered. Low energy photons (below 1 MeV), mainly interact via the photoelectic effect, sequentially exciting the material's electrons as the energy increases. At significantly higher energies, above a few MeV, the dominant interaction is electron–positron pair production via the contribution of the nuclear potential, which is required to exchange some of the angular momentum to make the process allowed. At these energies, which are the typical real of high energy physics experiments, the produced electron and position have enough energy to emit Bremstrahlung radiation, starting an electromagnetic shower until the energy is low enough for the photons and the electrons to be absorbed by the material. Taking advantage of electromagnetic showers particle detectors called *calorimeters* are able to detect electrons and photons and measure their energy. The third mechanism, which is dominant in intermediate energies, between a few hundred keV and a few MeV is the Compton effect, which is the phenomenon where the incident photon scatters against an electron present in the material and transfers part of its energy. Two examples of the photon cross section in carbon and lead are presented in Fig. 2.2. Additional details on the interaction of photons with matter can be found in Ref. [10].

2.2 Interaction of neutral hadrons with matter

The available descriptions of the interactions of neutral hadrons with matter are to a large extent derived from symmetry arguments with the respective charged hadron interactions, for which experimental data are significantly more abundant. A thorough collection of measured and derived cross sections for charged and neutral hadrons is presented in Ref. [11]. Another valuable resource to obtain a description of these interactions is the `GEANT4` physics reference manual [12]. `GEANT4` is a software package which can simulate the inteactions of all particles with matter and is used by nearly every generic particle physics simulation routine, including those used in Belle II.

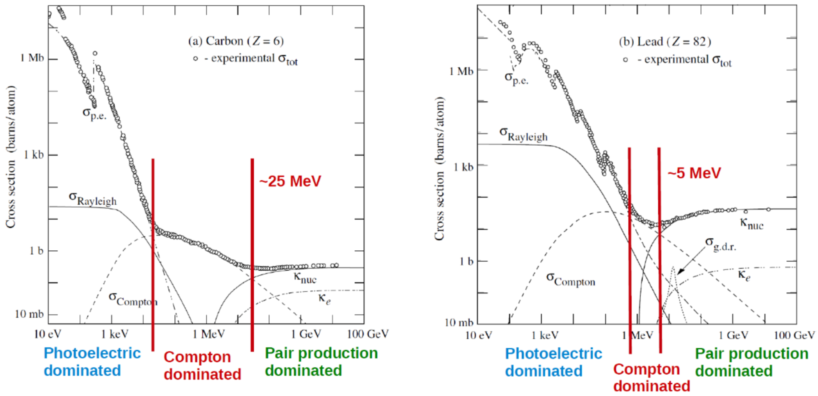


Figure 2.2: Photon cross section measurements in Carbon and Lead. From Ref. [10]

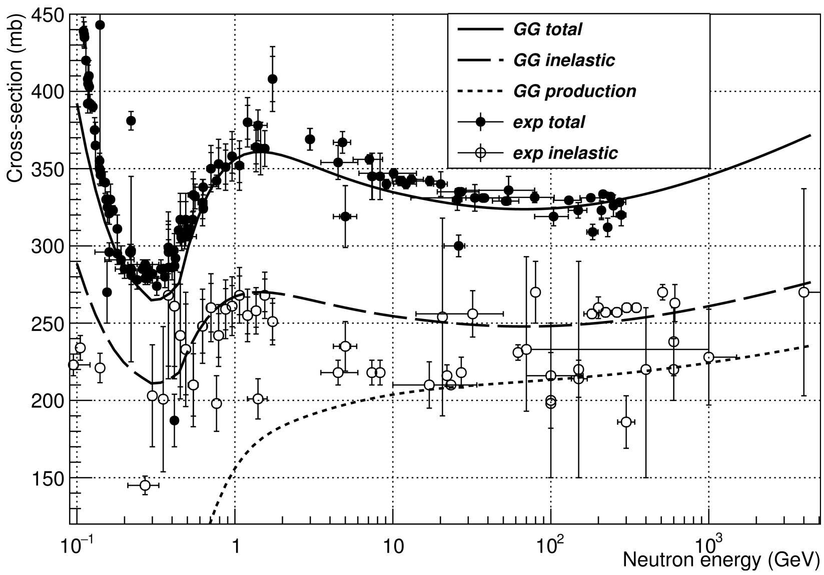


Figure 2.3: Neutron cross section measurements on a carbon target. From Ref. [12]. The lines correspond to the simplified Glauber-Gribov model [13].

In general, the interactions of all particles can be categorized in elastic and inelastic interactions, where in the latter, the number and species of partices can change after the interaction. An example of an elastic interaction is a neutron interacting with the nucleus giving it some of its energy, which is used to excite the nucleus. On the other hand, an example of an inelastic interaction is a neutron interacting with a nucleus with enough energy to produce a number of pions in the final state, effectively initiating a hadronic shower. An example of experimental measurements for the cross section of neutrons on a carbon target is presented in Fig. 2.3.

In contrast to the neutrons, anti-neutrons have an additional inelastic scattering mechanism, since they are anti-particles. An anti-neutron can annihilate with a neutron present in a material, releasing relatively large amounts of energy and producing a hadronic shower. The annihilation cross section drops as the anti-neutron energy becomes larger [11], meaning that at high enough energies, n and $\overline{n}$ interact in similar ways with matter.

The last neutral hadron stable enough to be of interest for studies in Belle II is the K-long meson (K_L^0). It consists of a superpostion of the two neutral strange meson flavor eigenstates [9] K^0 $\left(d\,\overline{s}\right)$ and $\overline{K}^0$ $\left(\overline{d}\,s\right)$. As a neutral hadron, its main interaction mechanisms are expected to be qualitatively similar to those of the neutron. It is possible that in high enough energies the $\overline{d}$ quark could annihilate with a d quark from a nucleon in the material, but no relevant literature about neutral kaon interaction with matter was found.

CHAPTER 3

SuperKEKB and the Belle II Experiment

The Belle II experiment is built around the interaction point (IP) of the SuperKEKB collider in Tsukuba, Japan. It is the direct successor of the Belle experiment, which operated at the KEKB accelerator from 1999 to 2010. Both experiments were designed to study the properties of B mesons via precision measurements and address fundamental questions in the flavour sector of particle physics, such as the origin of CP violation and the nature of dark matter. This chapter provides an overview of the SuperKEKB collider and the main components of the Belle II detector, which are the experimental setup for the studies presented in this thesis. For additional information, the reader is referred to Refs. [2, 14]

3.1 The SuperKEKB Collider

The SuperKEKB collider is an asymmetric-energy circular electron-positron collider located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. SuperKEKB is the upgraded version of the KEKB collider, which operated from 1998 to 2010 and achieved a world record luminosity of $2.1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Following the steps of its predecessor, SuperKEKB recently broke the world record for instantaneous luminosity, reaching $5.1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ in autumn 2024 [15], while the design target of SuperKEKB is to reach a peak luminosity of $6 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [15].

The electron beam is inside the high-energy ring (HER), rotating clockwise with an energy of 7 GeV, while the positron beam is inside the low-energy ring (LER), rotating in the opposite direction with an energy of 4 GeV, corresponding to a centre-of-mass energy of 10.58 GeV. The two beams collide at the interaction point (IP) with a crossing angle of 83 mrad. The energy asymmetry gives rise to an overall Lorentz boost in the electron direction, which increases the flight distance of the collision products, allowing for the precise measurement of the particle decay vertices. The nominal centre of mass energy of SuperKEKB is near the the rest mass of the $\Upsilon(4S)$ resonance, which according to the Particle Data Group (PDG) [5] has a mass of $(10.5794\ \pm 0.0012\,(\mathrm{tot.}))$ GeV and a width of $(20.5\ \pm 2.5\,(\mathrm{tot.}))$ MeV. According to the PDG, the cross section of $e^+e^- \to \Upsilon(4S)$ is approximately 1.1 nb and the branching fraction of $\Upsilon(4S)$ to decay to a B meson pair is at least 96 % at the 95 % confidence level.

Other parts of the SuperKEKB collider complex are presented in Fig 3.1 and include a linear accelerator (linac), which accelerates electrons and positrons to their respective energies before injecting them into the HER and LER, and a damping ring (DR) for the positrons, which reduces the beam emittance before injection into the LER.

The HER and LER each consist of four arcs and four straight sections. The arcs contain bending magnets and focusing quadrupole magnets to keep the beams on their circular paths, while the straight sections contain radio-frequency (RF) cavities to accelerate the beams. The beams are also focused vertically and horizontally using a series of final-focus quadrupole magnets located close to the IP, which are essential for achieving high luminosity.

Figure 3.1: Schematic view of the SuperKEKB collider complex. Figure from Tobias Blegsen.

3.2 The Belle II detector

The Belle II detector is a general purpose particle detector with a layered sub-detector structure hermetically installed around the SuperKEKB IP in order to cover as much of the 4π solid angle as possible. To that end, it has a cylindrical shape consisting of 3 parts: the backward endcap, the barrel and the forward endcap. The detector has an asymmetric design to account for the Lorentz boost given to the $e^+ e^-$ system and to allow for a higher resolution in the forward region of the barrel as well as the forward endcap, which is where the majority of the final state particles fly towards. A schematic view of the whole detector is shown in Fig. 3.2.

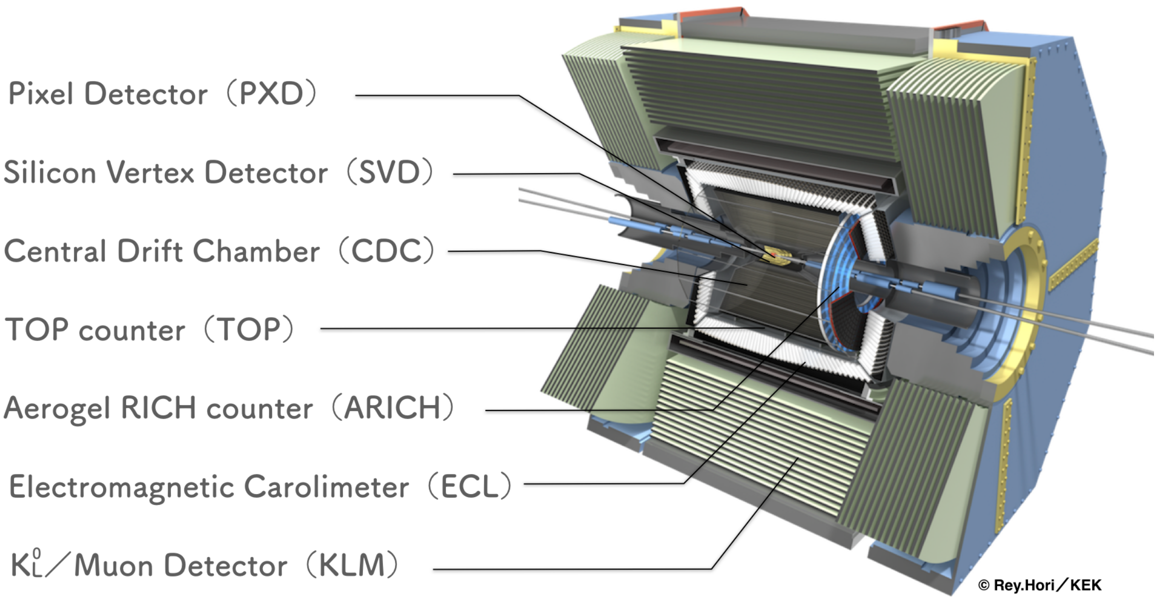


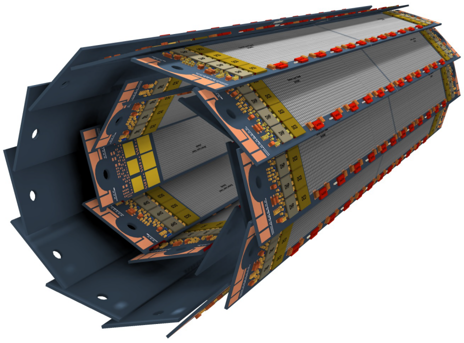
Figure 3.2: Schematic view of the Belle II detector. Figure from https://www.kek.jp/ja/about/pr/image/category/illustration.

3.2.1 The Vertex Detector (VXD)

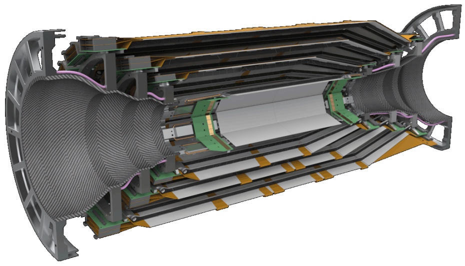
The Belle II Vertex Detector (VXD) is comprised of the Pixel Detector (PXD) and the Silicon Vertex Detector (SVD). They are located at the innermost part of Belle II, near the IP and are responsible for the precise reconstruction of the secondary decay vertices of B mesons, as well as other heavy particles that decay via the weak interaction such as the D mesons and the Λ_c^+ baryon.

The PXD is the closest detector to the IP, meaning it faces the most radiation (beam-related background). For this reason, the pixel design was preferred over semiconducting strips, allowing for less noise per detector module and enabling the decay vertex reconstruction. In contrast to the pixel detectors commonly found in experiments with much larger beam energies, such as those at the LHC, and are usually made of silicon, the PXD's active component is made from DEPFET (DEPleted Field Effect Transistor) pixels, which can be thinned to 50 µm and therefore contribute minimally to the material budget of the detector. Another advantage of this design is that the readout and cooling systems can be placed completely outside the detector acceptance, further reducing the PXD's interference with the incident particles.

The SVD is located between the PXD and the Central Drift Chamber (CDC) and is comprised of four layers of double-sided silicon strip detectors (DSSDs). It is responsible for the reconstruction of particles that decay outside the PXD radius, such as D mesons, and acts as the interface between the PXD and CDC allowing for track reconstruction across the whole detector.



(a) The PXD detector. Figure from Belle II Internal documentation.



(b) The SVD detector. Figure from https://web2.infn.it/Belle-II/images/SVD20spaccato.png.

Figure 3.3: Schematic views of the PXD and SVD detectors.

3.2.2 The Central Drift Chamber (CDC)

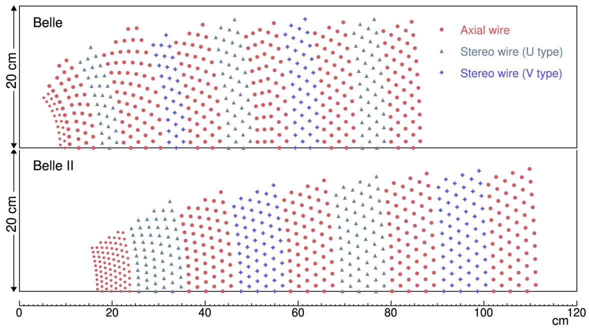


Figure 3.4: Schematic view of the layer structure in an angular slice of the CDC in Belle and Belle II. Figure from [16].

The CDC is the outermost tracking detector of Belle II. It consists of a gas-filled (50 % Helium, 50 % Ethane) cylindrical module with charged wires arranged on its inside. Its purpose is to reconstruct the trajectories of charged particles, which ionise the gas and produce charged avalanches which are collected by the wires. The wires are arranged in 9 super layers, while there are 3 super layer types: axial "A" and stereo "V" and "U". Having axial layers alone is not enough to be sensitive on the z axis position of a signal. To tackle this, problem, the stereo layers are introduced, which have their wires skewed with respect to the beam axis. The U-type stereo layers are tilted in a positive angle, while the V-type layers have a negative tilt. The 9 CDC layers are arranged in a AUAVAUAVA arrangement. Each super layer consist of 6 sense wire layers except the innermost axial super layer, which has 8 sense wire layers. The sense wires carry the high voltage responsible for collecting the ionization and each sense wire is surrounded by 8 field wires, responsible for shaping the electromagnetic field around it. A schematic of the layer structures in an angular slice of the CDC is presented in Fig. 3.4 as a comparison between the Belle and the Belle II designs.

Another significant role of the CDC is providing energy loss information for use in particle identification via the Bethe-Bloch equation 2.1. The energy loss information can be directly inferred from the average ionization for each track.

3.2.3 The Time-Of-Propagation (TOP) Counter and the Aerogel Ring-Imaging Cherenkov (ARICH) Detector

The TOP and ARICH detectors are Belle II's dedicated particle identification (PID) systems. The TOP counters are located in the barrel, covering the outer CDC wall, while the ARICH detectors cover the forward endcap. They are designed to reliably distinguish between pions and kaons, but they can also serve as complementary PID sources for other particles. Both detectors take advantage of the Cherenkov radiation produced when charged particles travel through a medium faster than light would. To identify the particles, only particle momentum (provided mainly by the CDC) and the angle θ_C between the incident particle and the Cherenkov photon are needed, since from the relativistic kinematics it is known that $\beta\gamma = p/m$ and $\beta = n\cos\theta_C$ [17], where n is the medium's refractive index. Combining the two expressions, and noting that $\gamma = 1/\sqrt{1-\beta^2}$, an expression for the particle's mass can be obtained: $m = p\sqrt{\frac{1}{n^2\cos^2\theta_C} - 1}$.

Each TOP counter module consists of a quartz bar in which the photons travel, an array of micro-channel plate (MCP) photomultiplier tubes (PMTs) to collect the light on one side and a focusing mirror to return the light flying in the opposite direction back to the PMTs. The TOP counters reconstruct the Cherenkov angle by using the position information from the modules and the timing information from the MCP PMTs.

The ARICH detector consists of an aerogel radiator and an array of position sensitive photon detectors capable of two dimensional photon position reconstruction along the forward endcap's surface. Between the two, there is a 20 cm gap which allows the Cherenkov photons to form rings on the photon detection matrix. The reconstruction of θ_C is then done by measuring the radius of the Cherenkov rings and projecting it back to the interaction point on the radiator.

3.2.4 The Electromagnetic Calorimeter (ECL)

The ECL is a homogeneous calorimeter consisting of 8736 CsI(Tl) crystals arranged along the barrel and both endcaps. The crystals are placed so that they approximately point towards the IP, and to that end, they have the general shape of a truncated pyramid. Specifically, 29 distinct shapes are needed for the barrel and 69 for the endcaps. The ECL has a polar angle acceptance of $\theta \in [32.2^\circ, 128.7^\circ]$ in the barrel, $\theta \in [12.4^\circ, 31.4^\circ]$ in the forward endcap and $\theta \in [130.7^\circ, 155.1^\circ]$ in the backward endcap.

The main purpose of the ECL is to induce electromagnetic showers from incoming electrons and photons, however other particles such as charged an neutral hadrons can interact with it. The ECL is the only way to measure photons (direction and energy) in Belle II, while it also helps with electron reconstruction. Since photons and electrons create very similar shower shapes when interacting with the ECL, the extrapolation of all CDC tracks to the ECL is necessary to make sure the photon showers are not matching a charged particle. In addition to that, electrons can emit one or more Bremstrahlung photons before interacting with the ECL, which are usually colinear and can therefore create bigger shower shapes when combined with the electron shower. Other charged particles ($\pi^\pm$, $K^\pm$, $\mu^\pm$) can also interact with the ECL, but they mostly interact elastically as MIPs, leaving small energy depositions. If they have an inelastic interaction, they produce irregular shower shapes which are further deformed by the irregular entry angle caused by the curvature of their path due to the magnetic field. Short-lived neutral hadrons such as the π^0 and η particles decay promptly to photons, leading (especially in the π^0 case) to partially or fully overlapping photon showers which need special reconstruction logic. Finally, the neutral hadrons stable enough to traverse the detector, namely n, $\overline{n}$ and K_L^0, can potentially interact with the ECL producing hadronic showers.

3.2.5 The Magnetic Solenoid

A sufficiently strong magnetic field is an essential part of any particle physics experiment like Belle II. In this experiment, a superconducting solenoid with size diameter $\times$ length = 3.4 m $\times$ 4.4 m creates an approximately uniform magnetic field of 1.5 T. The magnetic field is an essential part of charged particle trajectory reconstruction and momentum measurements. To be robust against deviations of the field from uniformity the field is simulated based on the exact detector geometry. To further calibrate the simulation, the collaboration took dedicated measurements before the start of operation. In addition, the magnetic field is regularly monitored based on real particle studies with cosmic rays, well modelled decays (such as $J/\psi \to \mu^- \mu^+$) and measurements with the magnet being turned off.

The return yoke of the magnetic field is the iron structure which holds the Belle II experiment in place, providing the necessary support for the sub-detectors.

3.2.6 The K-Long and Muon Detector (KLM)

The K_L^0 and Muon Detector (KLM) is located in the outermost part of Belle II. Its main purpose is to detect charge particles which have not created an EM shower in the ECL, which are mainly muons, as well as induce hadronic showers for neutral hadron detection. As its name states, the KLM was initially desinged to reconstruct hadronic showers from K_L^0 and muon tracks. However, the similarities of n interactions to those of K_L^0, make the KLM a more versitile system.

The KLM has a design resembling a sampling calorimeter. It consists of a set of thick iron plates alternating with resistive plate chambers (RPCs) which are gaseous detectors that can collect the ionization charge produced by passing particles. The iron plates are used to create hadronic showers, the products of which are detected by the RPCs, forming clusters. In addition, the same iron plates serve as the return yoke of the magetic field.

3.2.7 The Belle II Analysis Software Framework (`basf2`)

The Belle II Analysis Software Framework (`basf2`) [18] is the core software of the Belle II collaboration. It is written in `C++` with `ROOT` [19] and has a Python interface providing a plethora of convenience functions. `basf2` can handle a variety of tasks, namely: *generation*, *simulation*, *reconstruction* and *skimming*.

A generation task in `basf2` can create either a physics process generation, attempting to mimic nature or an uncorrelated particle generation task, mainly used for performance studies. To generate a physics process, the analyst provides `basf2` with a physics generator along with a configuration to allow the selected generator to output the desired process. An overview of the available generators in `basf2` is given in Ref. [20]. Generating uncorrelated particle samples can be achieved by the `ParticleGun` module, which allows generation of events with any number of particles with given momentum and anglular distributions. The most commom case of using `ParticleGun` samples is to generate events with a single known particle. Both approaches use Monte Carlo (MC) sampling algorithms to decide on the exact kinematics of each process, which use probabilities from quantum mechanical matrix elements or, in the case of ParticleGun, the desired distributions. In the generation step all unstable particles are usually instructed to decay into stable final state particles.

The simulation process is responsible for propagating each generated final state particle through the detector and outputs the simulated response of each detector compontent with which the particles interacted, which is called *digitisation*. From generation to digitisation, the software has managed to convert generated particles to the same type of output produced by real particles when the detector is operating. Thus, the reconstruction and skimming steps behave identically on simulated samples and on real data, except that generateed particle information is also available for simulated samples.

During the reconstruction step, all signals from the Belle II sub-detectors are converted into higher level objects, such as tracks, ECL clusters, KLM clusters, etc. Finally, the reconstructed files can be reprocessed with basf2 to obtain reconstructed particle objects, on which custom selections and corrections can be applied by the analysts. Using the reconstructed particle candidates, users can combine them to attempt to reconstruct the decay trees of unstable particles.

Track Reconstruction

A detailed description of the Belle II track finding algorithm is given in Ref. [21]. Track finding in Belle II needs to take into account the outputs of all three tracking sub-detectors (PXD, SVD and CDC). A sophisticated algorithm based on the Kalman-Filter method is used to fit a helical track through the hits in each detector. Since there can be an very large amount of detector hits, which creates an exponentially high number of potential tracks, there have been significant efforts to improve track finding and reconstruction. A notable modern approach is usign Graph Neural Networks (GNNs) to tackle this problem [22]. The reconstructed track objects can be defined by a set of 5 parameters, which are computed at the point of clisest approach (POCA) of said track to the origin of the Belle II coordinate system:

- d_0, the distance of the POCA in the transverse plane,
- z_0, the z coordinate of the POCA,
- ϕ_0, the angle defined by the x–axis and the transverse momentum at the POCA,
- ω, the inverse of the radius of the fitted helix, and
- $\tan\lambda$, the tangent of the angle between the momentum at the POCA and the xy plane.

It is important to note that track reconstruction makes no assumption on the particle species.

ECL Cluster Reconstruction

The reconstruction of ECL clusters from the energy depositions in the crystals is described in detail in Ref. [23], so only a brief description follows. In each crystal, the electrical waveform produced by the absorb photon conversion is fitted to a model from which an amplitude, a time and a status is saved. This set of information constitutes an ECLDigit and represents the uncalibrated energy deposit for a single crystal. This information is then calibrated, to form an ECLCalDigit object which contains the calibrated energy, time and time resolution information. Adjacent ECLCalDigits with deposits larger than ≈20 MeV are then grouped into connected regions (CRs), in which the largest local energy deposits are marked as "seeds". Each seed can eventually be identified as an ECL cluster or be merged with other seeds in order to form an ECL cluster. After identifying the seeds of each `ECLConnectedRegion` one of the following hypotheses must be attached to each CR, which depend on wether the CR has been matched to a track (labelled T1-T3) or not (labelled N1-N3):

- T1: Minimum Ionizing particles and, optionally, additional photons,
- T2: One charged hadron,
- T3: One $e^{\pm}$ and, optionally, additional photons,
- N1: Any number of photons,
- N2: One neutral hadron,
- N3: One merged π^0.

Then, following a clustering algorithm, each connected region is mapped to one (or more) `ECLShowers` which can partially overlap. In the case a track is matched to the CR, only one `ECLShower` for each `ECLConnectedRegion` is allowed to be matched to a track. When two or more `ECLShowers` overlap, the reconstructed energy of each `ECLCalDigit` needs to be split among the shower objects. Thus, for each `ECLShower` a set of weights w_i which sum to 1 are associated with each `ECLCalDigit`. Before storing the clusters to the final data format, a selection is applied, and passing `ECLShowers` are renamed to `ECLClusters`.

Monte Carlo Truth Matching

One of the most important insights one can have from studying MC samples is comparing reconstructed kinematic distributions with their generated counterpart. Doing so, analysts can estimate reconstruction efficiencies and resolutions, which are essential in producing accurate physics measurements that can also be compared with other experiments. It is therefore essential to be able to match each generated particle to a reconstructed high level object, which is usually a track, or a cluster. This process is usually referred as MC truth-matching, truth-matching or just MC matching.

The core concept of an MC Matching algorithm is to loop over all generated particles, compare their trajectories with the reconstructed tracks (only if charged) and/or compare the end of their trajectories with the positions of the reconstructed ECL or KML clusters, and assign the recosntructed object that best matches the generated particle to it. The Belle II MC matching algorithm is described in Ref. [24]. An updated algorithm for cluster MC matching, which is currently being developed is discussed in Ref. [25]. It is important to note that for a cluster with reconstructed energy E_{reco} to be able to be matched to a generated particle with energy E_{gen}, the following conditions must be satisfied for an `ECLCluster` with weight w [26]:

$$\frac{w}{E_{\mathrm{reco}}} > 0.2, \quad \text{and} \tag{3.2}$$

$$\frac{w}{E_{\mathrm{gen}}} > 0.3. \tag{3.3}$$

Particle identification (PID)

Using the energy loss information from the CDC and the outputs of the dedicated TOP and ARICH detector systems, as well as information about the size of the deposit in the ECL and KLM detectors, it is possible to calculate the likelihood for a candidate track to be a certain particle. By combining information from all sub-detectors, Ref. [27] describes this process in detail. At the analysis level, PID_X variables are provided for each reconstructed track, making it possible to select a *PID working point*, which influences the probability of the selected track to be a particle X. Such a selection is associated with an efficiency (true positive rate) and a mis-identification rate. These varialbes are only available for charged particles. The `clusterPulseShapeDiscriminationMVA` variable, described in Section 4.1, was created to categorize a reconstructed `ECLCluster` as originating from a photon or a neutral hadron, but it is not as well studied as the PID likelihoods for charged particles.

CHAPTER 4

Single Particle Simulation Studies

A general method to study the interaction of particles with the Belle II detector is to generate single particle samples using `basf2` [18]. In this way, it is possible to obtain samples that are independent of specific physics process assumptions, suitable for studying the interaction properties of each particle with the detector, as well as for general classification tasks. This chapter describes the generation process of such samples for studying the interaction rates of neutral particles with the Belle II ECL and use in neutron classification training.

4.1 Simulation Setup

Using the `ParticleGun` functionality of `basf2`, events with any number of particles with a specific angular and momentum distribution can be generated. For these and the following studies, samples are created uniformly in momentum, polar angle and azimuthal angle for three momentum ranges, one polar angle range and the whole azimuthal angle range $\phi \in [0^\circ, 360^\circ)$. No beam-induced background is overlaid to these samples during generation. The polar angle range for which the samples were generated is the ECL barrel polar acceptance, as defined in 3.2.4. The momentum ranges which will also be referred as "low", "central" and "high" momentum range or momentum bin are defined as:

- low: $p_{\text{gen}} \in [0.1, 0.5]$ GeV/c,
- central: $p_{\text{gen}} \in [0.5, 1.5]$ GeV/c, and
- high: $p_{\text{gen}} \in [1.5, 4.0]$ GeV/c.

For each momentum range, single particle samples of the neutral particles that do not decay before (possibly) interacting with the detector are neutrons (n), anti-neutrons ($\overline{n}$), K_L^0 and photons (γ) are generated. It is interesting to study the neutral hadrons, but since photons are the most common neutral particle produced in Belle II, it is helpful to compare their properties with those of neutral hadrons.

The motivation for selecting the above momentum ranges is as follows. For the low momentum bin, a threshold of 0.1 GeV was chosen, to sufficiently suppress beam background noise [28]. The 0.5 GeV threshold between the low and central momentum bin was chosen considering that, on average, particles with $p > 0.5$ GeV can reach the KLM [27], and have therefore interacted with at least one part of the detector. The central momentum bin was arbitrarily chosen to have a width of 1 GeV and allows for examining the transition of the interaction properties from low to hight momenta. Lastly, the high momentum bin essentially represents all particles with $p > 1.5\,\mathrm{GeV}$, for which a maximum value of 4.0 GeV is chosen to limit the dataset size.

Initially, 1 M candidates for each of the 12 samples ([3 momenta] × [4 particle species]) were generated, but it was observed that more candidates were needed in the low and central momentum ranges for the neutrons in order to overcome the forthcoming selection and interaction efficiencies. Therefore, for the neutron samples 5 M events with low momentum and 2 M events with intermediate momentum were generated instead.

For each sample, all ECL clusters were reconstructed under the "photon" hypothesis. The choice of this hypothesis for all particle species was made in order to keep the reconstruction logic simple and bias-free. In addition, the `basf2` MC matching module was run over each reconstructed cluster, which can return truth information on sufficiently energetic clusters.

The following cluster variables are saved and used in the following studies:

- `clusterE`: The calibrated energy of the cluster, in GeV.
- `clusterTheta`: The polar angle of the cluster, in radians.
- `clusterPhi`: The azimuthal angle of the cluster, in radians.
- `clusterE1E9`: The ratio of the energy deposited in the cluster seed crystal, divided by the energy sum of the 3×3 grid of crystals centered at the seed crystal.
- `clusterE9E21`: The ratio of the energy sum of the 3×3 grid of crystals centered at the seed crystal, divided by the energy sum of the 5×5 grid of crystals centered at the seed crystal, minus the 4 corners.
- `clusterLAT`: A number-valued representation of the lateral energy distribution of the cluster, defined by

$$\texttt{clusterLAT} = \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^{n} w_i E_i r_i^2}, \tag{4.4}$$

  where E_i are the energy-sorted `ECLCalDigit` energies in descending order, w_i is the `ECLCalDigit` weight and r_i is the transverse distance from the shower center to the i-th digit measured in a plane perpendicular to the shower axis, with $r_0 = 5\,\mathrm{cm}$ representing the spacing between neighboring crystals.
- `clusterNHits`: The sum of the `ECLCalDigit` weights w_i of the cluster, which corresponds to the number of crystals beloging to the clusters in the case where it is not overlapping with another cluster.
- `clusterHasPulseShapeDiscrimination`: A boolean variable indicating if the pusle shape discrimination (PSD) fit was successful.
- `clusterPulseShapeDiscriminationMVA`: The score ($\in [0, 1]$) of a multivariate classifier trained on PSD information to distinguish photon cluster (high score) from hadron clusters (low score).
- `clusterAbsZernikeMoment40`: The absolute value of the cluster Zernike moment Z_{40}, $|Z_{40}|$.

- clusterAbsZernikeMoment51: The absolute value of the cluster Zernike moment Z_{51}, $|Z_{51}|$.

Regarding the last two variables, the absolute Zernike moments are defined by

$$|Z_{nm}| = \frac{n+1}{\pi} \frac{1}{\sum_i w_i E_i} \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} w_i E_i. \tag{4.5}$$

where n, m are the integers, i iterates over the crystals in the shower, E_i denotes the energy of the i-th crystal in the shower, R_{nm} is a polynomial of degree n, ρ_i represents the radial distance of the i-th crystal in the perpendicular plane, and α_i signifies the polar angle of the i-th crystal in the perpendicular plane. Since a crystal may correspond to multiple showers, w_i represents the portion of the energy of the i-th crystal linked to the shower.

In addition, other variables such as clusterZernikeMVA, clusterTiming and clusterErrorTiming were also inspected, but their use was discouraged by the Belle II ECL software working group, who recommended restricting ourselves to the variables listed above.

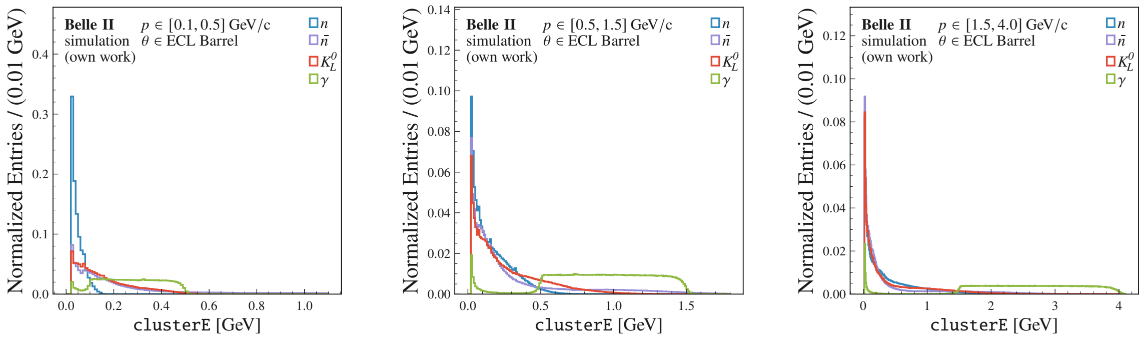
In Figures 4.1 to 4.3 the distributions of clusterE, clusterE1E9, clusterNHits and clusterPulseShapeDiscriminationMVA are presented. The rest of the variables are presented in Appendix A.1. These distributions represent all reconstructed clusters and no additional selections have been applied, including MC truth filtering, and are normalized to unit area.

Even without any pre-selection, the figures show differences between the hadronic distribution shapes, both among themselves and when compared to the photon distributions. These differences become less significant as the momentum range increases. In Fig. 4.3, the variable is missing if clusterHasPulseShapeDiscrimination = 0, meaning that not all reconstructed clusters are shown. The efficiency ϵ_{PSD} of the clusterHasPulseShapeDiscrimination = 1 selecton, given that at least one ECL cluster is reconstructed, is presented in Table Table 4.1.

Table 4.1: Efficiency of the clusterHasPulseShapeDiscrimination = 1 selecton, given that at least one ECL cluster is reconstructed, for all generated particles and momentum bins.

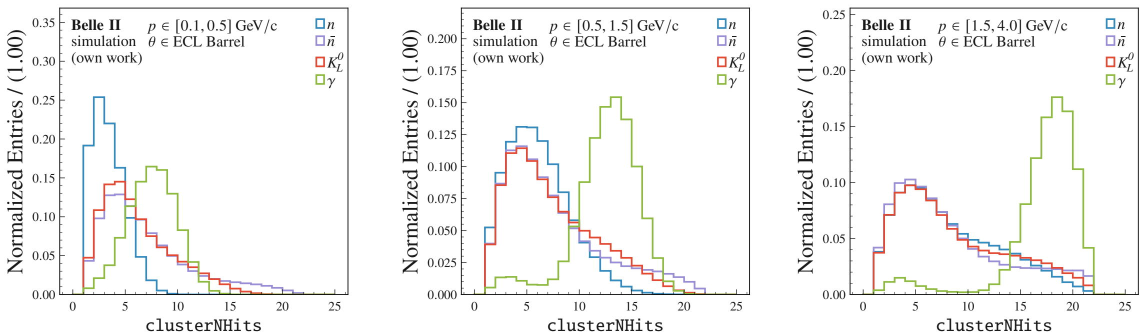
Momentum bin	n [%]	$\overline{n}$ [%]	K_L^0 [%]	γ [%]
low	50.42 ± 0.07	27.794 ± 0.025	54.11 ± 0.04	91.454 ± 0.028
central	70.86 ± 0.04	31.925 ± 0.030	52.20 ± 0.04	93.410 ± 0.025
high	51.05 ± 0.05	32.023 ± 0.031	42.01 ± 0.04	94.707 ± 0.022

Therefore, the correct interpretation of the figures is that they do not describe exactly the cluster variables of the generated particles, but they show the distributions of the cluster variables of all particles produced in these single particle events, given that they were reconstructed by the ECL. It is thus critical to understand the interaction rates of the different species with the ECL and craft suitable selections to make the distributions meaningful for further use in particle identification.



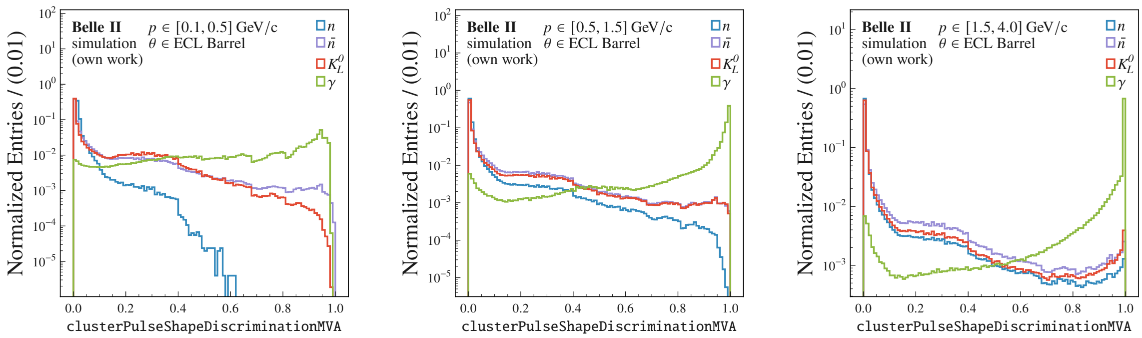
(a) Low (b) Central (c) High

Figure 4.1: `clusterE` distributions at low, central, and high momenta. The distributions are normalized to unit area.



(a) Low (b) Central (c) High

Figure 4.2: `clusterNHits` distributions at low, central, and high momenta. The distributions are normalized to unit area.



(a) Low (b) Central (c) High

Figure 4.3: `clusterPulseShapeDiscriminationMVA` distributions at low, central, and high momenta. The distributions are normalized to unit area.

4.2 ECL Interaction Rates

In order to study the interaction rates of neutral particles with the ECL, the single particle samples described in the previous section will be used. the *ECL interaction rate* for a particle can be defined to be the selection efficiency of requiring "$\geq$ 1 `ECLCluster`" which is equal to the fraction of single particle events resulting in at least one reconstructed `ECLCluster` divided by the total number of such generated events for that particle.

The ECL interaction rates for each generated momentum bin and particle species is presented in Fig. 4.4. Photons show an approximately constant and high interaction rate of ~ 90 %, which is sensible considering that they are the ECL's main target. The interaction rates have a similar shape for anti-neutrons and K_L^0, except for a shrinking offset of ~ 14 %, ~ 10 % and ~ 8 % in each momentum bin. Both the decresing rate and the offset can be explained by the observations in Section 2.2. The offset's origins is the additional interaction mechanism of anti-neutrons, annihilation, while the general trend follows the general expectations of hadronic cross sections in these energies.

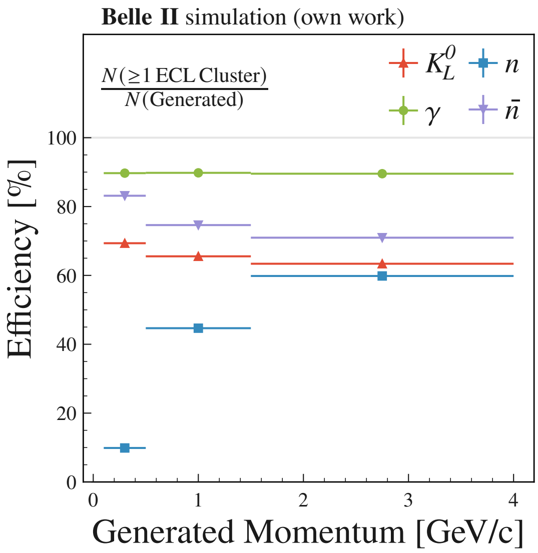


Figure 4.4: ECL interaction rates for each neutral particle as a function of the momentum bin. The horizontal error bars are used to show the momentum bin ranges. The vertical error bars are not visible at this scale.

Direct shape comparisons of interaction rates with the cross section plots in Section 2.2, are only valid assuming a constant efficiency of the clustering algorithm as a function of generated momentum. If that efficiency depends on momentum, the shape can significantly deviate from this expectation.

This could be the case for neutrons in Fig. 4.4. It would be expected that the neutron interaction rate behaves similarly to the K_L^0 sample, since both are neutral hadrons that cannot annihilate. However, the results show that the neutron interaction rate is ~ 5 % at the low momentum bin, rising to ~ 45 % in the central momentum bin, before almost reaching the same rate as K_L^0, ~ 60 % in the high momentum bin. The leftmost plot of Fig. 4.1 shows the cluster energy distribution for each particle in the low momentum range. It can be observed that neutrons tend to create clusters with significantly less energy than their generated momentum when compared to the distributions of the other hadrons in the same momentum bin. It is important to note that, this plot only includes the ~ 5 % of neutrons that created clusters able to be reconstructed. From this evidence, two conclusions can be drawn. First, the energy deposits which did not pass the criteria to become ECLClusters, were created by interactions of the

Further dedicated studies are needed to investigate the details of neutron interactions in Belle II between the GEANT4 simulation and before ECL cluster formation. Related studies are currently underway to improve the clustering algorithm [25] and to investigate the impact of different GEANT4 physics lists to the agreement of data with respect to the simulation for the ECL variables of neutral hadrons.

Given the calculated interaction rates, it is interesting to see the efficiencies of different selections for each particle. In Fig. 4.5, combinations of a cluster energy selection $E > 0.1$ GeV, an MC matching requirement for the clusters and the base selection $N_{\texttt{ECLClusters}} \geq 1$ are presented.

The cluster energy threshold was selected for the same reason as the lower limit of the low momentum bin, which was to filter a significant amount of beam-induced backgrounds. Another important selection to investigate is the requirement of an ECLCluster to be matched to a generated particle. Specifically, this selection requires an ECLCluster matched to the primary generated particle by the ParticleGun module. Both of the above selections only make sense in the context that there is at least one reconstructed ECLCluster in the event, which is why in Fig. 4.5 (a), (b) and (c) the efficiencies are shown conditionally to the $N_{\texttt{ECLClusters}} \geq 1$ selection. In sub-figure (d), the total efficiency is presented. All selections above are made on the event level. For example, $N(E > 0.1\,\text{GeV})$ is to be interpreted as the number of events with at least one ECLCluster having deposited at least 0.1 GeV in the ECL, and similarly for the other selections.

It is immediately noticeable that there are no MC matched events for neutron in the low momentum bin. As explained in Section 3.2.7, there is an energy threshold for a cluster to be MC matched to a particle, which the deposits from low momentum neutrons cannot overcome, causing the matching process to fail.

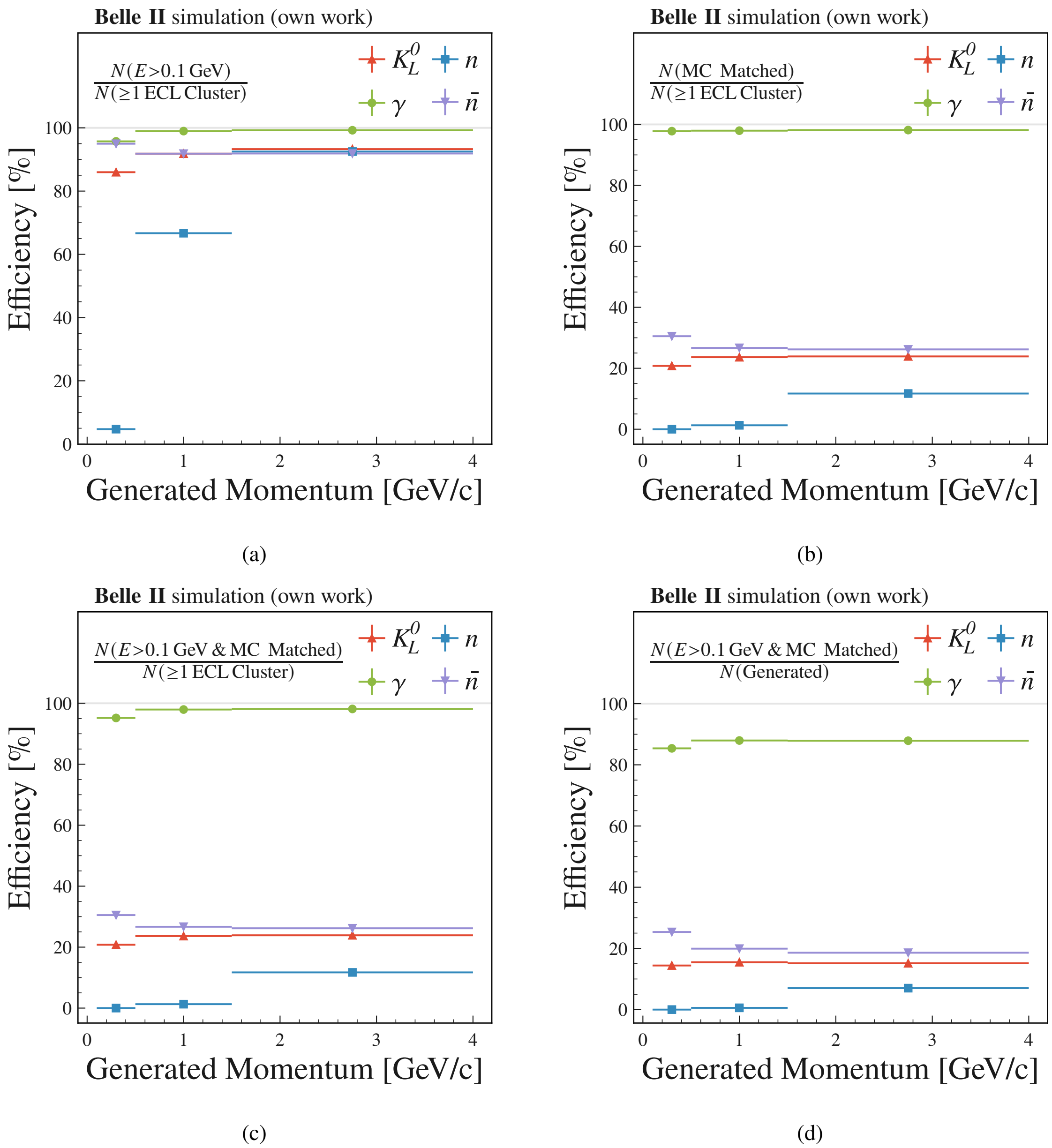


Figure 4.5: ECL interaction rates with different selections for each neutral particle as a function of the momentum bin. The efficiencies shown are (a) $\epsilon(E > 0.1\,\mathrm{GeV}| \geq 1\,\mathrm{ECL\ Cluster})$, (b) $\epsilon(\mathrm{MC\ Matching}| \geq 1\,\mathrm{ECL\ Cluster})$, (c) $\epsilon(E > 0.1\,\mathrm{GeV}\,\&\,\mathrm{MC\ Matching}| \geq 1\,\mathrm{ECL\ Cluster})$ and (d) $\epsilon(E > 0.1\,\mathrm{GeV}\,\&\,\mathrm{MC\ Matching}\,\& \geq 1\,\mathrm{ECL\ Cluster})$. The horizontal error bars are used to show the momentum bin ranges. The vertical error bars are not visible at this scale.

CHAPTER 5

Particle Abundance Studies

Studying the number of different particle species produced in generic Belle II collisions allows generic algorithms that operate on a per-particle basis, such as those for Tracking and PID, to make statements about their expected performance when applied to a real dataset. In this chapter, the relative abundances of charged and neutral particles in generic B meson pair production are presented. After a discussion on the sample generation process, the charged particle abundance will be compared to that of the Belle II Tracking paper [21] as a validation the calculation process. Then, the neutral particle abundances will be presented.

5.1 Sample Generation

A generic MC sample with 1 M $\Upsilon(4S)$ events to B meson pairs is generated with a 51.4 % probability for $\Upsilon(4S) \rightarrow B^0\overline{B}^0$ and a 48.6 % probability for $\Upsilon(4S) \rightarrow B^+B^-$. Since the particle abundance calculation needs the total number of particles produced for each species, the counting happens at the generator level and thus no simulation and reconstruction steps are included. The following particles are considered stable final state particles in the context of the Belle II experiment: γ, $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, K^0_L, p, $\overline{p}$, n and $\overline{n}$.

5.2 Charged Particle Abundance

The term *Charged Particle Abundance* refers to the frequency of production of each stable charged particle with respect to the total number of produced stable charged particles produced in a given sample.

An attempt to reproduce the particle abundances shown in Fig. 5.1 is made. The results are presented in Table 5.1, and are in good agreement, with a maximum underestimation of 1.5 % for the $\pi^{\pm}$ abundance.

To further validate this calculation, a comparison to the publication's generated p_{T} distribution for each particle is presented in Fig. 5.2. The general shapes of the distributions match to a satisfactory degree. The distributions in both plots are smoothed using kernel density estimation (KDE), which is one source of discrepancy between the plots.

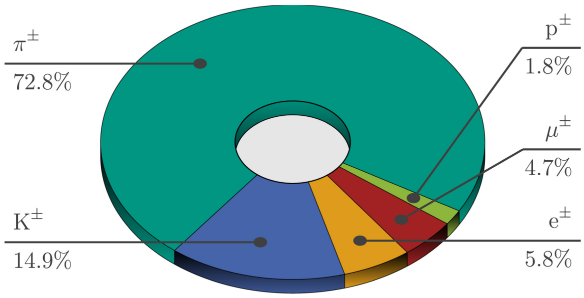


Figure 5.1: Charged particle abundance for generated p_T between 0.01–3 GeV. From Ref. [21].

Table 5.1: Charged particle abundance for generated p_T between 0.01–3 GeV, comparing this work to the reference paper.

Particle	Relative Abundance [%]	
	This work	Tracking paper
$e^\pm$	7.0	5.8
$\mu^\pm$	5.0	4.7
$\pi^\pm$	71.3	72.8
$K^\pm$	14.6	14.9
$p/\bar{p}$	2.1	1.8

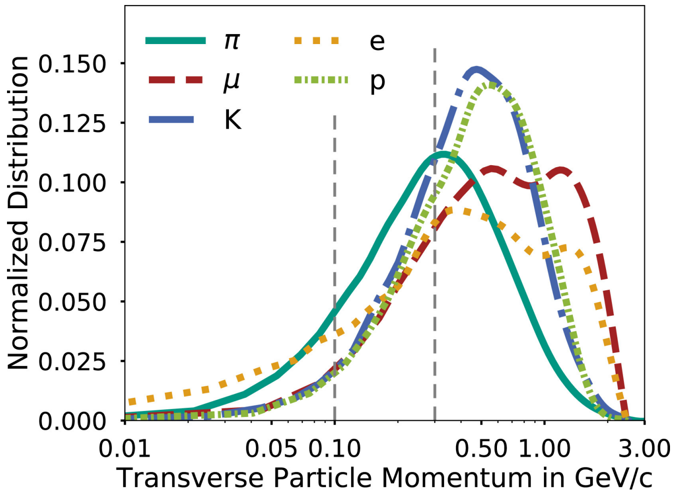
Given that the only information provided in the paper regarding the particle abundance calculation method is the generated p_T range and that they consider only $\Upsilon(4S)$ decays, and that since the time of publication there have been updates in the decay channels and decay rates in `basf2`, the agreement of this work's results to that of the reference is considered acceptable, and thus the same method is applicable to studying the neutral particle abundances.

5.3 Neutral Particle Abundance

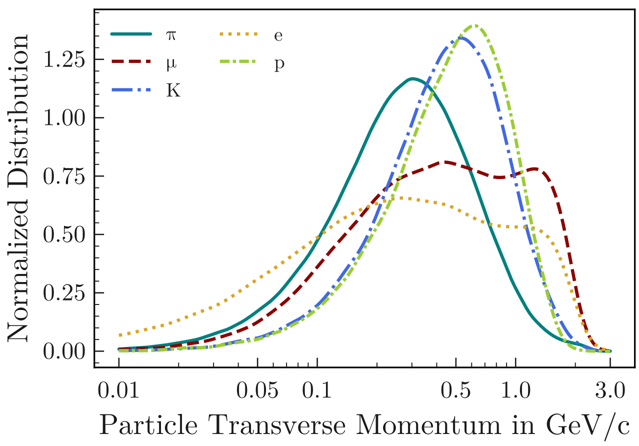
Similar to the previous section, the term *Neutral Particle Abundance* refers to the frequency of production of each stable neutral particle with respect to the total number of produced stable neutral particles produced in a given sample.

In order for these results to be connected to those of the previous section, as well as to be used in the following sections, the neutral particle abundance is calculated for different p_T and p ranges. Fig. 5.3 shows the normalized transverse momentum distributions of the produced neutral hadrons. It can be seen that all neutral hadrons are produces with essentially the same transverse momentum distribution, with a typical range from 0.1 to 2 GeV/c. On the other hand, the largest fraction of produced photons in B meson pair production is found to be produced with significantly lower transverse momenta, with its critical mass being in the range from 0.03 to 0.5 GeV/c.

Table 5.2 shows the relative abundance fractions of the neutral hadrons. More than 90 % of produced neutral particles are photons, while 5 % is K_L^0 mesons, and only ~ 1 % of the neutral particles produced in typical $\Upsilon(4S)$ decays are neutrons or anti-neutrons. As shown in Table 5.3, with the caveat that the



(a)



(b)

Figure 5.2: Charged particle generated p_{T} distribution from (a) Ref. [21] and (b) this work.

numbers reported are for generated p as opposed to generated p_T, the relative fraction of photons to K_L^0 decreases when moving to higher momentum bins, ultimately becoming almost equal for the high momentum bin. On the other hand, the ratio of neutrons or anti-neutrons with respect to the produced K_L^0 mesons, is approximately constant around 20 % (adding neutrons and anti-neutrons).

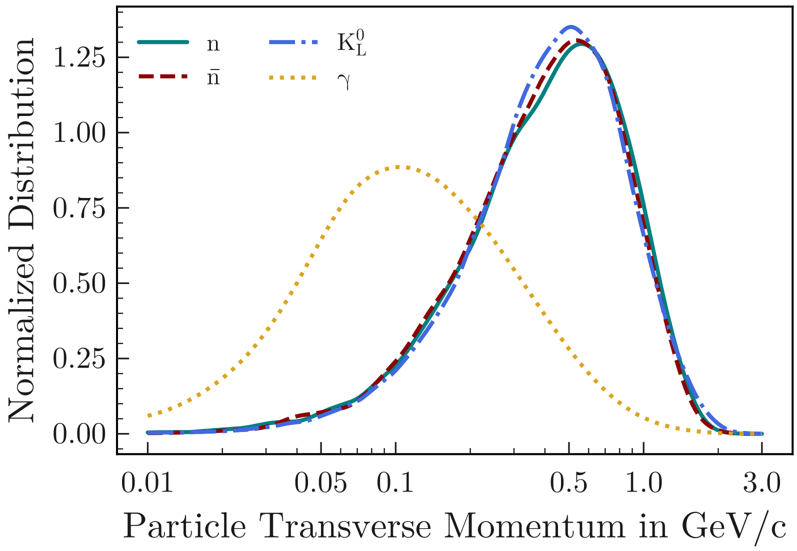


Figure 5.3: Neutral particle generated p_T distribution.

Table 5.2: Neutral particle abundance for generated p_T between 0.01–3 GeV.

Particle	Relative abundance [%]
n	0.6
$\bar{n}$	0.6
K_L^0	5.3
γ	93.5

Table 5.3: Neutral particle abundance for different generated p ranges.

Particle	Relative Abundance [%]			
	p = 0.01–0.1 GeV,	0.1–0.5 GeV,	0.5–1.5 GeV,	1.5–4.0 GeV
n	0.01	0.3	3.0	3.5
$\bar{n}$	0.01	0.3	3.0	3.5
K_L^0	0.15	3.4	25.0	44.0
γ	99.85	96.0	69.0	49.0

CHAPTER 6

BDT based neutron identification

Having studied the interaction properties of stable neutral particles in Belle II, their relative abundance, as well as their ECL variables, a classification algorithm to identify clusters originating from a neutron can be developed.

The amount of different variables and the large overlap of their distributions make this task highly non-trivial. To overcome this, Machine Learning (ML) methods can be used. In this prototype study, the Boosted Decision Tree (BDT) algorithms provided by the `XGBoost` (eXtreme Gradient Boosting – XGB) package [29] are used for the classification tasks.

6.1 Boosted Decision Trees

BDT algorithms are one of the most popular algorithms for classification tasks in HEP and beyond. An overview of the core elements of BDTs is given below, adapted from Ref. [30], starting with their building blocks, Decision Trees. The objective is to train one classifier for each generated momentum bin.

Decision Trees (DTs) consist of splitting nodes and output leafs, arranged in a rooted binary tree structure. Each splitting node contains a threshold for a specific feature. During training, an optimal threshold for an optimal feature is determined. The leaf outputs are also determined during training and can take different forms depending on the task. For (binary) classification tasks, they usually output a number between 0 and 1, while for regression tasks the output is the prediction for the function to be approximated. In the following, a binary classification task is assumed.

To train a DT, the following parameters must be provided:

- a *loss function*, which gives a measure of impurity for each split and acts as a local (for the node) optimization objective,
- an *objective function*, which is used to calculate the leaf outputs and acts as a global (for the tree) optimization objective,
- a *stopping condition*, which can be a maximum depth, a minimum number of candidates in the leaf node, insufficient improvement of the loss function, or perfect separation.

In addition, the analyst has to determine the set of input features over which the classifier will be trained, as well as how to normalize the number signal class entries to the number of background class entries. For the latter, the most common selection is to assign class weights such that the sum of weights across all classes is equal, leading to equal purity for each class in the root node.

The training of a DT is a recursive process, which creates the tree. A training sample must be given as an input, where each entry is a vector of the selected features. At each node, the following steps are executed:

1. If a stopping condition is fulfilled, the node is converted to a leaf and the iteration stops.
2. For each feature, sort all entries in ascending or descending order.
3. For each feature, determine the optimal threshold which creates the best separation between the two classes.
4. Select the feature whose threshold creates the optimal separation across all features.
5. Create two child nodes with the inputs being the entries which pass this nodes optimal threshold and those who do not.
6. On each created node, apply the above steps recursively.

A major advantage of DTs, is that they require little to no input feature preprocessing. This is in contrast to Neutral Networks (NNs) which require the input features to be ucorrelated and regularized, otherwise the redundant feature will prohibit proper convergence, while on the latter case, unregularized input features can significantly bias the training. The DT training algorithm does not use the values of the input features for training. In the case of correlated variables, the DT might choose either feature to assign to a node, which will only impact the computation time if many correlated features are supplied, without significant performance drop.

Even though DTs can be a powerful classification tool, they are exceptionally prone to overfitting the training sample if their stopping conditions are not properly chosen. To adress this major limitation, additional strategies such as *pruning* and *ensemble learning* can be used. One pruning strategy can start with a fully grown deep DT, likely to overfit the training set, and iteratively removes branches until the generalisation error reaches a minimum. Ensemble learning in the context of DTs, is a category of averaging methods, from which many DTs can be combined to create a more robust classifier when compared to a single tree. Some of these methods are *bagging*, *random forests* and *boosting*. BDTs take their name from the third category.

The core idea of the boosting algorithm is to combine many shallow (and thus more error-prone) DTs, called *weak classifiers*, into a new classifier with significantly better performance. The "boosting" part comes from fact that each subsequent tree is trained in a slightly different way which focuses on correctly predicting the cases where the previous trees failed. The original boosting algorithm, AdaBoost [31], acomplished this by assigning a larger weight to the misclassified entries. Gradient Boosting, which is used in this thesis, trains each successive DT to predict the gradients of the loss fucntion with respect to the previous model. In general, for a feature vector x, trained trees T_k and tree weights a_k, the output of a BDT is $F(x) = \sum_{k=1}^{N_{\text{trees}}} a_k T_k(x)$.

The main idea of gradient boosting can be described as follows. Conider that at the k^{th} iteration a classifier $F_k(x)$ is trained to predict a (truth) value y. The next step will be to add a new classifier $h_k(x)$, trained to predict the residual $y - F_k(x)$. Thus, the next classifier will be $F_{k+1}(x) = F_k(x) + \eta h_k(x)$. Here, the parameter η is the *learing rate* which controls the amount each successive correction contributes to the sum. The smaller it is, the finer the applied corrections will be. This can help controlling overfitting, but requires more iterations during training. The optimization objective is a loss function $L(x, y)$ which needs to be minimized. With careful selection of the loss function, the minimization of the global loss $J = \sum_i L(x_i, y_i)$ can lead to

$$\frac{\partial J}{\partial F_k(x_i)} = \frac{\partial L(x_i, y_i)}{\partial F_k(x_i)} = F_k(x_i) - y_i = -h_k(x_i) \tag{6.6}$$

For regression tasks, the most common choice is the *mean square error* loss $L_{\mathrm{MSE}}(x, y) = \frac{1}{2}\left(F_k(x) - y\right)^2$, while for classification tasks a common choice, and the one used in this thesis is the *binary cross entropy* loss

$$L(x, y) = -\left[y \log p(x) + (1 - y) \log\left(1 - p(x)\right)\right], \tag{6.7}$$

where $p(x) = \sigma\left(F_k(x)\right)$, $\sigma(x)$ is the sigmoid function and y is now the label to be predicted, either 0 or 1. The use of $p(x)$ is needed to normalize the BDT output between 0 and 1.

6.2 Input Features and Preprocessing

In order to train a BDT to distinguish neutrons from other neutral particles, the samples and most of the ECL variables defined in Section 4.1 are used. The distributions of the `clusterTheta` and `clusterPhi` variables are expected to the same among all particle species, which is the case according to Fig. A.1 and Fig. A.2, and are thus not included in the input feature set for the classification.

However, these samples can't effectively be used without some pre-selection. All input clusters are required to fulfill the following criteria:

1. `clusterE` > 0.1 GeV,
2. `clusterHasPulseShapeDiscrimination` $= 1$,
3. Cluster is MC matched to the generated particle, unless it is coming from low momentum neutron sample, for which there is no matching information available.

An energy threshold of 0.1 GeV is applied for the same reason as in the interaction rate studies. Then, all clusters are required to have valid PSD information so that the `clusterPulseShapeDiscriminationMVA` variable can be used. Finally, an MC matching requirement is applied to all clusters to verify that the input clusters are created from the correct particle. As discussed in Section 4.2 this information is not available for the low momentum neutrons. Thus, for this sample all reconstructed clusters satisfying the rest of the selections are considered. Since there is no truth information, it is not possible to determine exactly what other particles take part in the cluster creation, but it is reasonable to assume that a significant fraction of the non-neutron clusters comes from low energy photons. This introduces a bias in the low momentum bin, which must always be kept in mind.

Even though it is possible to generate samples with the same amount of events for each particle, the different interaction rates as well as the different cluster multiplicities per event can alter the number of reconstructed clusters for each particle type. It is thus much simpler to generate samples for each particle, with enough events to make sure that after all selections the surviving number of clusters will be statistically significant. Then, it is standard practice to apply *class weights* such that the sum of signal and background weights is equal. For this classification task, the signal class is defined to be the ECL clusters originating from a neutron, while the background class is defined to be the clusters originating from an anti-neutron, a K_L^0, or a photon. Since the background class consists of three sub-classes, one for each of the $\overline{n}$, K_L^0 and γ, this weighting strategy needs some modification. The recipe with which these non-trivial weights are computed will be referred to as the *weighting scheme*.

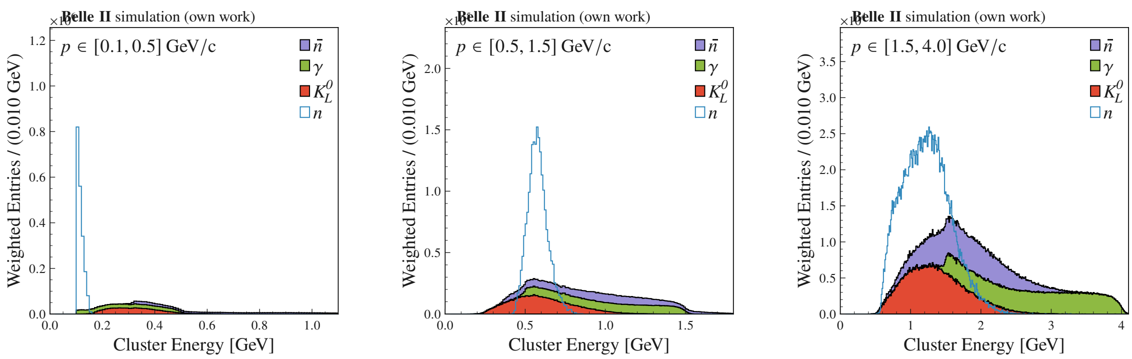
Let n_p, $p \in \{n, \overline{n}, K_L^0, \gamma\}$ be the number of particles in each sample after all pre-selections and p_0 a reference particle, which can be any of the above, but can be assumed to be from the background particles. Also, let $N_{\text{species}} = N^{\text{s}}_{\text{species}} + N^{\text{b}}_{\text{species}}$ be the number of particle species, where $N^{\text{s}}_{\text{species}}$ is the number of signal particle species and $N^{\text{b}}_{\text{species}}$ is the number of background particle species. Then the weight for a signal particle w^{s}_p and the weight for a background particle w^{b}_p are defined as

$$w^{\text{s}}_p = \frac{N^{\text{b}}_{\text{species}}}{N^{\text{s}}_{\text{species}}} \frac{n_{p_0}}{n_p} \quad \text{and} \tag{6.8}$$

$$w^{\text{b}}_p = \frac{n_{p_0}}{n_p}. \tag{6.9}$$

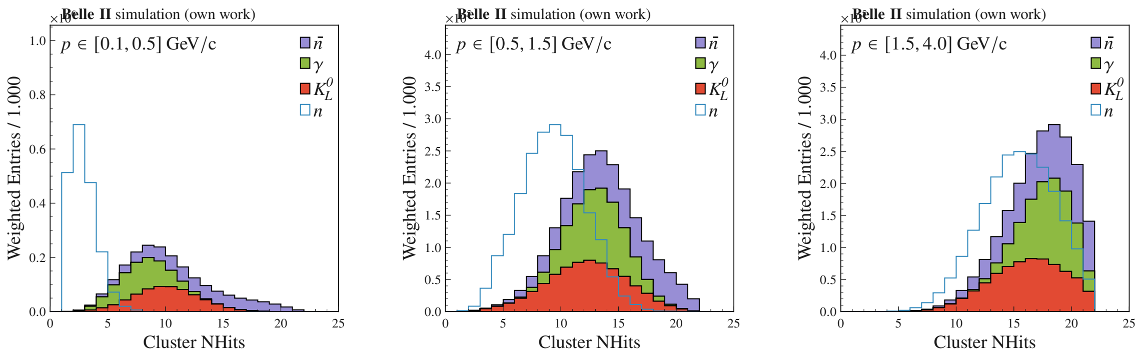
The weighting scheme described above can be called the *nominal weighting scheme* in the context of these studies. It is trivial to see that the weight of the reference particle is $w^{\text{b}}_{p_0} = 1$ and that the relative weight of the signal class to the background class is $\sum w^{\text{s}}_p / \sum w^{\text{b}}_p = \frac{N^{\text{b}}_{\text{species}}}{N^{\text{s}}_{\text{species}}}$ which is the ratio of the number of background particles to the number of signal particles.

Figures 6.1 to 6.3 (and Figures A.8 to A.14) show the input feature distributions in the nominal weighting scheme. The distributions of the background class features are shown as a stacked histogram, while the signal class distributions are shown as the unstacked component. Due to the uniform nature of the `clusterPhi` distribution in Fig. A.9 it can be used to visually verify that the weighting scheme is properly applied, even though the variable itself is not used as an input feature.



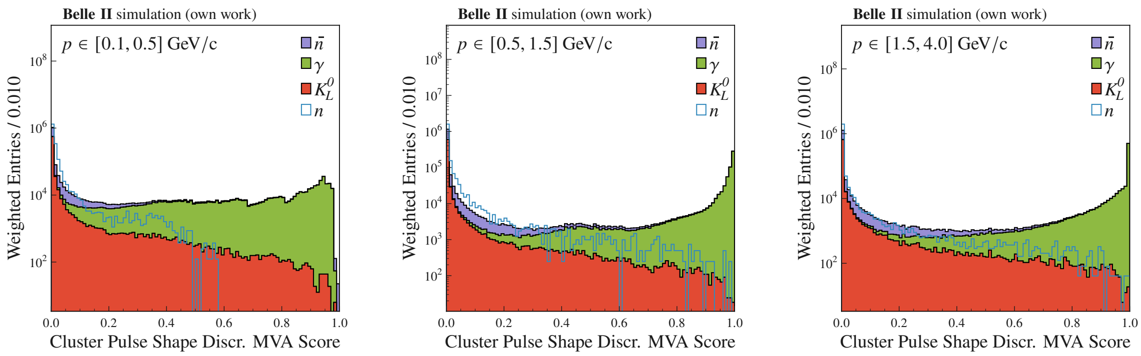
(a) Low (b) Central (c) High

Figure 6.1: Cluster E distributions at low, central, and high momenta.



(a) Low (b) Central (c) High

Figure 6.2: Cluster NHits distributions at low, central, and high momenta.

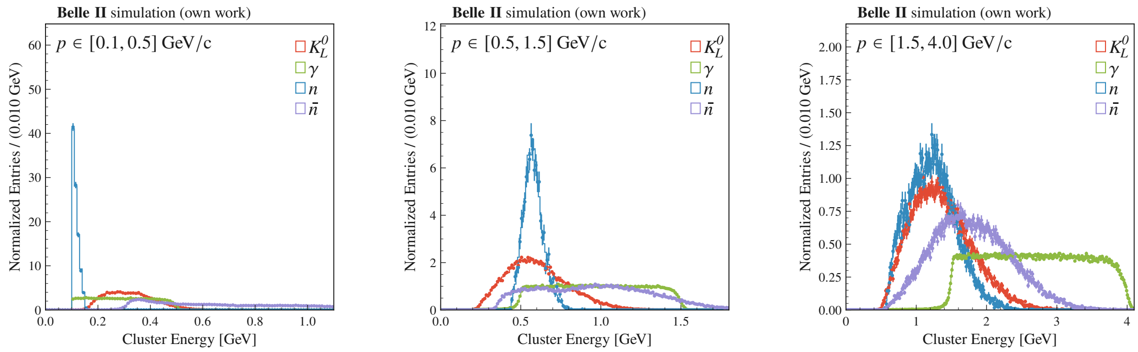


(a) Low (b) Central (c) High

Figure 6.3: Cluster PSD_MVA distributions at low, central, and high momenta.

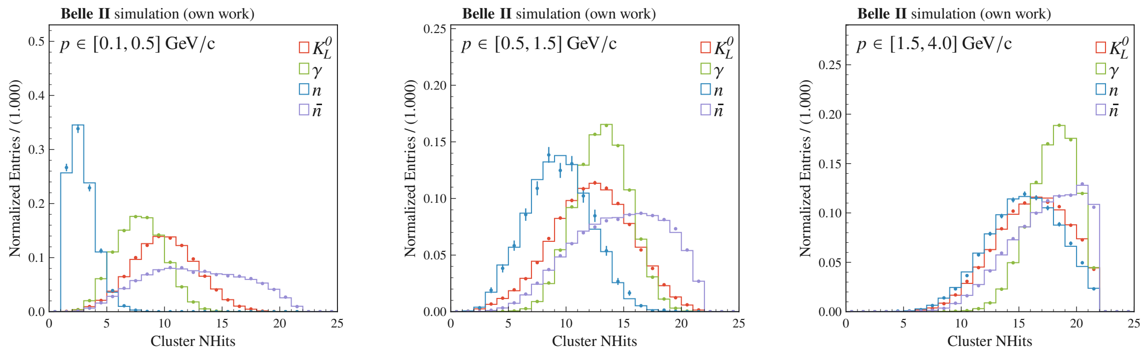
The last preprocessing step is to split the available samples in a training set and an independent testing set, which is standard practice in ML workflows and avoids biasing the evaluation metrics with the data on which the model is trained on. A splitting ratio of 75 % for training and 25 % for testing is used. The splitting happens randomly, to protect against biases regarding the order of the entries in the dataset. In addition, to be robust against the different sample sizes per particle, the splits are *stratified*, meaning that to get the overall training split, N_{species} sub-splits of the same percentage are first computed and are subsequently merged to create the full split. This startegy ensures that the relative frequencies of each class is preserved during splitting, without taking any weights into account, which decouples the weighting and splitting steps.

Figures 6.4 to 6.6 (and Figures A.15 to A.21) show the input feature distributions after all pre-selections have been applied. For each particle type, the histograms show the training set distributions, while the points with error bars show the testing set distributions. The normalized featrure distributions for each split per particle in each momentum range are identical, modulo the expected statistical fluctuations, which validates the splitting method.



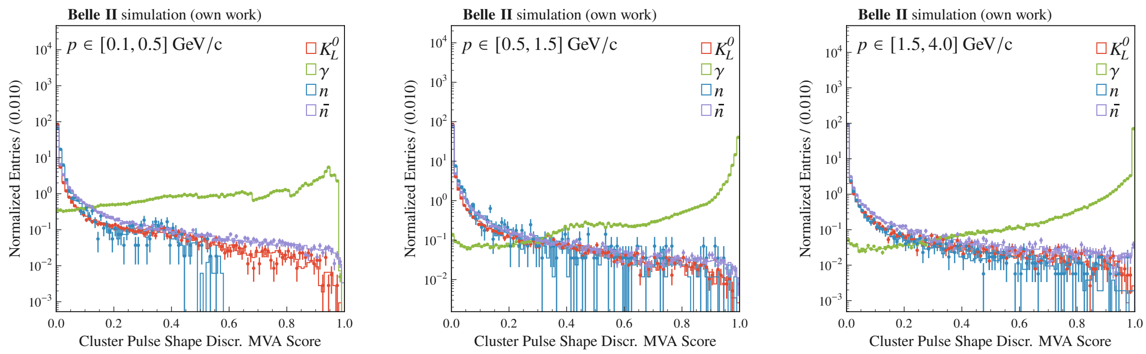
(a) Low (b) Central (c) High

Figure 6.4: Cluster E distributions at low, central, and high momenta.



(a) Low (b) Central (c) High

Figure 6.5: Cluster NHits distributions at low, central, and high momenta.



(a) Low (b) Central (c) High

Figure 6.6: Cluster PSD_MVA distributions at low, central, and high momenta.

6.3 Training the models

One BDT is trained for each generated momentum bin. All BDTs have the same hyperparameter configuration, except for the number of trees, which is selected based on an early stopping criterion, which uses *cross validation*. First, the training set is randomy split into N_{folds} sub-samples, called *folds*, of equal size, in a stratified manner. Then, the following steps are repeated while $N_{\text{trees}} < N_{\text{trees}}^{\text{max}}$, starting with $N_{\text{trees}} = 1$:

1. Train N_{folds} BDTs with the current N_{trees}, where each one is trained on a different set of $N_{\text{folds}} - 1$ folds.
2. Evaluate each BDT on the fold on which it was not trained on, and calculate an evaluation metric.
3. Average the N_{folds} evaluation metrics to get a mean and standard deviation.
4. If $N_{\text{trees}} > N_{\text{early stop}}$, check if the metric has improved with respect to the previous iteration.
    - If there is no improvement, exit the loop.
    - If there is improvement, increase N_{trees} by 1 and go to step (1).

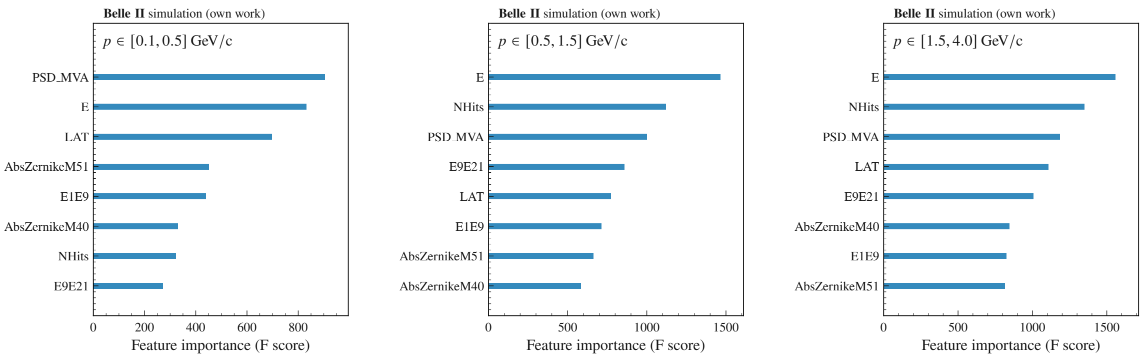
Table 6.1 show the values of the hyperparameters used during training. Besides the already introduced parameters, a value for the subsample ratio and γ have been selected for model *regularization*, which help reduce overfitting. Subsampling is a method where each new tree is trained using only part of the training set. In this case, each new tree is trained on half of the available training set, which when cross validation is considered, equals to $\sim 41.6\,\%$ of the whole training set. Hiding a large part of the training sample in this manner, is a way of reducing the dependency of the BDTs on the variance of the training set, which reduces overfitting. The parameter γ is an artificial threshold for the improvement caused by each split. In practice, it sets a minimum loss reduction that must be achieved for a proposed split to be accepted. Setting it to a moderate value of 0.1 is an attempt at limiting the number of insignificant splits, which only add to the complexity of the overall trees and show no real performance gain. An attempt to optimize these hyperparameters was made using the *Bayesian Hyperparameter Optimization* package `Optuna` [32], but there was no significant performance gain. To keep the worklow simpler and reduce the processing time, no hyperparameter optimization is performed.

Table 6.1: List of selected hyperparameters for BDT training.

Parameter	Value
Learning rate (η)	0.1
Max. tree depth	6
Subsample ratio	0.5
Gamma (γ)	0.1
N_{trees}^{max}	1000
N_{folds}	6
$N_{\text{early stop}}$	15

BDT feature importance

The `xgboost` library provides various methods to estimate how important each input feature is for the BDT. Fig. 6.7 shows the *feature importance by weight* for the trained BDTs in each momentum bin. This method calculates a score for each feature, referred to as *F score* which is analogous to the number of times each feature was used to split a tree node during training. Due to the nature of BDTs, this method is very sensitive to the particular training samples and to correlations between variables. For example, supplementary studies have shown that `clusterNHits` and `clusterLAT` are correlated, which means that the classifier can choose either of the two to use for a node, making the other appear less significant. Taking this limitation into account, it can be said that the most important features for the classification are `clusterE` and `clusterPSD_MVA` which consistently appear in the 3 most important features.



(a) Low (b) Central (c) High

Figure 6.7: Feature importance for each momentum bin, the higher the F score, the more times this feature was used to split a node.

6.4 Model Evaluation

Depending on the analysis objective, there are multiple ways to assess the performance of a trained classifier. In this thesis, the Area Under the Receiver Operating Characterisric (AUROC or AUC) curve is used, along with a custom Figure of Merit (FoM), which attempts to approximate the expected statistical significance which can be achieved if the classifier is to be evaluated on all neutral (i.e. not matched to a track) ECL clusters from generic $\Upsilon(4S) \to B\overline{B}$ decays with an integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 365.3\,\mathrm{fb}^{-1}$.

BDT Output Distribution

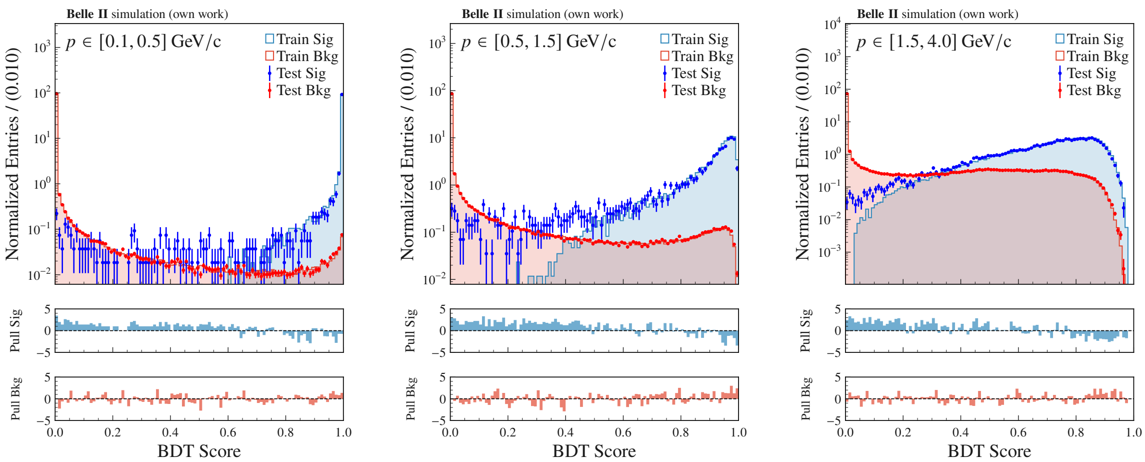
Besides the quantitative metrics such as the AUC and the FoM, an indispencible tool for judging the quality of a trained BDT is the distribution of its output, also called the BDT score distribution, which is defined as the distribution of BDT scores when evaluated on a sample. For a properly trained BDT, this distribution should be identical between the training an testing sets. In addition, the aggreement between training and testing sets should hold also when evaluating the BDT only on the signal or background class.

Fig. 6.8 shows the BDT score distributions for each momentum bin, separately for signal and background, and for the training and testing sample. The evaluation is done without considering the weights for each particle type. Below the score distribution, the *pulls* of the training distribution against the testing distribution are presented for signal and background respectively. The pull of a value $X = x \pm \sigma_x$ with respect to a value $Y = y \pm \sigma_y$ is defined to be

$$\text{Pull} = \frac{x - y}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \tag{6.10}$$

and is also referred to as the *normalized residual* if X is the prediction and Y is the target. Two distributions are considered to statistically aggree with each other, if their pulls are normally distributed [33]. In addition, the pulls as a function of BDT score should show no structure and should appear as random noise.

The distributions show clear separation potential in the low and central momentum bins, while for the high momentum samples the separation is worse. In addition, the effect of overfitting can be clearly seen in the difference between training and testing distributions, espectially towards lower BDT scores. The difference in the distributions becomes larger for lower generated momenta.

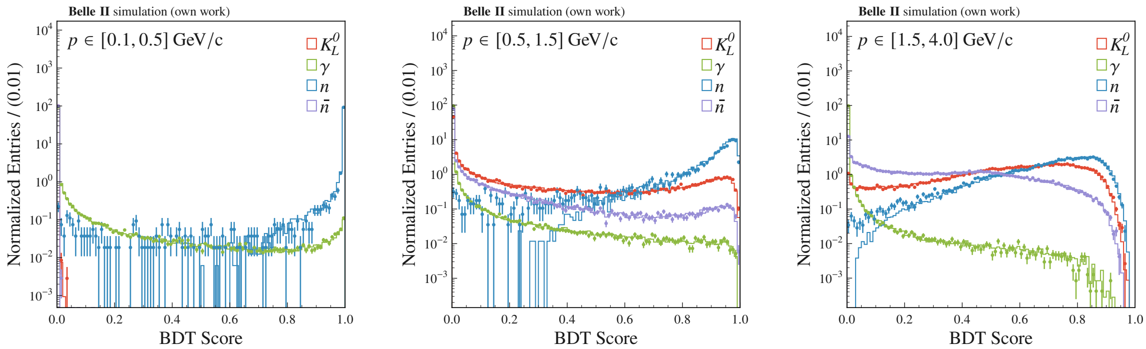


(a) Low (b) Central (c) High

Figure 6.8: BDT score distributions, for each momentum bin, separately for signal and background, for the training and testing samples. Below the score distribution are the pulls of the testing set with respect to the training set, for singal and background. The distributions are normalized to unit area.

Fig. 6.9 shows the same distributions as Fig. 6.8, but instead of evaluating the BDT on the background as a whole, it is evaluated for each particle type separately. This additional splitting of the BDT score can help interpret the overall distributions of Fig. 6.8. The first thing which can be observed is that the agreement between training and testing samples is good for all background particles in all momentum bins.

For the low momentum bin, the classifier maps the $\overline{n}$ and K_L^0 samples to scores near 0, while the photons strongly peak at zero and stretch until 1, where a much smaller signal-like peak is observed. The neutron distribution is strongly peaked at 1 for both the training and testing sets, but for the latter the scores reach all the way to zero. This, in combination with the shape of the photon distribution strongly suggests that a significant part of the neutron sample displays photon-like traits, which make the photon distribution peak at 1 and the neutron distributions reach towards zero. This feature of the low momenum bin can be attributed to the missing truth information for ECL Clusters created by low momentum neutrons. With an updated MC matching algorithm, which would make this information available, it could be possible to train a classifier without this bias.



(a) Low (b) Central (c) High

Figure 6.9: BDT score distributions, for each momentum bin and particle type, separately for signal and background, for the training and testing samples. The distributions are normalized to unit area.

For the central momentum bin, the picture is much different. Still all background distributions are strongly peaked at 0, but all background particles reach towards a BDT score of 1, with the neutral hadrons having small signal-like peaks. In this case, the photon distribution does not have a peaking structure at higher BDT scores, which means that there are no photon-like traits for neutrons in the central momentum bin. Besides its prominent peak near 1, the neutron distribution again extends all the way to 0, but here the cause of overfitting is that the neutral hadron distributions have started to become more similar to the neutron distributions, which inevitably can cause some of the testing set's neutron to be assigned background-like scores.

In the high momentum bin, the photon distributions is again showing pure background-like traits, while the background neutral hadrons are more complicated. The distribution of $\bar{n}$ is strongly peaked at 0, which then transitions to a plateau before starting to fall after a BDT score of ~ 0.6, making the overall shape background-like. On the other hand, the K_L^0 distribution starts with a small peak at 0 and then gains a rising trend creating a broad peak around a score of ~ 0.75. Neutrons also form a much broader peak, at lower BDT score values compared to the shape in the central momentum bin. This means that the classifier is less confident about its predictions with the signal, the features of which have great overlap with the features of the K_L^0 samples in the high momentum bin. Thus, the classifier is expected to have less separation power in this kinematic region.

Receiver Operating Characterisric (ROC)

The ROC curve is a function that connects the classifier's true positive rate (signal efficiency) to the true negative rate (background rejection). For a binary classifier, the action of selecting all entries with a score above some value uniquely determines an associated signal and background efficiency (ϵ_s and ϵ_b). The background rejection is then $1 - \epsilon_b$. Usually, the best classifier is the one that achieves a high signal efficiency and background rejection. Depending on the goal, a high signal efficiency might be desired in the case of rare signal, even though a significant part of the background might be present, while for cases where having large background is prohibiting, an analyst can choose to sacrifice some of the signal.

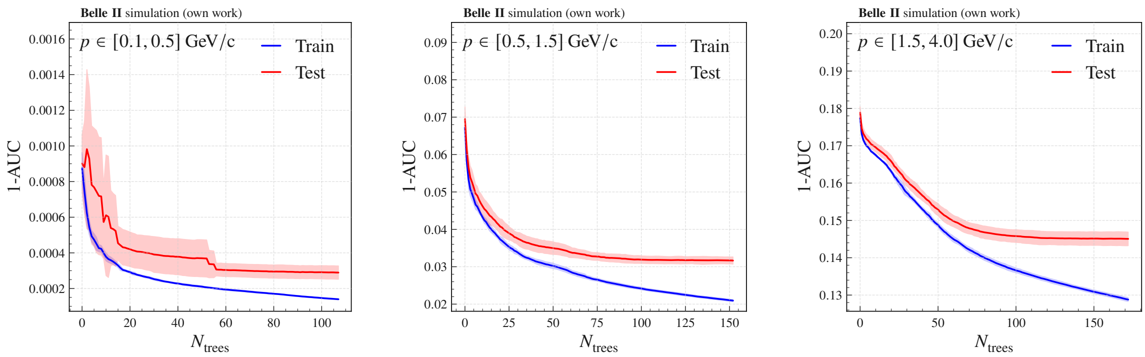
In any case, a universal metric that can quantify the quality of a classifier, is the Area Under the ROC curve (AUC), which for a perfect classifier equals 1 and for random guessing it equals 0.5. It is possible to construct classifiers with AUC < 0.5, but one has to actively penalize correct predictions during training, and is of no use in practice.

In this thesis, the AUC metric is used in the cross validation algorithm described in the previous section, to determine the optimal number of trees, above which there is no additional improvement. For this case, the AUC is evaluated only on the training set. Since BDTs are susceptible to overfitting, one way to visualise how much a model overfits is to comparte the ROC curves and AUC values between the training and the testing sets. If $\mathrm{AUC}_{\mathrm{train}} < \mathrm{AUC}_{\mathrm{test}}$, then the classifier overfits the training set, while if the inverse is true, the classifier is said to underfit. In the case where $\mathrm{AUC}_{\mathrm{train}} \approx \mathrm{AUC}_{\mathrm{test}}$, it cannot be said that there is no overfitting or underfitting present, but such a scenario is likely. Fig. 6.10 shows $1 - \mathrm{AUC}$ as a function of N_{trees}. As expected, the AUC is always approaching 1 for the training set, while for the testing sample it reaches a plateau. When the plateau is reached, no more improvement of the AUC can be achieved by increasing the N_{trees} for the testing sample. In the central and high momentum bins, there is a significant differenec in the final AUC values for the training and testing samples, which cannot be attributed to statistical fluctuation, as is the case with the low momentum bin. Even though this difference, which can quantify overfitting, is significant for almost all but the smallest values of N_{trees}, it is beneficial to continue adding trees until $1 - \mathrm{AUC}$ reaches a plateau or starts increasing. At that point, there is usually a finite difference in the AUC metric between testing and training.

Fig. 6.11 shows the final ROC curves after optimizing the number of trees. While the difference in AUC score is around 1 to 2 % for all momenta, the AUC value decreases for higher momenta. This means that the amount of overfitting is approximately constant and the separation power decreases for higher momenta. The optimal number of trees for each momentum range is summarized in Table 6.2

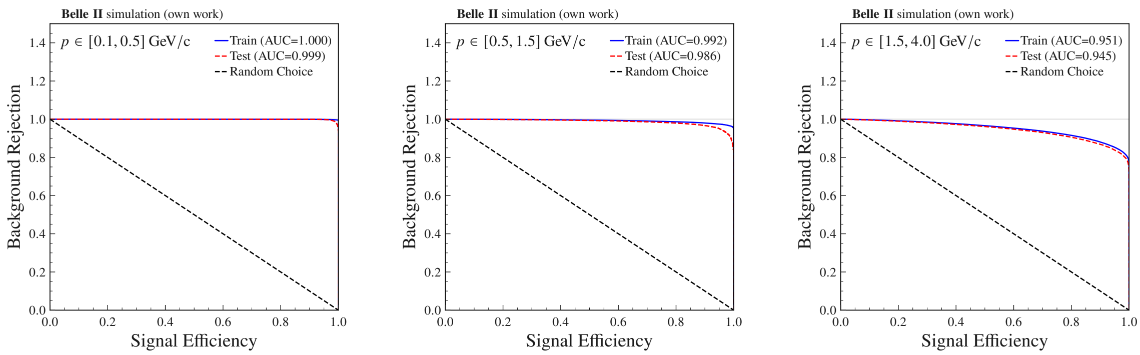
Table 6.2: Optimized number of N_{trees} for each momentum range.

Momentum Bin	Opt. N_{trees}
low	106
central	153
high	172



(a) Low (b) Central (c) High

Figure 6.10: $1 - \text{AUC}$ as a function of N_{trees} for each momentum bin.



(a) Low (b) Central (c) High

Figure 6.11: ROC Curve for each momentum bin.

Figures of Merit

In the context of a counting experiment where B background events are expected and $N > B$ events are observed, the number of potential signal can be estimated by $S = N - B$. The distribution of the obsrved number of events follows a Poisson distribution, which results in the uncertainty of the measurement to be $\sqrt{N} = \sqrt{S+B}$, assuming that the excess is interpreted as signal and not a fluctuation of the background. A way to quantify the relative uncertainty on the existence of the signal is to calculate the ratio $S/\sqrt{S+B}$. Another useful FoM is an approximation of the significance of the $S > 0$ hypothesis with respect to the $S = 0$ null hypothesis, which is $S/\sqrt{B}$. In practice, this approximation breaks down for very low expected backgrounds and $S/\sqrt{S+B}$ is preferred to as a FoM, even though they don not desrcibe the same quantity. If one insists on using the exact signal significance with respect to the null hypothesis the value is $\sqrt{2\left[(S+B)\ln\left(1+\frac{S}{B}\right) - S\right]}$ [34].

The above calculations require knowing (or assuming) the expected number of signal and background events. In the case where this is not known, either because an analysis is still in development or because the signal cross section is unknown, Ref. [35] shows that the signal efficiency with a backdround estimate are enought to approximate the significance.

In the context of this work, FoM $= S/\sqrt{S+B}$ is deemed satisfactory, since what is needed is a function which is maximized at a particular BDT output threshold ($\mathrm{BDT_{cut}}$). To estimate S and B as a function of $\mathrm{BDT_{cut}}$ the following formulas are used:

$$S\left(\mathrm{BDT_{cut}}\right) = N_{\Upsilon(4S)} \times \sum_{p\in\{\mathrm{signal}\}} \frac{n_p\left(\mathrm{BDT_{cut}}\right)}{N_p^{\mathrm{gen}}} a_p \tag{6.11}$$

$$B\left(\mathrm{BDT_{cut}}\right) = N_{\Upsilon(4S)} \times \sum_{p\notin\{\mathrm{signal}\}} \frac{n_p\left(\mathrm{BDT_{cut}}\right)}{N_p^{\mathrm{gen}}} a_p \tag{6.12}$$

where

- $N_{\Upsilon(4S)} = \sigma\left(e^+e^- \to \Upsilon(4S)\right) \times L_{\mathrm{Run\,1}}$ is the number of expected $\Upsilon(4S)$,
- $\sigma\left(e^+e^- \to \Upsilon(4S)\right) = 1.1\,\mathrm{nb}$ is the production cross section of $\Upsilon(4S)$,
- $L_{\mathrm{Run\,1}} = \int \mathcal{L}\mathrm{d}t = 365.3\,\mathrm{fb}^{-1}$ is the luminosity for Belle II Run 1,
- a_p is the abundance of particle p in generic $\Upsilon(4S) \to B\overline{B}$ decays, as calculated in Section 5.3,
- N_p^{gen} is the number of generated `ParticleGun` samples for particle p, and
- $n_p\left(\mathrm{BDT_{cut}}\right) = \int_{\mathrm{BDT_{cut}}}^{1} f_{\mathrm{BDT}}(x)\mathrm{d}x$ is the number of particles p to the right of the respective BDT score distribution $f_{\mathrm{BDT}}(x)$.

For each trained BDT, the optimal working point is considered to be the BDT score threshold which maximizes the FoM summarized above. The FoM, expected signal and expected background distributions are shown in Fig. 6.12. The optimal working point with the associated signal and background efficiencies are presented in Table 6.3.

Table 6.3: Results of neutron classification for each momentum range.

Momentum bin:	Low	Central	High
Optimal BDT threshold	0.999	0.907	0.566
Signal efficiency [%]	81.8	59.5	79.5
Background efficiency [%]	0.4	2.5	33.9
Test AUC [%]	99.9	98.6	94.5

The results of the FoM optimization further solidify the insights drawn from studying the BDT score distributions and the ROC curves. For neutrons with generated momenta between 0.1 to 0.5 GeV, neutrons can be well seperated with $\epsilon_s > 80\,\%$ and $\epsilon_b < 1\,\%$. However, this result is heavily biased from not having MC truth information available for neutrons. For neutrons with generated momenta between 0.5 to 1.5 GeV, the achieved separation can still be considered of a good level, keeping $\sim 60\,\%$ of neutrons and only 2.5 % of background. Finally, neutrons with generated momenta between

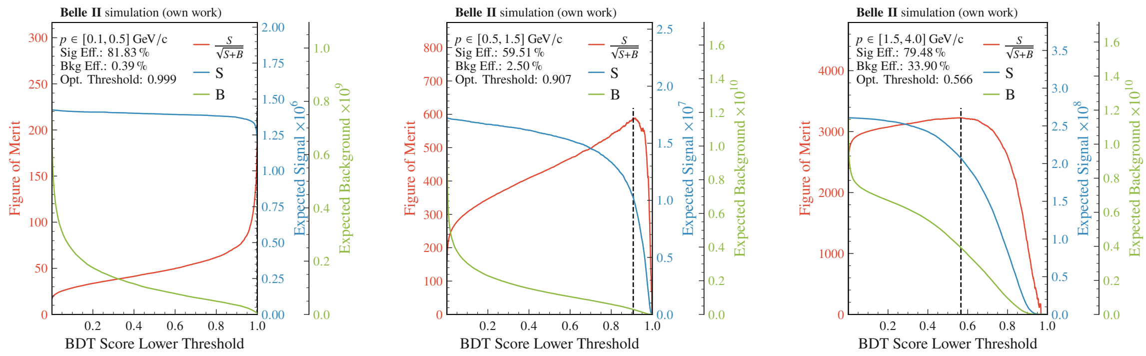


Figure 6.12: FoM for each momentum bin. The estimated values for S and B are calculated according to Equations (6.11) and (6.12).

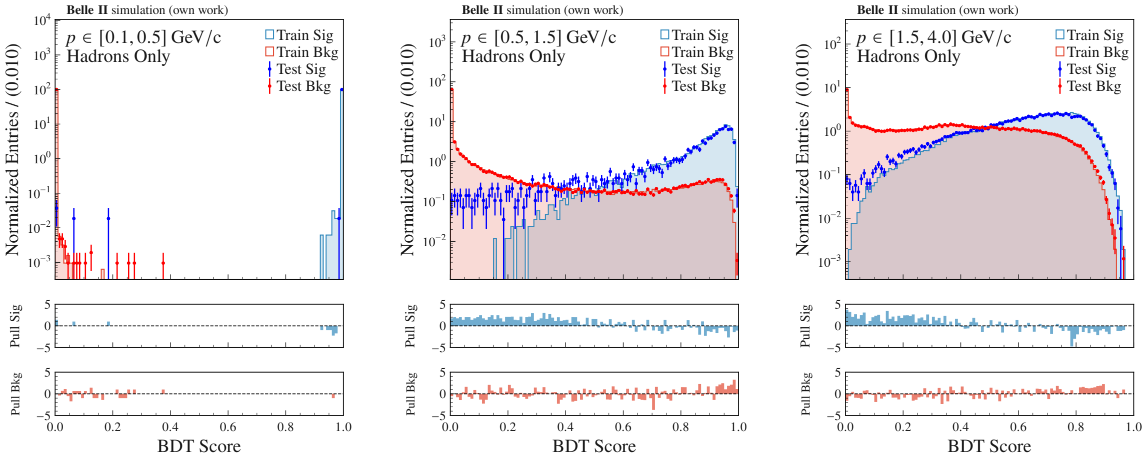
1.5 to 4.0 GeV can be correctly classified with a high efficiency of $\sim$ 80 % at the cost of keeping approximately one third of the background.

6.4.1 Pure Hadronic Classification

Since the PSD_MVA variable (Fig. 6.6) is by definition created to separate neutral clusters from photons, it is interesting to investigate the behavior of the BDTs when trained without this variable and removing the photon sample from the process. This leaves only neutral hadrons ($\overline{n}$ and K_L^0) in the background sample. Figures 6.13 to 6.17 show the respective plots that were discussed in the previous section, for the new training strategy. The main insights and trends remain the same as in the previous classifiers. However, since photons show the best separation with respect to neutrons, their removal is expected to degrade the overall performance of the hadronic classifier.

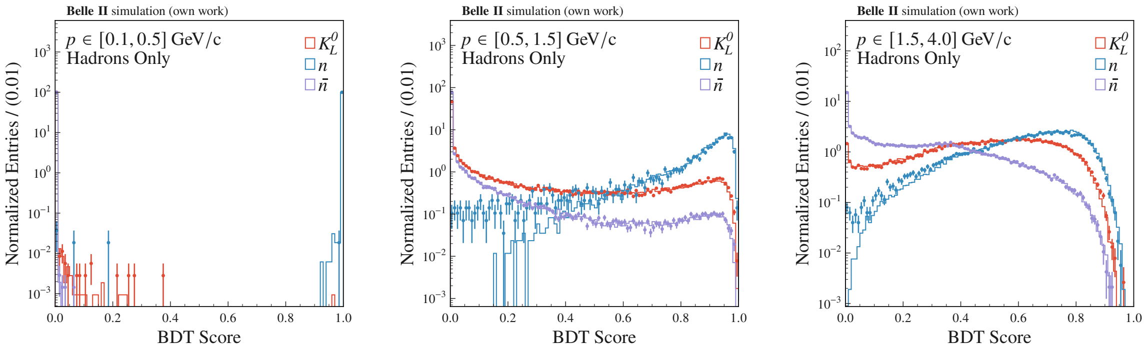
This is indeed the case for the central and high momentum bins, where both the AUC scores and the background efficiencies after FoM optimization are higher than for the nominal classifier, with the central momentum keeping 7.6 % and the high momentum almost 60 % of the background. The signal efficiencies remain quite similar, with a slight increase in the high momentum range of $\sim$ 4 %, at the cost of much higher background.

In the low momentum bin, an almost perfect separation is achieved. There are two components that can contribute to this result. The first is that low momentum neutrons have different interaction properties with respect to the other neutral hadrons. In addition, due to the lack of truth information, it is possible that the classifier leverages the photon-like component of the clusters contained in the sample and, at least partly, acts as a photon classifier with respect to $\overline{n}$ and K_L^0. A combination of these two components is likely to take place, as if only one of the two is the case, the classifier would not be that sensitive to the removal of photons. A comparison of the results of the nominal classifier with those of the hadronic classifier is presented in Table 6.4.



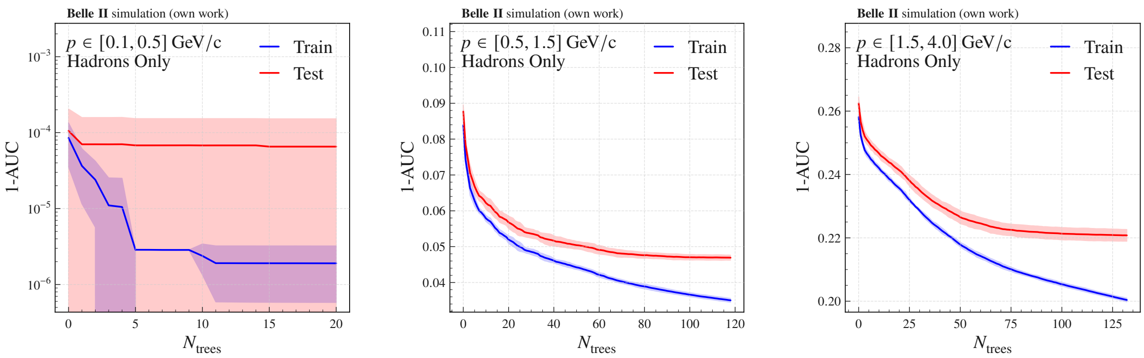
(a) Low (b) Central (c) High

Figure 6.13: BDT scores, hadrons only.



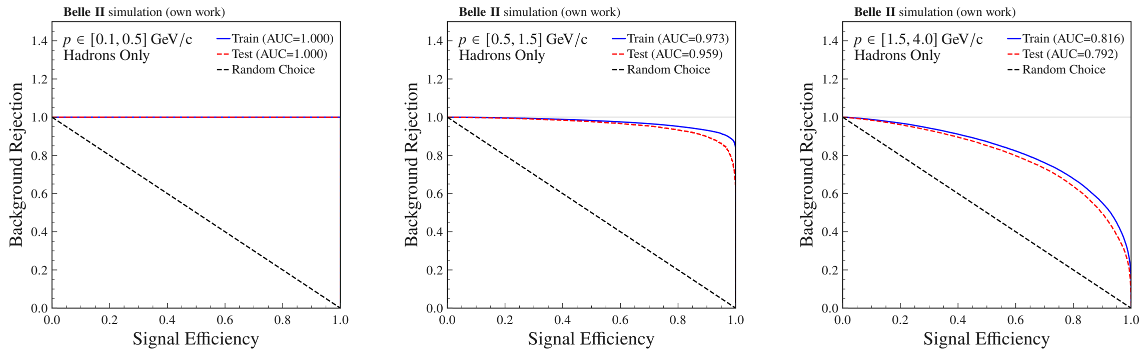
(a) Low (b) Central (c) High

Figure 6.14: BDT scores for each particle species, hadrons only.



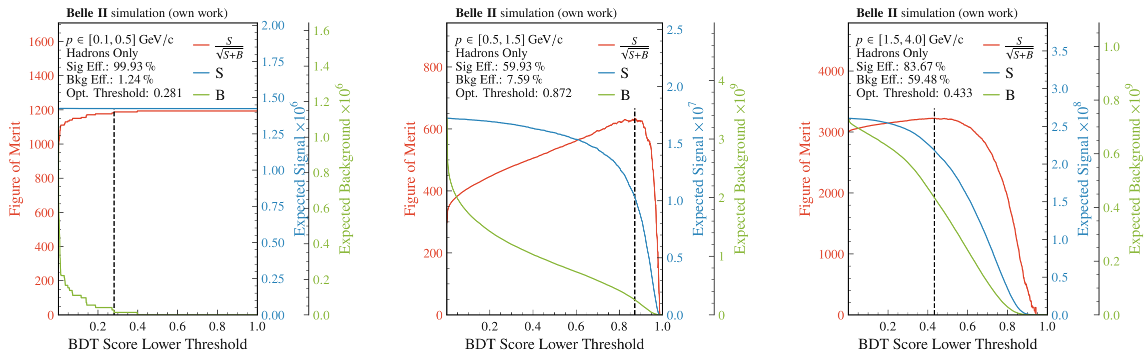
(a) Low (b) Central (c) High

Figure 6.15: $1 - \text{AUC}$ as a function of N_{trees} for each momentum bin, hadrons only.



(a) Low (b) Central (c) High

Figure 6.16: ROC Curve for each momentum bin, hadrons only.



(a) Low (b) Central (c) High

Figure 6.17: FoM for each momentum bin, considering only hadrons. The estimated values for S and B are calculated according to Equations (6.11) and (6.12).

Table 6.4: Comparion of the nominal and hadronic classifier results for each momentum bin.

Momentum bin	Low		Central		High	
Classifier	Nominal	Hadronic	Nominal	Hadronic	Nominal	Hadronic
Optimal BDT threshold	0.999	0.281	0.907	0.872	0.566	0.433
Signal efficiency [%]	81.8	99.93	59.5	59.9	79.5	83.7
Background efficiency [%]	0.4	1.24	2.5	7.6	33.9	59.5
Test AUC [%]	99.9	>99.95	98.6	95.9	94.5	79.2

CHAPTER 7

Neutron reconstruction in $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$

Measuring the performance of PID algorithms involves evaluating them on real data. For this to be possible, clean physics processes that can isolate the signature of the particle in question are used. For example, to evaluate the PID efficiency for (anti-)protons, the $\Lambda \rightarrow p\pi^-$ ($\overline{\Lambda} \rightarrow \overline{p}\pi^+$) decays are used [27].

The detector signatures of Λ baryons have very interesting propeties. They are short-lived neutral baryons with lifetimes that allow them to decay inside the CDC tracking volume producing two charged particles. Experimentally, this creates two oppositely charged particle tracks with a common origin displaced with respect to the interaction point. This feature is known as a V0 signature and it also appears in $K^0_S \rightarrow \pi^+\pi^-$ decays. Even though the produced particle tracks cannot be immediately identified as protons or pions, to determine if the mother particle was a Λ or a K^0_S, the fact that the former particle decays to two particles of different mass while the latter decays to two particles of equal mass, creates a kinematic assymetry in the longitudinal momentum components of the two daugthers with respect to the flight direction of the mother particle. This asymmetry is also referred to as Armenteros-Podolanski (AP) asymmetry or AP variable [36] and is calculated by

$$\alpha_{\mathrm{AP}} = \frac{p_L^+ - p_L^-}{p_L^+ + p_L^-}. \tag{7.13}$$

From momentum conservation during the particle decay, the momentum components transverse to the mother flight direction, q_{T} has to be equal in all V0 decays. Fig. 7.1 shows the two dimensional V0 canidate distribution in α_{AP} and q_{T} recorded by the ALICE experiment [37]. Note how each mother particle is associated with a particular region in this two dimensional space. Due to the definition of α_{AP}, the variable is sensitive to the charge of the particle with larger longitudinal momentum, which is also the particle with the higher mass. This allows for a clear separation of Λ and $\overline{\Lambda}$ baryons.

Λ ($\overline{\Lambda}$) baryons decay to $p\,\pi^-$ ($\overline{p}\,\pi^+$) with a probability of $\sim 64\,\%$. The other major decay channel, with a probability of $\sim 36\,\%$ is $\Lambda \rightarrow n\pi^0$ or $\overline{\Lambda} \rightarrow \overline{n}\pi^0$. These two decay channels could provide clean data samples with exactly one neutron or anti-neutron to be used for PID evaluation, provided that reliable reconstruction algorithms that can isolate the (anti-)neutron deposits are developed. In the following sections, a first iteration of such a reconstruction strategy is presented. It is worth mentioning that other channels are also being studied at the moment, such as the direct production of (anti-)neutrons along with a proton, or anti-proton in the case of neutron production, in $e^+\,e^-$

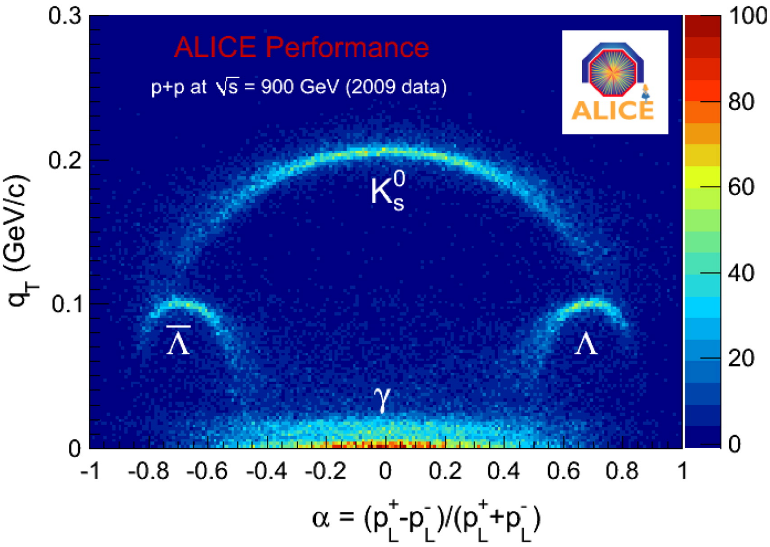


Figure 7.1: Reference Armenteros-Podolanski plot, from Ref [37].

collisions [38].

The physics process which will be analysed in the following sections is the Λ baryon pair production in Belle II, with initial state radiation (ISR): $e^+e^- \to \Lambda\bar{\Lambda}\gamma_{\mathrm{ISR}}$, where $\Lambda \to n\pi^0$ and $\bar{\Lambda} \to \bar{p}\pi^+$ and subsequently $\pi^0 \to \gamma\gamma$. Allowng for the additional γ_{ISR}, has the effect of giving more available phase space to the Λ baryon pair, which would otherwise be produced at the center-of-mass energy of SuperKEKB. More phase space for the Λ baryon pair also corresponds to a broader momentum spectrum for the daughter neutron.

In Section 7.1, the method used to generate MC samples for this process is described. The experimental signature consists of exactly two oppositely charged tracks with invariant mass near $m_\Lambda = m_{\bar{\Lambda}} = 1.116\,\mathrm{GeV}$, two ECL clusters with invariant mass near $m_{\pi^0} = 0.135\,\mathrm{GeV}$, a photon-like ECL cluster (from γ_{ISR}) and one leftover hadron-like ECL cluster which will be attributed to the neutron. However it is not trivial to obtain these reconstructed objects, especially in real Data samples. These nuances are discussed along with a suggested reconstruction strategy in Section 7.2. Finally, **??** contains the results of the reconstruction along with a comparison with simulation.

7.1 Physics Process Generation

To generate simulated samples for $e^+e^- \to \Lambda\left(\to \pi^0 n\right)\bar{\Lambda}\left(\to \bar{p}\pi^+\right)\gamma_{\mathrm{ISR}}$, the `generators` module of `basf2` is used. This module offers an interface to a wide selection of popular MC particle generators, such as `Pythia8` [39] and `EvtGen` [40, 41]. For this study, it is essential to choose a generator that includes precise calculations in the case of initial state radiation. `PHOKHARA` [42–45] is a particle generator which is available in `basf2` and can generate physics processes with a small number of particles in the final state along with ISR or final state radiation (FSR) photons from an e^+e^- initial state. The process $e^+e^- \to \Lambda\left(\to p\pi^-\right)\bar{\Lambda}\left(\to \bar{p}\pi^+\right)\gamma_{\mathrm{ISR}}$ is available for generation using `PHOKHARA>6.0`, but it is not possible to generate the respective process with the (anti-)neutron decay modes neither in PHOKHARA9.1 (used by basf2), nor in the latest version PHOKHARA10.0. In addition, no more updates are planned for the package, so this feature will also not be available in the future.

To get around this limitation, basf2 provides a function that combines PHOKHARA and EvtGen to allow for customized final states with ISR. This is achieved by first generating the $e^+e^- \to \mu^-\mu^+\gamma_{\mathrm{ISR}}$ process, which calculates everything related to the ISR photon production and as well as the four-momentum transfer q^2 of the virtual photon that then decays to the muon pair. The muons are then removed from the output and the virtual photon is with the calculated four-momentum is passed as an input to EvtGen which proceeds to calculate the subsequent decays of the desired final state.

The MC samples used in this study are generated using this strategy. After the PHOKHARA ISR calculation, the virtual photon is then decayed to a $\Lambda\,\overline{\Lambda}$ pair, with $\Lambda \to n\pi^0$ and $\overline{\Lambda} \to \overline{p}\pi^+$. Due to the SuperKEKB beam energy fluctuations, which are by default taken into account when generating MC samples with most generators in basf2, a minimum energy of 110 MeV is required for the produced ISR photon to satisfy the $E_{\mathrm{ISR}} > 0.0098E_{\mathrm{CMS}}$ which is enforced internally by PHOKHARA. The number of generated events used for this study is 1 million.

Finally, basf2's simulation module uses GEANT4 to simulate the interactions of final state particles with Belle II, the output of which is used as input for the reconstruction step.

7.1.1 MC Kinematic Distributions

It is interesting to look at the generation level true kinematic distributions of the signal channel, before simulation through the detector. Fig. 7.2 shows the distributions for the Λ baryon pair, the energy of the ISR photon and the neutron momentum and at the center-of-mass frame. The latter is also presented with different photon energy selections. The photon energy distribution is kinematically liked to the Λ baryon pair invariant mass.

It can be seen that most signal events are produced with the Λ pair gaining most of the available energy, leaving space only for additional photons with $E < 1$ GeV. However, near the production threshold of $2m_\Lambda = 2.232$ GeV the fraction of events increases once again. The change in the distribution of generated neutron momenta with respect to the application of a lower limit on the ISR photon energy shows that until requiring $E_{\mathrm{ISR}} \gtrsim 3$ GeV, only the very high momenta are affected. When this lower threshold becomes more than 4.5 GeV, which is equivalent to approaching the Λ baryon production threshold, events with lower neutron momenta also start to be affected.

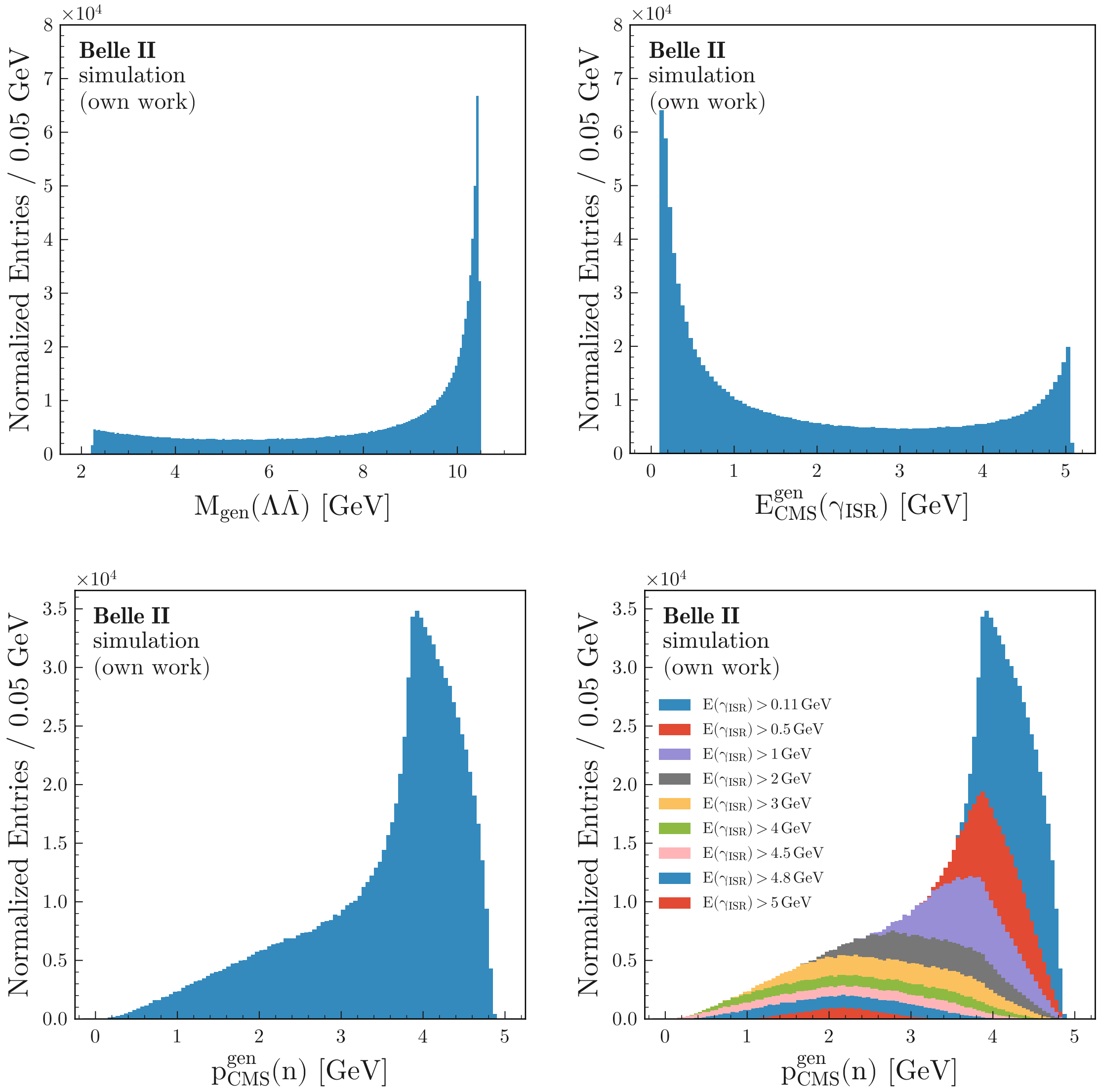


Figure 7.2: Generated Signal MC kinematic distributions in the center-of-mass reference frame.

7.2 Reconstruction Strategy

7.2.1 Data Sample

For each particle phyiscs analysis, the reconstruction strategy should be designed around the specific data sample it is intended to process, which consists of an initial state for the beam and pre-selections applied at the HLT level (referred to as *HLT skims*) and, if processing a skimmed sample, any additional selections that have been applied. Datasets containing only data quality criteria, without any physical pre-selections (*no-skim datasets*) are also available.

The initial state considered for this strategy is the nominal SuperKEKB center-of-mass energy of $\sim$ 10.58 GeV. Using a no-skim dataset ensures that there is no pre-selection that would bias the signal final state. The HLT skims described in Ref. [46] are imposing pre-selections that significantly bias, or select against the signal channel under study. The dataset used in this analysis is an updated reprocessing of the whole Belle II Run 1 period, located under the `/belle/collection/Data/proc16_4S_all_v1` directory on the Belle II distributed computing infrastructure.

7.2.2 Event Selection

Determining appropriate event selection criteria is essential for reducing the data volume for post-processing while retaining a high signal fraction. Identifying the high level characteristics of the desired signal process and imposing tight selections on them is a reliable way of achieving this. In addition, placing looser selections on lower level objects, such as clusters, tracks and their combinations can further increase the signal fraction contained in the output files. These selections are usually contained in the physics skims used for an analysis, but due to the uniqueness of the $e^+e^- \to \Lambda\left(\to \pi^0 n\right)\overline{\Lambda}\left(\to \overline{p}\pi^+\right)\gamma$ process a custom skim has to be created.

A defining characteristic of the chosen physics process is that there are exactly two oppositely charged tracks in the final state. Thus the first requirement for an event to be processed is that $N_{\text{tracks}} = 2$.

Since the ultimate goal is to study neutrons (and not anti-neutrons), it is required that the tracks originate from an $\overline{\Lambda}$ baryon which further needs the positively charged track to be a pion and the negatively charged track to be an anti-proton. Here, charged PID [27] can be applied to require the two tracks to satisfy $\text{PID}_p > 0.1$ and $\text{PID}_\pi > 0.1$ for the negatively and positively charged track respectively. These requirement will be tightened in post-processing, but this looser selection can significantly reduce the $e^+e^- \to e^+e^-$ Bhabha scattering background which is a large fraction of the no-skim dataset. The two tracks are required to be inside the CDC angular acceptance.

The next step in the event reconstruction process is to run a kinematic fit on the two tracks, with a constraint that their combined momentum vector originates from the IP and accept events for which the fit does not fail. For the kinematic fit, the `TreeFitter` [47] algorithm is used. In addition, the combined object, which corresponds to a $\overline{\Lambda}$ candidate is required to have an invariant mass $M(\overline{\Lambda})$ in the range of 1.10 to 1.13 GeV.

The remaining particles needed to be reconstructed are created from `ECLClusters`. All `ECLClusters` are required to have their polar angle inside the CDC acceptance. To protect against beam-induced background clusters, a requirement of `clusterE` > 0.1 GeV is applied in the center-of-mass reference frame.

To reconstruct the π^0 candidate all `ECLCluster` object pairs are considered, provided they satisfy the following criteria:

- `clusterNHits` > 1.5, and
- 0.2967 < `clusterTheta` < 2.6180 rad, and
- any one of
  - `clusterE` > 0.080 GeV if the cluster in the ECL forward region, or
  - `clusterE` > 0.030 GeV if the cluster in the ECL barrel region, or
  - `clusterE` > 0.060 GeV if the cluster in the ECL backward region.
- $0.120 < M(\gamma\gamma) > 0.145$ GeV.

These selections are provided by the Belle II Perfomance Group and were optimized in May 2020 to select π^0 candidates with an efficiency of 40 %. The quoted efficiency does not apply in this analysis since additional energy and polar angle requirements have been applied. In addition to these requirements, each photon pair is required to succeed a kinematic fit with an invariant mass constraint [48] imposing $m_{\pi^0} = 0.135$ GeV, and finally the candidate with the highest kinematic fit χ^2 is saved. A count of the number of π^0 candidates with a probability $P(\chi^2) > 0.1$ is kept to investigate the presense of additional π^0 mesons in the reconstructed sample.

The selected the ISR photon candidate corresponds to the highest energy `ECLCluster` not used to form the selected π^0 candidate, with a minimum energy requirement of `clusterE` > 0.11 GeV in the center-of-mass reference frame, which is needed to match the `PHOKHARA` ISR photon energy threshold.

Finally, a neutron candidate cluster can be selected by saving the `ECLCluster` with the highest `clusterPulseShapeDiscriminationMVA` score from the remaining clusters in the event.

When applying this strategy to the signal MC samples, MC matched truth information for all the aforementioned tracks and clusters is saved, if available.

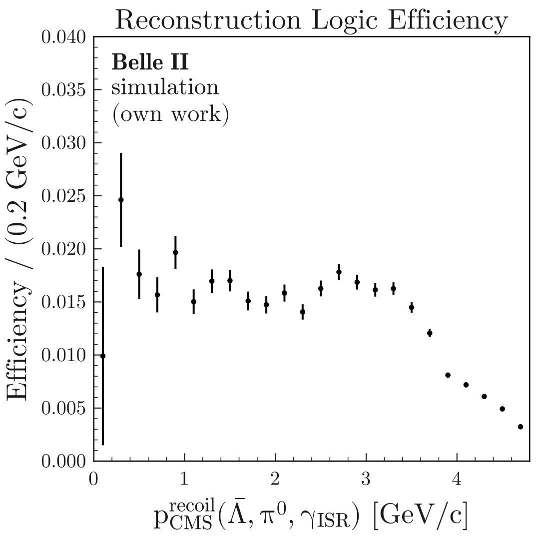
7.2.3 Reconstruction Efficiency

A way of evaluating a specific reconstruction strategy is the calculation of the reconstruction efficiency as a function of one or more variables of interest. Evaluation of the total efficiency is also referred to as *calibration* of the reconstruction strategy in the variables of interest because the efficiency values can be used to estimate the true number of produced events given a sample of reconstructed events.

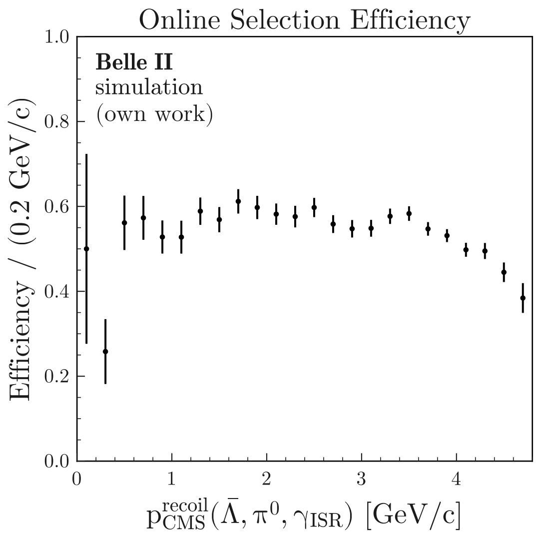
The variable of interest with respect to which the efficiencies for this study are calculated is the recoil momentum of the system of $\overline{\Lambda}$, π^0 and γ_{ISR} against the initial state of the beam in the center-of-mass frame. In the case that these particles are reconstructed correctly and that the only other particle in the event is the missing neutron, this recoil momentum, denoted $p^{\mathrm{recoil}}_{\mathrm{CMS}}(\overline{\Lambda}, \pi^0, \gamma_{\mathrm{ISR}})$, corresponds to the momentum of the neutron in the center-of-mass system.

The first efficiency of interest is the efficiency of the reconstruction logic, without any selections for the reconstructed particles. To evaluate this, the ratio of the generated neutron momentum $p^{\mathrm{gen}}_{\mathrm{CMS}}(n)$ to the reconstructed $p^{\mathrm{recoil}}_{\mathrm{CMS}}(\overline{\Lambda}, \pi^0, \gamma_{\mathrm{ISR}})$ is used and the result is presented in Fig. 7.3(a). It can be seen that even this relatively simple reconstruction logic discards more than 97 % of the signal. As a function of the neutron momentum, the reconstruction efficiency decreases from ~ 2.5 % for low momenta, to less than 0.5 % for the highest neutron momenta.

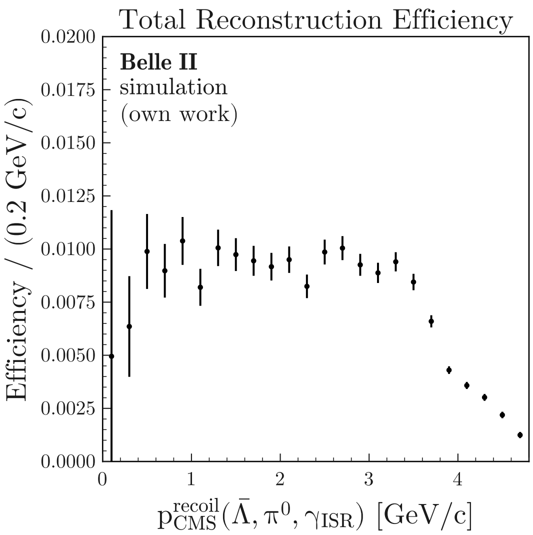
As described in the previous section, the reconstruction strategy is accompanied by a set of selections applied to the reconstructed objects, and the event as a whole. For the latter, the only event-level selection that is applied is that $N_{\text{tracks}} = 2$. The efficiency of this selection has been calculated to be 90 % for the signal MC sample. Taking all these selections into account, which are also referred to as *online selections*, the efficiency of the online selections given the described reconstruction logic is presented in Fig. 7.3(b). In this case, the efficiency is approximately constant across the central recoil momentum bins at just below 60 %. For momenta below 0.5 GeV/c and above 4 GeV/c, the online selection is less efficient.



(a) Reconstruction Logic Efficiency



(b) Online Selections Efficiency



(c) Total Reconstruction Efficiency

Figure 7.3: Reconstruction Efficiency.

Taking all of the reconstruction process into account yields a total reconstuction efficiency of $\sim 1\,\%$ for recoil momenta in the range of 0.5 to 3.5 GeV/c, as shown in Fig. 7.3(c). Considering that the goal of the reconstruction logic and the online selections is to reconstruct as much signal as possible and discard a large part of backgrounds, as well as that `offline selections` will be applied in order to further reduce the backgrounds, the total reconstruction efficiency for the signal, is much lower than what would be ideal.

The reconstruction logic, and not the online selections, has the most impact in the total reconstruction efficiency, which probably stems from the ISR photon selection. In the current implementation, the ISR photon candidate is the most energetic `ECLCluster` in the event, provided it is not used to form the π^0 candidate. This strategy heavily biases the reconstruction towards higher photon energies, even though Fig. 7.2 shows that most of the photons tend to have low energies. However, given that the event should have at least 4 `ECLClusters`, introducing such bias is a reliable way to make sure that the ISR photon is correctly reconstructed and is thus an essential component of the reconstruction strategy.

An additional irreducible bias that heavily affects the efficiency of the reconstruction logic is that when combining photon candidates to form π^0 candidates, `basf2` assumes that the π^0 was produced near the IP, with an uncertainty equivalent to the typical decay volume of B mesons. However, a large fraction Λ baryons is expected to decay outside this volume, creating a π^0 and its daughter photons displaced from where `basf2` is designed to reconstruct them. `ECLClusters` that are displaced in this way end up having incorrect angular variables, which ultimately leads to incorrect $M(\pi^0)$ estimation, causing the mass constrained kinematic fit to fail.

7.2.4 Background Reduction

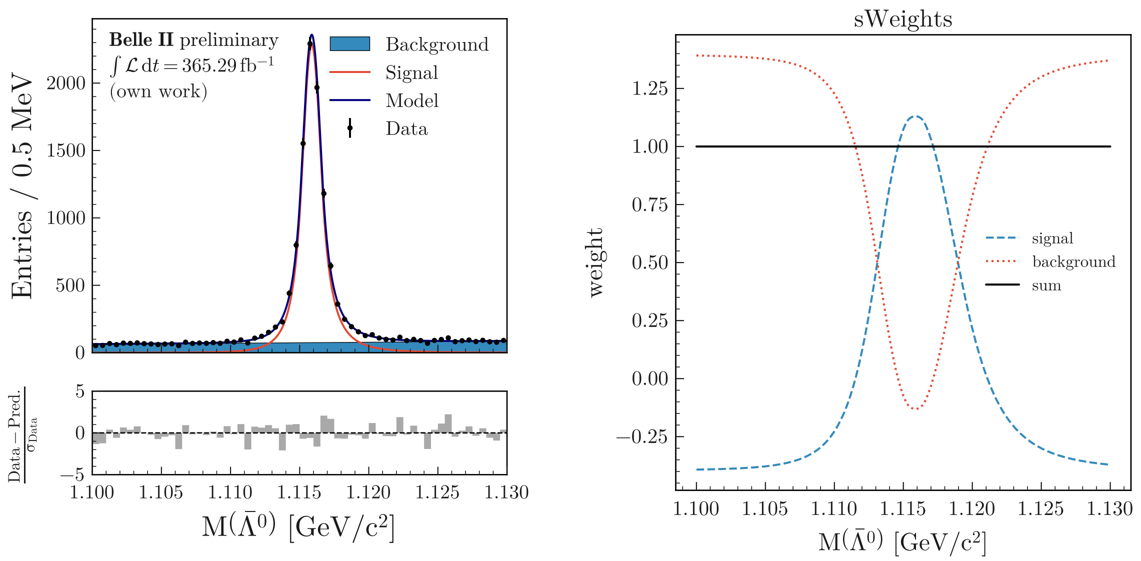
Physics processes that can fully or partially mimic the experimental signatures of the signal process can have multiple origins. One of the most important background processes contained the no-skim data samples, is Bhabha scattering ($e^+e^- \to e^+e^-$) which can also include additional photons. When additional photons are included, or if additional `ECLClusters` are created from beam-induced background processes in association with Bhabha scattering, two tracks and a number of clusters are produced, which could overcome the thresholds of the online reconstruction. To diminish such backgrounds, tight offline selections on the PID variables for the two reconstructed tracks are required. Thus, the negatively charged track is required to have $\mathrm{PID}_p > 0.9$ to increase the probability of reconstructing an $\bar{p}$, while the positively charged track is required to have $\mathrm{PID}_\pi > 0.8$ to increase the probability of reconstructing a π^-.

To reduce any leftover backgrounds of this origin, the sWeights method [49] is used. This method uses the results of a parametric fit to the distribution of a well modelled variable present in the reconstructed data sample to create event weights which can be applied to the data to recover the background reduced distributions of other variables of interest. The variable of interest must not be correlated with the fitted variable in order for this method to work. In what follows, the variables of interest are related to the ISR photon and recoil momentum distributions, which act as proxies for the neutron kinematics. These variables are uncorrelated to the $\bar{\Lambda}$ side of the signal process. Therefore, the invariant mass of the two charged tracks, which is also the invariant mass of the $\bar{\Lambda}$ candidate, $M(\bar{\Lambda})$, can be used as the fitting variable.

Fig. 7.4(a) shows the distribution of this variable for the processed data which includes the tight PID requirements described above. An extended parametric maximum likelihood fit [50] is performed on this distribution. The fitted model (blue line) is comprised of the sum of a double-sided crystal ball (DSCB) distribution, which models the signal component (red line) and an exponential distribution to describe the background (filled blue area). The component distributions are scaled with two additional parameters, N_{signal} and $N_{\text{background}}$ which will be fitted to extract the respective yields. All other parameters are free to float during the fitting process, besides the DSCB mean, which is constrained to be near between $m_{\Lambda} \pm 1$ MeV. The fitting process results in

$$N_{\text{signal}} = 9800 \pm 150 \quad \text{and} \tag{7.14}$$
$$N_{\text{background}} = 4500 \pm 130. \tag{7.15}$$

The extracted sWeights are presented in Fig. 7.4(b) as a function of the $\overline{\Lambda}$ candidate invariant mass. In order to recover the background supressed distributions of the variables of interst, each event in assigned a weight based on the value of $M(\overline{\Lambda})$ corresponding to that event. It is important to note that this background supression only targets background processes that do not contain real $\overline{\Lambda}$ baryons. All events that contain an actual $\overline{\Lambda}$ are given the same treatment as this analysis' signal.



(a) (b)

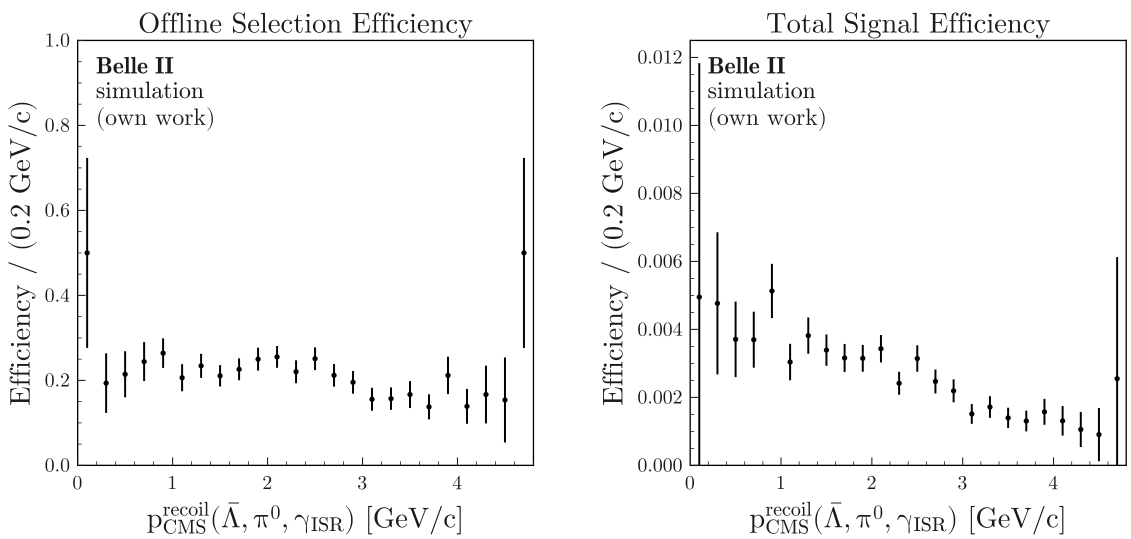
Figure 7.4: (a) $\overline{\Lambda}$ invariant mass fit which is used to extract the sWeights for background supression. (b) The extracted signal an background sWeights.

Applying the sWeights method has constrained the remaining relevant background processes to the form $e^+e^- \rightarrow \Lambda\overline{\Lambda}X$. Baryon number conservation requires one Λ to be present, which when combined with the requirement of the event containing exactly two tracks, has to decay to $n\,\pi^0$. In addition, any number of additional neutral particles is allowed, such as photons, neutral pions or neutral kaons, provided there is enough phase space to produce them. According to Ref. [51], in which a measurement of the cross section for $e^+e^- \rightarrow \Lambda\overline{\Lambda}\gamma$ using the charged Λ baryon decays is performed, the main sources of background with a Λ pair are: $e^+e^- \rightarrow \Lambda\overline{\Lambda}\pi^0$, $e^+e^- \rightarrow \Lambda\overline{\Lambda}\gamma\gamma$ and $e^+e^- \rightarrow \Lambda\overline{\Lambda}\pi^0\gamma$, since producing additional strange particles, such as neutral kaons, has a much lower probability of occuring.

Therefore, additional offline selections are required to restrict the contributions of such processes before examining the variables of interest. First, a requirement for exactly one π^0 candidate with a kinematic fit probability greater than 10 % is imposed. In addition, an upper limit on the number of up to 10 `ECLClusters` in the event is require, in order to restrict the number of additional neutral particles. Ideally, a tighter selection would be used, but it was found that the remaining number of events is too small for the analysis to continue if the threshold is lowered.

In order to reduce the bias that the reconstuction logic introduces to the analysis regarding the choice of the ISR photon, $E_{CMS}(\gamma_{ISR})$ is required to be between 3 to 5.05 GeV. The upper limit is inspired by the $E^{gen}_{CMS}(\gamma_{ISR})$ distribution in Fig. 7.2 and protects against photons with energies not prediced by the simulation. Finally, a loose requirement of `clusterPulseShapeDiscriminationMVA` < 0.5 for the neutron candidate cluster is applied, which biases the selection to include clusters that tend to be hadron-like.

The efficiency of these additional requirements with respect to the total reconstruction efficiency is presented in Fig. 7.5(a). The total efficiency including the reconstruction logic, the online and the offline selections for the signal MC sample, is presented in Fig. 7.5(b). The efficiency of the offline selections is mostly flat as a function of the recoil momentum, with a value of ~ 20 %. The largest efficiency reduction comes from constraining $E_{CMS}(\gamma_{ISR})$. The total efficiency results in keeping less than 0.5 % of the generated signal, which greatly hinders the conclusions that can be drawn from this analysis, and should be a major target for improvement in any future studies of this channel.



(a) Offline Selections Efficiency

(b) Total Signal Efficiency

Figure 7.5: Signal reconstruction efficiencies after all selections.

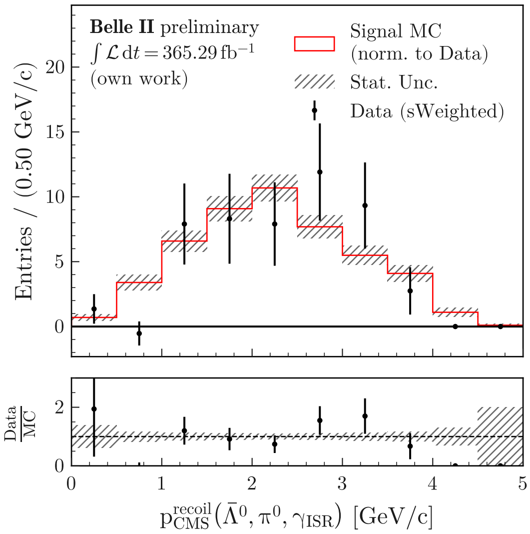
7.2.5 Comparison of Data and MC distributions

Having applied all necessary selections to reduce the backgrounds, it is possible to attempt to compare the processed and sWeighted data with the signal MC simulation. Since the cross section for

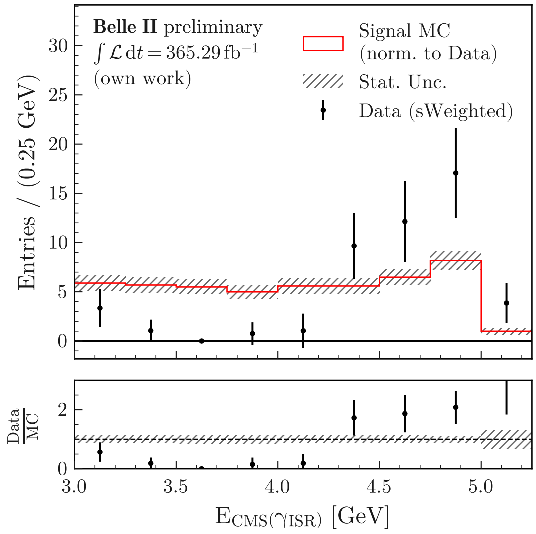
$e^+e^- \to \Lambda\overline{\Lambda}\gamma$ has not been experimentally measured for the phase space targeted by this analysis, it is not possible to scale the signal MC distributions to the expected number of candidates according to the expected cross section and branching fractions. Therefore, the signal MC distribution is normalized to the sum of signal sWeights for the data distribution. This has the consequence that the Data and MC distributions can aggree if and only if all background is removed by the selections (within the uncertainties) or if the signal and background distributions have similar shapes such that a normalization factor can make the total data distribution aggree with the signal MC.

In Fig. 7.6, a comparison between the (sWeighted) data and the signal MC simulation is presented for the recoil momentum, the ISR energy and the recoil mass distributions. The first observeation is that the number of surviving events is so small that definitive conclusions are not possible to be formulated. Taking this into account, it can be seen that the recoil momentum distributions shows good agreement, while for the other two variables the aggreement is significantly poorer. However, the expected increase in event count towards the end of the ISR photon energy distribution is indeed observed, but in simulation the shape of the distribution is less peaked.

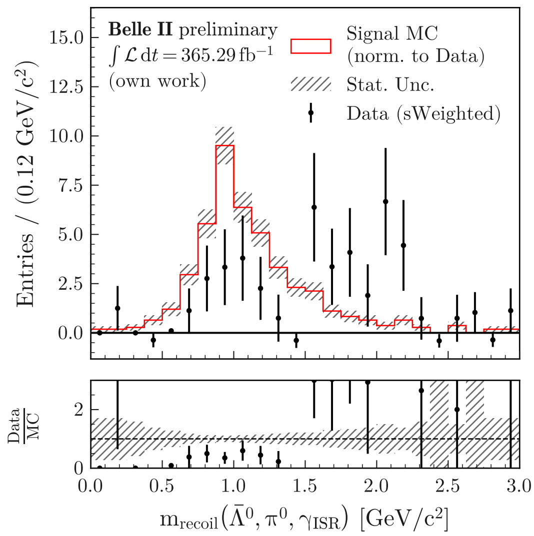
Finally, the recoil mass distribution displays a rather poor aggreement between the data and signal MC samples. The distribution for signal MC shows a clear peaking structure with a maximim slightly below 1 GeV, which is exactly where the real neutron mass of ~ 0.94 GeV is located, albeit with a quite poor resolution of ~ 0.5 GeV. In (sWeighted) data, the distribution shows two distinct features. The first is that between 0.7 and 1.2 GeV, a peaking structure around the neutron mass, with a similar resolution to simulation can be observed, if one ignores the precence of the large statistical uncertainties. In addition, between 1.5 and 2.5 GeV, an unexpected excess of data is observed, which can most likely be interpreted as leftover background events that managed to pass all previous selections.



(a)



(b)



(c)

Figure 7.6: Data MC sWeighted

CHAPTER 8

Discussion and Outlook

In this thesis, several studies around the neutral particles considered stable in Belle II, with an emphasis on neutrons, were presented.

In Chapter 4, the interaction rates of n, $\overline{n}$, K_L^0, and γ were calculated for 3 momentum ranges, using dedicated single particle simulations. The interaction rates were found to be in reasonable agreement with the expectation, when one considers the general interaction properties of these particles with matter, as well as the specific details of the experimental setup of Belle II. During these studies, the lack of MC truth information for low momentum neutrons was discovered.

In Chapter 5, the expected neutral particle abundances in generic $\Upsilon(4S) \rightarrow B\overline{B}$ decays were calculated for the aforementioned neutral particles. The method was also applied to charged particles and a comparison with the "Track Finding at Belle II" publication [21] was made in order to validate the calculation.

Chapter 6 discussed the design and evaluation process of 3 BDT classifiers, one for each of the same momentum ranges for which the interaction rates were calculated. The classifiers were designed to perform binary classification with a neutron signal and a background which is comprised of equal parts $\overline{n}$, K_L^0 and γ. A significant bias in the low momentum classifier is introduced from the lack of MC truth information for low momentum neutrons. An updated cluster MC matching algorithm, such as the one being discussed in Ref. [25], could significantly improve the reliability of the low momentum classifier, and potentially improve the performance of all classifiers. An expected performance for the BDTs is calculated by maximizing a Figure of Merit calculated using the neutral particle abundances found in the previous chapter. For each momentum range, an expected signal and background efficiency, under the hypothesis of a certain amount of $\Upsilon(4S)$ decays, is calculated. It is found that that in higher momenta neutrons are difficult to distinguish from K_L^0 mesons, while for the rest of the particles and momentum ranges neutron classification shows promising results. In future studies, different classifier architectures that can take advantage of the low level structure of the ECL, such as Graph Neural Networks, should be considered. In addition, using information from additional detector systems, like the KLM could improve the overall performance of the classifier.

Lastly, in Chapter 7 a prototype analysis to reconstruct the $e^+e^- \rightarrow \Lambda\left(\rightarrow \pi^0 n\right)\overline{\Lambda}\left(\rightarrow \overline{p}\pi^+\right)\gamma$ process is presented. The results are statistically insignificant given the low overall efficiency of the reconstruction and further selections to reduce the backgrounds. Estimates of the efficiencies at each stage of the analysis were presented in order to identify the cause of low efficiency. The reconstruction logic, without any further selections, is found to have the most negative impact in the overall efficiency on the signal MC sample. In addition, the offline selections imposed in order to reduce the additional backgrounds that can appear for a process with a Λ baryon pair, also have a significant impact in the analysis. Finally, not having access to an experimental measurement for the differential cross section of $e^+e^- \to \Lambda\overline{\Lambda}\gamma$ production prohibits proper scaling of the simulation to the data.

To get around these limitations, measuring the differential cross section of $e^+e^- \to \Lambda\overline{\Lambda}\gamma$ in the decay channel where both Λ baryons decay to charged particles should be prioritised. Conducting this study, the related neutral particle background could be studied and the insights that will be drawn would be highly beneficial to redesigning a reconstruction strategy for the neutron decay mode. Once these improvements are achieved, it will be possible to use this channel to calibrate a neutron classifier.

Bibliography

[1] *LHC Machine*, JINST **3** (2008) S08001, ed. by L. Evans and P. Bryant (cit. on p. 1).

[2] T. Abe et al., *Belle II Technical Design Report*, (2010), arXiv: 1011.0352 [physics.ins-det] (cit. on pp. 1, 9).

[3] W. Altmannshofer et al., *The Belle II Physics Book*, PTEP **2019** (2019) 123C01, ed. by E. Kou and P. Urquijo, [Erratum: PTEP 2020, 029201 (2020)], arXiv: 1808.10567 [hep-ex] (cit. on p. 1).

[4] I. Adachi et al., *Evidence for $B^+ \to K^+ \nu\bar{\nu}$ decays*, Phys. Rev. D **109** (11 2024) 112006, URL: https://link.aps.org/doi/10.1103/PhysRevD.109.112006 (cit. on p. 1).

[5] S. Navas et al., *Review of particle physics*, Phys. Rev. D **110** (2024) 030001 (cit. on pp. 1, 3, 9).

[6] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579; A. Salam, "Weak and Electromagnetic Interactions", *Elementary particle theory. Relativistic groups and analyticity. Proceedings of the Eighth Nobel Symposium*, ed. by N. Svartholm, Stockholm: Almquist & Wiksell, 1968 367; S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, cit. on p. 3.

[7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 3).

[8] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the color octet gluon picture*, Physics Letters B **47**.4 (1973) 365, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/0370269373906254; D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (26 1973) 1343, URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343; H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (26 1973) 1346, URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1346, cit. on p. 3.

[9] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, Wiley, 1984, ISBN: 9780471887416 (cit. on pp. 3, 7).

[10] Particle Data Group, "Passage of Particles Through Matter", *Review of Particle Physics*, ed. by S. N. *et al.*, vol. 110, 3, Section: Passage of Particles Through Matter, 2024 030001, URL: https://pdg.lbl.gov/2024/reviews/rpp2024-rev-passage-particles-matter.pdf (cit. on pp. 4–6).

[11] V. S. Barashenkov and V. M. Maltsev, *Cross Sections for Elementary Particle Interactions*, Fortschritte der Physik **9** (1961) 549 (cit. on pp. 5, 6).

[12] Geant4 Collaboration, *Geant4 Physics Reference Manual (PDF version)*, Accessed: 11th May 2026, CERN, 2024, URL: https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsReferenceManual/fo/PhysicsReferenceManual.pdf (cit. on pp. 5, 6).

[13] V. M. Grichine, *A simplified Glauber model for hadron–nucleus cross sections*, The European Physical Journal C **62** (2009) 399, ISSN: 1434-6052, URL: https://doi.org/10.1140/epjc/s10052-009-1033-z (cit. on p. 6).

[14] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **907** (2018) 188, Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn), ISSN: 0168-9002 (cit. on p. 9).

[15] H. Aihara et al., *The Belle II Detector Upgrades Framework Conceptual Design Report*, (2024), arXiv: 2406.19421 [hep-ex] (cit. on p. 9).

[16] T. V. Dong et al., *Calibration and alignment of the Belle II central drift chamber*, Nucl. Instrum. Meth. A **930** (2019) 132 (cit. on p. 12).

[17] H. Kolanoski and N. Wermes, *Particle Detectors*, Oxford University Press, 2020, ISBN: 978-0-19-885836-2 (cit. on p. 13).

[18] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth and N. Braun, *The belle II core software*, en, Comput. Softw. Big Sci. **3** (2019) (cit. on pp. 14, 19).

[19] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81, ed. by M. Werlen and D. Perret-Gallix (cit. on p. 14).

[20] T. Ferber and P. Urquijo, *Overview of the Belle II Physics Generators*, tech. rep. BELLE2-NOTE-PH-2015-006, Belle II Internal Document, 2015, URL: https://pubdb.belle2.org/pub_data/documents/444/ (cit. on p. 14).

[21] V. Bertacchi et al., *Track finding at Belle II*, Computer Physics Communications **259** (2021) 107610, ISSN: 0010-4655 (cit. on pp. 15, 27–29, 59).

[22] L. Reuter et al., *End-to-end multi-track reconstruction using Graph Neural Networks at Belle II*, en, Comput. Softw. Big Sci. **9** (2025) (cit. on p. 15).

[23] T. Ferber and C. Hearty, *Design of the ECL Software for Belle II*, tech. rep. BELLE2-NOTE-TE-2016-001, Belle II Internal Document, 2016, URL: https://pubdb.belle2.org/pub_data/documents/658/ (cit. on p. 15).

[24] Y. Sato, S. Cunliffe, F. Meier and A. Zupanc, *Monte Carlo matching in the Belle II software*, EPJ Web Conf. **251** (2021) 03021 (cit. on p. 16).

[25] J. Eppelt and T. Ferber, *MCParticle Matching concerning the ECL*, tech. rep. BELLE2-NOTE-TE-2025-020, Belle II Internal Document, 2025, URL: https://pubdb.belle2.org/pub_data/documents/4641/ (cit. on pp. 16, 24, 59).

[26] Belle II Collaboration, *Truth-matching: MC matching / Photon matching*, https://software.belle2.org/release-09-00-06/sphinx/analysis/doc/MCMatching.html#photon-matching, Accessed: 2025-11-17, 2025 (cit. on p. 16).

[27] B. I. Collaboration et al., *Charged-hadron identification at Belle II*, European Physical Journal C **85** (2025) 1237 (cit. on pp. 17, 19, 47, 51).

[28] P. Cheema, *Search for the lepton flavour violating decay $B^0 \to \ell\tau$ at Belle II*, PhD thesis, 2025, URL: https://hdl.handle.net/2123/33763 (cit. on p. 19).

[29] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, San Francisco, California, USA: ACM, 2016 785, ISBN: 978-1-4503-4232-2, URL: http://doi.acm.org/10.1145/2939672.2939785 (cit. on p. 31).

[30] P. Calafiura, D. Rousseau and K. Terao, *Artificial intelligence for high energy physics*, en, Singapore, Singapore: World Scientific Publishing, 2022, ISBN: ISBN: 978-981-12-3403-3 (cit. on p. 31).

[31] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences **55** (1997) 119, ISSN: 0022-0000 (cit. on p. 32).

[32] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework", *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019 (cit. on p. 37).

[33] G. Cowan, *Statistical data analysis: with applications from particle physics*, eng, Repr, Oxford science publications, Oxford: Clarendon Press, 1998, ISBN: 978-0-19-850155-8 978-0-19-850156-5 (cit. on p. 39).

[34] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) 1554, ISSN: 1434-6052, URL: https://doi.org/10.1140/epjc/s10052-011-1554-0 (cit. on p. 42).

[35] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, ed. by L. Lyons, R. P. Mount and R. Reitmeyer, arXiv: physics/0308063 (cit. on p. 42).

[36] J. Podolanski and R. Armenteros, *III. Analysis of V-events*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **45** (1954) 13, URL: https://doi.org/10.1080/14786440108520416 (cit. on p. 47).

[37] C. Lippmann, *Particle identification*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **666** (2012) 148, Advanced Instrumentation, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900211005419 (cit. on pp. 47, 48).

[38] S. Longo, *Neutron and Antineutron Performance Studies*, tech. rep. BELLE2-NOTE-TE-2025-012, Belle II Internal Document, 2025, URL: https://pubdb.belle2.org/pub_data/documents/4539/ (cit. on p. 48).

[39] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 48).

[40] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152, ed. by S. Erhan, P. Schlein and Y. Rozen (cit. on p. 48).

[41] A. Ryd et al., *EvtGen: A Monte Carlo Generator for B-Physics*, (2005) (cit. on p. 48).

[42] G. Rodrigo, H. Czyż, J. H. Kühn and M. Szopa, *Radiative return at NLO and the measurement of the hadronic cross-section in electron–positron annihilation*, The European Physical Journal C - Particles and Fields **24** (2002) 71, ISSN: 1434-6052, URL: https://doi.org/10.1007/s100520200912 (cit. on p. 48).

[43] H. Czyż et al., *PHOKHARA version 9.1 — Manual*, version 9.1, Available at https://looptreeduality.csic.es/phokhara/phokhara9.1.pdf, Institute of Nuclear Physics PAN / CSIC, Kraków / Madrid, 2014 (cit. on p. 48).

[44] H. Czy ż, J. H. Kühn and S. Tracz, *Nucleon form factors and final state radiative corrections to $e^+e^- \to \overline{p}p\gamma$*, Phys. Rev. D **90** (11 2014) 114021, URL: https://link.aps.org/doi/10.1103/PhysRevD.90.114021 (cit. on p. 48).

[45] H. Czy ż, A. Grzeli ńska and J. H. Kühn, *Spin asymmetries and correlations in Λ-pair production through the radiative return method*, Phys. Rev. D **75** (7 2007) 074026, This version first introduced Λ baryon decays, URL: https://link.aps.org/doi/10.1103/PhysRevD.75.074026 (cit. on p. 48).

[46] K. Trabelsi, *HLT Skims*, tech. rep. BELLE2-NOTE-TE-2020-018, Belle II Internal Document, 2020, URL: https://pubdb.belle2.org/pub_data/documents/559/ (cit. on p. 51).

[47] J.-F. Krohn et al., *Global decay chain vertex fitting at Belle II*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **976** (2020) 164269, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900220306653 (cit. on p. 51).

[48] P. Avery, *Applied fitting theory III: non-optimal least squares fitting and multiple scattering*, CLEO Note CBX91-74, CLEO (1991) (cit. on p. 52).

[49] H. Dembinski, M. Kenzie, C. Langenbruch and M. Schmelling, *Custom Orthogonal Weight functions (COWs) for event classification*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1040** (2022) 167270, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900222006076 (cit. on p. 54).

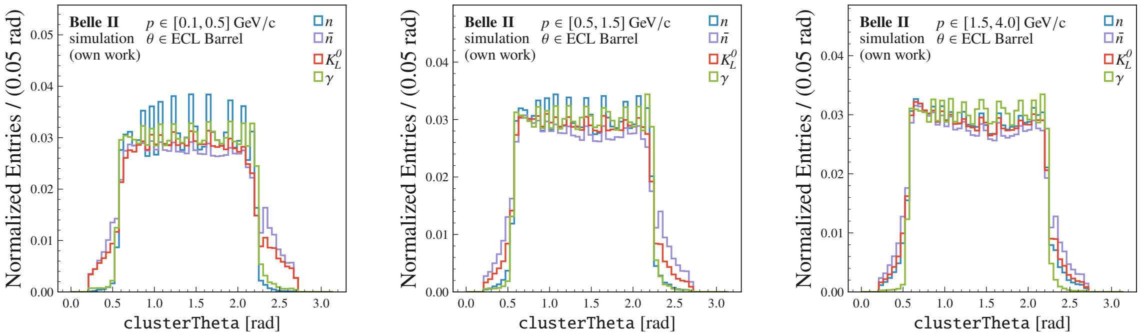
[50] O. Behnke and L. Moneta, "Parameter Estimation", *Data Analysis in High Energy Physics*, John Wiley & Sons, Ltd, 2013, chap. 2 27, ISBN: 9783527653416, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527653416.ch2, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527653416.ch2 (cit. on p. 54).

[51] B. Aubert et al., *Study of $e^+e^- \to \Lambda\overline{\Lambda}$, $\Lambda\overline{\Sigma}^0$, $\Sigma^0\overline{\Sigma}^0$ using initial state radiation with BABAR*, Phys. Rev. D **76** (9 2007) 092006, URL: https://link.aps.org/doi/10.1103/PhysRevD.76.092006 (cit. on p. 55).

APPENDIX A

Additional Material

A.1 Particle Gun Variable Distributions



(a) Low (b) Central (c) High

Figure A.1: `clusterTheta` distributions at low, central, and high momenta. The distributions are normalized to unit area.

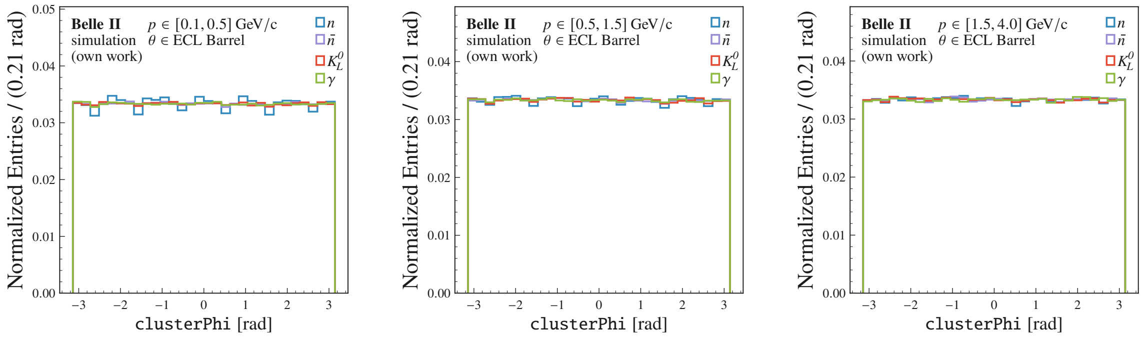
A.2 ECL Interaction Rate Table

A.3 Particle Abundance

A.4 BDT input features

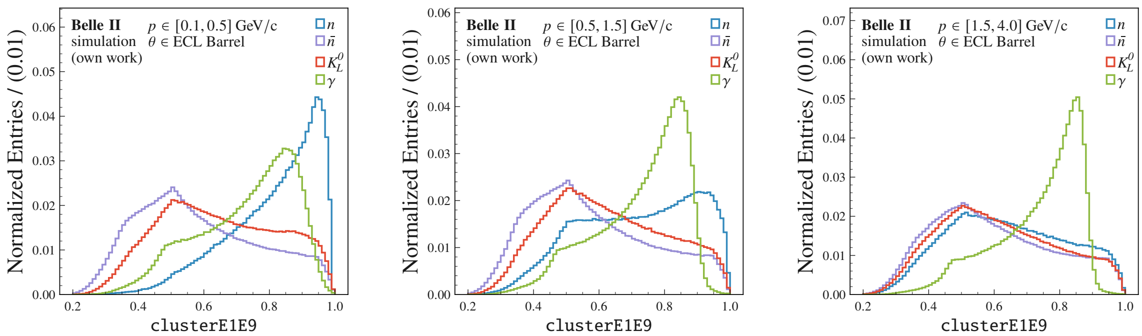
A.4.1 Weighted

A.4.2 Unit area normalized



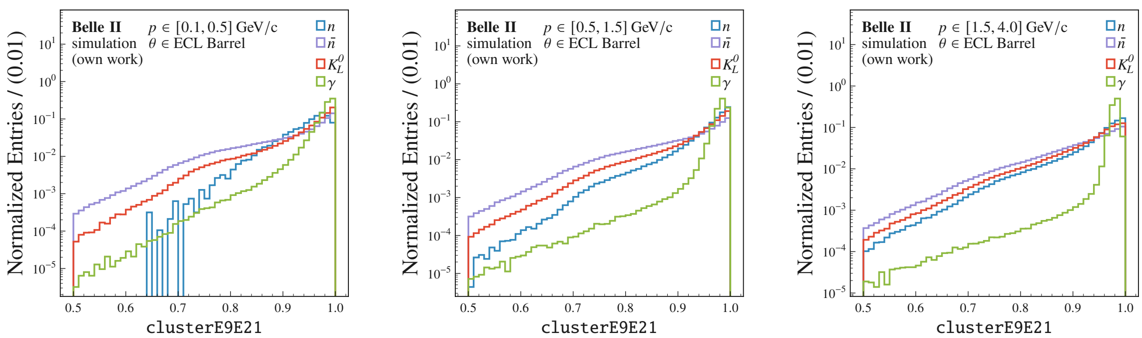
(a) Low (b) Central (c) High

Figure A.2: `clusterPhi` distributions at low, central, and high momenta. The distributions are normalized to unit area.



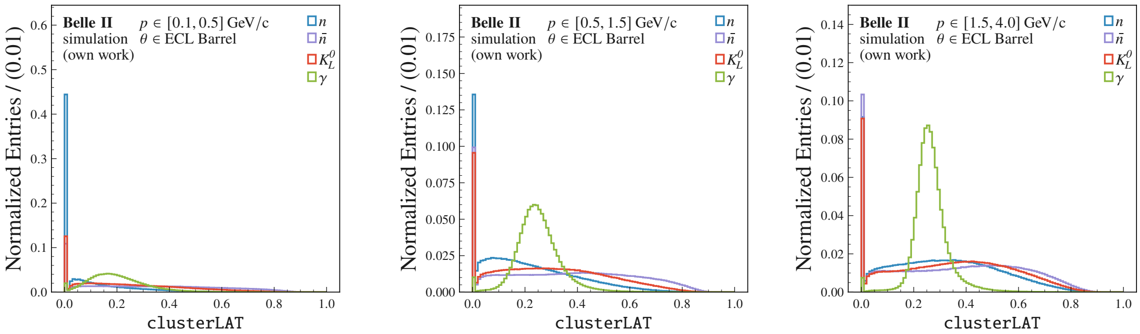
(a) Low (b) Central (c) High

Figure A.3: `clusterE1E9` distributions at low, central, and high momenta. The distributions are normalized to unit area.



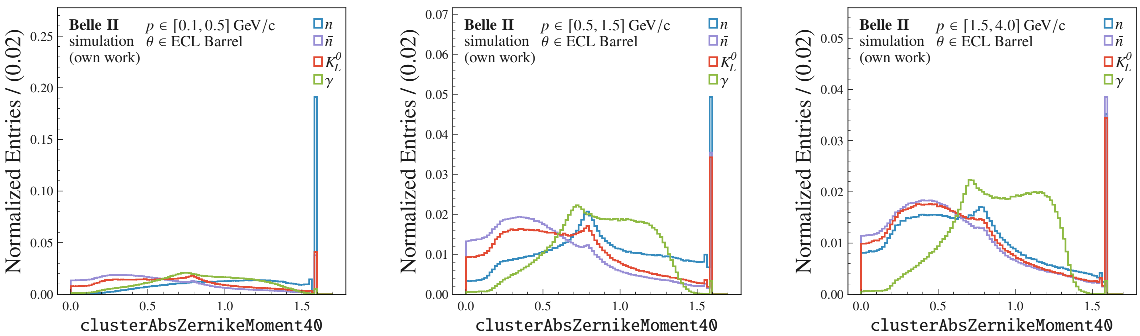
(a) Low (b) Central (c) High

Figure A.4: `clusterE9E21` distributions at low, central, and high momenta. The distributions are normalized to unit area.



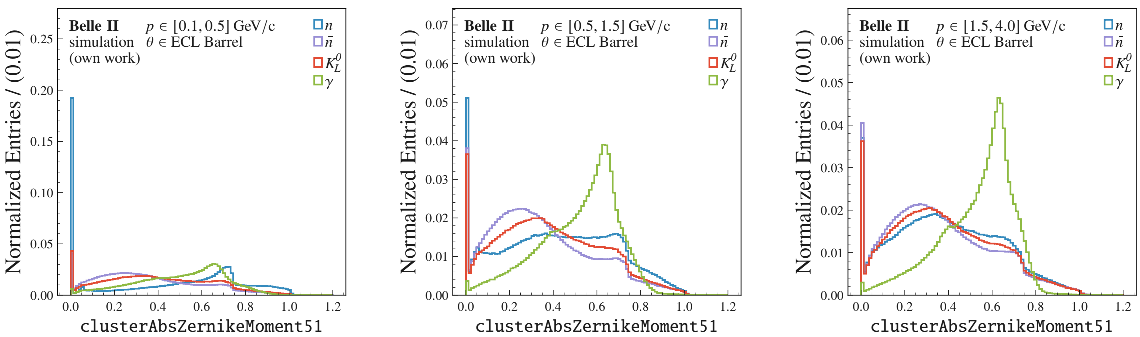
(a) Low (b) Central (c) High

Figure A.5: `clusterLAT` distributions at low, central, and high momenta. The distributions are normalized to unit area.



(a) Low (b) Central (c) High

Figure A.6: `clusterAbsZernikeMoment40` distributions at low, central, and high momenta. The distributions are normalized to unit area.



(a) Low (b) Central (c) High

Figure A.7: `clusterAbsZernikeMoment51` distributions at low, central, and high momenta. The distributions are normalized to unit area.

Table A.1: Efficiency of different selections for each particle type in the three momentum bins. Two types of selections are presented: $\epsilon(A)$ refers to the efficiency of A with respect to the generated number of events, whereas $\epsilon(A|B)$ refers to the efficiency of A given that selection B is already applied

Low momentum bin

Efficiency(Selections)	n [%]	$\bar{n}$ [%]	K_L^0 [%]	γ [%]
$\epsilon(\geq 1$ ECL Cluster)	9.872 ± 0.013	83.14 ± 0.04	69.35 ± 0.05	89.703 ± 0.030
$\epsilon(E > 0.1$ GeV$\| \geq 1$ ECL Cluster)	4.716 ± 0.030	94.974 ± 0.024	85.97 ± 0.04	95.723 ± 0.021
ϵ(MC Matching$\| \geq 1$ ECL Cluster)	0.0 ± 0	30.52 ± 0.05	20.78 ± 0.05	97.790 ± 0.016
$\epsilon(E > 0.1$ GeV & MC Matching$\| \geq 1$ ECL Cluster)	0.0 ± 0	30.52 ± 0.05	20.78 ± 0.05	95.186 ± 0.023
$\epsilon(E > 0.1$ GeV & MC Matching & ≥ 1 ECL Cluster)	0.0 ± 0	25.37 ± 0.04	14.409 ± 0.035	85.384 ± 0.035

Central momentum bin

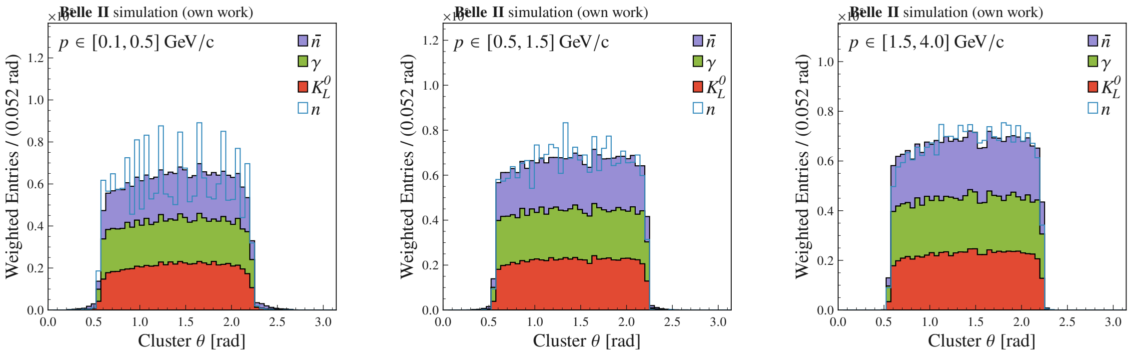
Efficiency(Selections)	n [%]	$\bar{n}$ [%]	K_L^0 [%]	γ [%]
$\epsilon(\geq 1$ ECL Cluster)	44.666 ± 0.035	74.61 ± 0.04	65.54 ± 0.05	89.791 ± 0.030
$\epsilon(E > 0.1$ GeV$\| \geq 1$ ECL Cluster)	66.68 ± 0.05	91.837 ± 0.032	91.838 ± 0.034	98.965 ± 0.011
ϵ(MC Matching$\| \geq 1$ ECL Cluster)	1.288 ± 0.012	26.69 ± 0.05	23.62 ± 0.05	97.951 ± 0.015
$\epsilon(E > 0.1$ GeV & MC Matching$\| \geq 1$ ECL Cluster)	1.288 ± 0.012	26.69 ± 0.05	23.62 ± 0.05	97.951 ± 0.015
$\epsilon(E > 0.1$ GeV & MC Matching & ≥ 1 ECL Cluster)	0.576 ± 0.005	19.92 ± 0.04	15.48 ± 0.04	87.952 ± 0.033

High momentum bin

Efficiency(Selections)	n [%]	$\bar{n}$ [%]	K_L^0 [%]	γ [%]
$\epsilon(\geq 1$ ECL Cluster)	59.83 ± 0.05	70.96 ± 0.05	63.39 ± 0.05	89.541 ± 0.031
$\epsilon(E > 0.1$ GeV$\| \geq 1$ ECL Cluster)	92.499 ± 0.034	91.885 ± 0.032	93.276 ± 0.031	99.239 ± 0.009
ϵ(MC Matching$\| \geq 1$ ECL Cluster)	11.71 ± 0.04	26.19 ± 0.05	23.89 ± 0.05	98.145 ± 0.014
$\epsilon(E > 0.1$ GeV & MC Matching$\| \geq 1$ ECL Cluster)	11.71 ± 0.04	26.19 ± 0.05	23.89 ± 0.05	98.145 ± 0.014
$\epsilon(E > 0.1$ GeV & MC Matching & ≥ 1 ECL Cluster)	7.006 ± 0.026	18.59 ± 0.04	15.14 ± 0.04	87.880 ± 0.033

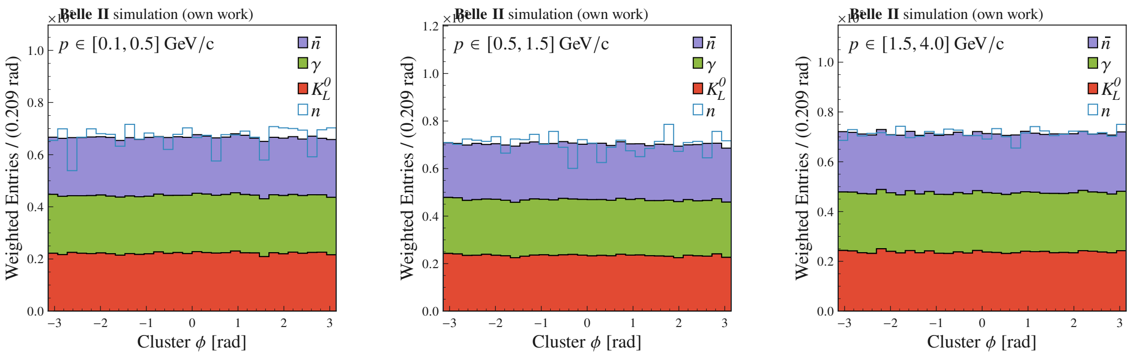
Table A.2: Neutral particle abundance for generated p_T between 0.01–3 GeV.

Particle	Relative abundance [%]
$e^\pm$	3.28
$\mu^\pm$	2.38
$\pi^\pm$	33.64
$K^\pm$	6.88
$p/\,\overline{p}$	0.11
$n/\,\overline{n}$	0.64
K_L^0	2.79
γ	50.38



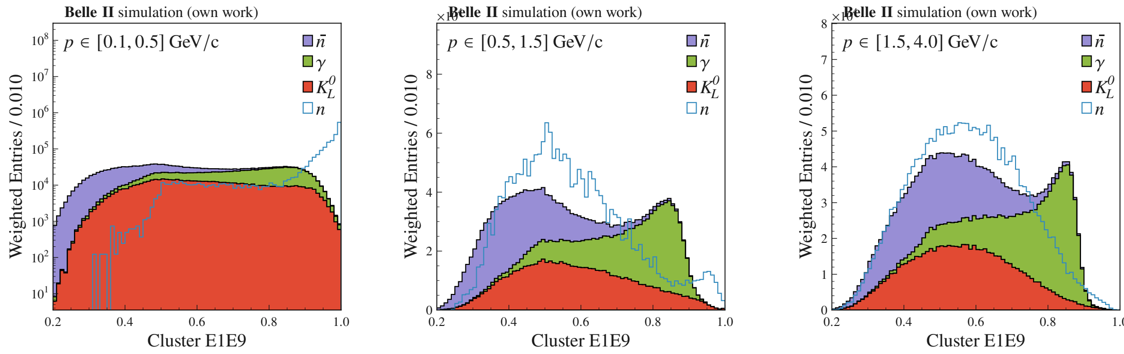
(a) Low (b) Central (c) High

Figure A.8: Cluster θ distributions at low, central, and high momenta.



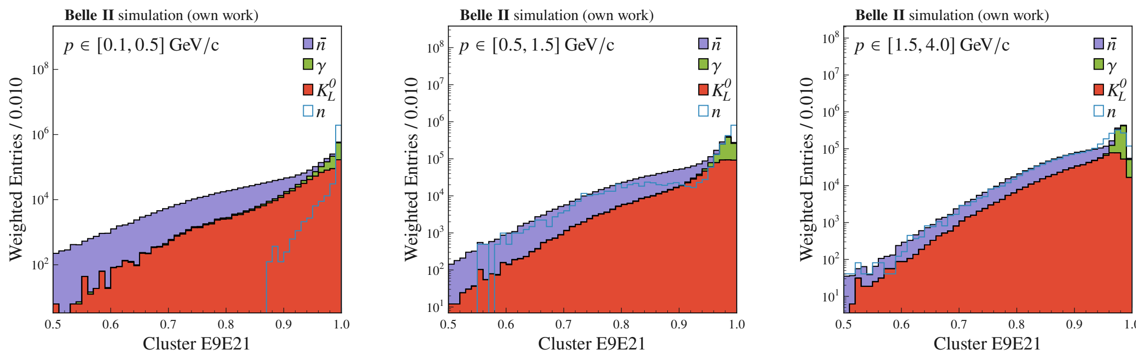
(a) Low (b) Central (c) High

Figure A.9: Cluster ϕ distributions at low, central, and high momenta.



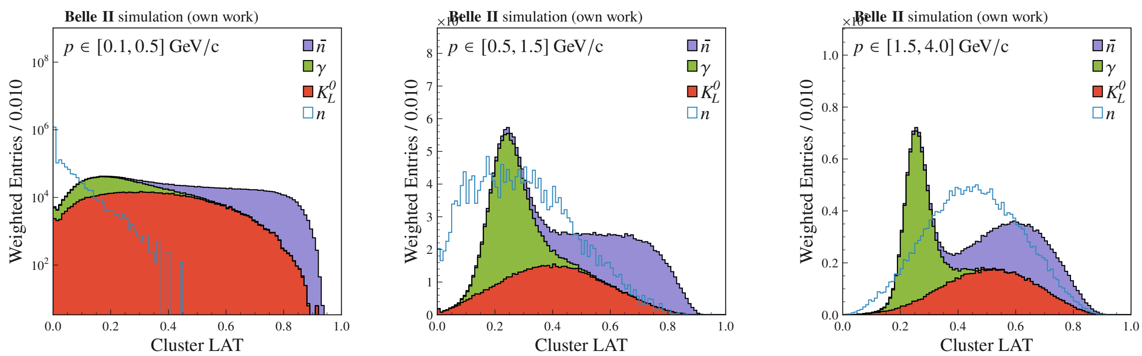
(a) Low (b) Central (c) High

Figure A.10: Cluster E_1/E_9 distributions at low, central, and high momenta.



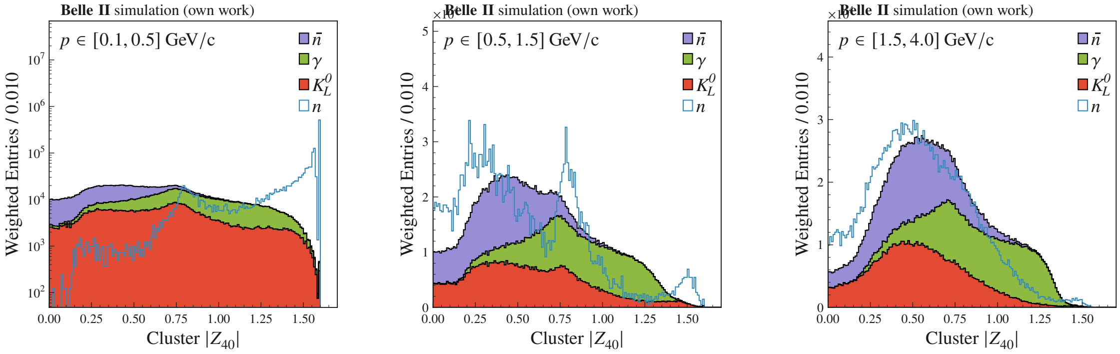
(a) Low (b) Central (c) High

Figure A.11: Cluster E_9/E_{21} distributions at low, central, and high momenta.



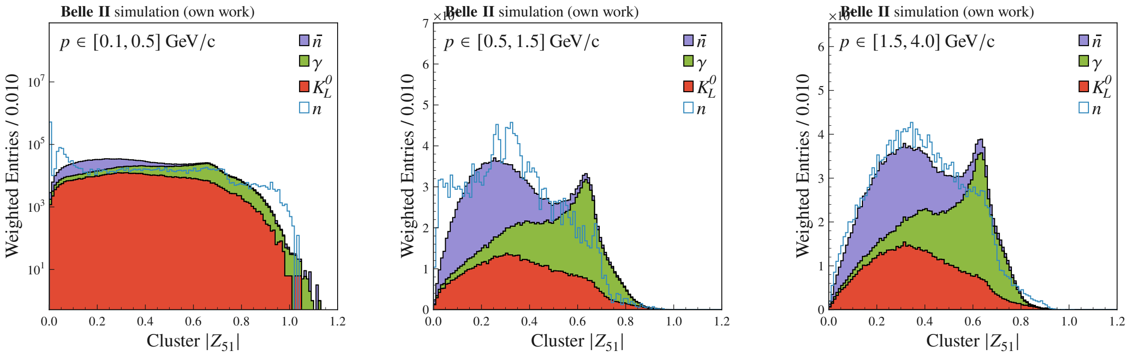
(a) Low (b) Central (c) High

Figure A.12: Cluster LAT distributions at low, central, and high momenta.



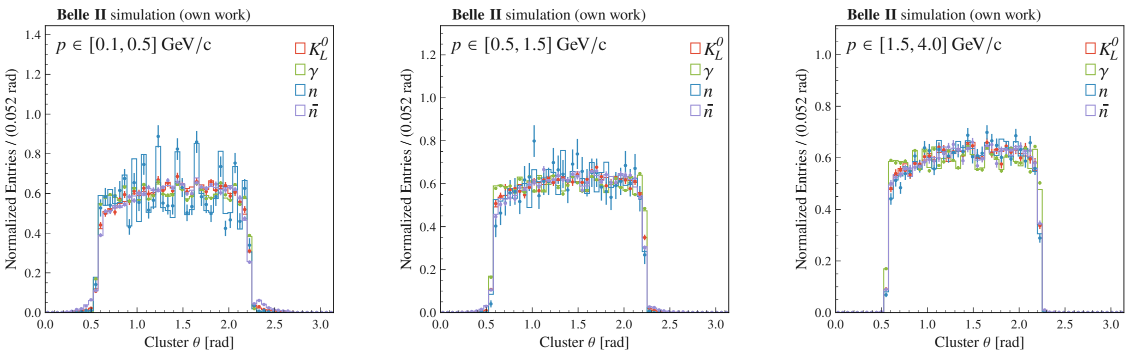
(a) Low (b) Central (c) High

Figure A.13: Cluster $|Z_{40}|$ distributions at low, central, and high momenta.



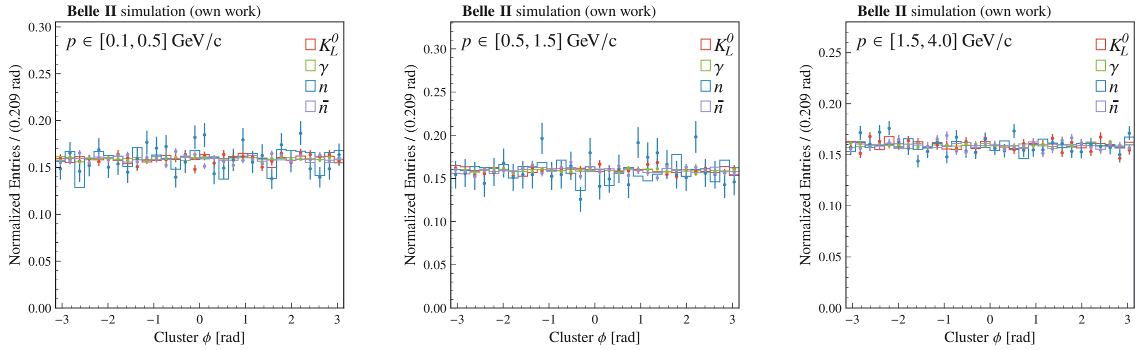
(a) Low (b) Central (c) High

Figure A.14: Cluster $|Z_{51}|$ distributions at low, central, and high momenta.



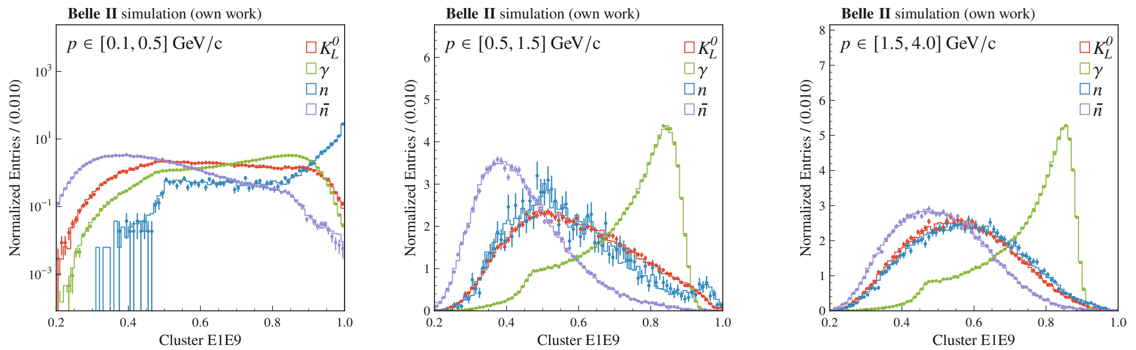
(a) Low (b) Central (c) High

Figure A.15: Cluster θ distributions at low, central, and high momenta.



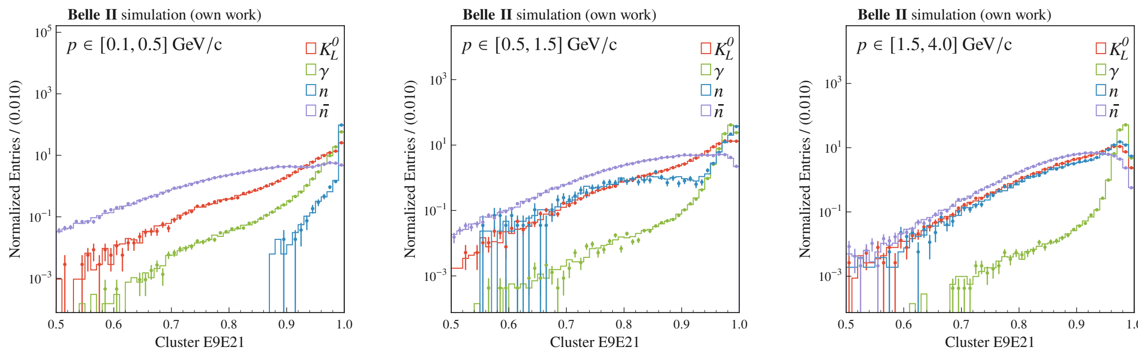
(a) Low (b) Central (c) High

Figure A.16: Cluster ϕ distributions at low, central, and high momenta.



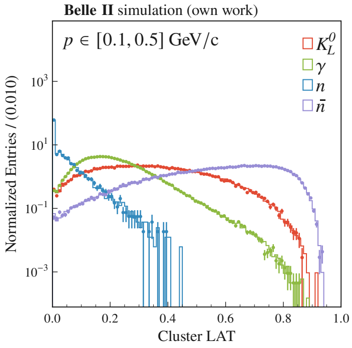
(a) Low (b) Central (c) High

Figure A.17: Cluster E_1/E_9 distributions at low, central, and high momenta.

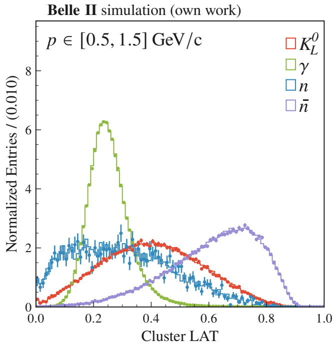


(a) Low (b) Central (c) High

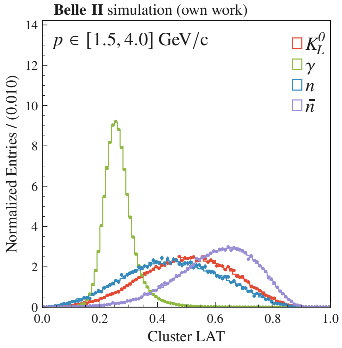
Figure A.18: Cluster E_9/E_{21} distributions at low, central, and high momenta.



(a) Low

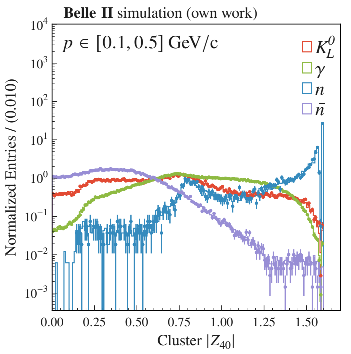


(b) Central

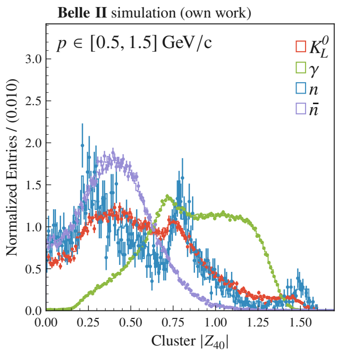


(c) High

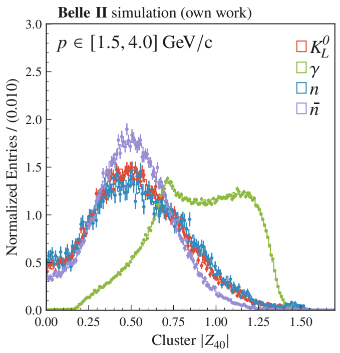
Figure A.19: Cluster LAT distributions at low, central, and high momenta.



(a) Low

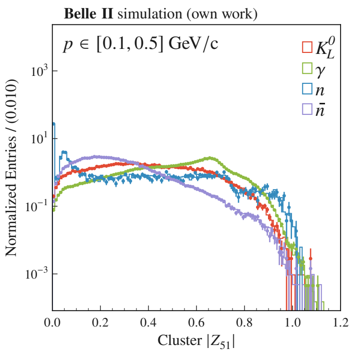


(b) Central

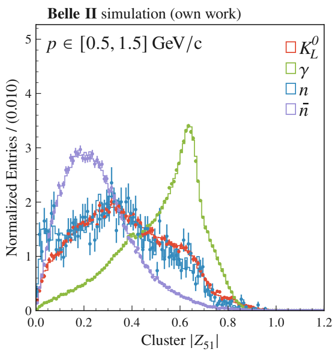


(c) High

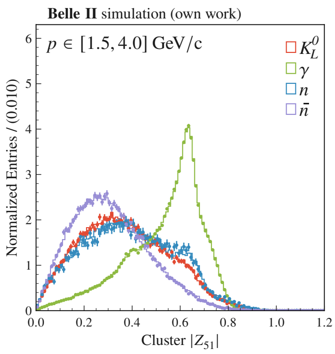
Figure A.20: Cluster $|Z_{40}|$ distributions at low, central, and high momenta.



(a) Low



(b) Central



(c) High

Figure A.21: Cluster $|Z_{51}|$ distributions at low, central, and high momenta.

List of Figures

List of Tables