

Photon Reconstruction of Axion-Like Particles with Graph Neural Networks at Beamdump Experiments

Kylian Schmidt

Masterthesis

2nd March 2024

Institute of Experimental Particle Physics (ETP)

First Reviewer:	Prof. Dr. Markus Klute
Second Reviewer:	Prof. Dr. Torben Ferber
Advisor:	M.Sc. Alexander Heidelberg

Editing time: 2nd March 2023 – 2nd March 2024

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 2. März 2024

.....
(Kylian Schmidt)

Abstract

Axion-Like Particles (ALPs) are hypothetical weakly interacting light particles predicted by theories Beyond the Standard Model which could be mediators between a dark sector and the Standard Model. Some of these theories predict light ALPs which decay into two photons and could be detected at future beamdump experiments such as LUXE - New Physics search at Optical Dump.

To investigate the properties of such ALPs, an accurate reconstruction of the common photon vertex from the hits measured in the detector can aid the search significantly. For this purpose, the photon shower direction needs to be reconstructed precisely, combining techniques from cluster and track reconstruction. This task is a prime candidate for modern methods of reconstruction based on Machine Learning such as Graph Neural Networks.

This thesis presents a Graph Neural Network composed of GravNet and GarNet, which is able to reconstruct the decay vertex of ALPs from the sparse detector hits of the two photon showers. The performance of the network is assessed on a data-set simulated with a high granularity calorimeter.

Software

This work relies heavily on software developed by other people which are acknowledged in the following. The BEAMDUMP TRACKING CALORIMETER simulation software used to produce the Monte Carlo data is developed by Alexander Heidelberg [1]. I wrote the python interface for this project with the help of Giacomo de Pietro using the PYBIND library [2]. The simulation of all samples used in this thesis was done by me. The output pipeline of the simulation was also written by me using the AWKWARD package [3]. The initial idea for the Graph Neural Network used in this thesis was proposed by Jan Kieseler, who also provided extensive help and access to his Machine Learning framework [4]. The network architecture was created by me in TENSORFLOW and KERAS [5, 6]. The custom implementation of the GarNet graph layer was made by me, based on the original paper [7]. Every plot in this thesis was produced by me with the MATPLOTLIB library [8]. The HGCal analysis was done by me with GEANT4 using software provided by Jan Kieseler and the KAFE2 software [9]. Every person mentioned by name is a member of the Institute of Particle Physics (ETP) at KIT.

Contents

1. Introduction	1
2. Theoretical Foundations	3
2.1. Chiral Anomaly and the Strong CP Problem	3
2.2. The QCD Axion	4
2.3. ALP Decay into Two Photons	5
3. Detector Systems	7
3.1. New Physics at Colliders	7
3.2. Beamdump Experiments	7
3.3. LUXE-NPOD	8
4. Simulations	11
4.1. Framework	11
4.2. Geometry	12
4.3. Output and Readout	13
4.4. Event Generation	15
4.5. Selection Criteria	17
5. Graph Neural Networks	21
5.1. Machine Learning and Computer Vision	21
5.2. GravNet and GarNet Layers	23
5.3. Training Targets and Loss Function	26
5.4. Architecture for Photon Reconstruction	26
5.5. Model Training	28
6. Photon Reconstruction with GNNs	31
6.1. Event View	31
6.2. Accuracy Measure	34
6.3. Radial Accuracy	34
6.4. Angular Accuracy	35
6.5. Vertex Accuracy	36
6.6. Summary	37
7. Energy Resolution of Calorimeters	41
7.1. Derivation of the Energy Resolution Equation	41
7.2. Energy Resolution of the CMS HGCal	42

8. Conclusion	49
9. Outlook	51
A. Appendix	61
A.1. Wald distribution	61
A.2. Additional Figures for Reconstruction with GNNs	62

1. Introduction

The Standard Model (SM) of particle physics is the most precise and complete theory describing the interaction of matter at the smallest scale [10]. Still, there are experimental observations that the SM is simply unable to explain. The most notable example known today is the existence of Dark Matter (DM), which composes most of the mass in the universe but remains undetected with current experimental methods [11]. Moreover, the SM has many free parameters whose values can only be determined from experiment [12]. This apparent fine-tuning calls into the question the naturalness of the SM. Therefore, extensions of the SM try to describe these phenomena in the form of New Physics (NP) models that may address these shortcomings. Some models predict the existence of new particles that could be observed with modern particle detectors. Though their signatures might potentially be hard to discern from regular SM processes.

When looking for a needle in a haystack, it is best to bring a fine comb. Accordingly, current experiments collect and scrutinize vast amounts of data in search of deviations from the SM [13]. New experiments are being built to probe unexplored regions of the parameter space for NP models. One example is the proposed LUXE-NPOD experiment at Deutsches Elektronen-Synchrotron which enables the search for feebly interacting particles that decay into two photons [14]. At this experiment, the signature in the detector materializes as two high energy photons originating from a common vertex. In order to separate this potential NP phenomenon from the SM background, the vertex and direction of the resulting photons must be well understood.

The available detection methods for neutral particles are highly invasive and require for the original particle to shower into lower-energy measurable particles. In contrast to the tracks left by charged particles, finding the direction of neutral particles from their stochastic showers is challenging. Most experiments have therefore put an emphasis on the reconstruction of the energy of the neutral particle. This thesis explores an approach for the direction reconstruction of photons, which is made possible by advancements in software tools.

Computationally expensive simulations allow scientists to model the detector with great precision, including before the construction of the physical experiment. New techniques are continuously developed in parallel to account for the large amounts of simulated data. Paradigms such as Machine Learning (ML) are able to efficiently learn complex non-linear relations from large data-sets. A Graph Neural Network (GNN), a modern ML architecture, can work directly on raw detector data. With these modern tools, irregular detector

geometries and variable input shapes no longer present an issue. This work utilizes an application of a GNN composed of GravNet and GarNet graph layers to reconstruct the direction of two photon showers [7].

Chapter 2 gives a short introduction to the strong CP problem, which is one of the unresolved questions of the SM. In this Chapter, the dynamic solution of the QCD axion is discussed, along with its generalisation in the form of Axion-Like Particles. These potentially long-lived particles can be searched at beamdump experiments, which is the topic of Chapter 3. Chapter 4 shows the full simulation of a highly granular calorimeter, and outlines the generation of Monte Carlo (MC) data used in this work. The architecture of the GNN is detailed in Chapter 5, which comprises a full description of GravNet and GarNet. Chapter 6 applies the GNN on the MC data-set and studies its performance based on relevant physical parameters. Before reaching the conclusion, Chapter 7 details a standalone study of the impact of the manufacturing tolerances of absorber plates on the new CMS High Granularity Calorimeter (HGCAL). Finally, Chapter 8 gives a summary on the capabilities of the GNN approach for photon direction reconstruction. Possible directions for further studies with this powerful tool are proposed in Chapter 9.

2. Theoretical Foundations

This chapter introduces the Axion-Like Particle (ALP) as a kind of new physics Beyond the Standard Model (BSM). Section 2.1 discusses the chiral anomaly in Standard Model QCD and how it could generate a CP violating term. It also addresses the non-observation of the neutron electric dipole moment and how it forms the strong Charge Parity (CP) problem. Section 2.2 provides a possible explanation for the strong CP problem by adding a new light neutral scalar field and lists the properties of its associated excitation, the axion. Finally, Section 2.3 expands on the idea of the axion with ALPs coupling to photons.

2.1. Chiral Anomaly and the Strong CP Problem

This section introduces the theoretical motivation for the chiral anomaly in Quantum Chromodynamics (QCD) and is based on [15]. The spectrum of light hadrons involving only up and down quarks can be explained well by the chiral perturbation theory of QCD (in the approximation of vanishing quark masses, as $m_u, m_d \ll \Lambda_{QCD}$). The chiral symmetry group at the UV scale is the composition of right and left handed fermion Lie groups, $U(2)_L \times U(2)_R$. In the low energy regime, the physical quark condensate has ground states $\langle \bar{u}u \rangle, \langle \bar{d}d \rangle$ which spontaneously break down the chiral symmetry into

$$U(2)_L \times U(2)_R = SU(2)_V \times SU(2)_A \times U(1)_V \times U(1)_A \xrightarrow{SSB} SU(2)_I \times U(1)_B. \quad (2.1)$$

Here, the original chiral symmetry is rotated into a basis with an axial $SU(2)_A$ and a vector $U(1)_V$ symmetry. The resulting subgroups $SU(2)_I$ and $U(1)_B$ are identified as the isospin and the baryon number respectively, which are both good symmetries of nature. The breakdown of $SU(2)_A \times U(1)_A$ should create four Nambu-Goldstone bosons according to the Goldstone theorem [16]. However, only the remnants of $SU(2)_A$ in the form of the three pions are observed in nature. The closest SM candidate is the η meson but it is too heavy to be the fourth Goldstone boson. This leads to the conclusion that $U(1)_A$ must be an anomalous symmetry of QCD, broken by the quantum effects of the vacuum. The resulting Noether current of the form $J^{\mu 5} = \bar{\psi} \gamma^\mu \gamma^5 \psi$ is not conserved, since it couples at loop level with the vector current $J^\mu = \bar{\psi} \gamma^\mu \psi$, as can be seen in Fig. 2.1. The divergence of the axial current is

$$\partial_\mu J^{\mu 5} = \frac{g^2}{16\pi^2} G_a^{\mu\nu} \tilde{G}_{\mu\nu a}, \quad (2.2)$$

where g is the strong coupling constant, $G_a^{\mu\nu}$ are the gluon field strengths and $\tilde{G}_{\mu\nu a} = \frac{1}{2} \epsilon_{\mu\nu\alpha\beta} G_a^{\alpha\beta}$ their duals. In fact, by using the appropriate gauge, one can show that $U(1)_A$

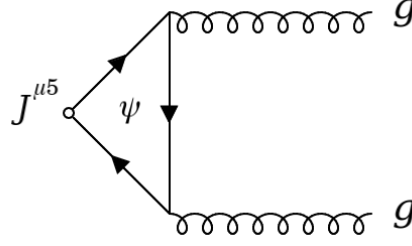


Figure 2.1.: Feynman diagram of the axial current of QCD coupling to two gluons via a quark loop.

is not a symmetry of QCD at the UV scale either. This leads to an extra term in the QCD Lagrangian which depends on the phase $0 \leq \theta \leq 2\pi$ of the true θ -vacuum of QCD,

$$\mathcal{L} \supset \bar{\theta} \frac{g^2}{32\pi^2} G_a^{\mu\nu} \tilde{G}_{\mu\nu a}, \quad (2.3)$$

with $\bar{\theta} = \theta + \arg \det(M)$. For completeness, the mass matrix $M = \text{diag}(m_u, m_d)$ of the light quarks is added as a correction from the CP violation of weak interactions. This new term in the Lagrangian should generate an electric dipole momentum for the neutron of magnitude $d_n = (2.4 \pm 1.0 \times 10^{-16} \text{ e cm}) \bar{\theta}$. However, the experimental upper bound of $|d_n| < 1.8 \times 10^{-26} \text{ e cm}$ [10] indicates that $|\bar{\theta}| < 10^{-10}$. This apparent fine-tuning of $\bar{\theta}$ is called the "strong CP problem". A CP violation in the strong sector should be manifested in nature unless it is suppressed by a yet unknown effect, e.g. protected by an unaccounted symmetry.

2.2. The QCD Axion

The axion solution was put forward in order to solve the strong CP problem by adding a new global chiral $U(1)_{PQ}$ (Peccei-Quinn) symmetry, which is anomalous under QCD [17, 18]. This symmetry spontaneously breaks down in the same way as $U(1)_A$ and contributes to the CP violating term. In this way, it adds the missing degree of freedom necessary to dynamically rotate away the θ -term akin to an unphysical phase. The spontaneous breaking of this symmetry will also create a Nambu-Goldstone boson named the axion. This scalar axion field $a(x)$ is invariant under $U(1)_{PQ}$,

$$a(x) \rightarrow a(x) + \alpha f_a, \quad (2.4)$$

where $0 \leq \alpha \leq 2\pi$ is a phase and f_a denoted the PQ scale. The chiral anomaly of $U(1)_{PQ}$ under QCD generates a Noether current

$$\partial_\mu J_{PQ}^\mu = \frac{g^2}{16\pi^2} G_a^{\mu\nu} \tilde{G}_{\mu\nu a}. \quad (2.5)$$

The Lagrangian now includes the combination

$$\mathcal{L} \supset \left(\bar{\theta} + \frac{a}{f_a} \right) \frac{g^2}{32\pi^2} G_a^{\mu\nu} \tilde{G}_{\mu\nu a}, \quad (2.6)$$

and at leading order the effective axion potential has the form,

$$V(a) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4m_u m_d}{(m_u + m_d)^2} \sin^2 \left(\frac{a}{2f_a} \right)}, \quad (2.7)$$

where the axion shift $a \rightarrow a - f_a \bar{\theta}$ was used. The potential is now minimized at $\langle a \rangle = 0$, cancelling its contribution to the Lagrangian. Another way to formulate the axion solution is that it adds the additional degree of freedom f_a that allows the QCD potential to reach a CP conserving minimum.

The mass of the axion is given by the quadratic term in the expansion of the potential in Eq. (2.7) and current computations with chiral perturbation theory [10] yield a value of

$$m_a = 5.691 \pm 0.051 \left(\frac{10^9 \text{ GeV}}{f_a} \right) \text{ meV}. \quad (2.8)$$

The relation between m_a and f_a is necessary to cancel the CP violating phase and defines the QCD axion. In this model, the coupling to the SM is dictated by axion-meson mixing with π^0 and η . This leads to the axion-diphoton coupling as seen in Fig. 2.2. However, models with $f_a = v_{EW}$ at the electroweak scale have long been excluded, and models with $f_a \gg v_{EW}$ are extremely weakly coupled and currently undetectable [19].

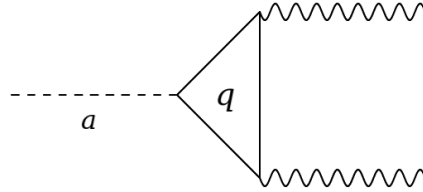


Figure 2.2.: Classical QCD axion to two photon coupling via a quark loop. This diagram is inherited from the anomalous coupling of axions to gluons and the subsequent mixing with light mesons. The loop can be expressed by the effective axion-photon coupling $g_{a\gamma\gamma}$.

2.3. ALP Decay into Two Photons

Extensions of the SM can include light pseudoscalar fields that behave in similar ways to the QCD-axion. These more general fields have an unconstrained $m_a f_a = \text{const.}$ phase space. As such, they do not solve the strong CP problem and are labelled Axion-Like Particle (ALP). Current experiments set stringent bounds on the masses and the coupling of these particles [20]. This work focuses on ALPs which couple predominately to two photons as seen on Fig. 2.3. Although large, the phase space spanned by m_a and $g_{a\gamma\gamma}$ can be probed by dedicated experiments with purpose-built geometries.

An interesting byproduct of the ALP-two-photon diagram is the Primakoff effect, where one of the two photon is a virtual photon provided by an external magnetic field. This

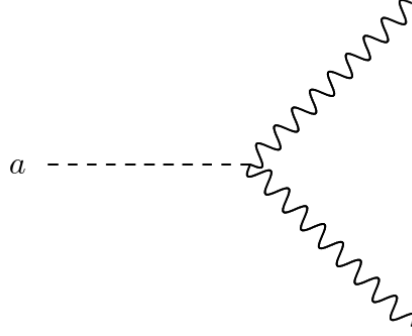


Figure 2.3.: ALP decaying into two photons with direct coupling $g_{a\gamma\gamma}$.

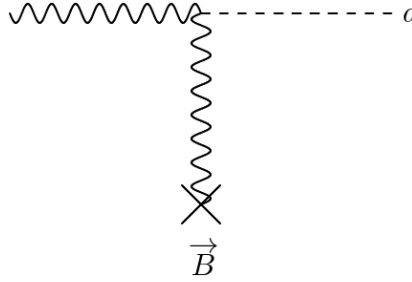


Figure 2.4.: Primakoff effect: the incoming photon interacts with a virtual photon from an external magnetic field \vec{B} and produces an ALP. This magnetic field can also be provided by the nuclei of a dense material.

process can be directly seen from the low energy effective coupling of ALPs and photons,

$$\mathcal{L} \supset -\frac{g_{a\gamma\gamma}}{4} F_{\mu\nu} \tilde{F}^{\mu\nu} = g_{a\gamma\gamma} a \vec{E} \cdot \vec{B}, \quad (2.9)$$

where $F_{\mu\nu}$ is the electromagnetic field strength, $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$ its dual. Effectively, the ALP is able to source the second photon directly from the magnetic field, as is shown on Fig. 2.4. Since the coupling remains very small, it is necessary to use a large source of photons in order to produce a detectable flux of ALPs. The sun or other stars have both strong magnetic fields and high flux of photons and could be source for stellar ALPs. Another

The decay width of an ALP with predominant coupling to photons is given by $\Gamma_{a\rightarrow\gamma\gamma} = g_{a\gamma\gamma}^2 m_a^3 / (64\pi)$ [14]. Based on this, the probability for the ALP to decay after the distance $x \geq x_0$ is given by [20],

$$P(x) = \exp\left(-\frac{m_a^4 g_{a\gamma\gamma}^2}{64\pi |\vec{p}_a|} x\right), \quad (2.10)$$

where \vec{p}_a denotes the momentum of the ALP.

3. Detector Systems

The aim of this chapter is to introduce beamdump experiments in the context of New Physics (NP) searches. Section 3.1 gives a brief summary of current approaches in the search for new particles at collider experiments. Section 3.2 then describes the principles of beamdumps and more specifically optical dumps for ALP searches. Finally, Section 3.3 details one such experiment, LUXE-NPOD, and a possible detector design.

3.1. New Physics at Colliders

Detecting a new particle at a collider experiment is a two-step process. First, the particle has to be created by colliding two beams of highly energetic particles. The crossing of the two beams at the interaction point allows the production of particles with masses limited by the center of mass energy $E_{\text{cme}} = \sqrt{E_1 E_2}$ of the beams [10]. The resulting particles can then be reconstructed by a cylindrical detector which is wrapped around the beam axis. However, the production rate of a signal event N is related to the cross-section σ of the fundamental process at play and the instantaneous luminosity of the beam. It is given by

$$L\sigma = \frac{dN}{dt}, \quad (3.1)$$

where the instantaneous luminosity L , given in $\text{cm}^{-2}\text{s}^{-1}$, is a proportionality factor that depends on the beam parameters. The integrated luminosity $\int L dt$ given in fb^{-1} is a measure for the collected data over a period of time. Increasing the instantaneous luminosity of a collider yields a higher rate of signal events and by extension improves the statistics used for background rejection.

The elusive nature of new physics processes can be explained either by a large mass scale ($m_{\text{NP}} \geq \text{TeV}$) or very low cross-sections. The former case motivates the construction of new machines with colliding beams that push the energy frontier above $\mathcal{O}(10)\text{TeV}$ [21]. The latter case however must be approached by experiments with increased instantaneous luminosity.

3.2. Beamdump Experiments

Beamdumps trade one of the beams with a fixed target to increase the luminosity at the cost of a lower $E_{\text{cme}} = \sqrt{E_b}$. Beamdump experiments search for stable long-lived SM or

unstable NP particles by stopping the high flux beam with a thick high density material [22]. Due to the short interaction length of the dump material, most particles created in the subsequent shower will be completely absorbed. Only feebly interacting particles are able to cross the remaining dump material. Historically, beamdumps were used to search for new neutrino flavors [23]. In recent years, they have experienced a resurgence in interest in the search for new physics [24, 25]. Indeed, beamdumps experiments are usually low cost and can be added to existing collider facilities with little interference.

Beamdump experiments are able to cover the very low opening angles with respect to the beam axis that are otherwise inaccessible to cylindrical detectors. They also benefit from an increase in instantaneous luminosity by using a dense fixed target instead of an opposing particle beam. This makes them specially suitable for searches of weakly interacting particles that have very small production cross sections.

Photon dumps are a potential source of ALPs as described in Section 2.3. An incoming photon is converted via the Primakoff effect with the dump material into an ALP, for which the remaining dump material will be translucent. The decay volume placed after the dump allows the potentially long-lived ALP to decay into two detectable photons. A part of the decay volume is usually submerged in a magnetic field to bend away charged background particles. These experiment types can be considered largely background free, with only a low amount of neutral particles reaching the detector in addition to the ALP decay products. The expected ALP signal comes in the form of two simultaneous photons originating from a common vertex. The non-observation of this signal can exclude masses and couplings with a confidence level dictated by the integrated luminosity, the selection efficiency and reconstruction uncertainties. The main challenge of these experiments is achieving a perfect background rejection. Unfavorable events include one or two neutrons misidentified as photons, two photons originating from different vertices and photons created before the beam dump.

3.3. LUXE-NPOD

The Laser Und XFEL Experiment (LUXE) is a planned experiment at the European X-Ray Free-Electron Laser (eu.XFEL) facility of the Deutsches Elektronen-Synchrotron (DESY) in Hamburg (English: German Electron-Synchrotron research centre). LUXE phase-0 will investigate electron-light collisions between the 16.5 GeV eu.XFEL beam and a pulsed 40 TW laser in order to probe strong-field QED [26, 27]. These collisions will produce a large number of multi-GeV photons along the electron beam axis predominantly through the non-linear Compton effect amongst other processes.

LUXE - New Physics search at Optical Dump (LUXE-NPOD) is a proposed extension of the LUXE project which aims to probe new physics using the resulting high-intensity hard photon stream [14]. The proposal includes a photon beam dump placed downstream of the LUXE collision point. This beam dump is followed by a decay volume of and finally the detector. Two ALP production mechanisms are considered at LUXE-NPOD, which are shown on Fig. 3.1. The main mode, named *secondary production*, arises from the Primakoff effect in the dump and can potentially create ALPs with masses in the range $50 \text{ MeV} \leq m_a \leq 250 \text{ MeV}$ region. The other mode, *primary production* emerges directly

from the LUXE electron-laser interaction via ALP-electron coupling, with a reach up to the $m_a \leq 100$ keV mass range. SM background processes include neutral long-lived kaons K_L^0 that decay into three photons and neutrons.

The total expected signal decreases linearly from 3×10^3 ALP candidates at $m_a = 50$ MeV down to only 1 candidate at $m_a = 250$ MeV [14]. Therefore, the capability of LUXE-NPOD to harness the physics potential provided by the LUXE photon beam depends largely on the detector performance. A considerable challenge is the rejection of background events in the form of photons or neutrons generated from outside the decay volume. One approach is to find the point of origin of a particle recorded in the detector. A proposal for a high granularity detector design is composed of two subsystems:

1. Tracking system: provides the necessary spatial information for the reconstruction of the detected particle's origin. Silicon semiconductor sensors are used to record the passing of charged particles and the absorption of low-energy photons through the photoelectric effect. A dense material is mounted before the Silicon sensors in order to shower the high energy photons produced at LUXE-NPOD.
2. Calorimeter: responsible for the measurement of the photon energy. The uncertainties on the ALP invariant mass depend mainly on the calorimeter energy resolution. In addition, the previous absorber and tracker layers decrease the energy resolution and must be accounted for by calibration.

While *tracks* are usually limited to charged particles, photon tracks refer in the following to the direction of the photons that is gathered from the detector information. The detector presented here is a simplified design. Section 4.2 describes the complete geometry of the detector simulated in this work.

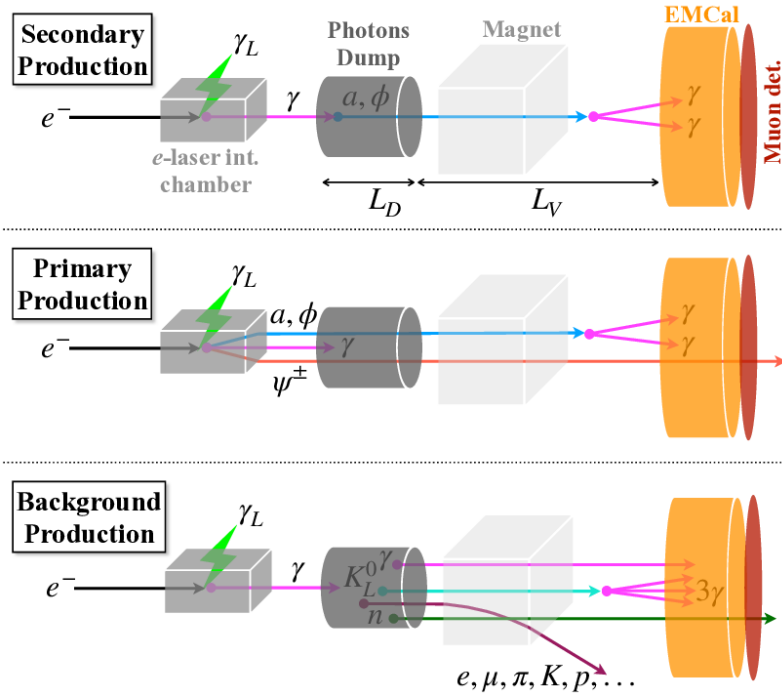


Figure 3.1.: Production mechanisms at the proposed LUXE-NPOD beamdump experiment. During the secondary production, a photon produces an ALP via the Primakoff effect. Primary production creates an ALP via its coupling to electrons. Other SM processes are responsible for background, with only neutral particles reaching the detector. Muons are vetoed by a dedicated muon detector. From [14].

4. Simulations

Monte Carlo (MC) methods supported by high performance computing play a key role in the accurate modelling of detectors due to the stochastic nature of HEP processes. The design of future experiments like LUXE-NPOD also benefits from the modelling of the detector before their construction. In this way, different design choices can be compared and optimized. This chapter describes the simulation used for the LUXE-NPOD detector. Sections 4.1 to 4.3 give a detailed view of the software itself. Section 4.4 shows a simple example for the MC simulation of two photons originating from a common vertex.

4.1. Framework

The detector simulation used in this work is built with GEANT4, which is a versatile software toolkit that models the response of particles interacting with matter [28]. GEANT4 is written in C++ and supports multi-threading for fast parallel simulations on many CPUs. It provides a general framework on which to build a specific software. This is achieved through object-oriented programming and the ability to overload the default GEANT4 classes. The detector simulation software used in this work is named BEAMDUMP TRACKING CALORIMETER (BDTC) [1].

The GEANT4 toolkit allows the user to interact with the simulation at different stages of the process. By default this is done either using shell commands that can be entered in the terminal or by executing the program with a macro file. Alternatively, it is possible to add a Graphical User Interface (GUI), which is controlled by these shell commands. In either case, the user inputs are handled by the GEANT4 messenger classes. Using these functions can create a certain overhead since they must be handled manually by the user. Modern experiments that use GEANT4 access the simulation software with Python [29, 30]. For the BDTC project, this is done with the PYBIND11 library [2]. PYBIND is used to expose the main executable and many of the simulation parameters to Python. A comprehensive list of these parameters is given in Section 4.2.

The Python bindings in BDTC are structured as shown on Fig. 4.1. The central object of the bindings is a singleton class that stores all the values used in the simulation. A singleton is a class that allows only one instance of itself at any time and is useful in the context of multi-threading. However, directly exposing this class to Python fails to safeguard its singleton properties, leading to race conditions during the output writing process. Instead, a proxy Accessor class must be exposed with PYBIND that bridges the gap between the

single-threaded Python and the multi-threaded C++ code. This Python interface allows fast and scalable access to the simulation with parallel processing.

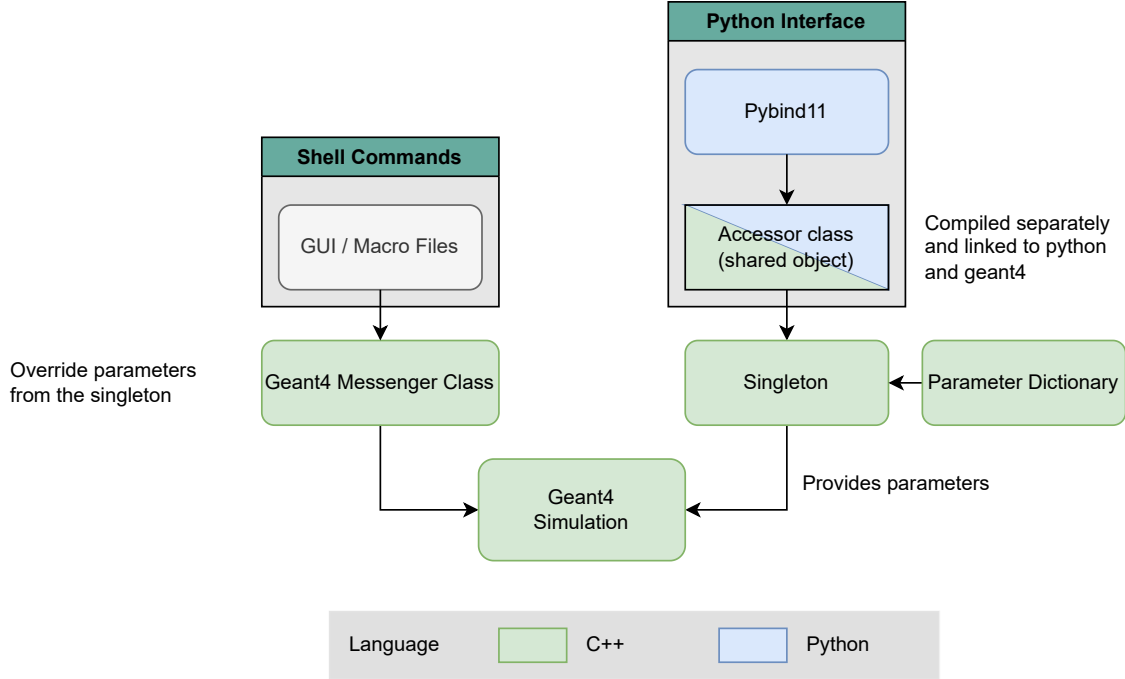


Figure 4.1.: Control flow of the BDTC simulation software. The default parameters are stored in a dictionary and are read by the singleton instance. The values in this singleton can be changed at runtime in Python using the separate Accessor class. The singleton scheme supports the multi-threading capabilities of the simulation and is compatible with the existing GEANT4 GUI and macro file interface.

4.2. Geometry

The aim of the BDTC is to provide a flexible simulation environment. For example, it is of interest to investigate the effects of different detector geometries and materials on the energy resolution of the calorimeter or on the tracking performance. The detector design of LUXE-NPOD is in an early development stage at this point and will be subjected to changes. For this reason, the position, dimensions and materials of the detector components are all variable. In the following, the *default* configuration refers to a basic detector that fulfills all the requirements of a beamdump detector similar to the one described in Section 3.3. In general, the detector is a cuboid composed of layers placed in series, with a single layer being the lateral arrangement of individual cells in the form of a grid. Layers can be one of three types:

- Absorber (layer type 0, passive material). Responsible for showering the highly energetic particles such that they can be recorded afterwards. The thickness of the absorber layer is an important parameter that impacts the showers of primary particles.

- Tracker (layer type 1, active material). Each cell is further subdivided into a grid of sensors which provide the bulk of the information for the reconstruction.
- Calorimeter (layer type 2, active material). Records the energy of the remaining showered particles. Commonly, the calorimeter is the most important detector component at a beamdump experiment. In this work, the focus is shifted to the high resolution tracker instead.

An Absorber-Tracker Layer (AT-layer) consists of one absorber layer placed in front of one tracker layer. The default longitudinal configuration of the beamdump detector has two such AT-layers, placed consecutively with a small gap between each of the four layers and with the calorimeter located after the last tracker layer. All parameters presets are summarized in Table 4.1. Fig. 4.2 shows a typical event with a 1 GeV photon moving along the z -axis and subsequently showering in the calorimeter.

Table 4.1.: Default detector geometry parameters for the beamdump simulation. The absorber thickness is at least one mean free path length of a 1 GeV photon. The tracker parameters are motivated by the state-of-the-art tracking sensors of the HGCAL [31]. The calorimeter has a similar configuration to the H1 lead/scintillator-fibre calorimeter at HERA [32].

Cell parameter	Absorber	Tracker	Calorimeter
Thickness (mm)	10	0.32	250
Side length (mm)	40	40	40
Material	Pb	Si	Pb
Cell arrangement	5x5	5x5	5x5
Subdivisions per cell	1x1	20x20	1x1
Lateral distance (mm)	1	1	1
Distance behind (mm)	1	1	1

4.3. Output and Readout

The output of the simulation is written to ROOT files [33, 34] for fast and compact readout. It includes the detector geometry parameters, the MC information about the primary particles and all recorded hits.

A hit is defined as any deposition of energy within a detector component. Each hit contains the information $h = [\text{layerType}, \text{cellID}, x, y, z, E, \text{sourcePID}]$, where the coordinates are the center of the affected cell. The cellID is a serial number assigned to each cell inside of a given layer. The sourcePID documents which primary particle lead to the hit and is specific to MC generated events. This additional MC information is important to distinguish hits belonging to each primary particle and for the analysis of the detector response. The structure of the hits in the output of the simulation is shown in Fig. 4.3.

The indexation of the cells with layerType and cellID allows the filtering of hits depending on where they are recorded. For example, hits in the absorber with layerType = 0 should

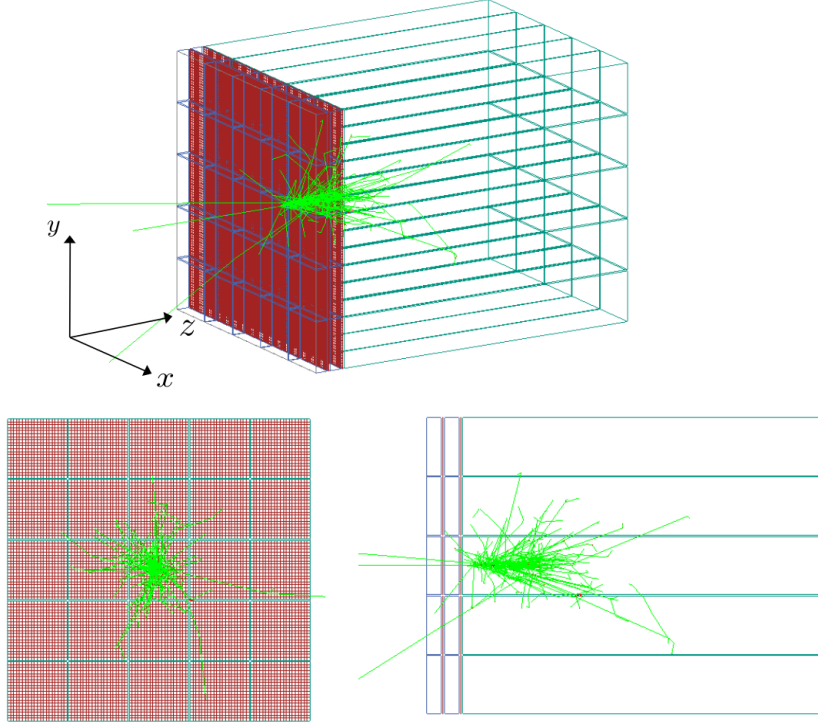


Figure 4.2.: Orthogonal (top), front (left) and side (right) view of the beamdump detector geometry with default parameters as per Table 4.1. In this configuration, there are two AT-layers (absorber in dark blue, tracker in red), followed by the calorimeter (cyan), aligned along the z axis. The cells are arranged in a 5×5 grid in the xy -plane. A single photon impinges perpendicularly on the detector with an energy of 1 GeV. The resulting shower is marked in light green (this includes photons, electrons and positrons).

be removed during the analysis since it is considered a passive material. Also, the notion of a deposited energy per cell is idealized and does not include digitization effects. Each cell sums over the deposited energy attributed to the same sourcePID. In reality, detector cells can only record the total energy deposited during a given time interval and are generally not able to differentiate energy depositions from different particles.

An event is recorded as a sparse collection of hits, typically in the order of $\mathcal{O}(100)$ hits per event, along with the details about the original primary particles. For the purposes of this work, the relevant information are the initial energy, momentum, point of origin and the particle type. Finally, events are grouped into runs, which always have a fixed detector geometry initialized before starting the simulation. A typical run of 10^4 events with the default detector geometry has a disk size of 135 MB.

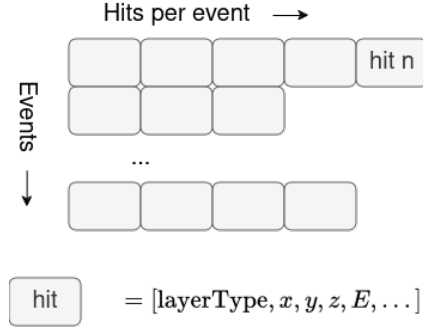


Figure 4.3.: Structure of the detector hits as a sparse array. A hit contains the information about the type of layer, the position of the center of the cell and the recorded energy without digitization.

4.4. Event Generation

The overarching goal of this thesis is to reconstruct ALPs at beamdump experiments from their signature in the detector. As such, the signature of an ALP decay is plainly recreated by producing two initial photons sharing a common vertex. Afterwards, the two photons shower in the detector and their hits are recorded. This section details the parameters used during the simulation and the a priori assumptions made on some of the generator distributions for the momenta and origin of the photons. In order to investigate the impact of the distance between AT-layer on the reconstruction, this section considers two data-sets, one with a TA distance of 30 mm and another one with a distance of 100 mm. The following description of the simulation applies to both data-sets.

The vertex position of the ALP decay is sampled from a uniform distribution along the z -axis of the detector $V_0 = [0, 0, V_z]$, where $V_{\max} < V_z < V_{\min}$. For the momentum direction \vec{p} of the first photon, the cylindrical coordinates (r, ϕ, z) are best suited, with r the radial distance from the z axis and ϕ the rotation along z . The angular component p_ϕ is sampled uniformly from $[0, 2\pi]$, while the radial component r is sampled from a Wald distribution, $\mathcal{W}(\mu, \lambda)$ (see Appendix A.1). Finally, the Cartesian components of \vec{p} are scaled appropriately such that the photon always hits the detector surface,

$$p_x = r \sin(\phi) \frac{d_x}{2V_z}, \quad p_y = r \cos(\phi) \frac{d_y}{2V_z}, \quad (4.1)$$

where d_x, d_y are the height and width of the detector respectively. The second photon is simply reflected along the z -axis with momentum $\vec{p}_2 = [-p_x, -p_y, p_z]$. In this way, the two photons share the same vertex V_0 , have the same opening angle with respect to the z axis and lie on a plane which intersects with the z axis. The resulting distribution of photons crossing the boundary of the detector (the first absorber layer) at $z = 0$ can be seen in Fig. 4.4. The actual four-momentum is then normed by GEANT4 and rescaled by the energy of the particle,

$$p^\mu = \left[m \quad \frac{\vec{p}}{|\vec{p}|} E \right], \quad (4.2)$$

as such, the direction of the particle is dictated by \vec{p} , but the total energy is always equal to E .

All these generating distributions for \vec{p} and V were chosen in order to facilitate the reconstruction of the trajectory for the neural network afterwards. They represent a particle decaying along the z axis with momentum conservation along the x and y components. The choice of the Wald distribution ensures that there are no events where the angle of the photons is too small and have too much overlap.

In both data-sets used for the performance demonstration, the two photons originate from a uniform distribution with $-1400 \leq V_z \leq -10$ mm in front of the detector. The transverse momentum is sampled from a Wald distribution $\mathcal{W}(0.5, 1)$. This results in the xy -distribution at the detector surface shown in Fig. 4.4. The energy of both photons is 2 GeV each, which is at the higher end of the expected energy range of an ALP at LUXE-NPOD.

The detector geometry is built as described in Table 4.1 with 5 AT-layer and a gap between each AT-layer (named Tracker-Absorber (TA) distance in the following). Data-set A has a Tracker-Absorber (TA) distance of 30 mm and data-set B a TA distance of 100 mm. An exemplary event in both cases can be seen in Fig. 4.5 and Fig. 4.6 respectively.

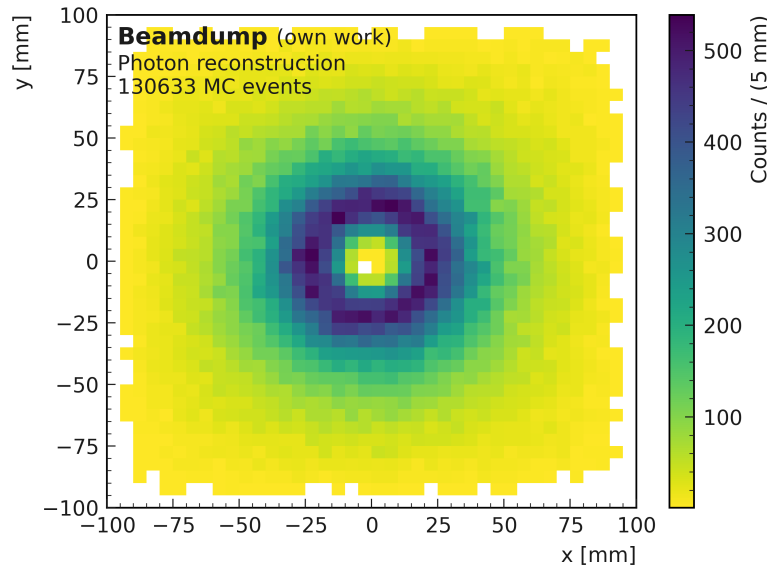


Figure 4.4.: Distribution in the xy -plane of particles crossing the surface of the detector in Model A, according to the generation from Section 4.4. The central region is avoided during MC generation in light of the event selection criteria $\theta \leq 2^\circ$ of Section 4.5.

4.5. Selection Criteria

After the simulation, a series of selection criteria are applied in order to have a homogeneous dataset with two well resolved photon showers. Events with only one showered photon, too few hits or only hits in the calorimeter are excluded from the data-sets, as they will negatively impact on the performance of the reconstruction algorithm. The followings steps are performed before the selection:

1. Hits in the absorber are ignored.
2. All hits in the same calorimeter cells have their energies summed over. This means that the sourcePID information is lost in this step.
3. Only keep events with at least one hit.

The first selection step then removes events with total energy lower than $E < 200 \text{ MeV}$ in order to reduce the number of edges cases where for example one of the photons does not shower in the detector. Afterwards, events where the two photons have an opening angle smaller than $\theta < 2^\circ$ are removed. The opening angle is defined as

$$\theta = \arccos \left(\frac{\vec{p}_1 \cdot \vec{p}_2}{|\vec{p}_1| |\vec{p}_2|} \right), \quad (4.3)$$

where \vec{p}_1 and \vec{p}_2 denote the momenta of photon 1 and 2. The shape of the Wald distribution helps to reduce the number of events that are produced with very shallow angles. Finally, events with less than 20 hits are discarded, as they provide insufficient information for the reconstruction. The acceptance on both data-sets A and B can be seen in Table 4.2. The corresponding distribution of the vertex position along the z axis of the detector are shown in Fig. A.3 and Appendix A.2.

Table 4.2.: Number of events passing the selection criteria of Section 4.5 for Model A (TA distance of 30 mm) and Model B (TA distance of 100 mm). The low number of events that pass the minimal angle selection in Model B is linked to the minimal energy requirement. Events that originate closer to the detector with a large opening angle are more prone to lateral shower leakage. Events originating from further away with a shallow opening angle have showers that are more likely to be fully contained.

Selection	Training set		Testing set	
	Model A	Model B	Model A	Model B
No cuts	200000	200000	40000	40000
$E_{\min} > 200 \text{ MeV}$	159816	150182	32096	31487
$\theta_{\min} > 2^\circ$	130846	102157	26515	21593
$N_{\text{hits}} > 20$	130633	102029	26455	21553
Number hits	11572033	8623723	2321682	1808487
Acceptance	65.32%	51.01%	66.14%	53.88%

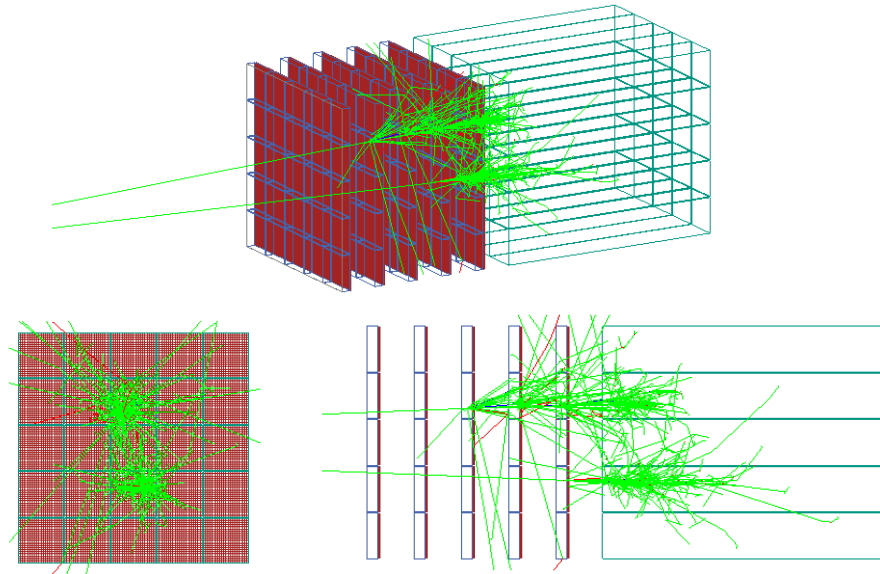


Figure 4.5.: Model A: Two photons with 1 GeV each originating from a common vertex point $V_z = -660$ mm. The detector has 5 Absorber-Tracker Layers and there is a gap of 30 mm behind each AT-layer. In this configuration, the total detector length is 465.6 mm.

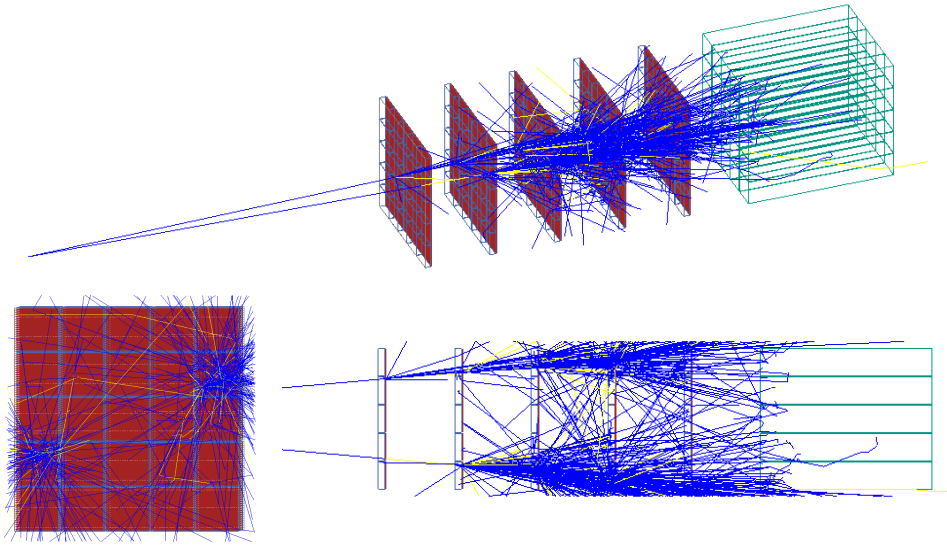


Figure 4.6.: Model B: Two photons with 1 GeV each originating from a common vertex point $V_z = -644$ mm. The detector has 5 Absorber-Tracker Layer (AT-layer)s and there is a gap of 100 mm behind each AT-layer for a total detector length of 806.6 mm. In contrast to the case with a Tracker-Absorber distance of 30 mm, there is considerable lateral shower leakage.

5. Graph Neural Networks

Machine Learning (ML) has experienced a steady increase in popularity in HEP due to its capacity to extract high-level information from large data sets with little human intervention. This chapter describes the aspects of machine learning relevant for this thesis in the context of particle reconstruction algorithms. Section 5.1 gives a short introduction to the representation of detector hits as a weighted graph. Section 5.2 explains the working principles of two graph neural network layers, namely GravNet and GarNet. Section 5.3 shows the implementation of a loss function which enables the identification of two objects and any number of high level properties. Finally, Section 5.4 details the neural network architecture used in this thesis for the purposes of reconstructing photons.

5.1. Machine Learning and Computer Vision

A comprehensive introduction to ML and Graph Neural Network (GNN) is given in [35, 36], which this Section is based on. Machine learning algorithms are generally a set of parameterized functions which are tasked with reproducing a target distribution based on training data. During the training phase, the algorithm adjusts these parameters and aims to improve a certain metric. This metric called the loss function is calculated after each step based on the output of the algorithm at that time. During the back-propagation phase, the parameters are updated with the goal of improving the loss in the next cycle. A second validation set can be used to quantify the current performance of the algorithm on previously unseen data. Once the algorithm is trained to a satisfying degree, its performance is gauged on a third testing data set.

The goal of training a machine learning algorithm is to learn the general characteristics that explain the features seen in the training data. However, after training for too long, the algorithm may also learn the statistical fluctuations of the training sample and lose its predictive power on unseen testing data. This phenomenon is called over-training and must be mitigated by stopping the training phase early, decreasing the number of trainable parameters or by increasing the size of the data set, amongst other methods. A fully trained machine learning algorithm can then be used at will to infer the properties of new data sets. This provides large scalability as the inference step is computationally inexpensive.

The use of machine learning in high energy particle physics is strongly motivated by the ability of Monte Carlo (MC) methods to produce large sets of exactly labelled training data. These methods allow changes in both the underlying physics models and the modelling of

the detector itself. Moreover, the rate of data produced by large experiments are quickly outpacing traditional computational methods. For example, the High Luminosity Large Hadron Collider (HL-LHC) upgrade is expected to produce five times more data per year and will require four times the projected available CPU resources in the next three years [37].

Traditional approaches like fully connected or Convolutional Neural Networks are often confined by the structure of input data. The sparse nature of detector hits in turn favors a geometric approach with Graph Neural Network (GNN)s. Indeed, an event can in general be represented as a cloud of points, with each hit being defined by a point with certain properties. For example, their spacial relation is given by the 3D coordinates of the tracker sensors or calorimeter cells. A graph network associates each hit with a node in the graph and connects all nodes with edges quantifying the information exchange between them. Layers in the network then accumulate information either locally or globally and reweight the edges of the graph after each learning step. This back-propagation step boosts the exchange of information along edges that impact the network output and reduces the exchange along unimportant edges.

While having a complex structure, GNNs are usually based on a large set of simple dense layers. A dense layer is a machine learning block which fully connects a set of nodes (e.g. the output of a previous layer) to a set N of new nodes, with linear edge weights $y = ax + b$. The output of the layer is then passed through an activation function that acts as a threshold for the node. Activation functions are used to introduce non-linearity to the layer and dynamically suppress the output of nodes or restrict their codomain. A common example is the rectified linear unit (ReLU) [38, 39], which is $f(x) = x$ for $x > 0$ and $f(x) = 0$ otherwise. This allows the layer to constrain its output to the positive real domain and prevents vanishing gradients during back-propagation. The Exponential Linear Unit (ELU), defined as $f(x) = \alpha(e^x - 1)$ for $x \leq 0$ and $f(x) = x$ for $x > 0$, provides some computational advantages over ReLU [40]. The inclusion of negative values in the ELU function, although suppressed, contributes to the batch normalization of the layer and is shown to improve the stability of the training. Most often, $\alpha = 1$ is used to ensure a smooth gradient at $x = 0$, which prevents statistical noise when the node output oscillates around the threshold value at $x = 0$.

Hyperparameters are global variables that dictate the shape of the layers of the network. For example, the activation parameter α of the previous dense layer can be considered a hyperparameter of the whole network. Other examples include the number of nodes per layer, the total depth of the network or optimization algorithm. Choosing an adequate set of hyperparameters is paramount for an optimal network performance. They can either be chosen by hand or fine-tuned by a dedicated hyperparameter optimization program [41]. Furthermore, the number of training epochs, the learning rate and batch sizes are also hyperparameters. Most often, the training set is divided into subsets called batches in order to reduce memory consumption. They also allow the network to train faster by updating the weights more often during training. Batch norm layers re-normalize the input data after each step for a more stable training [42].

5.2. GravNet and GarNet Layers

GravNet and GarNet are two GNN architectures first that were first introduced in the context of energy clustering for the High Luminosity Large Hadron Collider (HL-LHC) [7]. Their implementation relies on building a new graph based on the features learned from the input graph. In this way, the new graph is a representation of the original graph in a learned space, and is more efficient at pooling together nodes that have a large information exchange.

Both layers are tailored to data in the form of tensors of rank (E, H, F) , where E denotes the event index, H is the hit index in each event and has a dimension of variable length (named *ragged* in the following), and F is the index labelling the features contained in each hit. F can be the coordinates of the sensor cell, the deposited energy or additional information of each hit. In the following, the symbols \otimes and \oplus denote the tensor outer product and concatenation respectively.

GravNet

With GravNet, a new graph is formed based on the input data that aims to cluster together nodes with larger information exchange. The learned features are then reweighted by a potential of the distance between nodes from the graph. The information flow in the GravNet layer can be seen on Fig. 5.1. In detail, GravNet takes as input a tensor $F_{\text{in}} (E, H, F)$ and proceeds as follows:

1. The spacial representation $F_S (E, H, S)$ is created by a dense layer of S output nodes, which can be identified as the coordinates of a new graph in an abstract space. The number of dimensions of the new graph S is a hyperparameter of the layer.
2. The learned representation $F_{\text{LR}} (E, H, L)$ is produced by a dense layer with L output nodes, based on the features F of the input graph.
3. An unconnected graph is built by placing nodes at the coordinates S with their associated features L .
4. The graph is then connected by weighting the edges with $w = \exp\left(-cd_{ij}^2\right) (E, S_i, S_j)$, where d_{ij} is the euclidean distance between two points i, j in S space and $c = 10$ is a scaling coefficient dictating the strength of the potential. The number of connected neighbors k of each node is a hyperparameter of the layer. This step is performed by a k -Nearest-Neighbor (KNN) algorithm, which is explained in more detail at the end of this Section.
5. The learned features are scaled by the edge weights $\tilde{F} = w \otimes F_{\text{LR}} (E, H, L, S)$.
6. The new set of features $\tilde{F}_{\text{LR}} = F_{\text{in}} \oplus \tilde{F}_{\text{mean}} \oplus \tilde{F}_{\text{max}} (E, H, F + 2LS)$ condenses the local and global information from the graph.
7. A final dense layer $F_{\text{out}} (E, H, O)$ with \tanh activation and O output filters yields the output of the layer.

GarNet

The GarNet layer as detailed here is a custom implementation based on the non-ragged version from the Qasim et al. paper. The original version of the layer is restricted to regular fully padded arrays. Since the data structure of this work is highly irregular, the GarNet was rewritten completely from the ground up for sparse arrays. In contrast to GravNet, the input nodes only exchange information with a new set of nodes named aggregators. In the first step, a dense layer learns a weight w for the edge between the node $i \in F_{\text{in}}$ and the aggregator $s \in S$. Visually, this can also be seen as a distance-weighted potential similar to GravNet. The amount of aggregators is an important hyperparameter and can be used to funnel the input data into a desired shape. Next, a set of learned features F_{LR} are generated by a dense layer. These features are weighted by w , pooled together and exchanged with S . The result is then processed by a final dense layer F_{out} .

Let $F_{\text{in}}(E, H, F)$ be the tensor representing the input data as introduced at the start of the section. The GarNet algorithm then proceeds as follows:

1. The distance $d(E, H, S)$ is learned by the first dense layer of S nodes, with S being the number of aggregators.
2. The edge weights $w(E, H, S)$ scale d by a gravitational potential $w = \exp(-d^2)$.
3. The learned representation $F_{LR}(E, H, P)$ is generated by the second dense layer of P propagators nodes.
4. The information from w and F_{LR} for each event are combined into $\tilde{F}_{LR} = w \otimes F_{LR}(E, H, S, P)$
5. \tilde{F}_{mean} and $\tilde{F}_{\text{max}}(E, S, P)$ then take the mean and maximum of \tilde{F}_{LR} along the H index in each event.
6. The updated features are reweighted $F' = w \otimes (\tilde{F}_{\text{mean}} \oplus \tilde{F}_{\text{max}})(E, H, S, 2P)$.
7. The last two indices are contracted to $F'(E, H, 2PS)$.
8. The third and final dense layer of O nodes takes as input $F_{\text{in}} \oplus F'(E, H, F + 2PS)$ and outputs $F_{\text{out}}(E, H, O)$ with tanh activation function.

The variables S , P and O are hyperparameters of the layer.

K-Nearest Neighbor Algorithms

A k -Nearest-Neighbor (KNN) algorithm considers the k nearest points in the vicinity of an input point according to a given distance metric, usually the euclidean L^2 distance. It then takes a weighted aggregation operation of all the features from these points, such as an average or the maximum value. After calculating the aggregation value for each point in the graph, the resulting graph quantifies the local information present around each point. This type of algorithms performs best in scenarios where the distance between two points is strongly correlated with the features of the nodes, which means that points that are closer together have similar properties. This assumption works well in many applications but can be improved further by considering the type of data that is fed to the network. For example,

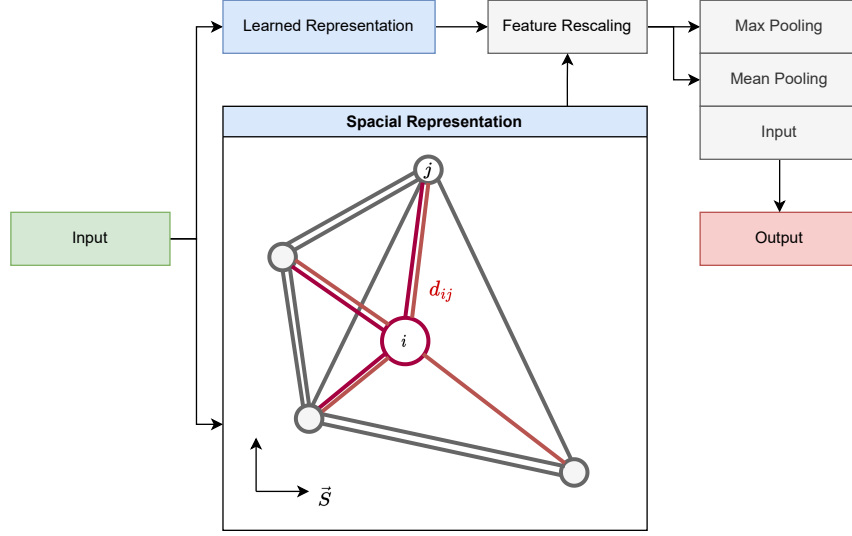


Figure 5.1.: Information flow in the GravNet layer. The input graph connects to two dense layers: the first is a new spacial representation of the graph that is fully learnable. The second layers is a learned representation of the features of the input graph. During the feature rescaling process, the learned representation is clustered by a KNN. The edge weights used by the KNN are the distances in the spacial representation. A final max and mean pooling reduce the dimension of the graph into the output shape.

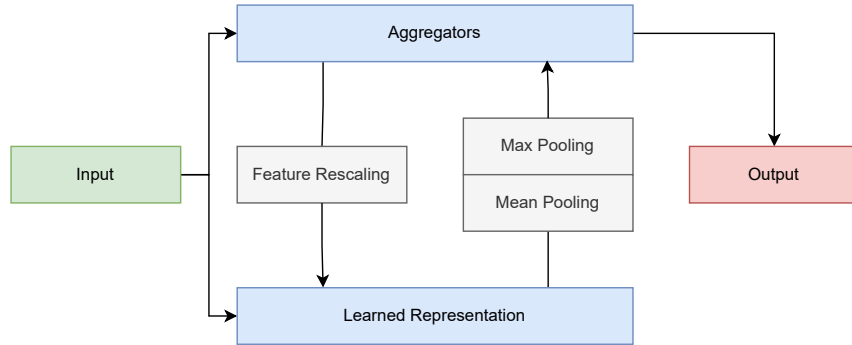


Figure 5.2.: Information flow in the GarNet layer. Two dense layers are connected to the input graph. The first is a learned representation, similarly to Fig. 5.1. The second layers learns the distance between each node and a set of new aggregator nodes. The features of the aggregator are trained based on the learned representation nodes. Again, a max and mean pooling condense the information of the graph into the output shape.

the background noise in the detector should be efficiently excluded from the aggregation step of any KNN algorithm, as it will otherwise propagate into reconstruction errors. The same applies to overlapping clusters of hits, which are spatially near but otherwise share no information, and pollute each others data points. GarNet learns the edge weights with a dense layer instead of using a KNN algorithm, but essentially performs the same clustering step. GravNet's key improvement is to learn the coordinates of a new graph and then to cluster with a KNN in this new representation space.

5.3. Training Targets and Loss Function

The loss function $L : \mathbb{R}^N \rightarrow \mathbb{R}^+$ quantifies the deviation of the current output $p(E, N)$ of the network with the target $t(E, N)$, where N is the dimension of the target distribution. In reconstruction tasks, the loss function is often a variation of the L^1 or L^2 distances between prediction and target. An important aspect when reconstructing two separate objects is that the arrangement in which hits are labelled as belonging to photons 1 and 2 will change from event to event. In mathematical terms, the true features are the concatenation of the features of photon 1 and 2, $t^{12} = t^1 \oplus t^2$. The order of the predicted features however is unknown, and the loss must be compared between both cases $p^{12} = p^1 \oplus p^2$ and $p^{21} = p^2 \oplus p^1$. The L^2 loss under consideration of photon permutation is

$$L = \sum_e \min \left(\sum_f (p^{12} - t^{12})^2, \sum_f (p^{21} - t^{12})^2 \right). \quad (5.1)$$

The training targets t^i for each photon are three spatial coordinates

- V_i : Vertex position
- A_i : Impact point of the photon with the first absorber layer (at the $z = 0$ mm plane, $A = [a_x, a_y, 0]$).
- B_i : Intersection of the photon path with the center of the calorimeter. Since the position of the calorimeter depends on the thickness of the preceding layers, this is defined as the plane at $z = 300$ mm ($B = [b_x, b_y, 300]$).

The total training target is $t^{12} = [V_1, A_1, B_1, V_2, A_2, B_2]$, with a total of 18 variables. This is the shape of the final layer of the network. The output shape could be reduced to 12 by omitting the point B_i at the calorimeter center as they are linearly dependent on V_i, A_i . However, they provide an additional visual control if V_i, A_i and B_i are not aligned for each photon i .

The two samples A and B generated in Section 4.4 has several more constraints, with the two photons sharing a common vertex and plane on the z axis. This further reduces the number of independent parameters in the loss function to $L = [v_z, a_{1,x}, a_{1,y}]$ at almost no extra cost.

5.4. Architecture for Photon Reconstruction

The reconstruction algorithm was implemented with the CMS-PEPR/HGCALML [4] software framework which builds on the DEEPJETCORE (DJC) [43] machine learning

package. DJC is a KERAS-based python package [6] with a TENSORFLOW C++ backend [5], designed to provide fast algorithms for Graph Neural Networks on GPUs. The segmentation of training data sets into GPU-readable batches is done automatically. DJC includes custom implementations of k -Nearest-Neighbor (KNN) algorithms, which are essential for the information clustering steps of GravNet.

The network used in this work is a deep learning neural network with a total depth of eleven layers. It is composed of two GravNet and three GarNet blocks. The combination of these layers is motivated by the reconstruction goals. The GravNet layers are shown to be excellent at gathering information from large sets of data. However, the number of clustering points is generated dynamically for each graph. This is a feature that can be exploited further using object condensation [44]. However, the large computational overhead added by object condensation is better suited to events with a high track multiplicity, which is not expected at LUXE-NPOD. Another solution is to use a GarNet layer where the number of aggregators S is specifiable. Whether the aggregator $s_i \in S$ can be directly identified with either of the photons in the loss function remains open for further studies. Nonetheless, the inclusion of GarNet layers in the network leads to better performance.

The network starts with a batch norm layer and is followed by two GravNet blocks. Each GravNet block is composed sequentially of two dense layers, a GravNet layer and a batch norm. These blocks mirror the layout detailed in the original paper. The number $N = 100$ of nearest neighbors in both GravNet layers is based on the expected number of hits in LUXE-NPOD detector in the order of $\mathcal{O}(100)$. The remaining hyperparameters of the layer are motivated by the original paper.

Next, there are the three sequential GarNet blocks each composed of one GarNet layer and a batch norm. The number $S = 2$ of aggregators in the GarNet layer is based on the expected number of physics objects, in this case two photons. The number of propagators $P = 128$ is chosen to be large, in order to keep the most information in each aggregator object. Notably, the output of all three GarNet blocks are combined into a final output list as in the original paper. This ensures that *coarse* information gathered by block 1 is not diluted by blocks 2 and 3.

The complete layout of the network can be seen in Fig. 5.3 and is listed as follows:

1. Batch Normalization
2. Two sequential GravNet blocks with each:
 - a) Two Dense layers with 64 nodes each and ELU activation
 - b) GravNet layer with $N = 100$ neighbors, $S = 4$ dimensions, $L = 64$ learned features, $O = 64$ output filters and ELU activation for the learned representation F_{LR} layer.
 - c) Batch Normalization
3. Three sequential GarNet blocks with each:
 - a) GarNet layer with $S = 2$ aggregators, $O = 64$ output features and $P = 128$ propagators.

b) Batch Normalization

The final output is a list with the output of blocks one, two and three concatenated together.

4. Dense layer with 128 nodes and linear activation
5. Dense layer with 18 outputs and linear activation (final output layer)

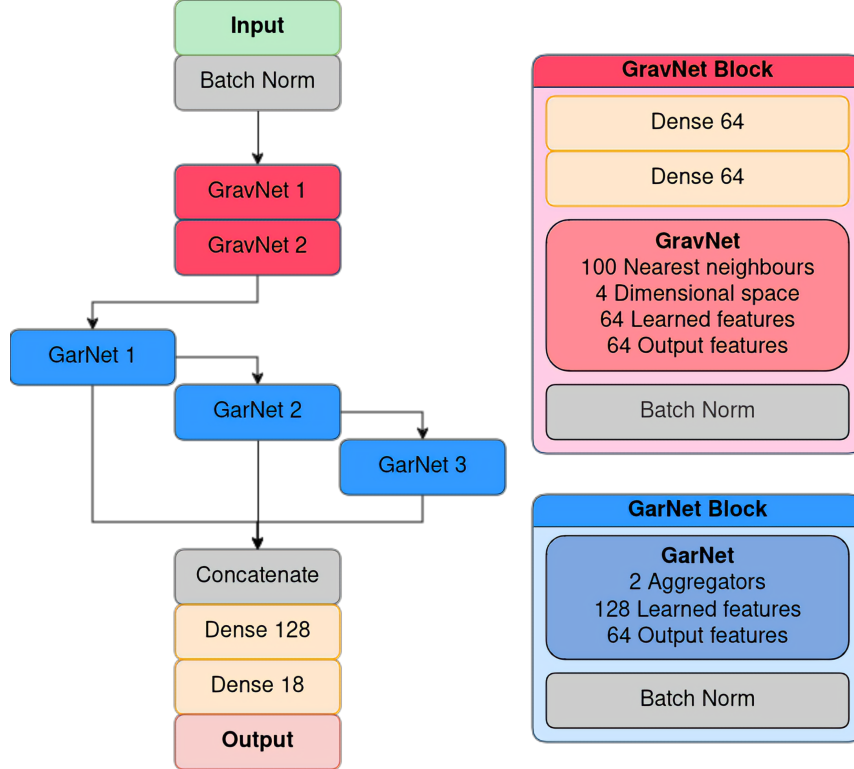


Figure 5.3.: Visualisation of the information flow for the neural network used for photon reconstruction. White background indicates a non-trainable layer, orange are dense layers and light blue are the graph network layers. The initial data is fed twice through the GravNet block. Afterwards, the output of each GarNet block is passed to the next block and all three are concatenated into a list.

5.5. Model Training

The network is trained on an NVIDIA TITAN X GPU with 12 GB memory supported by an Intel Xeon E5-2630v4 CPU. The loss is minimized with stochastic gradient descent using the Adam optimizer [45]. With the DJC software, 20% of the events from the training data are used as the validation set.

Two separate models are trained for the data-sets generated in Section 4.4. Model A and B are trained on data-set A and B with a Tracker-Absorber distance of 30 mm and 100 mm respectively. In order to investigate the effects of the GNN training on the reconstruction performance, models A and B have slight differences in their training.

The two models are first compiled with a learning rate of 10^{-3} . Afterwards, there are five training sections with decreasing learning rate. This evolution promotes the fast convergence to the region of the minimum at first and subsequently slows down in order to accurately find a minimum in the loss space. The complete training process for both models is detailed in Table 5.1. The weights of the model are recorded every 10 epochs and the version with the lowest training loss is kept as the best model. The loss progression is shown in Fig. 5.4.

Table 5.1.: Training steps of Model A (with a Tracker-Absorber distance of 30 mm and Model B (TA distance of 100 mm). Model A is trained for a total 410 epochs with a large batch size. Model B is trained on 360 epochs but with a constant and smaller batch size.

Epochs	Learning rate	Batch size	
		Model A	Model B
10	10^{-3}	5×10^3	5×10^3
50	10^{-4}	10^4	5×10^3
50	5×10^{-5}	10^4	5×10^3
100	10^{-5}	10^4	5×10^3
150	10^{-6}	5×10^3	5×10^3
50	10^{-6}	5×10^3	

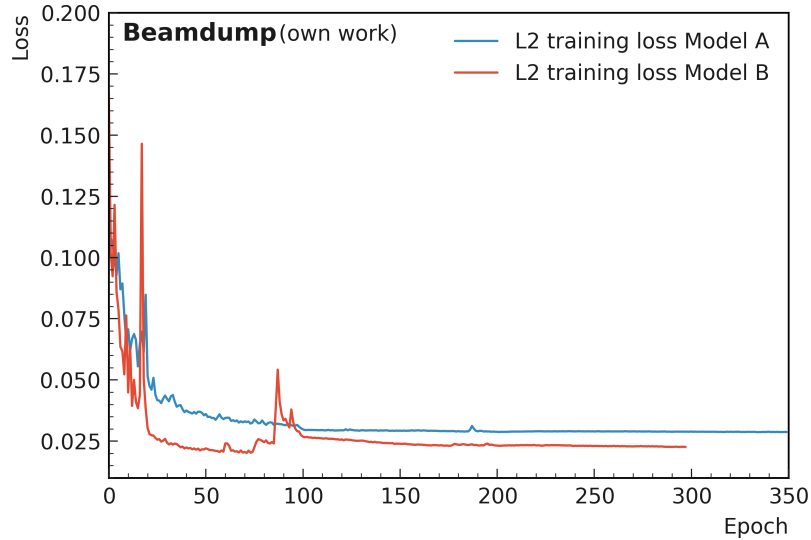


Figure 5.4.: Training loss progression of the last 350 epochs of Model A and 300 epochs for Model B. The larger batch size of Model A leads to a more stable training with minimal jumps in the loss. On the other hand, Model B with the smaller batch sizes has several jumps.

6. Photon Reconstruction with GNNs

This chapter presents the results of the reconstruction algorithm on two testing MC data-sets generated according to the process detailed in Section 4.4. The information gathered from the detector hits is used to find the common vertex of two photons. The goal of this chapter is to show the performance of the network in a proof of concept scenario. Section 6.1 shows the reconstruction output for a single event. In order to measure the performance over a whole testing data-set, Section 6.2 introduces a metric for the resolution. This is followed by Sections 6.3 to 6.5 that show the resolution of the radial, angular and vertex accuracy of the network, respectively. Each section also addresses the limitations of the data-sets and proposes adjustments for further studies. Section 6.6 concludes the chapter with a summary of all the results.

6.1. Event View

This section provides a qualitative view of the reconstruction of two photons with a common vertex. The two version of the network are detailed in Table 5.1 and are trained on two different geometries of the detector introduced in Chapter 4.

Model A is the result of training the GNN with a large batch size for 410 epochs on the data-set generated with a spacing of 30 mm between each high resolution layer, named the TA distance in the following. Model B is trained with a smaller batch size but with a larger TA distance of 100 mm. Both models are explained in more detail in Section 5.5.

Two exemplary visual displays of the predicted and true trajectories of two photons from the testing data-sets of Models A and B are shown in Fig. 6.1 and Fig. 6.2. The figures show each hit as a colored point in the tracking system (blue) and the calorimeter cells (red). The size of each hit scales with the recorded energy fraction with respect to the maximum energy of its layer. In both figures, the true MC and predicted trajectories are shown to be in good agreement.

Several observations are made:

1. The energy of each hit is taken into account by the GNN during clustering.
2. The GNN correctly emphasizes hits in the tracker over those in the calorimeter. This balancing would be difficult to achieve with a classical algorithm.
3. Some outlier hits originate from showering particles that back-scatter from the absorber plates.

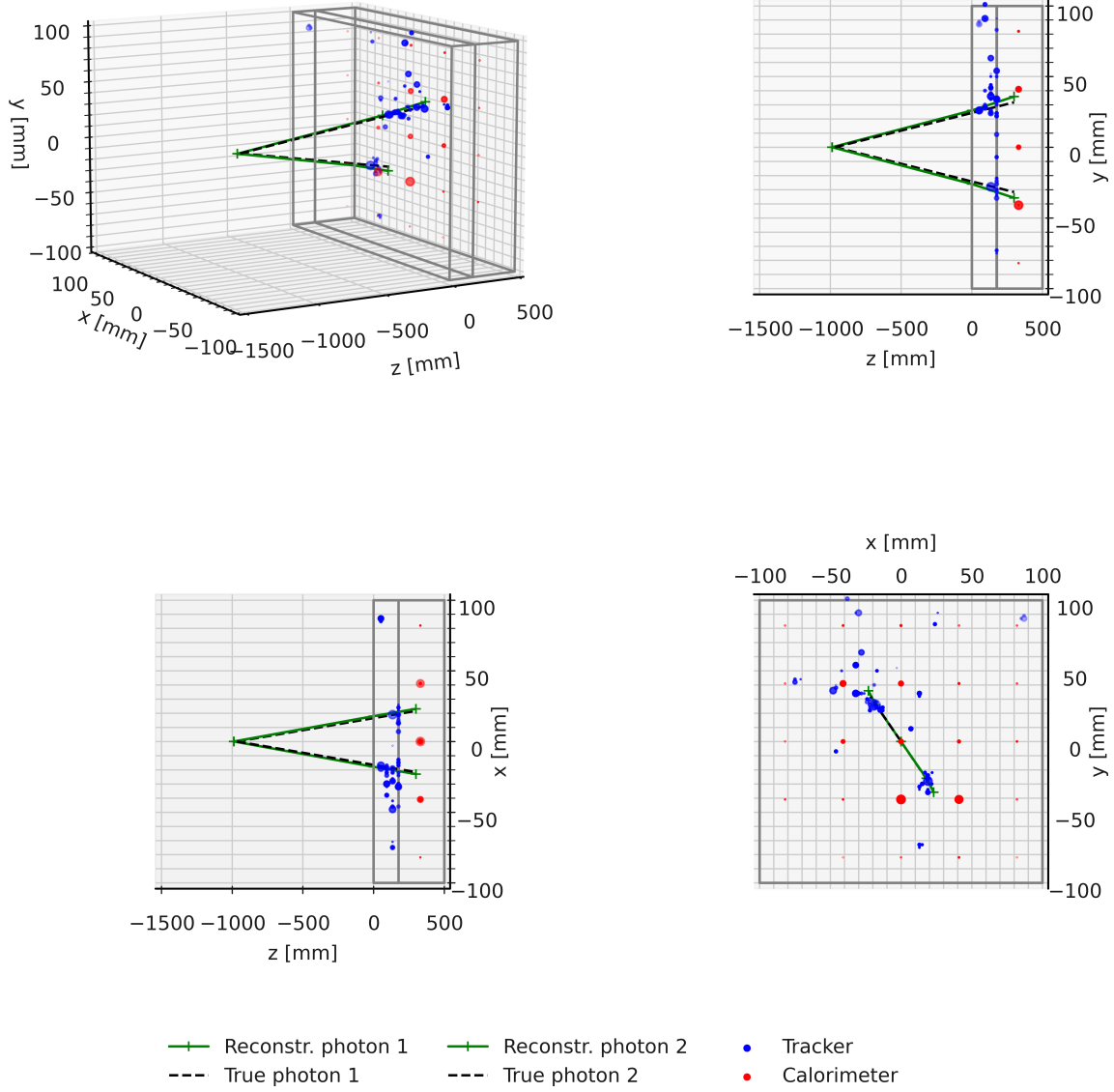


Figure 6.1.: Model A: Successfully reconstructed event from the Monte Carlo (MC) testing data-set with Tracker-Absorber (TA) distance of 30 mm. The true trajectories of both photons are displayed with the dashed lines and the MC information is $V_z = -965.73$ mm and $\theta = 3.47^\circ$. The predicted tracks of the Graph Neural Network are shown in green and are based on $V_z = -986.92$ mm and $\theta = 3.66^\circ$. The x and y axis are scaled by a factor of 5 in order to properly show the detector hits. The detector itself is shown schematically with the two grey boxes (the first being the tracking system and the second the calorimeter).

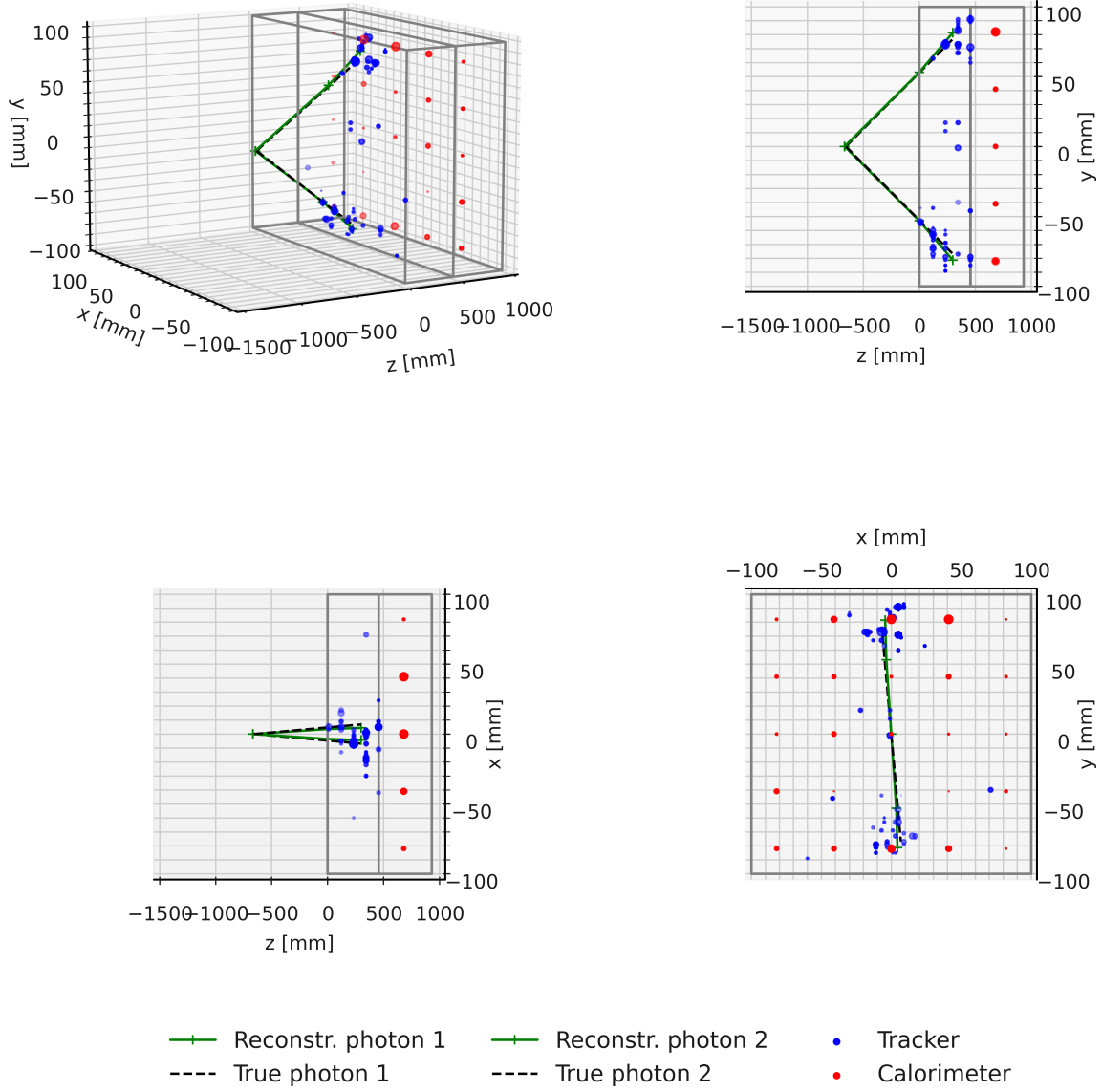


Figure 6.2.: Model B: Successfully reconstructed event from the Monte Carlo testing dataset with Tracker-Absorber distance of 100 mm. The true trajectories of both photons are displayed with the dashed lines and the MC information is $V_z^{\text{true}} = -655.82$ mm and $\theta^{\text{true}} = 9.23^\circ$. The photon directions predicted by the Graph Neural Network are shown in green and are based on $V_z^{\text{true}} = -666.32$ mm and $\theta^{\text{true}} = 9.12^\circ$.

6.2. Accuracy Measure

The regression accuracy of an output variable s by the GNN can be defined based on the residual $\Delta s = s^{\text{true}} - s^{\text{pred}}$ of the deviation of the true MC information from the predicted value. This distribution is in general the convolution of all the output distributions of each layer in the network weighted by the corresponding activation function. It is therefore best to characterize the resolution of the network by a fixed percentile instead of the standard deviation. The resolution of a variable s is defined as

$$r(\Delta s) = q_{68\%}(|\Delta s - q_{50\%}(\Delta s)|). \quad (6.1)$$

Where $q_a(s)$ is the a -th percentile of the distribution. This corresponds approximately to the 1σ standard deviation of the normal distribution.

6.3. Radial Accuracy

The impact point of both photons with the detector surface is described by the radial distance to the z axis,

$$R = \sqrt{x^2 + y^2}|_{z=0}. \quad (6.2)$$

Due to the symmetric nature of the demonstration data-set introduced in Section 4.4, this distance R is the same for both photons in each event, as they are reflected along the z axis. It does not coincide with the start of the shower as recorded by the tracker. Indeed, the shower already originates in the absorber and this behaviour should be accounted for by the GNN.

The residuals between the true and predicted radius at the detector surface are shown in Fig. 6.3a for Model A and in Fig. 6.3b for Model B. The overall radial resolution of both models is $r(R) = 2.04 \pm 0.01$ mm for Model A and $r(R) = 1.83 \pm 0.01$ mm for Model B. The dependence of $r(R)$ on V_z^{true} and R^{true} are shown in Figs. A.8 and A.9b.

A notable feature of both models is that the R distribution has a slight offset by $R_{50\%} = -1.1 \pm 0.1$ mm with Model A and $R_{50\%} = -0.8 \pm 0.1$ mm with Model B. This indicates a tendency for the GNN to, on average, overestimate the radial distance of the photons at the detector surface. This could be attributed to a fraction of the shower being reflected and back-scattered by the absorber plate, as seen qualitatively in Figs. 4.5 and 4.6. These particles may then be recorded by the preceding tracker layer. The resulting hits are potentially considered to belong to the primary shower, which will in turn influence the perceived direction of the photons. This behaviour clearly poses a challenge for the GNN and should be considered in the design of the detector. For example, thinner absorber plates could mitigate this back-scattering effect. Inversely, this might reduce the total number of showered particles and recorded hits in the tracker. Since the two Models are trained differently, it is not possible to clearly associate the improvement in radial accuracy of Model B over Model A solely to the increase in the distance between AT-layers.

The region with $R < 5$ mm is not covered in the scope of this work due to the angle selection of Section 4.5. This prevents the events from having overlapping clusters of hits that are more difficult to separate. Events far away from the detector with shallow opening angles

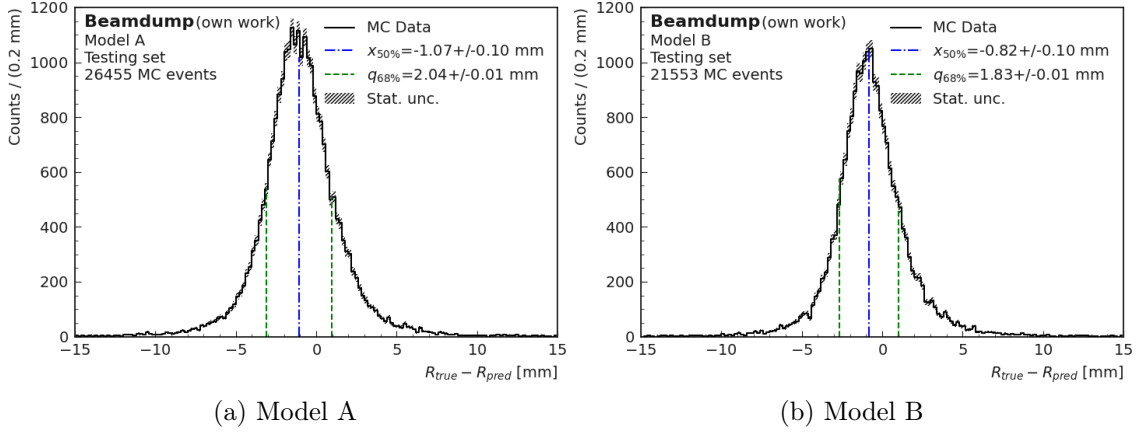


Figure 6.3.: Reconstruction accuracy of the predicted and true radius R from Section 6.3 of the photon trajectory at the detector surface. This metric displays the performance of the GNN to recognize the impact point of the photon with the first absorber plate.

are the most difficult to reconstruct, as even small reconstruction errors in the angle and in the radial distance will compound. Further studies should investigate the changes in performance of the GNN on samples with a large shower overlap. It is worth to mention that other applications of the GravNet layer already work well with overlapping objects [7].

6.4. Angular Accuracy

For Model A, the relation between the true opening angle θ^{true} and the predicted angle θ^{pred} can be seen in Fig. 6.4a. Most events have a small opening angle $2^\circ \leq \theta \leq 15^\circ$ which is accurately predicted by the GNN. However, in the region located closer to the detector, some of the events can have a large opening angle and still pass the angle selection criteria, this is shown in Fig. 6.5a. These edge cases are less frequent, but stand out in Fig. 6.4a due to the logarithmic scale. The decrease in reconstruction performance for these larger angles may be attributed to their low incidence. Fig. 6.6a shows the angular accuracy of the reconstruction which has a resolution of $1.69 \pm 0.01^\circ$. This is equivalent to $29.5 \pm 0.2 \text{ mrad}$, which surpasses the requirements listed in [14].

For Model B, the range of θ^{true} is more limited due to the detector geometry, as seen in Fig. 6.4b and Fig. 6.5b. Events with $\theta^{\text{true}} > 25^\circ$ do not have fully contained showers since the AT-layers are further apart. In the region $14^\circ \leq \theta^{\text{true}} < 25^\circ$, the opening angle is predominantly overestimated by the GNN. The primary photons of these events have peripheral collisions with the detector at a steep angle. If the shower develops mainly outside of the detector, the recorded energy and the number hits will be low. This leads to most of these events to be removed by the selection process. ML algorithms also have diminished predictive power for data-points at the boundaries of their training space, which is exacerbated by the lack of events in the training set in this region. The angular accuracy of Model B is shown in Fig. 6.6b and over the whole range of θ^{true} , the resolution is $r(\theta) = 0.98 \pm 0.01^\circ$.

The opening angle θ of both photons is bisected by the z axis, as described in Section 4.4. This symmetry is an assumption made for the demonstration data-sets and does not accurately depict the general behaviour of a three-body decay. In this sense, it is a simplification that enables the GNN to have twice the information per event regarding the opening angle. A data-set composed of MC generated ALP samples will likely experience a diminished reconstruction performance. In general, their angles with respect to the z axis are not symmetric.

6.5. Vertex Accuracy

The main focus of the reconstruction algorithm is to find the common vertex of the two photons. Since the vertex position $V = [V_x, V_y, V_z]$ is a direct output of the network, it is the most natural metric out of the three considered in this Chapter. The network correctly predicts $V_x = 0$ and $V_y = 0$, which is in agreement with the true MC information and the following discussion will only consider V_z . The vertex position is geometrically the most susceptible to errors and there are also effects that arise solely from the acceptance of the detector, as explained next.

For Model A, the true vertex position along the z axis and the predicted position by the GNN are shown in Fig. 6.7a. The total resolution of Model A is shown in Fig. 6.8a and amounts to $r(V_z) = 172 \pm 1$ mm. This is equivalent to 12% of the whole decay volume length but the actual resolution is highly dependent on the distance from the detector. From Fig. 6.7a, there are two regions of interest with different vertex resolution. The region close to the detector with $-500 \text{ mm} < V_z^{\text{true}}$ is, as expected, the best reconstructed. Events in this region have a wide range of opening angles and are easily reconstructed. A slight smearing can be seen at $V_z^{\text{true}} < -100$ mm due to its location at the boundary of the training space and the low number of events (see Fig. A.3). The vertex resolution in this region is $r(V) = 10.0 \pm 1.0$ cm as shown in Fig. A.6, but decreases with increasing distance to the detector. This is expected, as the angle of incidence of both photons flattens and predicting the vertex position with precision becomes increasingly prone to error propagation. A noticeable change in behaviour arises approximately at $V_z^{\text{true}} = -500$ mm. As can be seen in Fig. 6.5a, the opening angle starts to be constrained by the selection criterion $\theta^{\text{true}} < 2^\circ$. This leads to an abrupt change in the vertex reconstruction in Fig. 6.7a. Contrary to the near region, the resolution in the far region does not increase any further with the distance to the detector.

In the far region with $-1400 \text{ mm} \leq V_z^{\text{true}} \leq -500$ mm, the opening angle is bound upwards by the acceptance of reaching the detector edges and downwards by the selection criterion $\theta^{\text{true}} < 2^\circ$, see Fig. 6.5a. Due to the limited range in the opening angle, events in this far region behave similarly. This effect is amplified by the batch size for Model A. As listed in Table 4.2, the training set contains 1.157×10^7 hits. This equates to 1158 total batches of 10^4 hits each. Since the average event has 88 hits (see Appendix A.2), this in turn indicates that a training step in Model A is performed on 113 events. Such a large batch size leads to a smoother training gradient but also considerably smears the output of the network. Due to this smearing, the network is not able to accurately discern the fine features of events in the far region. As a result, the distribution of events in the far region is strongly skewed towards $V_z^{\text{true}} = -900$ mm.

In contrast, for Model B, the far region with smearing starts further away from the detector at $V_z^{\text{true}} < -850$ mm, as shown in Fig. 6.7b. This is due to the smaller batch size of 5×10^3 hits per batch used in the training process as listed in Table 4.2. This amounts to 1725 batches of 59 events on average per batch, since the number of hits per event is approximately the same as with Model A (see Appendix A.2).

The vertex resolution over the whole testing set for Model B is $r(V_z) = 140 \pm 1$ mm (see Fig. 6.8b). In the far region with $-1400 \text{ mm} < V_z^{\text{true}} < -800$ mm, the resolution is $r(V_z) = 164 \pm 1$ mm, while in the near region the resolution improves to $r(V_z) = 106 \pm 1$ mm (see Fig. A.7).

Based on the results of the vertex accuracy for both models, it is clear that the near region is a more representative sample. The selection criterion $\theta^{\text{true}} < 2^\circ$ constrains the far region too much and the training sample is less representative as in the near region. This can also be seen in Fig. 6.9, where the vertex resolution of both models displays a discontinuity at the boundary where the selection criterion begins to apply to the data-set. This occurs at $V_z^{\text{true}} \approx -500$ mm for Model A and at $V_z^{\text{true}} \approx -850$ mm for Model B. The restriction for the opening angle is intended to prevent overlapping showers. A better approach for this purpose would be to constrain the radial distance at the detector surface instead.

6.6. Summary

In conclusion, it is very feasible to find the direction of photons based only on the detector hits of their showers. This type of reconstruction is made possible by advancements in the representation of sparse input data as graphs. The clustering algorithms used in the GNN group together the information from the important tracks and are able to discard the spacial outliers in each event. By extension, the resulting predicted photon direction is close to optimal and utilizes all the available information of the data.

This Chapter has identified two phenomena that must be avoided in further studies. First, the training data has to be homogeneous and selection cuts should not over-constrain the opening angle. Homogeneity refers to an equal distribution of events in all relevant parts of the parameter space. Second, it is preferable to use smaller batch sizes, potentially at the cost of longer and less stable training. The parameter space for very shallow angles was left out of this analysis but should be addressed in the future. An extensive discussion about possible improvements for further studies is deferred to Chapter 9.

While most of this chapter was dedicated to analysing the shortcomings of the simulation data-set, the main result draws a very positive picture. The primary goal of this thesis is reached and the network shows an impressive performance. In the regions unaffected by artefacts originating from the samples, the network exceeds expectations. Not only is it able to consistently find the two photon showers, it is also very accurate in finding their point of origin. The analysis of the network's performance further away from the detector is found to be inconclusive.

The main take-away of this Chapter is that reconstruction algorithms with GNNs are promising a new tool and should be trained on broad and representative data-sets.

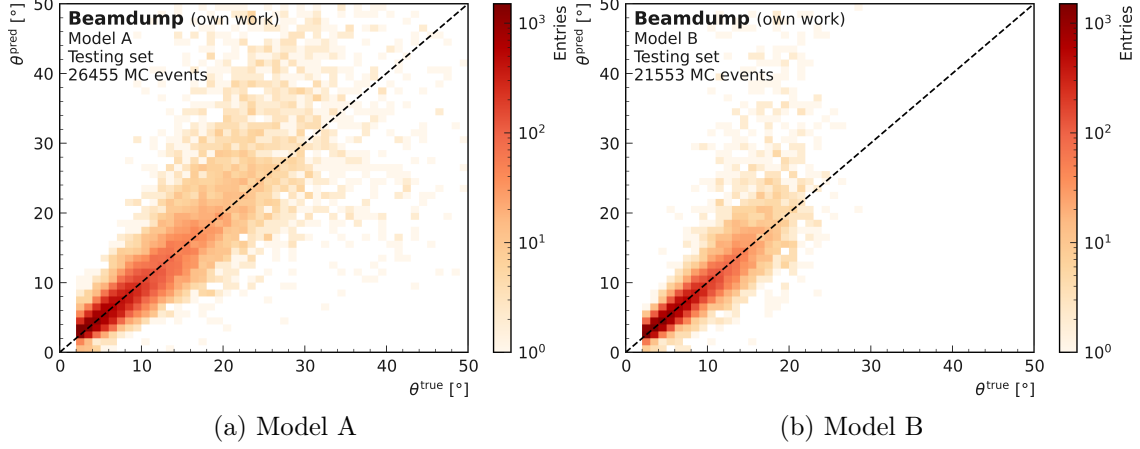


Figure 6.4.: 2D histogram of the true opening angle of the two photons θ^{true} and the predicted opening angle θ^{pred} . In both Models, the performance of the network for shallow opening angles is considerably better than for large angles. The GNN tends to overestimate θ^{pred} at larger θ^{true} . For very shallow angles under $\theta_{\text{pred}} < 5^\circ$, the prediction of the GNN slightly diffuses into the excluded $\theta < 2^\circ$ region. In Model B, due to the longer detector, events must have smaller angles in order to pass the selection criteria.

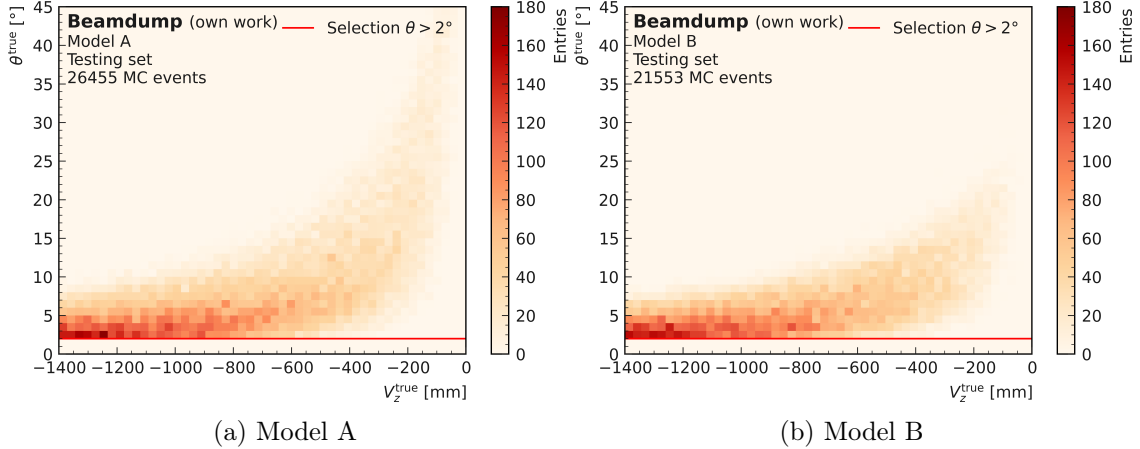


Figure 6.5.: 2D histogram of the dependence of the opening angle θ^{true} on the vertex position V_z . The region further away from the detector is dominated by shallow opening angles while events have increasingly larger opening angles closer to the detector. This is the result of the acceptance dictated by Eq. (4.1).

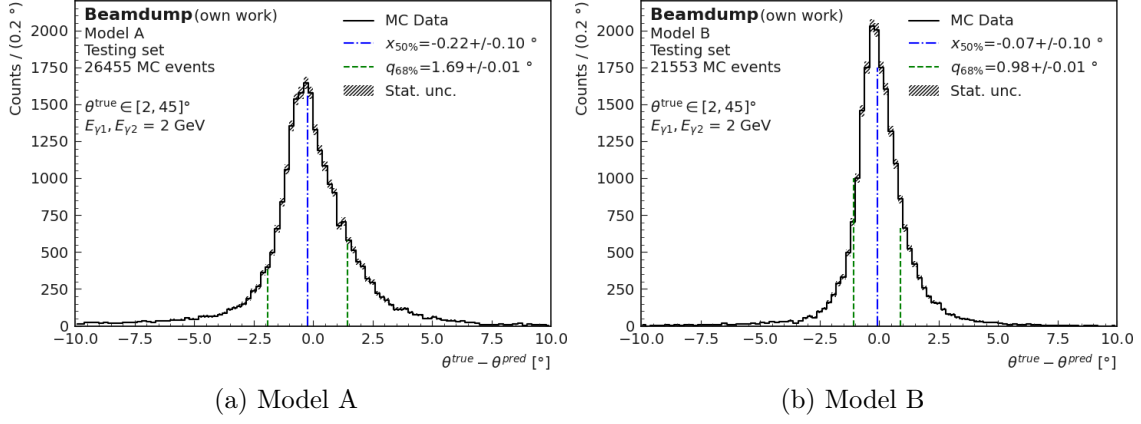


Figure 6.6.: Reconstruction accuracy of the predicted and true opening angle θ^{pred} and θ^{true} . The angular resolution amounts to $r(\theta) = 1.69 \pm 0.01^\circ$ for Model A and $r(\theta) = 0.98 \pm 0.01^\circ$ for Model B over the whole data-set.

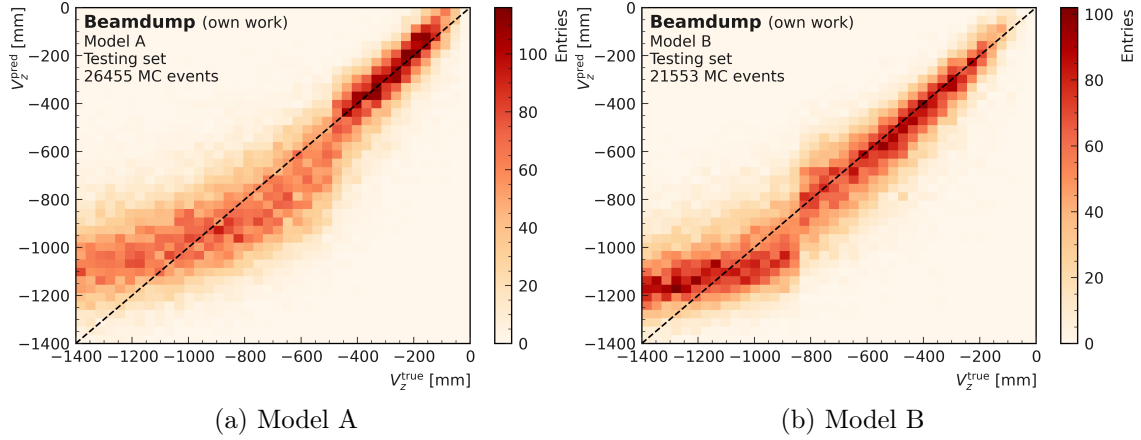


Figure 6.7.: 2D histograms of the true and predicted vertex position V_z^{true} and V_z^{pred} . In both models, the performance of the network is affected by the features of their respective training data-set. The discontinuity appears at $V_z^{\text{true}} = -500 \text{ mm}$ in Model A and at .

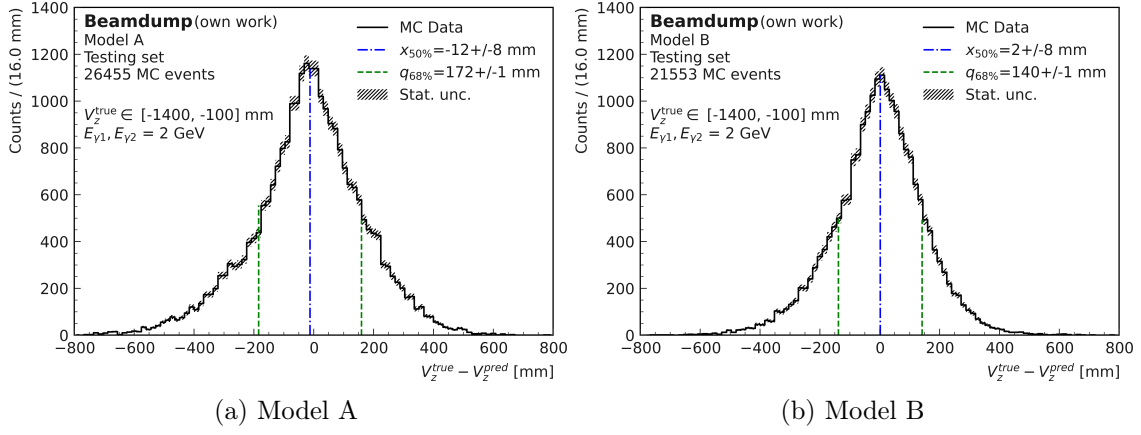


Figure 6.8.: Reconstruction accuracy of the predicted and true vertex position V_z^{pred} and V_z^{true} along the detector axis. The average vertex resolution is 172 ± 1 mm for Model A and 140 ± 1 mm for Model B.

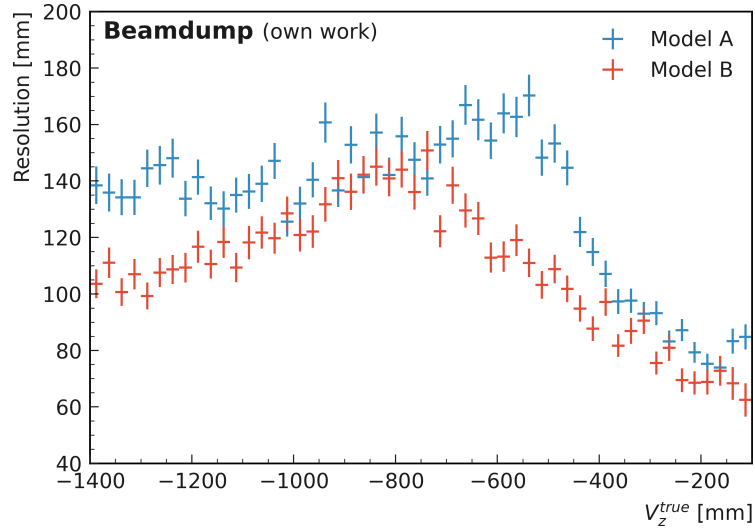


Figure 6.9.: Resolution of the vertex accuracy of the GNN over the range of V_z^{true} for Models A and B.

7. Energy Resolution of Calorimeters

Calorimeters are detector components designed to measure the energy of particles. Generally, this is achieved in two steps: Firstly, the incoming particle interacts with a dense medium which produces a shower of lower energy particles. Secondly, the readout electronics convert the recorded electric charge or number of photons into a signal voltage. Calorimeters can have different shapes and material composition depending on their physics targets. Homogeneous calorimeters have the same material acting as the absorber and the sensitive material. In heterogeneous (or sampling) calorimeters, the absorber material is placed in front of the sensitive material, usually with multiple layers of this arrangement.

Electromagnetic calorimeters rely on electromagnetic interactions between incoming particle and the electron clouds in the absorber to produce photons, electrons and positrons. The physical process involved depends on the energy of the incoming particle: at high energies, pair production dominates and leads to a cascade of gradually lower energy products, while at low energies, particles gradually lose energy via Compton scattering and the photoelectric effect [46].

In hadronic calorimeters, the shower process of the hadrons is more complex and depends on their electric charge. Charged hadrons can have electromagnetic interactions and lose their energy by ionization with the atoms in the absorber material. Neutral hadrons interact with the nuclei in the absorber material and eventually produce neutral pions which decay instantly into two photons. These photons are then recorded by the sensitive material. These effects make hadronic showers more difficult to calibrate for due to their composite nature. Part of the energy is also lost in the form of neutrinos that escape the calorimeter undetected.

7.1. Derivation of the Energy Resolution Equation

The energy resolution is a measure of the accuracy with which the energy of the original particle can be reconstructed. Due to several sources of statistical fluctuations, the readout will vary from event to event. This becomes a limiting factor when determining the invariant masses of particles. This section is based on [47] and re-derives the equation for the energy resolution of a calorimeter subject to noise from physical fluctuations in the shower, electronic noise and the modelling of the detector geometry.

The probability of a shower particle crossing the active material is statistically independent of the initial particle type, provided the preceding absorber layer is thick enough in terms

of radiation lengths. The radiation length is a measure for the energy attenuation of photons (via pair production) and electrons (via Bremsstrahlung) in a material [20]. This means that a particle producing a number n_S of recorded signal electrons through the photoelectric effect obeys a Poisson distribution with standard deviation $\sigma(S) = \sqrt{n_S}$. Since these signal electrons share the total energy E of the initial particle, the relative standard deviation of measuring the mean energy $\sigma(E)/E$ is a function of $\sigma(S)/n_S$. In the case of linear calorimeters, defined by having a constant response $n_S \propto E$, the energy resolution due to statistical fluctuations in shower development is $\sigma(E)/E = a/\sqrt{E}$, where a is a proportionality factor given in $\text{GeV}^{1/2}$.

Electronic noise, also called Johnson-Nyquist noise [48, 49], originates from thermal fluctuations of the electrons in the readout electronics and materializes in the form of a white noise underground. It is independent of the energy of the incoming particle and as such its contribution to the energy resolution is $\sigma(E)/E = b/E$, where b is the magnitude of thermal noise in GeV. For reference, the CMS ECAL endcaps have a thermal noise of 200 MeV in the endcaps which amounts to a power of 32 pW [50]. This source of noise can be efficiently mitigated via cryogenic cooling.

Lastly, the constant term is the dominating source of noise at high energies and stems from the geometry of the calorimeter itself. This can occur from lateral shower leakage. In sampling calorimeters, it appears from the effects of the chosen geometry on the sampling fraction, such as the thickness of each absorber plate and the total number of layers. Irregularities in the detector such as varying densities or misaligned components are also sources of constant noise. This noise scales linearly with the energy of the incoming particle and adds to the energy resolution a term $\sigma(E)/E = c$, where c is a dimensionless constant for any given detector geometry.

Additional sources of noise that occur in a real calorimeters but are not considered here are: radiation damage, temperature gradients, detector aging and structural non-uniformity. Since the three terms are statistically independent, their relative standard deviations can be added quadratically, giving the final form for the energy resolution:

$$\frac{\sigma(E)}{E} = \sqrt{\left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2} \quad (7.1)$$

7.2. Energy Resolution of the CMS HGCal

The High Granularity Calorimeter (HGCal) is the planned Phase 2 upgrade to the CMS Electromagnetic Calorimeter (ECAL) endcaps (EE) which cover the pseudo-rapidity range $1.5 < |\eta| < 3.0$ of the detector. The current ECAL is a compact homogeneous PbWO_4 crystal calorimeter located after the Silicon Tracker, see Fig. 7.1. Its excellent energy resolution was decisive in the search for resonant Higgs decays [51]. For the High Luminosity Large Hadron Collider (HL-LHC) however, the desired luminosity of $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and up to 200 interactions per bunch crossing will require a completely new calorimeter, capable of withstanding the large amounts of radiation and still provide fast triggering. The new HGCal proposal [52] satisfies these requirements and will additionally provide enhanced lateral and longitudinal shower resolution.

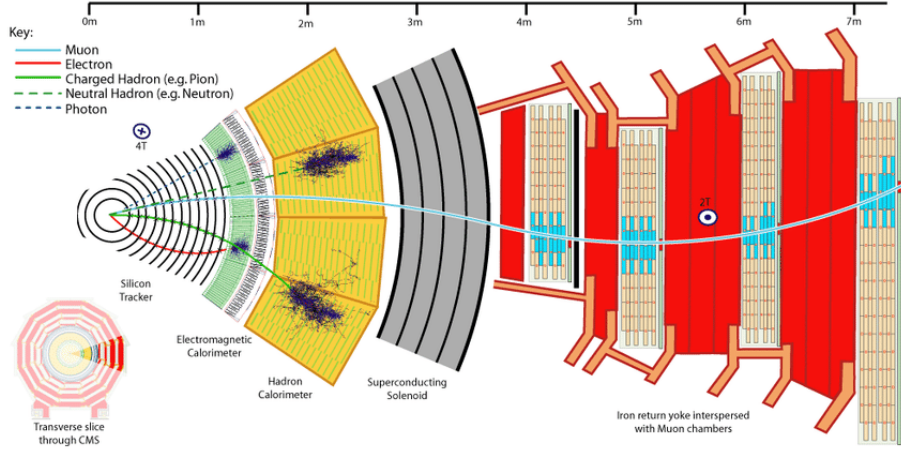


Figure 7.1.: Transverse cross-sectional view of the current CMS detector. The trajectories of particles originating from the interaction point are bent by the superconducting solenoid magnet, which allows accurate transverse momentum measurements in the Silicon Tracker. The electromagnetic and hadronic calorimeters measure the energy through showering and absorption and provide radiation shielding for the magnet coils. Lastly, the large Muon System records the momentum of muons which do not interact with the previous detector layers using drift tubes and cathode strip chambers. From [53].

Setup

The HGCAL is a sampling calorimeter composed of an Electromagnetic Compartment (CE-E) and a Hadronic Compartment (CE-H). The CE-E section is 34 cm thick and is composed of 28 longitudinal cassettes of alternating absorber and active material. The specific longitudinal composition for each cassette of the CE-E is shown in Table 7.1. The CE-H consists of 12 brass and copper absorber plates interleaved with Si and plastic scintillator trackers for a total thickness of 161 cm. A cross-sectional view of the HGCAL can be seen on Fig. 7.2. The CE-E is segmented into six wedges in the direction perpendicular to the beam axis. Each wedge is composed of a series of 15 cm large hexagonal tracking sensors as depicted on Fig. 7.3.

Each CE-E cassette has two layers of 1.4 mm thick 25Cu75W absorber plates. Their purpose is to provide structure to the central part of the cassette, transfer heat from the readout circuit boards to the copper cooling plate and serve as additional absorber material. This section investigates the influences of manufacturing tolerances of the CuW absorber plates thickness on the energy resolution of the CE-E. For this purpose, a single longitudinal block of the CE-E was simulated in GEANT4. The expected tolerances for the CuW absorber plate thickness lies in the range 0 – 5%. In this analysis, tolerances δ_{CuW} of 0% (reference case), 2%, 5%, 10% and 20% are considered. The last two cases are included for better visualisation of the effects on the energy resolution. For each case, energies ranging from 5–1000 GeV are each simulated in 10 batches of 100 events. For each batch, the thickness of the absorber plate in each layer is sampled from a normal distribution with $\mu = d_{CuW}$ and $\sigma = \delta_{CuW} d_{CuW}$, where d_{CuW} is the nominal thickness and δ_{CuW} is the thickness variation.

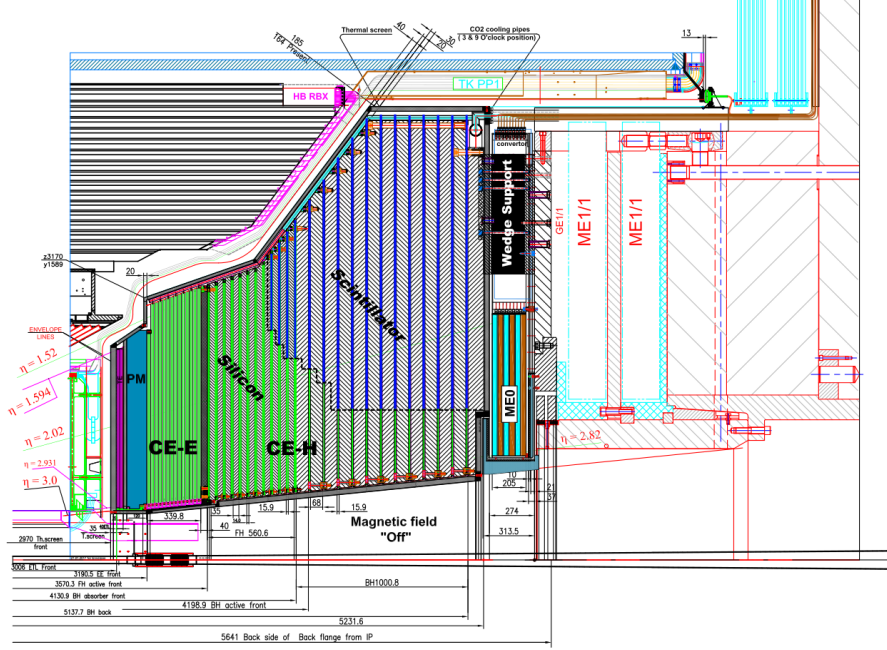


Figure 7.2.: Longitudinal cross-sectional view of the top part of the proposed HGCal. From left (close to the interaction point) to right (away): the endcap timing layer (TE), neutron moderator (PM), electromagnetic compartment (CE-E) and hadronic compartment (CE-H). Green and dark blue indicate sections with Si-based and plastic scintillator tracking layers respectively. From [52].

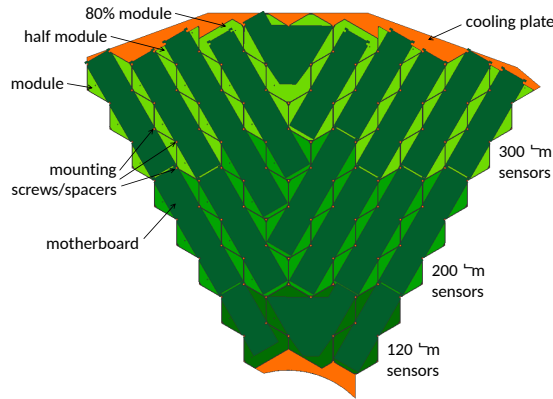


Figure 7.3.: Lateral segmentation of a single CE-E 60° wedge with Si sensor modules. Denser modules with a thickness of $120\ \mu\text{m}$ and 432 channels are placed near to the high-fluency area closer to the beam line. From [52].

Table 7.1.: Longitudinal composition of a single CE-E cassette, as used in the GEANT4 simulation. Layer 1 has a thickness of 3.6 mm for cassette 1, 5.8 mm for cassettes 2 to 9 and 9.1 mm for cassettes 10 to 13. Layers 3 and 11 have a variable thickness of $3.6 - \sigma$ mm and $8.7 - \sigma$ mm in order to compensate for the variability of layers 6 and 8 respectively. This insures that the total length of the detector remains constant.

Layer	Material	Thickness (mm)	Properties
0	Stainless Steel	0.3	
1	Pb	3.6-9.1	Layer dependent
2	Stainless Steel	0.3	
3	Air	3.6	Variable thickness
4	Si	0.2	Active layer
5	PCB	0.2	
6	25Cu75W	$1.4 \pm \sigma$	Variable thickness
7	Cu	6.05	
8	25Cu75W	$1.4 \pm \sigma$	Variable thickness
9	PCB	0.2	
10	Si	0.2	Active layer
11	Air	8.7	Variable thickness

Calibration

The energy E_{dep} deposited in each layer n is scaled by a per-layer calibration ω_n [54]. This per-layer calibration compensates the non-linear detector response due to sampling effects. A global calibration is performed to find the baseline response for the reference case $\delta = 0\%$,

$$E_{\text{true}} = k \sum_{n \in N} \omega_n E_{\text{dep},n} \quad (7.2)$$

The calibration sweeps over energies from 5–1000 GeV in the range of energies expected at the HL-LHC and the fit of k is shown in Fig. 7.4.

Results

The relative reconstructed energy $E_{\text{rec}}/E_{\text{true}}$, where $E_{\text{rec}} = kE_{\text{dep}}$, is fitted with a normal distribution for each d_{CuW} thickness and each sampled energy E_{true} . Afterwards, the tails outside the 90%-percentile are cut away in order to mitigate effects of shower leakage at higher energies, which influence the skewness of the distribution. The fit is repeated and the resulting $\sigma_{90\%}$ standard deviation is recorded. The standard deviation is a measure of the resolution of the calorimeter for each E_{true} . The energy resolution $\sigma(E)/E$ then follows Eq. (7.1). The resulting curves can be seen in Figs. 7.5a to 7.5e for each δ_{CuW} .

The case $d_{CuW} = 1.4$ mm with no spread $\delta_{CuW} = 0$ is treated as the reference energy resolution. Therefore, changes in the relative performance can only be attributed to an increase in the spread of absorber thickness $\delta_{CuW} > 0\%$. As mentioned in Section 7.1, the constant term c in the energy resolution quantifies the stochastic noise due to the

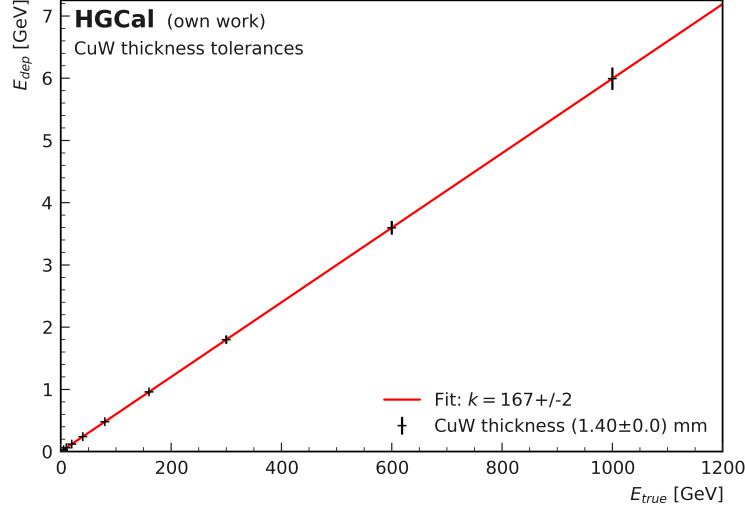


Figure 7.4.: Calibration of the energy response of a section of the CE-E. The figure shows a linear relation between the total reconstructed energy in the tracking layers E_{rec} and the initial photon energy E_{true} .

detector geometry. Fig. 7.6 then shows the constant term $c(\delta_{CuW})$ of the energy resolution fit as a function of δ_{CuW} . From this, it is apparent that increased variations in thickness of the absorber plates will lead to a constant smearing in the energy resolution. In the realistic region mentioned at the start of this section, the increase is minimal, from $c(\delta_{CuW} = 0\%) = 0.32 \pm 0.03\%$ up to $c(2\%) = 0.34 \pm 0.03\%$ and is subject to large uncertainties. Already at $c(5\%) = 0.46 \pm 0.03\%$, this represents an increase of 43% compared to the baseline. Accurate measurements of the actual thickness during the construction phase of HGCAL should provide the necessary information to calibrate each layer and compensate digitally for any deviations from the nominal thickness.

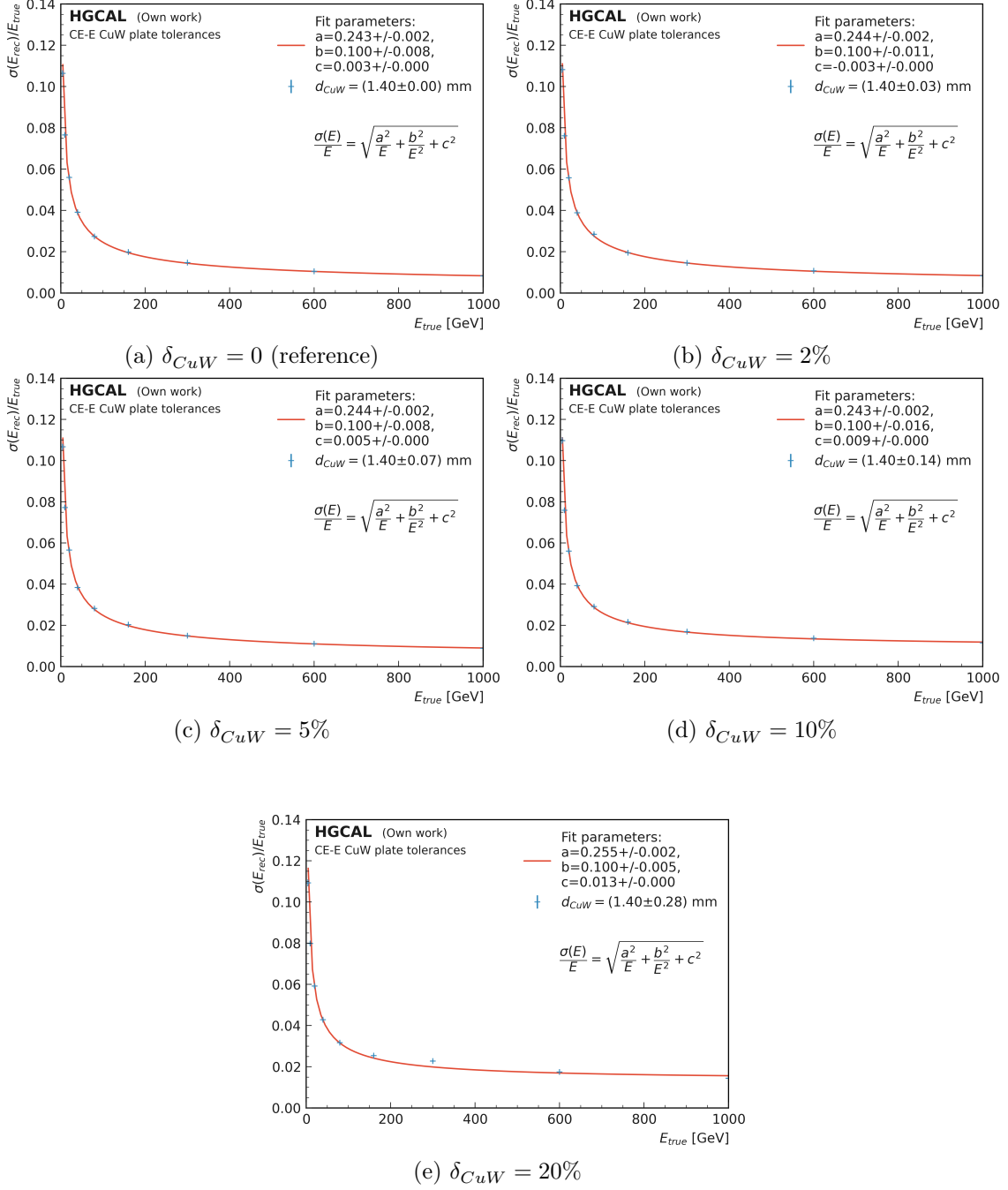


Figure 7.5.: Energy resolution of HGCal with increasing tolerance on the thickness of the CuW absorber plates. All data-points are Monte Carlo simulations.

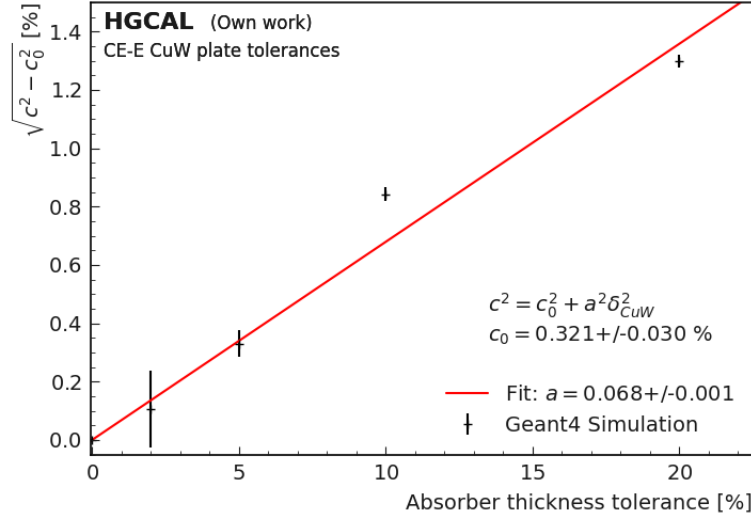


Figure 7.6.: Constant term c versus the variation δ_{CuW} in the thickness of the CuW absorber plates in HGCAL Electromagnetic Compartment. In the region 0 – 3%, there is no noticeable increase that can be distinguished from the uncertainties. As the tolerances increase, the effects are evident and follow a square law.

8. Conclusion

This thesis shows the feasibility of reconstructing the common vertex of two photons based only on their shower shape using a composite Graph Neural Network (GNN) algorithm. The network utilizes the low-level hit from the detector to predict variables of interest such as the vertex position and the direction of the primary photons.

The capabilities of the GNN are tested on a sample motivated by the decay of a long-lived Axion-Like Particle (ALP) into two photons $A \rightarrow \gamma\gamma$. Beamdump experiments, such as LUXE-NPOD, can potentially produce highly boosted ALPs that decay into photons with energies in the $\mathcal{O}(1)\text{GeV}$ range. In the approach used in this thesis, the high energy photons shower in a dense absorber material. The subsequent shower is recorded by a high granularity calorimeter. Chapter 4 covers the simulation of the proposed detector setup which records the 3D spacial information of the shower through the use of stacked absorber and sensor layers.

In order to train the network, two data-sets are simulated where each event consists of two photons arising from a common vertex. The simulation of the first and second data-set is performed on a short and long version of the detector respectively. These events are motivated by a 4 GeV ALP that decays on axis between 1.4 m and 10 cm in front of the detector into two photons with equal momenta and angle with respect to ALP direction. A series of selection criteria ensure that the events in the data-set have two well-resolved showers. Each event is composed of a variable number of hits and each hit in turn contains the information about the type of sensor, its location and the recorded energy. This unprocessed information is passed as input to the GNN.

In Chapter 5, a geometric deep learning approach is proposed in the form of GNNs, which provides several key capabilities for this task. First, the network uses a graph representation that effortlessly adapts to any irregular detector geometry. This representation also handles variable sized inputs with ease and has less computational overhead in comparison to conventional Machine Learning architectures. Second, the network can accommodate for different input information, as well as arbitrarily many output features. An important feature of this work is the use of stacked GarNet layers, which funnel the coarse information clustered by the previous GravNet layers into separate, information-dense output nodes. These GarNet layers are custom-built in this work for the sparse data structures of the detector output.

The performance of the network is investigated in Chapter 6 with two models trained on the different detector geometries with separate training protocols. In both models, the radial,

angular and vertex accuracy of the network is shown to be excellent, with the performance being dependent on the distance between the initial vertex and the detector. The model trained on a more spaced out detector outperforms the more compact geometry based on the previous criteria. Over the considered parameter space, the radial distance of the impact point of the particles is reconstructed with a resolution of 2.04 ± 0.01 mm and 1.83 ± 0.01 mm with the short and longer detector respectively. The opening angle of both photons is reconstructed with a resolution of $1.69 \pm 0.01^\circ$ and $0.98 \pm 0.01^\circ$. Finally, in the model with a longer detector, the algorithm is able to reconstruct a vertex located 90 cm away from the detector with a resolution of 13.6 ± 0.6 cm.

Chapter 7 investigated the effects of manufacturing tolerances on the thickness of CuW absorber plates in the CMS High Granularity Calorimeter (HGCAL). The simulation of a segment of the Electromagnetic Compartment (CE-E) concludes that a tolerance of 5% leads to a 43% increase in the constant term of the energy resolution, from the baseline $c(0\%) = 0.32 \pm 0.03\%$ to $c(5\%) = 0.46 \pm 0.03\%$. Tighter tolerances are found to be within the uncertainties of the simulation.

9. Outlook

The reconstruction algorithm presented in this work is able to reconstruct photons and their common origin in a simple proof-of-concept case. This tool opens new avenues for particle reconstruction from complex detector data. There are several open questions that should be addressed in further studies involving this algorithm.

A study of the performance of this algorithm in the same region but with shallower angles and possibly overlapping showers is the first recommendation. This would validate the capability of the algorithm to cover the range of vertex positions and opening angles of interest at LUXE-NPOD. Furthermore, a full hyperparameter optimization will increase the performance of the network further.

The motivation for this work is to reconstruct the decay of an ALP from its two-photon signature. The next step is then to test the algorithm in-situ with a MC sample depicting the physics process of ALP decays into two photons. It is possible that the acceptance of this sample will decrease in cases where the decay is highly asymmetric, leading to possibly only one fully resolved shower. Since the present algorithm expects two photon shower, these events might produce unpredictable results.

There are multiple approaches to address this challenge. First, the GNN could be trained to gauge its own capability to reconstruct a given event. This task is highly complex and is a type of self-supervised ML. Adding such an estimator to the network would also partially address the long-tails problem which most regression algorithms have. As discussed in Chapter 6, there are very rare cases where the prediction is completely erroneous. Identifying these events and labelling them as faulty is of paramount importance for eventual deployment on experimental data. Second, a classical or ML algorithm could identify energy clusters and determine whether a given event constitutes a possible two-photon event.

An interesting addition to the present algorithm would be to include a soft-max layer to act as a binary classifier. For LUXE-NPOD, neutrons are expected to be responsible for most of the background. A classifier layer could give the probability of each shower to have originated from a neutron or a photon, based on the shower shape.

The use of two different models in this work already hinted at the possibility to compare different detector geometries. In this strategy, the training process of both models is kept the same and only one parameter in the simulation is changed. The resulting difference in reconstruction performance can then be solely attributed to this change in geometry, up to uncertainties in the resolution. By sampling over a range of parameter values, it might even

be possible to find an optimal detector geometry. The only costs arising from this approach would be related to the simulation and retraining of the network for each sample point.

Glossary

BDTC	BEAMDUMP TRACKING CALORIMETER
ALP	Axion-Like Particle
AT-layer	Absorber-Tracker Layer
BSM	Beyond the Standard Model
CE-E	Electromagnetic Compartment
CE-H	Hadronic Compartment
CMS	Compact Muon Solenoid
CP	Charge Parity
DESY	Deutsches Elektronen-Synchrotron
DJC	DEEPJETCORE
DM	Dark Matter
ECAL	Electromagnetic Calorimeter
ELU	Exponential Linear Unit
eu.XFEL	European X-Ray Free-Electron Laser
GNN	Graph Neural Network
GUI	Graphical User Interface
HEP	High Energy Physics
HGCal	High Granularity Calorimeter
HL-LHC	High Luminosity Large Hadron Collider
KNN	k -Nearest-Neighbor
LUXE	Laser Und XFEL Experiment
LUXE-NPOD	LUXE - New Physics search at Optical Dump
MC	Monte Carlo
ML	Machine Learning

NP	New Physics
PCB	Printed Circuit Board
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
SM	Standard Model
TA	Tracker-Absorber

Bibliography

- [1] Alexander Heidelberg. *Beamdump Tracking Calorimeter Simulation*. Feb. 7, 2024. URL: <https://gitlab.etp.kit.edu/aheidelberg/beamdumptrackcalo>.
- [2] Wenzel Jakob et al. *pybind/pybind11: Version 2.11.1*. Version v2.11.1. July 17, 2023. DOI: [10.5281/ZENODO.5807779](https://doi.org/10.5281/ZENODO.5807779). URL: <https://zenodo.org/record/5807779> (visited on 02/12/2024).
- [3] Jim Pivarski et al. *Awkward Array*. Version v2.5.2. Jan. 12, 2024. DOI: [10.5281/ZENODO.10498548](https://doi.org/10.5281/ZENODO.10498548). URL: <https://zenodo.org/doi/10.5281/zenodo.10498548> (visited on 02/29/2024).
- [4] Jan Kieseler. *CMS-PEPR/HGCalML*. URL: <https://github.com/cms-pepr/HGCalML>.
- [5] *tensorflow*. URL: <https://www.tensorflow.org/>.
- [6] *KERAS*. URL: <https://keras.io/>.
- [7] Shah Rukh Qasim et al. „Learning representations of irregular particle-detector geometry with distance-weighted graph networks“. In: *The European Physical Journal C* 79.7 (July 2019), p. 608. ISSN: 1434-6044, 1434-6052. DOI: [10.1140/epjc/s10052-019-7113-9](https://doi.org/10.1140/epjc/s10052-019-7113-9). arXiv: [1902.07987](https://arxiv.org/abs/1902.07987)[hep-ex, physics:physics, stat]. URL: <http://arxiv.org/abs/1902.07987> (visited on 11/02/2023).
- [8] Thomas A Caswell et al. *matplotlib/matplotlib: REL: v3.7.5*. Version v3.7.5. Feb. 16, 2024. DOI: [10.5281/ZENODO.592536](https://doi.org/10.5281/ZENODO.592536). URL: <https://zenodo.org/doi/10.5281/zenodo.592536> (visited on 02/29/2024).
- [9] Johannes Gäßler et al. *PhiLFitters/kafe2: v2.8.0 - Summer Term 2023*. Version v2.8.0. June 15, 2023. DOI: [10.5281/ZENODO.7086793](https://doi.org/10.5281/ZENODO.7086793). URL: <https://zenodo.org/record/7086793> (visited on 02/29/2024).
- [10] R. L. Workman et al. „Review of Particle Physics“. In: *PTEP* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [11] Vera C. Rubin. „Dark Matter in the Universe“. In: *Proceedings of the American Philosophical Society* 132.4 (1988). Publisher: American Philosophical Society, pp. 434–443. ISSN: 0003049X. URL: <http://www.jstor.org/stable/986963> (visited on 02/29/2024).
- [12] Peter Athron and D. J. Miller. „A New Measure of Fine Tuning“. In: *Physical Review D* 76.7 (Oct. 17, 2007), p. 075010. ISSN: 1550-7998, 1550-2368. DOI: [10.1103/PhysRevD.76.075010](https://doi.org/10.1103/PhysRevD.76.075010). arXiv: [0705.2241](https://arxiv.org/abs/0705.2241)[hep-ph]. URL: <http://arxiv.org/abs/0705.2241> (visited on 02/29/2024).

- [13] Lyndon Evans and Philip Bryant. „LHC Machine“. In: *Journal of Instrumentation* 3.8 (Aug. 14, 2008), S08001–S08001. ISSN: 1748-0221. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001). URL: <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001> (visited on 02/28/2024).
- [14] Zhaoyu Bai et al. „LUXE-NPOD: new physics searches with an optical dump at LUXE“. In: *Physical Review D* 106.11 (Dec. 28, 2022), p. 115034. ISSN: 2470-0010, 2470-0029. DOI: [10.1103/PhysRevD.106.115034](https://doi.org/10.1103/PhysRevD.106.115034). arXiv: [2107.13554](https://arxiv.org/abs/2107.13554)[hep-ex, physics: hep-ph]. URL: <http://arxiv.org/abs/2107.13554> (visited on 01/08/2024).
- [15] Markus Kuster, Georg Raffelt, and Berta Beltrán, eds. *Axions: Theory, Cosmology, and Experimental Searches*. Red. by R. Beig et al. Vol. 741. Lecture Notes in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-73517-5 978-3-540-73518-2. DOI: [10.1007/978-3-540-73518-2](https://doi.org/10.1007/978-3-540-73518-2). URL: <http://link.springer.com/10.1007/978-3-540-73518-2> (visited on 01/08/2024).
- [16] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. „Broken Symmetries“. In: *Physical Review* 127.3 (Aug. 1, 1962), pp. 965–970. ISSN: 0031-899X. DOI: [10.1103/PhysRev.127.965](https://doi.org/10.1103/PhysRev.127.965). URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965> (visited on 01/08/2024).
- [17] R. D. Peccei and Helen R. Quinn. „CP Conservation in the Presence of Pseudoparticles“. In: *Physical Review Letters* 38.25 (June 20, 1977), pp. 1440–1443. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.38.1440](https://doi.org/10.1103/PhysRevLett.38.1440). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.38.1440> (visited on 01/08/2024).
- [18] Steven Weinberg. „A New Light Boson?“. In: *Physical Review Letters* 40.4 (Jan. 23, 1978), pp. 223–226. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.40.223](https://doi.org/10.1103/PhysRevLett.40.223). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.40.223> (visited on 01/08/2024).
- [19] R. L. Workman and (Particle Data Group). *Searches for Axions (A^0) and Other Very Light Bosons*. 083C01. Prog.Theor.Exp.Phys, 2022. URL: <https://pdg.lbl.gov/2022/listings/rpp2022-list-axions.pdf>.
- [20] Particle Data Group et al. „Review of Particle Physics“. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 8, 2022), p. 083C01. ISSN: 2050-3911. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097). URL: <https://academic.oup.com/ptep/article/doi/10.1093/ptep/ptac097/6651666> (visited on 02/05/2024).
- [21] Rebeca Gonzalez Suarez. *The Future Circular Collider (FCC) at CERN*. May 25, 2022. DOI: [10.48550/arXiv.2204.10029](https://doi.org/10.48550/arXiv.2204.10029). arXiv: [2204.10029](https://arxiv.org/abs/2204.10029)[hep-ex]. URL: <http://arxiv.org/abs/2204.10029> (visited on 03/01/2024).
- [22] F. Dydak. „Beam-Dump Experiments“. In: *Neutrino Physics and Astrophysics*. Ed. by Ettore Fiorini. Boston, MA: Springer US, 1982, pp. 341–360. ISBN: 978-1-4757-0521-8 978-1-4757-0519-5. DOI: [10.1007/978-1-4757-0519-5_17](https://doi.org/10.1007/978-1-4757-0519-5_17). URL: http://link.springer.com/10.1007/978-1-4757-0519-5_17 (visited on 01/25/2024).
- [23] G. Danby et al. „Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos“. In: *Physical Review Letters* 9.1 (July 1, 1962), pp. 36–44. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.9.36](https://doi.org/10.1103/PhysRevLett.9.36). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.9.36> (visited on 01/25/2024).

- [24] SHIP Collaboration. *The SHiP experiment at the proposed CERN SPS Beam Dump Facility*. May 20, 2022. arXiv: [2112.01487\[hep-ex,physics:physics\]](#). URL: <http://arxiv.org/abs/2112.01487> (visited on 01/25/2024).
- [25] FASER Collaboration et al. *Letter of Intent for FASER: ForwArd Search ExpeRiment at the LHC*. Nov. 26, 2018. arXiv: [1811.10243\[hep-ex,physics:hep-ph,physics:physics\]](#). URL: <http://arxiv.org/abs/1811.10243> (visited on 01/25/2024).
- [26] H. Abramowicz et al. *Letter of Intent for the LUXE Experiment*. Sept. 2, 2019. arXiv: [1909.00860\[hep-ex,physics:physics\]](#). URL: <http://arxiv.org/abs/1909.00860> (visited on 01/26/2024).
- [27] Halina Abramowicz et al. „Conceptual Design Report for the LUXE Experiment“. In: *The European Physical Journal Special Topics* 230.11 (Oct. 2021), pp. 2445–2560. ISSN: 1951-6355, 1951-6401. DOI: [10.1140/epjs/s11734-021-00249-z](#). arXiv: [2102.02032\[hep-ex,physics:hep-ph,physics:quant-ph\]](#). URL: <http://arxiv.org/abs/2102.02032> (visited on 01/26/2024).
- [28] S. Agostinelli et al. „GEANT4—a simulation toolkit“. In: *Nucl. Instrum. Methods Phys. Res. A* 506 (2003), pp. 250–303. DOI: [10.1016/S0168-9002\(03\)01368-8](#).
- [29] T. Kuhr et al. „The Belle II Core Software: Belle II Framework Software Group“. In: *Computing and Software for Big Science* 3.1 (Dec. 2019), p. 1. ISSN: 2510-2036, 2510-2044. DOI: [10.1007/s41781-018-0017-9](#). URL: <http://link.springer.com/10.1007/s41781-018-0017-9> (visited on 02/12/2024).
- [30] The Belle II Collaboration. *Belle II Analysis Software Framework (basf2)*. Version release-06-00-09. Aug. 1, 2022. DOI: [10.5281/ZENODO.5574115](#). URL: <https://zenodo.org/record/5574115> (visited on 02/12/2024).
- [31] Christophe Ochando and CMS Collaboration. „HGCAL: A High-Granularity Calorimeter for the Endcaps of CMS at HL-LHC“. In: *Journal of Physics: Conference Series* 928 (Nov. 2017), p. 012025. ISSN: 1742-6588, 1742-6596. DOI: [10.1088/1742-6596/928/1/012025](#). URL: <https://iopscience.iop.org/article/10.1088/1742-6596/928/1/012025> (visited on 01/08/2024).
- [32] R.-D. Appuhn et al. „The H1 lead/scintillating-fibre calorimeter“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 386.2 (Feb. 1997), pp. 397–408. ISSN: 01689002. DOI: [10.1016/S0168-9002\(96\)01171-0](#). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168900296011710> (visited on 02/15/2024).
- [33] F. Rademakers et al. *"ROOT"*. Version v6.24/00. 2020. DOI: [10.5281/zenodo.848818](#). URL: <https://doi.org/10.5281/zenodo.848818>.
- [34] R. Brun and F. Rademakers. „ROOT: An object oriented data analysis framework“. In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: [10.1016/S0168-9002\(97\)00048-X](#).
- [35] Alexander Jung. *Machine Learning: The Basics*. Jan. 29, 2022. arXiv: [1805.05052\[cs,stat\]](#). URL: <http://arxiv.org/abs/1805.05052> (visited on 03/01/2024).

- [36] Javier Duarte and Jean-Roch Vlimant. „Graph Neural Networks for Particle Tracking and Reconstruction“. In: Mar. 2022, pp. 387–436. DOI: [10.1142/9789811234033_0012](https://doi.org/10.1142/9789811234033_0012). arXiv: [2012.01249](https://arxiv.org/abs/2012.01249)[[hep-ex](#), [physics:hep-ph](#), [physics:physics](#)]. URL: <http://arxiv.org/abs/2012.01249> (visited on 11/02/2023).
- [37] The HEP Software Foundation et al. „A Roadmap for HEP Software and Computing R&D for the 2020s“. In: *Computing and Software for Big Science* 3.1 (Dec. 2019), p. 7. ISSN: 2510-2036, 2510-2044. DOI: [10.1007/s41781-018-0018-8](https://doi.org/10.1007/s41781-018-0018-8). URL: <http://link.springer.com/10.1007/s41781-018-0018-8> (visited on 01/24/2024).
- [38] Vinod Nair and Geoffrey Hinton. „Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair“. In: *Proceedings of ICML*. Vol. 27. June 2010, pp. 807–814.
- [39] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. „Deep Sparse Rectifier Neural Networks“. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Apr. 11, 2011, pp. 315–323. URL: <https://proceedings.mlr.press/v15/glorot11a.html>.
- [40] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. Feb. 22, 2016. arXiv: [1511.07289](https://arxiv.org/abs/1511.07289)[[cs](#)]. URL: <http://arxiv.org/abs/1511.07289> (visited on 01/22/2024).
- [41] *optuna*. URL: <https://optuna.org/>.
- [42] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. Mar. 2, 2015. arXiv: [1502.03167](https://arxiv.org/abs/1502.03167)[[cs](#)]. URL: <http://arxiv.org/abs/1502.03167> (visited on 01/29/2024).
- [43] Jan Kieseler. *DeepJetCore*. URL: <https://github.com/DL4Jets/DeepJetCore>.
- [44] Shah Rukh Qasim et al. *Multi-particle reconstruction in the High Granularity Calorimeter using object condensation and graph neural networks*. June 2, 2021. arXiv: [2106.01832](https://arxiv.org/abs/2106.01832)[[physics](#)]. URL: <http://arxiv.org/abs/2106.01832> (visited on 02/01/2024).
- [45] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Jan. 29, 2017. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980)[[cs](#)]. URL: <http://arxiv.org/abs/1412.6980> (visited on 01/18/2024).
- [46] Michele Livan and Richard Wigmans. *Calorimetry for Collider Physics, an Introduction*. UNITEXT for Physics. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-23652-6 978-3-030-23653-3. DOI: [10.1007/978-3-030-23653-3](https://doi.org/10.1007/978-3-030-23653-3). URL: <http://link.springer.com/10.1007/978-3-030-23653-3> (visited on 01/08/2024).
- [47] Christian W. Fabjan and Fabiola Gianotti. *Calorimetry for Particle Physics*. CERN-EP/2003-075. CERN, 1211 Geneva 23, Switzerland: CERN, Oct. 31, 2003, p. 96. URL: <https://cds.cern.ch/record/692252/files/ep-2003-075.pdf> (visited on 08/01/2024).
- [48] J. B. Johnson. „Thermal Agitation of Electricity in Conductors“. In: *Physical Review* 32.1 (July 1, 1928), pp. 97–109. ISSN: 0031-899X. DOI: [10.1103/PhysRev.32.97](https://doi.org/10.1103/PhysRev.32.97). URL: <https://link.aps.org/doi/10.1103/PhysRev.32.97> (visited on 01/08/2024).

- [49] H. Nyquist. „Thermal Agitation of Electric Charge in Conductors“. In: *Physical Review* 32.1 (July 1, 1928), pp. 110–113. ISSN: 0031-899X. DOI: [10.1103/PhysRev.32.110](https://link.aps.org/doi/10.1103/PhysRev.32.110). URL: <https://link.aps.org/doi/10.1103/PhysRev.32.110> (visited on 01/08/2024).
- [50] CMS Collaboration. *CMS Electromagnetic Calorimeter Technical Design Report*. CMS TDR 4 CERN/LHCC 97–33. CERN, Dec. 15, 1997, p. 389. URL: https://cds.cern.ch/record/349375/files/ECAL_TDR.pdf.
- [51] Cristina Biino. „The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections“. In: *Journal of Physics: Conference Series* 587 (Feb. 13, 2015), p. 012001. ISSN: 1742-6588, 1742-6596. DOI: [10.1088/1742-6596/587/1/012001](https://iopscience.iop.org/article/10.1088/1742-6596/587/1/012001). URL: <https://iopscience.iop.org/article/10.1088/1742-6596/587/1/012001> (visited on 01/08/2024).
- [52] CMS Collaboration. *The Phase-2 Upgrade of the CMS endcap calorimeter Technical Design Report*. CMS-TDR-019 CERN-LHCC-2017-023. CERN, Sept. 4, 2018. URL: <https://cds.cern.ch/record/2293646/files/CMS-TDR-019.pdf>.
- [53] Stefanos Dris. „Performance of the CMS Tracker Optical Links and Future Upgrade Using Bandwidth Efficient Digital Modulation“. In: (2010). Publisher: arXiv Version Number: 1. DOI: [10.48550/ARXIV.1004.5574](https://arxiv.org/abs/1004.5574). URL: <https://arxiv.org/abs/1004.5574> (visited on 01/08/2024).
- [54] Pedro Silva. *Private conversation*.
- [55] J. L. Folks and R. S. Chhikara. „The Inverse Gaussian Distribution and its Statistical Application - A Review“. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 40.3 (1978), pp. 263–289.

A. Appendix

A.1. Wald distribution

The Wald distribution is an asymmetric function defined on the support $x \in [0, \infty[$ and features a steep peak with a long tail [55]. Its probability density function (pdf) is

$$\mathcal{W}(\mu, \lambda; x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x - \mu)^2}{2x\mu^2}\right). \quad (\text{A.1})$$

The pdf of $\mathcal{W}(3, 2; x)$ is shown in Fig. A.1.

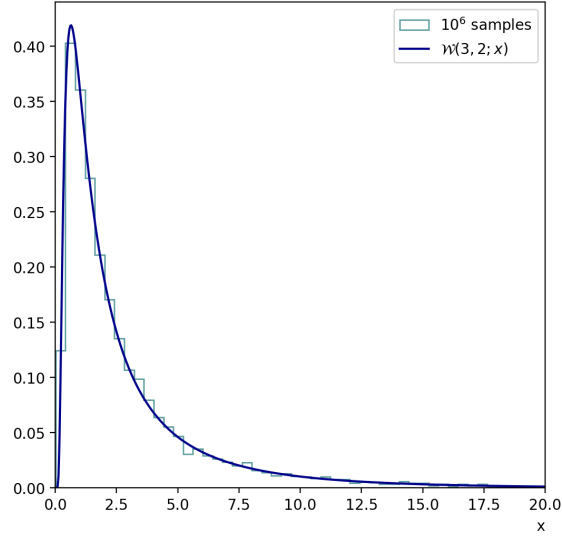


Figure A.1.: Wald distribution used to generate the transverse momenta of the photons in Section 4.4.

A.2. Additional Figures for Reconstruction with GNNs

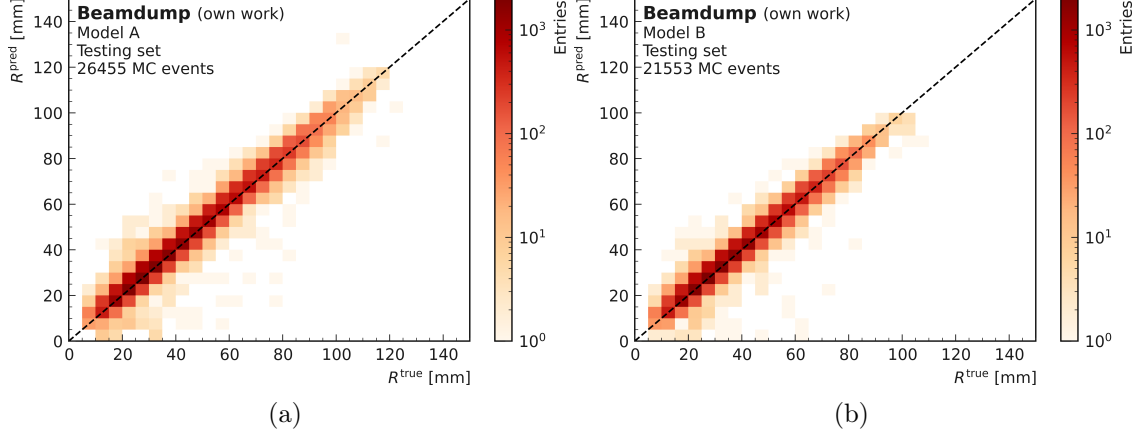


Figure A.2.: 2D histogram of the radial distance R^{true} of the true photon from the detector center and the predicted distance R^{pred} for Model A (a) and B (b).

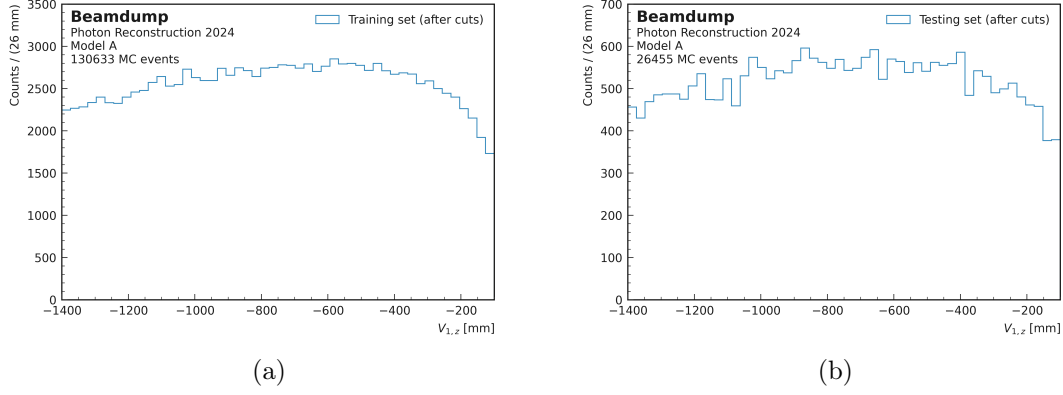


Figure A.3.: Distribution of the vertex position V along the z axis of the detector (the detector is located at $z = 0$) for the training (a) and testing (b) set of Model A. The initial vertex position is uniformly distributed between $[-1400, -100]$ mm. In both models, the selection criteria predominantly removes events close to detector as they are less likely to develop full showers in the detector if their opening angle is large. On the other hand, events originating from further away may have a shallower angle than the required $2\theta > 2^\circ$.

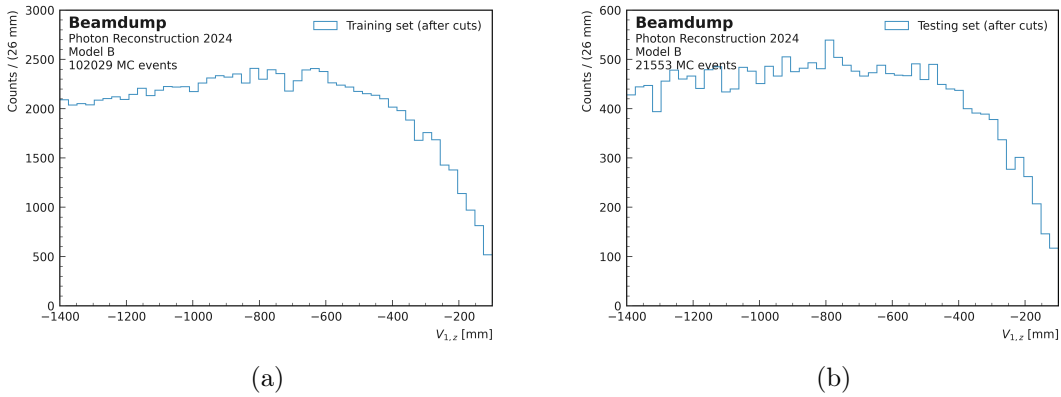


Figure A.4.: Distribution of the vertex position V along the z axis of the detector for the training (a) and testing (b) set of Model B. See also Fig. A.3.

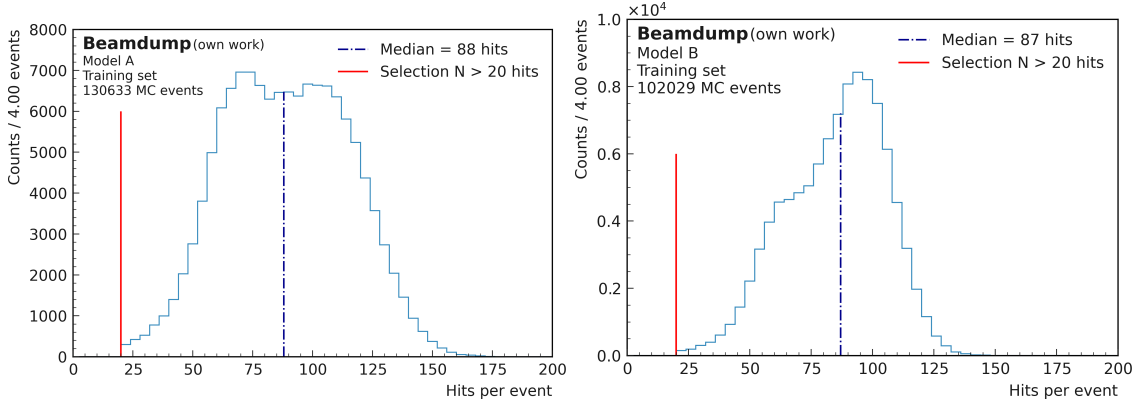


Figure A.5.: Number of hits per event in the training set of Model A (a) and B (b). Events with less than 20 hits are removed by the selection criteria of Section 4.5.

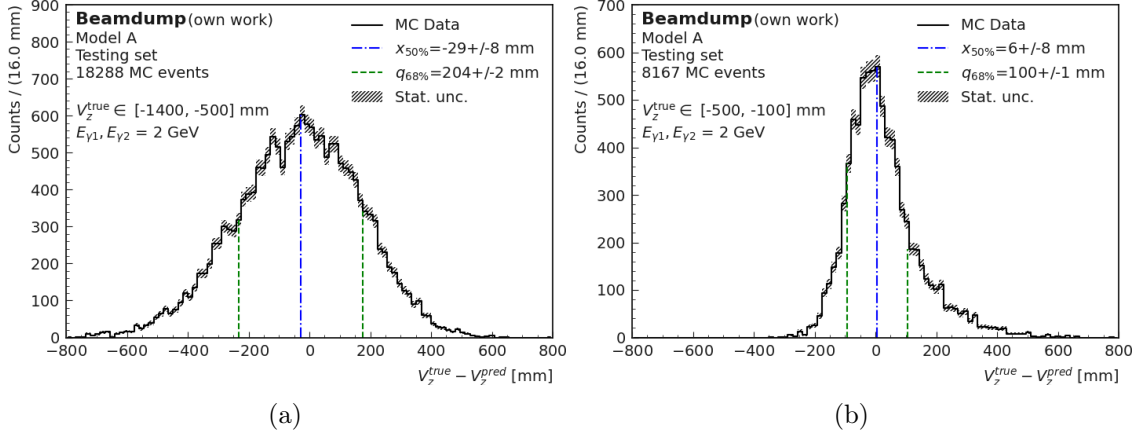


Figure A.6.: Reconstruction accuracy of Model A in the far region with $V_z^{\text{true}} < -500$ mm (a) and in the near region of the detector with $-500 \text{ mm} < V_z^{\text{true}}$ (b).

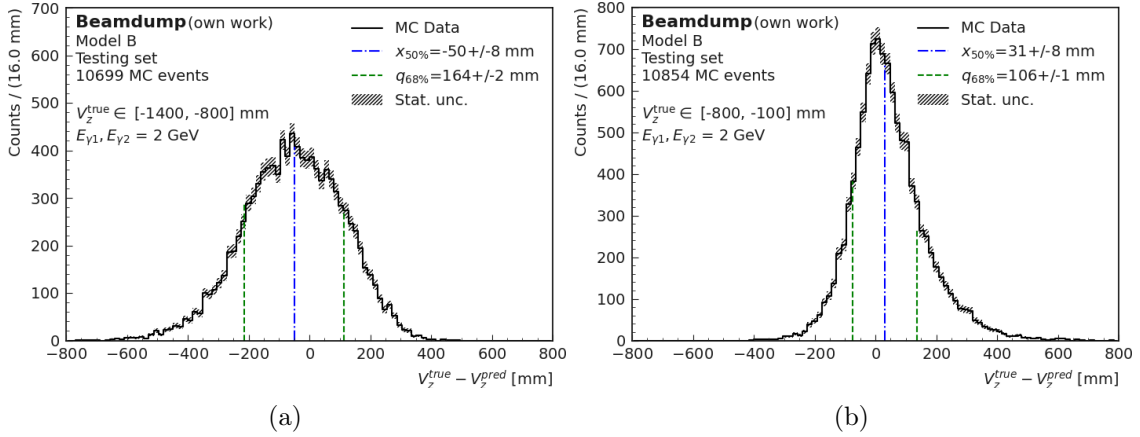


Figure A.7.: Reconstruction accuracy of Model B in the far region of the detector with $V_z^{\text{true}} < -800$ mm (a) and in the near region with $-800 \text{ mm} < V_z^{\text{true}}$ (b).

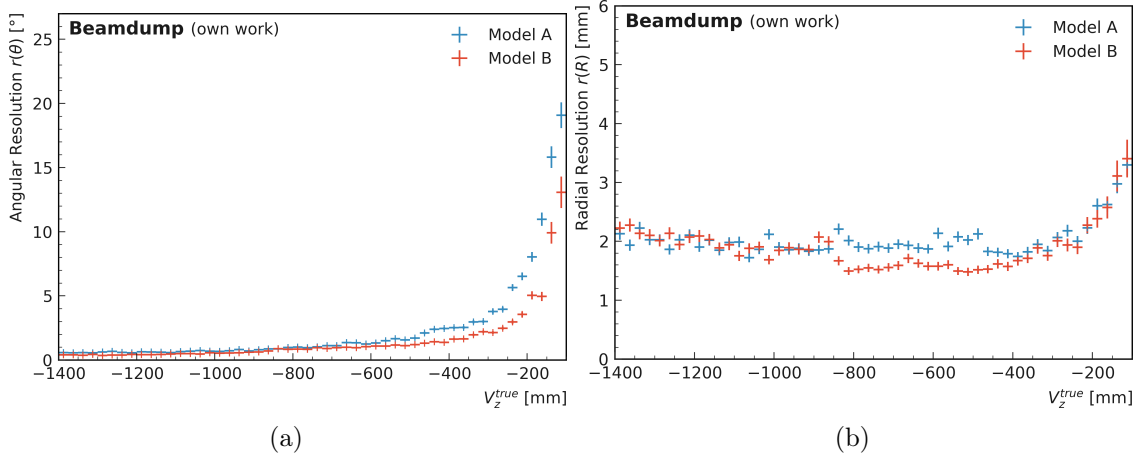


Figure A.8.: Evolution of the angular (a) and radial (b) resolution of both models with respect to V_z^{true} .

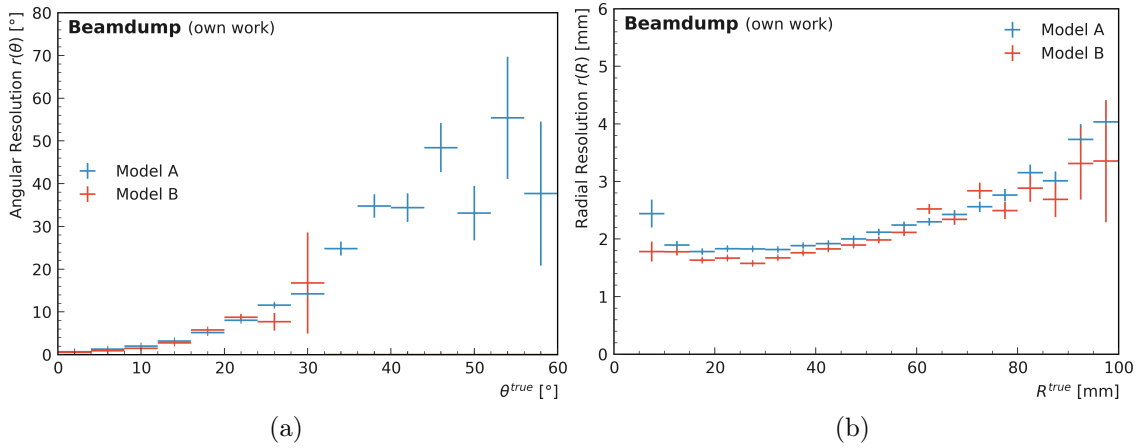


Figure A.9.: Evolution of the angular (a) and radial (b) resolution of both models with respect to their true MC information.

Acknowledgements

I would like to express my sincerest gratitudes to Prof. Markus Klute and Prof. Torben Ferber for the amazing chance of working in their group. I am deeply thankful to my supervisor Alexander Heidelberg for his expertise, his encouragements and most of all his incredible patience when working with me, you truly are the best. Special appreciation goes to Jan Kieseler for all his assistance and knowledge that went into my work.

I extend my most earnest gratitude to every member of the Belle II group at the ETP, for welcoming me in their team and for proof-reading the present rag. Without any particular order, thanks to Moritz, Giacomo, Isabel, Jonas, Lennard, Lea, Patrick, Greta and Sally. Speaking of the ETP, I would like to thank everyone on the 9th floor (sorry 8th floor, I don't know you that well) for their joy and bountiful cakes. I also want to acknowledge the support of my research group, specially Nicolás, who gifted so much of his time.

Thanks to our beverage admin for always providing refreshments in times of need. And before going on, I would like to thank the sponsor of this master thesis, BAfÖG Amt Karlsruhe, for generously bankrolling this whole endeavour. Now back to topic, I thank the Fachschaft Physik for being the backbone of the faculty and for being such an accommodating space. I want to thank Antonio Radić and Thibault Duplessis for providing the resources that kept me focused on my goals.

Thanks to my dear friends, Paul, Vincenz and Oliver for showing me the power of teamwork (yes, this is a joke about CS). Merci à toute ma famille, sans qui je ne serais jamais arrivé là, donc merci Mélya, Théo, Maman und Papa! Danke Hannah für deine unendliche Liebe und Aufmerksamkeit, die mich immer voran gebracht hat.

Last but not least, I want to thank me, for doing all the hard work.

