Event Reconstruction for an Opaque Liquid Scintillator Detector

---

Kitzia Michelle Hernandez Curiel

March 26, 2025
Version: Final

FB08 / Institute of Physics
FB08
AG Weber

Master Thesis

Event Reconstruction for an Opaque Liquid Scintillator Detector

Kitzia Michelle Hernandez Curiel

1. Reviewer Prof. Alfons Weber
Institute of Physics
Johannes Gutenberg University Mainz

2. Reviewer Dr. Cloé Girard-Carillo
Institute of Physics
Johannes Gutenberg University Mainz

Supervisors Prof. Alfons Weber

March 26, 2025

Kitzia Michelle Hernandez Curiel
Event Reconstruction for an Opaque Liquid Scintillator Detector
Master Thesis, March 26, 2025
Reviewers: Prof. Alfons Weber and Dr. Cloé Girard-Carillo
Supervisors: Prof. Alfons Weber

Johannes Gutenberg University Mainz
AG Weber
FB08
Institute of Physics
Staudingerweg 9
55128 Mainz

Abstract

Opaque liquid scintillators represent a novel approach to particle detection. This technology uses Mie scattering to confine the scintillation light around its interaction point, conserving the topology of the event in comparison with classical transparent scintillators. However, the opacity of these detectors and their energy deposition topologies represent a new paradigm opening the way to new reconstruction methods.

Here, we explore the adaptation of the Cambridge-Aachen jet clustering algorithm, traditionally used in High-energy physics, for event reconstruction in opaque liquid scintillator detectors. By clustering optical photons in the (x, y) plane and incorporating timing information, this method can effectively reconstruct event position, energy, and particle type. Furthermore, it provides a robust framework for particle discrimination and is a reference point for comparing with more sophisticated approaches using Graphic Neural Networks.

Contents

1. Introduction **1**

2. A Brief Introduction to Neutrino Physics **3**
- 2.1. Neutrino discovery 3
- 2.2. Neutrino oscillation 5
- 2.3. Current Open Questions 6
  - 2.3.1. Double Beta Decay 8

3. Liquid Scintillators **12**
- 3.1. Inorganic Scintillators 12
- 3.2. Organic Liquid Scintillators 12
- 3.3. Novel Techniques 14
  - 3.3.1. Hybrid Liquid Scintillators 15
  - 3.3.2. Opaque Liquid Scintillators 16

4. Application in Neutrino Detectors **19**
- 4.1. The CLOUD experiment 19
- 4.2. The NuDoubt^{++} experiment 21
  - 4.2.1. Event Simulation in an Opaque Scintillator Detector . 23
  - 4.2.2. Fiber Propagation and SiPM Detection Model 25

5. Cambridge-Aachen Algorithm for Opaque Scintillators **29**
- 5.1. Jet Clustering Algorithms 29
- 5.2. Implementation in Opaque Scintillator Detectors 32
  - 5.2.1. Adaptation of the Clustering Algorithm 32
  - 5.2.2. Algorithm's Implementation 33
  - 5.2.3. Distance Threshold Optimization 34
  - 5.2.4. Cluster Separation 37

5.2.5. Computational Performance 42

6. Event Reconstruction 45
6.1. Position and Energy Reconstruction 45
6.1.1. Estimating the Cluster Z-Position using Timing Information 46
6.1.2. Reconstruction of the Interaction Position 52
6.1.3. Energy Reconstruction 56
6.2. Particle Identification 58
6.2.1. Binary classification 62
6.2.2. Random Forest Classifier 63

7. Conclusion and Outlook 66

Bibliography 69

List of Figures 77

List of Tables 84

A. Appendix 85
A.1. Implementation of the C-A algorithm 85

1

Introduction

Liquid scintillators are widely used in particle physics experiments due to their ability to efficiently convert ionizing radiation into detectable light signals. Traditionally, these materials are transparent, allowing for straightforward photon collection and well-established reconstruction techniques. However, as experiments demand greater sensitivity and improved background rejection, new scintillator designs have been explored.

Opaque liquid scintillators are a recent advancement that differ from conventional transparent scintillators by incorporating scattering materials to modify light transport. Instead of traveling long distances, the light produced undergoes multiple scattering, effectively localizing its emission near the interaction point. This property enhances spatial resolution, making these scintillators particularly attractive for applications requiring precise event location.

These unique characteristics necessitate dedicated reconstruction strategies tailored to their response. This context has motivated studies into classical and machine-learning-based methods for event reconstruction. While recent advances in machine learning have highlighted the potential of Graph Neural Networks (GNNs) for event reconstruction in complex detector environments, their practical implementation often depends on the availability of extensive, high-quality training datasets and significant computational resources. These requirements can limit their feasibility and quick implementation. Additionally, to better understand the capabilities of GNNs, it is essential to compare them with more established reconstruction methods.

This thesis presents the development and implementation of a reconstruction method specifically designed for opaque liquid scintillator detectors focusing on a classical approach using jet clustering algorithms, providing a baseline to compare with machine-learning techniques.

The structure of the thesis is organized as follows. Chapter 2 provides an overview of neutrino physics, detailing the key properties of neutrinos and outlining current open questions in the field. Chapter 3 presents a concise background on liquid scintillators, discussing their detection mechanisms with particular attention to recent innovations in opaque liquid scintillator technology.

Chapter 4 introduces two neutrino experiments that plan to utilize opaque liquid scintillator detectors, with a detailed focus on the NuDoubt^{++} project. This chapter also describes the simulation framework and detector response model used to generate representative data from an opaque scintillator detector configuration.

Chapter 5 introduces jet clustering algorithms within the context of an opaque scintillator detector, providing an in-depth discussion of the Cambridge-Aachen algorithm, its implementation, performance, and characterization. Chapter 6 then applies the results obtained from the clustering analysis to the overall event reconstruction, evaluating the efficacy of this method in determining the positions of the energy deposits and in performing particle identification.

Finally, the conclusion and outlook are presented in Chapter 7.

2

A Brief Introduction to Neutrino Physics

2.1 Neutrino discovery

The history of neutrinos begins with the study of beta decay, first observed by J. Chadwick in 1914 [1]. Initially, beta decay was thought to be a two-body process, where a nucleus emits an electron and transforms into a daughter nucleus (Eq. 2.1).

$$ {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \rightarrow {}_{\mathrm{Z-1}}^{\quad\mathrm{A}}\mathrm{Y} + {}^{\;0}_{-1}\beta \tag{2.1} $$

However, the observed continuous energy spectrum of the emitted electrons presented a significant challenge to existing theories. According to classical physics, the energy of the emitted electrons should have been discrete, but experiments showed a continuous range. This discrepancy indicated the need for an additional particle to account for the missing energy and momentum.

$$ {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \rightarrow {}_{\mathrm{Z-1}}^{\quad\mathrm{A}}\mathrm{Y} + {}^{\;0}_{-1}\beta + {}^{0}_{0}\nu \tag{2.2} $$

In 1930, W. Pauli proposed the existence of a new particle to resolve the energy discrepancy in beta decay [2]. He hypothesized that this particle, initially called *neutron*, was electrically neutral, had a small mass, and possessed a spin of $1/2$. In 1933, E. Fermi renamed Pauli's particle the *neutrino* to distinguish it from Chadwick's neutron. Fermi also developed the modern theory of beta decay [3], proposing that neutrinos were emitted alongside electrons or positrons in reactions such as:

$$\beta^- \text{ decay: } n \to p + e^- + \bar{\nu}_e \tag{2.3}$$

$$\beta^+ \text{ decay: } p \to n + e^+ + \nu_e \tag{2.4}$$

This framework explained the continuous energy spectrum of beta decay and established the role of neutrinos in nuclear processes.

Despite the theoretical predictions, detecting neutrinos proved an immense challenge due to them being electrically neutral and only interacting weakly. In 1934, the inverse beta decay reaction (IBD) (Eq. 2.5) was proposed as a potential method for neutrino detection. However, its cross-section ($\sim 10^{-44}$ cm^2) made detection seem nearly impossible at the time [4].

$$\bar{\nu}_e + p \to n + e^+ \tag{2.5}$$

It was not until the 1950s that technological advances in nuclear physics allowed for the design of neutrino detectors [5]. In 1953, F. Reines and C. L. Cowan conducted the first successful neutrino detection experiment near the Hanford nuclear reactor, using a liquid scintillator doped with cadmium [6, 7]. Their detector identified neutrinos through scintillation light from positron annihilation and neutron capture on cadmium nuclei. Although their initial results were not statistically significant, a repetition of the experiment in 1956 near the Savannah River Plant yielded positive results, confirming the existence of neutrinos [8].

Further milestones came in the 1990s when experiments at CERN's Large Electron-Positron Collider (LEP), such as ALEPH [9], DELPHI [10], L3 [11], and OPAL [12], analyzed the decays of the Z^0 boson, a neutral mediator of the weak interaction. The Z^0 boson can decay into various particle-antiparticle pairs, including neutrino-antineutrino pairs. Although neutrinos can not be detected directly, their presence can be inferred through "missing energy" in the decay processes. By measuring the visible-to-invisible Z^0 decay ratio, the experiments determined three types of neutri-

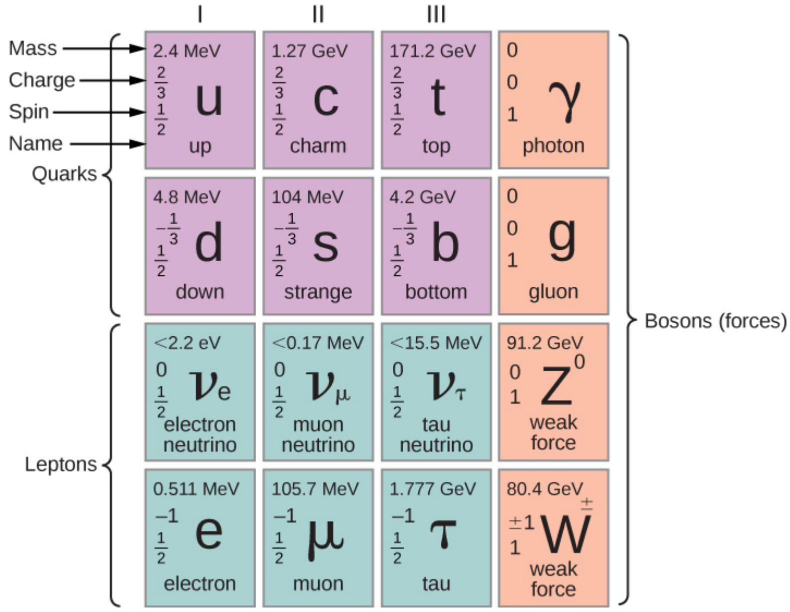


Fig. 2.1.: The Standard Model of Particle Physics. Neutrinos belong to the lepton family and interact only by the weak force through the Z and W bosons [13].

nos corresponding to the three charged leptons: e, μ, and τ. This discovery confirmed neutrinos as integral components of the Standard Model (Figure 2.1).

2.2 Neutrino oscillation

The LEP results confirmed the number of light neutrino flavors predicted by the Standard Model. However, within the formulation of the Standard Model, they are also assumed to be massless due to symmetry constraints [14]. Consequently, the discovery of neutrino oscillations [15], a quantum mechanical phenomenon in which neutrinos change flavor states as they propagate, questioned this assumption. This implies neutrinos have distinct, nonzero masses, indicating physics beyond the Standard Model.

The theoretical foundation for neutrino oscillation was laid in 1957 by B. Pontecorvo [16], who proposed that neutrinos could transform between different states. This idea was later formalized by Z. Maki, M. Nakagawa, and S. Sakata [17], whose work introduced the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix (Eq. 2.6). This unitary matrix describes the mixing between flavor (ν_α, $\alpha = e, \mu, \tau$) and mass eigenstates ($\nu_i, i = 1, 2, 3$).

$$\nu_\alpha = \sum_{i=1}^{3} U_{\alpha i} \nu_i \tag{2.6}$$

Where U is a 3×3 unitary matrix parameterized by a mixing angle (θ_{12}, θ_{13} and θ_{31}) and a CP-violating phase (δ_{CP}). Understanding this CP violation is critical for testing theories of leptogenesis, which link neutrino properties to the baryon asymmetry of the universe [18].

Experimental confirmation of neutrino oscillations came from studies of solar and atmospheric neutrinos. The Homestake experiment in 1968 first measured a deficit in solar neutrino flux compared to theoretical predictions, known as the solar neutrino problem [19]. Subsequent experiments at Kamiokande [20] and SAGE [21] corroborated this deficit. Atmospheric neutrino anomalies were observed by Super Kamiokande in 1998 [15], providing direct evidence of flavor oscillations by measuring differences in neutrino fluxes based on propagation distance. These findings established that neutrinos oscillate and must have mass.

2.3 Current Open Questions

The study of neutrino properties remains a central focus in theoretical and experimental physics, with multiple experiments focusing on different aspects of neutrinos. The discovery of neutrino oscillations had different implications. As mentioned, oscillations demonstrated that neutrinos possess mass, and the Standard Model needs to be modified to account for it. This introduces key considerations, such as determining the absolute mass

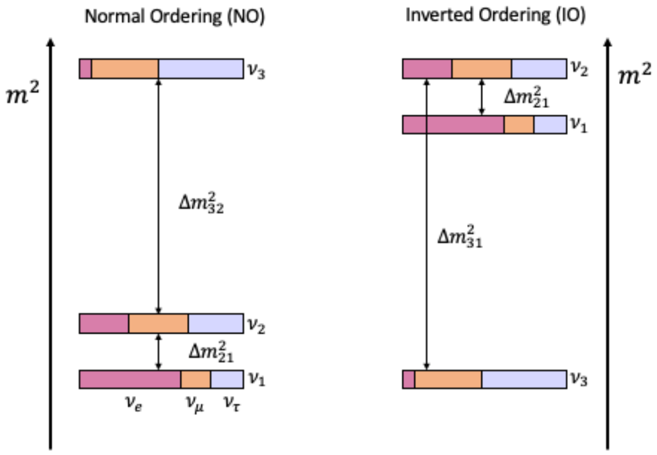


Fig. 2.2.: Two possible mass ordering of neutrinos. Normal ordering and inverted ordering.

of neutrinos, understanding the mass ordering among the three flavors, and identifying the mechanism responsible for their mass.

While oscillations are sensitive to differences in the squared masses of the three neutrino mass eigenstates, known as mass-squared differences (Δm^2), the exact mass ordering of these eigenstates is still unknown. It can follow a *normal ordering* ($m_1 < m_2 < m_3$), or an *inverted ordering* ($m_2 < m_3 < m_1$) [22].

The origin of the neutrino mass is still unknown. Neutrinos could have a Dirac mass, like other fermions, where neutrinos and antineutrinos are distinct particles. Alternatively, they might have a Majorana mass, where neutrinos are their own antiparticles. The latter would violate the lepton number conservation and could help explain the universe's matter-antimatter asymmetry. Processes like neutrinoless double-beta decay (Figure 2.3) are key to testing this possibility [23, 24].

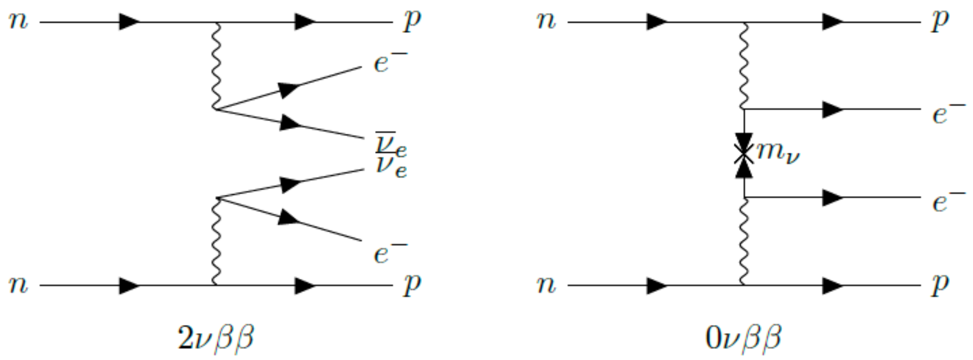


Fig. 2.3.: Feynman diagrams illustrating two double beta decay processes. The diagram on the left depicts the two-neutrino double beta decay ($2\nu\beta\beta$), where two electrons and two corresponding antineutrinos are emitted, reflecting the behavior of Dirac fermions. The diagram on the right shows the neutrinoless double beta decay ($0\nu\beta\beta$), in which the two neutrinos annihilate, leaving only two electrons in the final state, consistent with neutrinos being Majorana particles [25].

2.3.1 Double Beta Decay

Double beta decay is a rare nuclear process in which two neutrons within an atomic nucleus simultaneously convert into two protons (or two protons convert into two neutrons) emitting two (anti) neutrinos. This decay pathway allows unstable nuclei to achieve a more stable configuration in cases where single beta decay is energetically forbidden. The rarity of this process is highlighted by the exceptionally long half-lives of isotopes undergoing double beta decay, ranging from 10^{18} to 10^{24} years [26]. The decay produces a characteristic energy release, known as the Q-value, defined as:

$$Q = (M_{\text{parent}} - M_{\text{daughter}})c^2 \tag{2.7}$$

where M represents the nuclear mass and c the speed of light. The Q-value is shared as the kinetic energy among the emitted particles, making it a critical parameter for detecting and analyzing the decay.

The concept of double beta decay was first introduced by M. Goeppert-Mayer in 1935 as a theoretical extension of beta decay within the framework of nuclear physics [27]. However, experimental verification of this phenomenon proved elusive for decades due to the extreme rarity of the process and the technical challenges associated with detecting the emitted particles. Indirect evidence of double beta decay first surfaced in 1966 for $^{130}\mathrm{Te}$ [28], followed by further findings in 1968 [29]. Direct observation of the process was achieved much later in 1987 when two-neutrino double beta decay ($2\nu\beta\beta$) of $^{82}\mathrm{Se}$ was successfully detected [30]. Double beta decay manifests in several distinct modes (Eq. 2.8), classified by the particles emitted during the decay. The most studied is the $2\beta^-$ process, where two neutrons in the parent nucleus are converted into two protons. This can occur via the Standard Model-allowed two-neutrino mode ($2\nu2\beta^-$) or via the hypothesized neutrinoless mode ($0\nu2\beta^-$). In the $2\nu2\beta^-$ decay, two electrons and two antineutrinos are emitted; the available energy is distributed among these particles, producing a continuous electron energy spectrum. This decay has been experimentally confirmed in several isotopes, including $^{136}\mathrm{Xe}$, $^{130}\mathrm{Te}$, and $^{82}\mathrm{Se}$ [32, 26]. In contrast, the $0\nu2\beta^-$ decay, which has not yet been observed, would emit only two electrons without accompanying neutrinos. Such a decay would violate lepton number conservation, implying the existence of new physics if detected. The absence of neutrinos

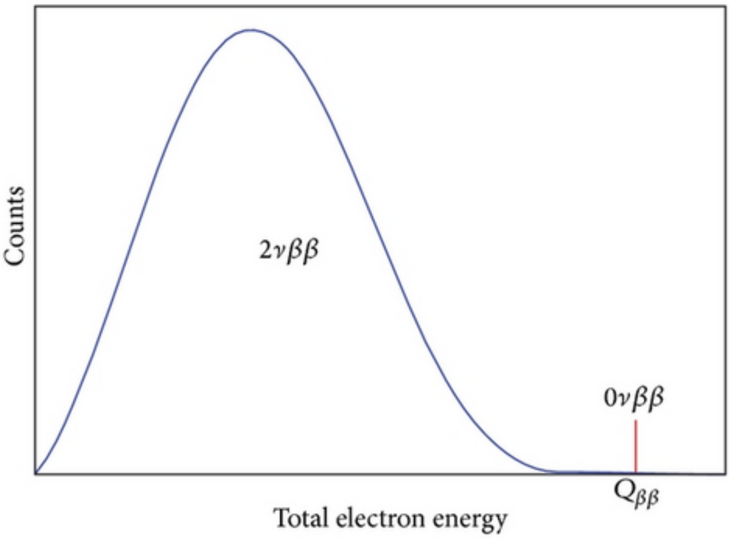


Fig. 2.4.: Schematic view of the $2\nu\beta\beta$ and the $0\nu\beta\beta$ energy deposition. For the latter, a monochromatic peak at the Q-value is expected [31].

would result in the electrons carrying the full Q-value of the decay, producing a sharp peak in the energy spectrum. Detecting this signature is a primary goal for most double beta decay experiments [33]. A summary of the experiments and which isotopes are being used is shown Table 2.1.

Isotope	$Q_{\beta\beta}$ [keV]	nat. ab. [%]	$T^{2\nu}_{1/2}$ [10^{20} y]	Experiments
^{48}Ca	4273.7	0.187	0.44	CANDLES [34]
^{76}Ge	2039.1	7.8	15	GERDA [35], Majorana[36].
^{82}Se	2995.5	9.2	0.92	SuperNEMO [37]
^{100}Mo	3035.0	9.6	0.07	AMoRe [38]
^{116}Cd	2809.1	7.6	0.29	Cobra [39]
^{130}Te	2530.3	34.5	9.1	CUORE [40]
^{136}Xe	2457.8	8.9	21	EXO [41], NEXT [42]
^{150}Nd	3367.3	5.6	0.08	SNO+ [43]

Tab. 2.1.: Q-values and natural abundance for isotopes that can go through $\beta\beta$ decays. Half-life of $2\nu\beta\beta$ process for those isotopes and some of the experiments performing $2\nu\beta\beta/0\nu\beta\beta$ searches with them [44].

$$\begin{aligned}
2\nu 2\beta^- &: (A, Z) \rightarrow (A, Z+2) + 2e^- + 2\bar{\nu} \\
0\nu 2\beta^- &: (A, Z) \rightarrow (A, Z+2) + 2e^- \\
2\nu 2\beta^+ &: (A, Z) \rightarrow (A, Z-2) + 2e^+ + 2\nu \\
0\nu 2\beta^+ &: (A, Z) \rightarrow (A, Z-2) + 2e^+ \\
2\nu EC\beta^+ &: (A, Z) + e^- \rightarrow (A, Z-2) + e^+ + 2\nu \\
2\nu 2EC &: (A, Z) + 2e^- \rightarrow (A, Z-2) + 2\nu
\end{aligned} \tag{2.8}$$

Other decay modes include the $2\beta^+$ process, where two protons convert into two neutrons, emitting two positrons and either two neutrinos ($2\nu 2\beta^+$) or no neutrinos ($0\nu 2\beta^+$). Additionally, hybrid modes such as electron capture with positron emission (ECβ^+) and double electron capture (2EC) exist [45]. These modes are less energetically favorable and occur less frequently, with longer expected half-lives than the $2\beta^-$ mode. In positron-emitting modes, the emitted positrons annihilate with electrons, producing two pairs of 511 keV gammas, which can serve as an experimental signature. While extremely rare, double electron capture may experience resonance effects that enhance its decay rate under specific conditions [46], making it a topic of interest for theoretical studies and experimental exploration.

3

Liquid Scintillators

Scintillators are materials that emit light when ionizing radiation interacts with them. This light emission occurs when the scintillator absorbs the energy from charged particles or electromagnetic radiation and re-emits it as photons in the visible or ultraviolet spectrum. Scintillators can be broadly classified into inorganic and organic types, distinguished by their material composition and scintillation mechanisms.

3.1 Inorganic Scintillators

Inorganic scintillators are usually ionic crystals, such as sodium iodide (NaI) or cesium iodide (CsI), doped with activator centers to enhance scintillation efficiency. Their detection mechanism involves the absorption of radiation energy, which excites electrons from the valence band to the conduction band of the crystal lattice. Excited electrons eventually return to the valence band via recombination at activator centers, emitting photons. The wavelength of the emitted light depends on the energy difference between the conduction band and the activator states [47].

3.2 Organic Liquid Scintillators

Organic scintillators are carbon-based materials composed of aromatic molecules with conjugated π-electron systems, such as benzene rings, that exhibit scintillation properties. When exposed to ionizing radiation, these molecules absorb energy, promoting pi-electrons from the ground state (S_0) to higher electronic states, such as the first excited singlet state (S_1) or higher triplet states (S_3). The scintillation process involves rapid, non-radiative relaxation to the lowest vibrational level of S_1 (10^{-12} s), followed by radiative decay back to S_0, emitting ultraviolet or visible light (350-750nm). This light emission, called fluorescence, occurs on a nanosecond timescale [48]. In contrast, triplet states decay more slowly via phosphorescence, often over milliseconds. The time-dependent luminescence dynamics of scintillators are typically modeled by a bi-exponential function [49]:

$$N(t) = A_p e^{-t/\tau_p} + A_d e^{-t/\tau_d} \tag{3.1}$$

where $N(t)$ is the number of excited molecules at time t, A_p and A_d are the amplitudes of the prompt and delayed components, and τ_p and τ_d are the respective decay time constants. To optimize performance, organic scintillators are doped with fluorophores, which enhance light yield and emission efficiency. Excited states, known as excitons, formed within the scintillator matrix transfer energy to primary fluorophores via the FRET

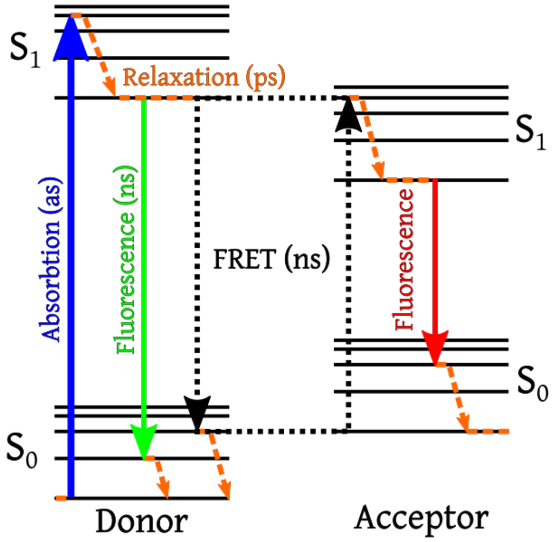


Fig. 3.1.: Jablonski diagram illustrating the Fluorescence Resonance Energy Transfer (FRET) mechanism in organic scintillators. The diagram illustrates the electronic transitions between donor and acceptor molecules, along with their typical timescales. Credits: A. Mooney [50].

mechanism [51], a non-radiative mechanism governed by dipole-dipole interactions. The efficiency of FRET scales as R^{-6}, where R is the distance between donor (matrix molecule) and acceptor (fluorophore) molecules. Secondary fluorophores are often added to shift the emission spectrum to longer wavelengths, minimizing re-absorption and extending the attenuation length [52].

The light yield of organic scintillators, typically 10000 photons/MeV, quantifies the number of photons emitted per unit of energy. However, quenching effects, especially at high ionization densities, reduce the light yield due to exciton-exciton annihilation and other non-linear processes [49]. These effects are described semi-empirically by Birk's law:

$$\frac{dL}{dx} = \frac{S(dE/dx)}{1 + kB(dE/dx)} \tag{3.2}$$

where dL/dx is the light yield per unit path length, S is the scintillation efficiency, dE/dx is the energy deposition per unit path length, and kB is Birk's constant which depends on the scintillating material.

3.3 Novel Techniques

The performance of liquid scintillators relies heavily on their optical transparency. This transparency ensures long scattering and absorption lengths, allowing light to propagate efficiently through large detector volumes. Impurities, temperature variations, and re-absorption can degrade transparency, affecting light collection and energy resolution. New technologies have been developed in the last few years to tackle some of these limitations [53, 54].

3.3.1 Hybrid Liquid Scintillators

Within scintillating materials, in addition to the primary scintillation light, Cherenkov radiation can be produced when charged particles pass through at speeds exceeding the phase velocity of light in the medium [55]. Unlike the isotropic scintillation emission, Cherenkov radiation is emitted in a cone along the particle's trajectory (Figure 3.2), offering directional information. The opening angle θ_c of the Cherenkov cone is determined by

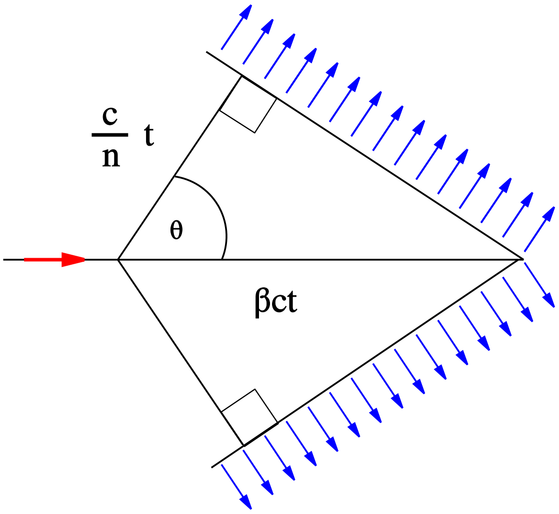


Fig. 3.2.: Schematic of the light cone emitted by Cherenkov radiation. Credit: A. Horváth [56].

$$\cos\theta_c = \frac{1}{n\beta} \tag{3.3}$$

where n is the refractive index of the medium, and $\beta = v/c$ is the ratio between the speed of the particle and the speed of light.

Hybrid liquid scintillators represent a novel detection technique by combining traditional organic scintillators with complementary detection mechanisms [54, 57].

There are two approaches of interest for hybrid scintillators: a Water-Based hybrid and a Hybrid-slow. The Water-Based technique combines

1-10% liquid scintillator with water, reducing scintillation light to separate it from Cherenkov light. It is cost-effective but has a lower light yield. In the Hybrid-slow case, fluors are added to delay scintillation light, producing a prompt Cherenkov light signal and a delayed scintillation light signal. Comparing the ratios of the two signals enables particle identification, as different particles generate these signals in varying proportions. Therefore, by analyzing the timing and intensity of both signals, particle types can be distinguished [58]. This technology provides full scintillation yield and good energy resolution, though it may have reduced spatial resolution.

3.3.2 Opaque Liquid Scintillators

Opaque liquid scintillators, pioneered by the LiquidO consortium [59], represent an innovative approach to particle detection [53]. These scintillators consist of a traditional liquid scintillator doped with different components, such as wax or metals, to form a diffusive medium for light propagation. Whereas classical liquid scintillators depend on long scattering and absorption lengths to efficiently collect light over large distances, a defining characteristic of opaque liquid scintillators is their short scattering length,

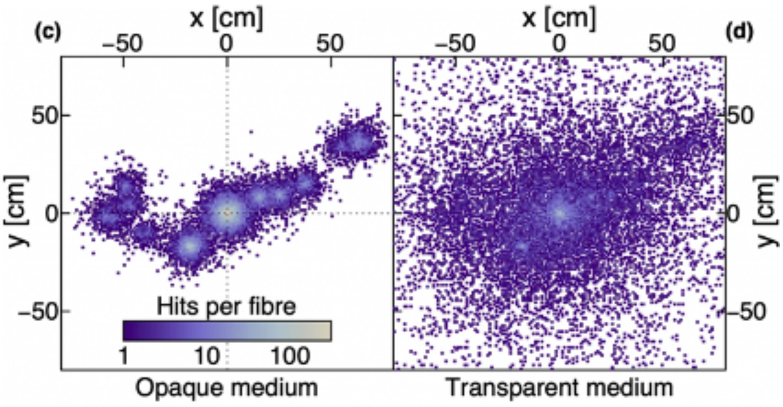


Fig. 3.3.: Example of a 1 MeV e^+ simulated event in an opaque scintillator (left) and in a transparent scintillator (right). Credit: LiquidO Consortium [59].

typically on the order of a few millimeters [53], which effectively traps scintillation light within a small region around the interaction point through multiple Mie scattering events, forming a localized "light ball" often referred as blob (bulky light object) that can then be collected using optical fibers. Mie scattering occurs when the size of the scattering particles is comparable to the wavelength of the light, causing the light to be confined and isotropically propagate within the medium. This design enhances spatial resolution and particle identification by preserving the details of the energy deposition patterns [59]. For instance, electrons produce a single localized blob with a characteristic size of only a few centimeters. Meanwhile, gammas generate a series of depositions as they undergo multiple Compton scatterings before a final photoelectric interaction. Positrons exhibit a hybrid pattern: an initial ionization deposition similar to electrons, followed by the annihilation process, which produces two back-to-back

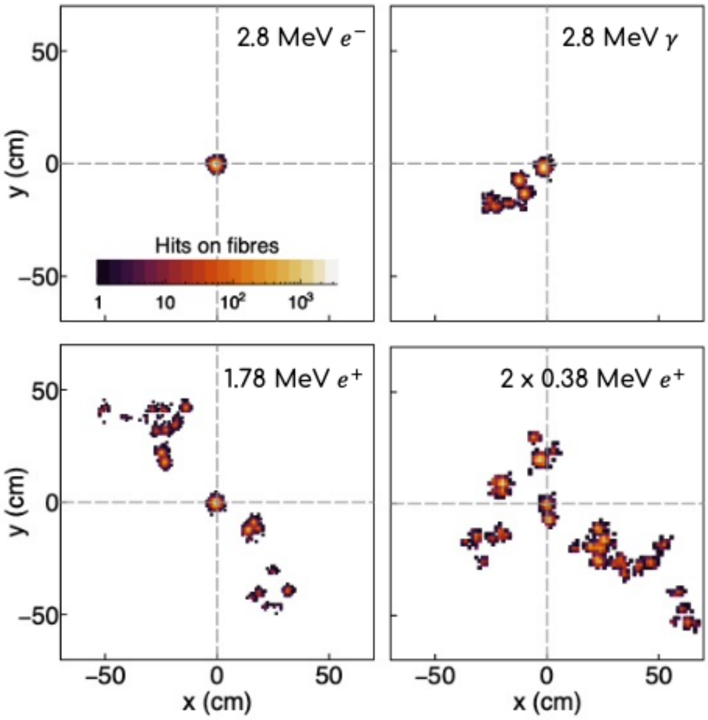


Fig. 3.4.: Example of different energy and particle depositions in an opaque scintillator. It is observed that the topology of the event is conserved which allows for particle discrimination [60].

gammas. These gammas subsequently create their Compton-scattering signatures. An example of this at the simulation level is shown in Figure 3.4.

Moreover, the reliance on short light travel distances reduces the need for stringent absorption length requirements. Thus, the optical properties of the scintillator can be tailored for various applications by adjusting the concentration and size of the dopants. In general, the configuration of a detector using an opaque liquid scintillator should consist of a scintillator volume, wavelength-shifting fibers to collect and transport the light, and photosensors to measure and convert the photons into electrical signals. In the context of this work, two relevant experiments that are planning to use simple or hybrid opaque liquid scintillator detectors in the context of neutrino physics are presented in the next chapter.

Application in Neutrino Detectors

4.1 The CLOUD experiment

The Chooz LiquidO Ultra-near Detector (CLOUD) experiment will be the third-generation neutrino experiment at the Chooz nuclear power plant, following the CHOOZ[61] and Double Chooz[62] experiments. CLOUD will exploit the particle discrimination capabilities offered by the LiquidO technology to study reactor neutrinos at a new level of precision. The CLOUD

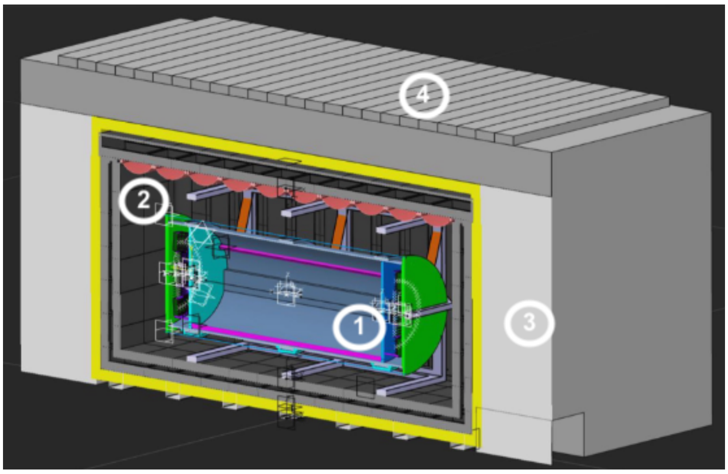


Fig. 4.1.: First proposed design of the CLOUD experiment. The inner volume (1) will contain 5-10 tons of opaque liquid scintillator instrumented with approximately 10,000 wavelength-shifting fibers running through the volume connected to SiPMS on both extremities. A veto volume (2) consisting of transparent liquid scintillator and PMTs. Surrounding this, a water tank, a concrete bunker (3) and a steel layer (4) are included for shielding.

detector (Figure 4.1) will be positioned at the Ultra-near site at the Chooz power plant, approximately 30 meters from the reactor core, with about 3 meters of water equivalent (m.w.e) overburden and an expected rate of around $\sim 10^4$ inverse beta decay events per day. Being installed at the surface level presents significant challenges due to background noise from cosmic and atmospheric neutrons [63]. The proposed design includes an inner detector, an outer veto detector, and shielding. The inner detector will feature a cylindrical volume approximately 1 meter in radius and 4 meters long, containing 5 to 10 tonnes of opaque scintillator. This scintillator will be instrumented with 10,000 wavelength-shifting fibers arranged optimally and connected to silicon photomultipliers (SiPMs) at both ends. Surrounding the inner detector will be a veto volume made of transparent scintillator and photomultiplier tubes (PMTs), encased within a water tank and a concrete bunker for additional shielding. An extra 15 cm steel layer will further reduce background contamination. The first phase of the CLOUD physics program will focus on reactor antineutrino detection, aiming to accurately measure the reactor antineutrino flux via inverse beta decay interactions [63]. In this process, an antineutrino interacts with a proton, producing a positron and a neutron. The positron generates an initial scintillation burst through ionization and annihilates with an electron, emitting two 511 keV gamma photons. The neutron thermalizes before being captured by a nucleus, resulting in delayed gamma emission. This sequence of prompt and delayed signals provides a unique temporal signature for neutrino interactions [64]. The properties of the opaque scintillator will enable CLOUD to distinguish these antineutrino signals, characterized by positron annihilation, from background events dominated by electron and gamma interactions, ensuring a high signal-to-background ratio (in the order of 100).

4.2 The NuDoubt^{++} experiment

NuDoubt^{++} [60] is an experiment under development designed to measure double weak decay processes, including both two-neutrino and the hypothesized neutrinoless positive beta decays (see Subsection 2.3.1). The positive beta decay channel is specifically targeted because it produces positrons, which are very challenging to detect through traditional methods. By employing a novel detector technology that combines a hybrid-slow and opaque scintillator resulting in enhanced particle identification (PID), NuDoubt^{++} will be optimized to reliably identify these positrons. Combining the two scintillators by introducing opacity through wax loading to the hybrid-slow scintillator provides both a prompt Cherenkov and a delayed scintillation signal. Analyzing the Cherenkov-to-scintillation light ratio allows for the temporal separation of the signals, which can be used for particle discrimination [58]. Together with the cm-scale topological properties of the opaque scintillator, particle identification, and energy

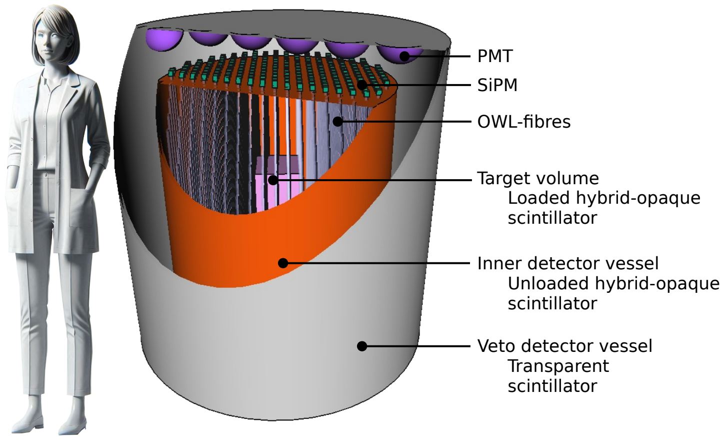


Fig. 4.2.: Proposed detector design of NuDoubt^{++}. It consists of a target volume containing an isotope-loaded opaque scintillator placed inside the inner volume. This is filled with a hybrid-opaque scintillator and instrumented with OWL fibers along the detector's vertical axis. Dedicated SiPMs and electronics will be at both ends of each fiber. Finally, a veto detector vessel equipped with a transparent scintillator and PMTs.

measurement are improved, providing a powerful tool for background discrimination crucial for this kind of experiment.

In the original proposal for experiments utilizing the LiquidO technology, light read-out is performed using commercial wavelength-shifting fibers; however, NuDoubt^{++} will employ optimized wavelength-shifting fibers known as OWL fibers [65]. These fibers are based on advancements in Wavelength-shifting Optical Module (WOM) technology [66], which was initially developed for the IceCube Upgrade [67]. Made of polystyrene, they are coated on the outer surface with a wavelength-shifting paint layer. This increases photon capture efficiency through total internal reflection, especially in the ultraviolet range, compared to commercial fibers, expecting a trapping efficiency of up to 38% [65].

The first design proposal (Figure 4.2) features a primary volume configured as a cylinder with a diameter and height of 110 cm, containing one tonne of a specially formulated hybrid-slow opaque scintillator, an embedded cubical target volume for isotope loading, and a grid of OWL fibers. These fibers, each 3 mm in diameter and arranged with a 1 cm pitch in a hexagonal grid running parallel to the axis of the detector (Figure 4.3), are coupled to dedicated silicon photomultipliers (SiPMs) with associated electronics at both ends. The fiber layout and pitch are critical parameters that directly influence detection efficiency and the resulting light patterns, necessitating further detailed studies. Surrounding this inner volume is an active veto, a cylinder of 155 cm in diameter and height, filled with a transparent scintillator and equipped with photomultiplier tubes (PMTs) at the top and bottom.

This work will focus on reconstructing events based on a NuDoubt^{++} configuration. Therefore, the following section presents the foundation for simulating and processing the expected events in this detector.

4.2.1 Event Simulation in an Opaque Scintillator Detector

A Geant4 [68] simulation was developed to model particle interactions within an opaque liquid scintillator. Optical photons are produced via scintillation or Cherenkov processes; these photons subsequently propagate through the detector volume. Their behavior is governed by material-specific optical properties such as the light yield, the scattering length, absorption length, refractive index, and the time decay constants of the scintillator τ_1 and τ_2 (see Subsection 3.2). However, these optical properties provide only a first-order approximation, as they are assumed to be constant across all wavelengths. In reality, their values may vary with wavelength, requiring a more detailed model in future studies. The exact constant values used for the simulation are presented in Table 4.1. To reduce computational cost, photon propagation within the optical fibers is

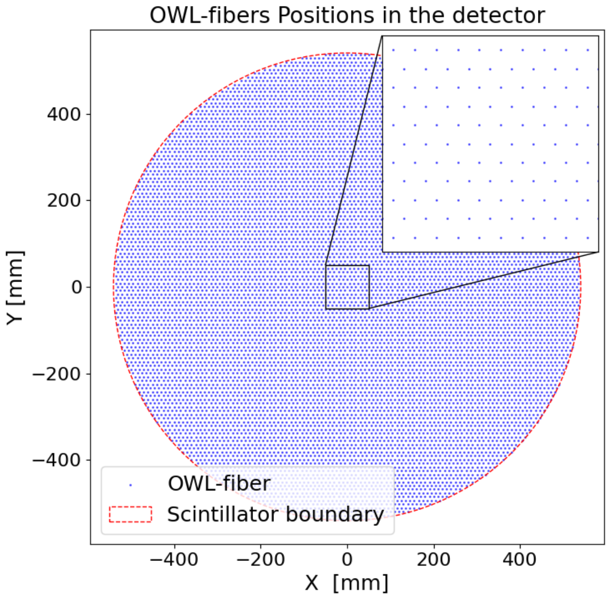


Fig. 4.3.: Top view (x-y) of the inner detector volume illustrating the positions of OWL-fibers in the simulation. The fibers are arranged in a parallel hexagonal distribution along the z-axis (HexZ) with a 1 cm pitch, providing coverage for optimal light collection.

not explicitly simulated. Instead, photons are removed from the simulation as soon as they interact with a fiber rather than tracked until detected by the SiPM. This approximation significantly reduces memory usage while still providing an estimate of photons collected by the light-readout system (fibers and SiPMs). Thus, a more detailed study of fiber and SiPM response will be necessary to accurately characterize the full detection process. A different set of simulations was produced to benchmark the detector's performance. These simulations included events such as photon bombs, simulated bursts in which a discrete number of photons is released instantaneously from a point source to mimic a concentrated energy deposition, as well as events generating 1 MeV electrons, positrons, and gammas. All these scenarios were generated at the center of the detector (Figure 4.4) to evaluate its response to typical energy deposition patterns. Additionally, the simulation provides access to true information, recording individual interactions produced by the initial simulated particle, including their positions and deposited energy. This serves as a reference for event reconstruction, enabling a comparison between reconstructed and actual interaction data. This will be further discussed in Chapter 6.

Scintillator Properties						
Composition	L.Y [p.e/MeV]	τ_1 [ns]	τ_2 [ns]	Scattering length	Abs. length	R.I.
98% LAB + 2% paraffin wax	9000	13	26.3	2 mm	2 m	1.48
Fiber Properties						
Composition	Diameter	τ_{fiber} [ns]	ϵ_{fiber} [%]	$1/v_{\text{fiber}}$ [mm/ns]	l_{att} [m]	
Polystyrene	3 mm	2.8	7.20	160	2.5	

Tab. 4.1.: Scintillator and fiber properties used in the detector simulation and readout model. L.Y. is the light yield in photoelectrons per MeV, τ_1 and τ_2 are the fast and slow decay times of the scintillator, the scattering and absorption lengths characterize light propagation, and R.I. is the refractive index. For the fibers, the composition (Polystyrene) is not explicitly simulated, τ_{fiber} is the fiber decay time, ϵ_{fiber} is the efficiency (accounting for trapping, quantum efficiency, and optical coupling), t_{fiber} average time for a photon to travel 1 m of fiber, and l_{att} is the attenuation length (m).

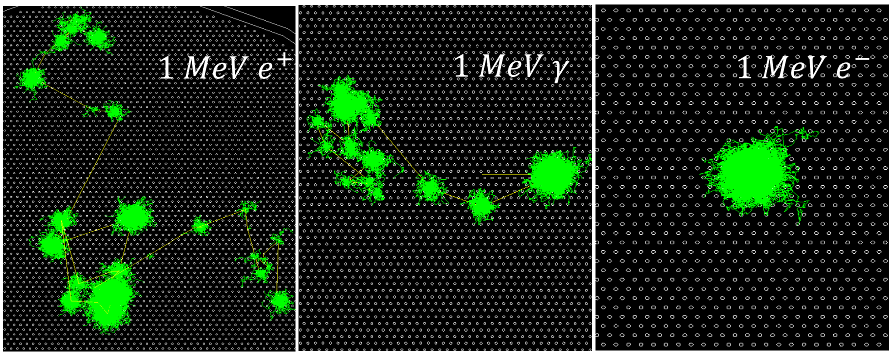


Fig. 4.4.: Geant4 visualization of 1 MeV positron, gamma and electron depositing energy in the detector, and optical photons propagation. The particles are generated at the center of the detector.

4.2.2 Fiber Propagation and SiPM Detection Model

As an output of the simulation, various physical quantities are obtained, including the event number, the three-dimensional position (x, y, z) of each photon when it hits a fiber, the time (t) at which it interacts with the fiber, and the $(x_{\text{fiber}}, y_{\text{fiber}})$ coordinates of each fiber within the detector. To provide a more realistic approximation of the detector response, initial modeling of fiber propagation and subsequent arrival to the SiPM is implemented. This requires incorporating key fiber properties such as decay time (τ_{fiber}), fiber transit time (*tts*), average time for a photon to travel 1 m of fiber (t_{fiber}), attenuation length (l_{att}), and fiber efficiency (ϵ_{fiber}). The decay time of the wavelength-shifting fiber is characterized by a probability distribution of the form:

$$p(t) \sim e^{-t/\tau_{\text{fiber}}} \tag{4.1}$$

As mentioned, a photon is considered to belong to a fiber if it has been absorbed during its interaction with that fiber. However, this interaction can occur at any position around the fiber's boundary, and this information is not accessible in reality. Therefore, as a first step in the process, each optical photon is assigned to a fiber based on the spatial proximity of its position to the center of the fiber. This assignment is determined by calculating the Euclidean distance:

$$d_{\text{center}} = \sqrt{(x - x_{\text{fiber}})^2 + (y - y_{\text{fiber}})^2} \tag{4.2}$$

Following this, the position of the photon will be simply the position $(x_{\text{fiber}}, y_{\text{fiber}})$ of its fiber. Then, a 50% probability is assigned for the photon traveling to either the top or bottom of the detector, where the SiPMs are located. If the photon is directed towards the top (or bottom) SiPM, its traveled distance is determined as:

$$d_{\text{travel}} = \frac{L_{detector}}{2} - Z_{\text{event}} - z \tag{4.3}$$

where Z_{event} represents the initial event generation position (e.g, Z_{event} = 0 at the center of the detector). After that, the detection efficiency (ϵ_{det}) is calculated as:

$$\epsilon_{\text{det}} = \epsilon_{\text{fiber}} \cdot e^{-d_{\text{travel}}/l_{\text{att}}} \tag{4.4}$$

Based on this efficiency, a random probability is assigned to determine whether the photon is detected. Finally, if the photon is detected in the SiPM, its arrival time is computed as:

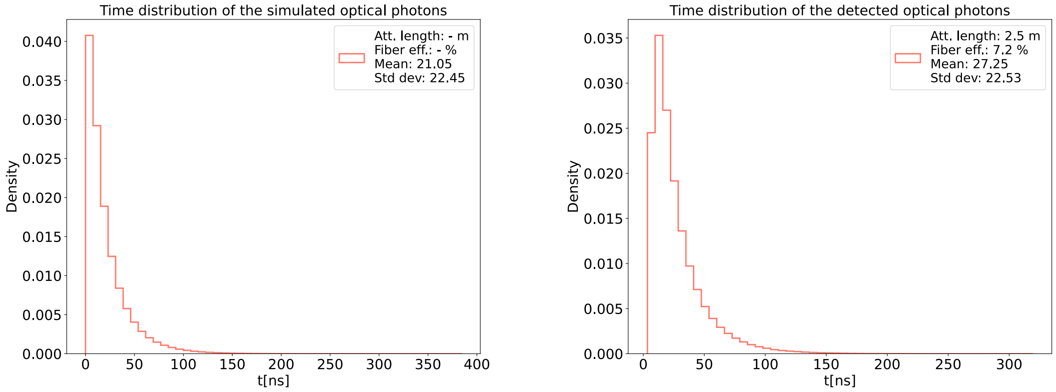
$$t_{\text{arrival}} = t + t_{\text{WLS}} + S + (d_{\text{travel}} \cdot 1/v_{\text{fiber}}) \tag{4.5}$$

where the spread (S) accounts for the transit time spread (*tts*) that characterize how much the arrival times of the photons vary as they travel along the fiber, quantifying how much timing uncertainty is introduced per unit length in the fiber. This variation arises from factors such as path length differences, scattering and other optical effects in the fiber. The spread is sampled from an uniformly distributed function from $[-0.5, 0.5]$ to introduce a symmetric random variation. Thus, defined as:

$$S = (X - \frac{1}{2}) \cdot d_{\text{travel}} \cdot \text{tts} \tag{4.6}$$

with $X \sim U(0, 1)$ a random variable following a uniform distribution on the interval $[0, 1]$. This approach is a simplified way to account for the fact that not all photons will arrive at the exactly same time to the SiPMs. Figure 4.5

shows the photon time distributions for 1 MeV e^-; when it hits the fiber (Figure 4.5a) and when it arrives to the SiPM (Figure 4.5b). After this step, each



(a) Original time distribution of the total simulated optical photons produced by a 1 MeV e^-.

(b) Arrival time to the SiPM distribution of the detected optical photons produced by a 1 MeV e^-.

Fig. 4.5.: Time distributions before and after the fiber propagation and SiPM model.

detected photon is characterized by its fiber-assigned $(x_{\text{fiber}}, y_{\text{fiber}})$ position, the z-coordinate where it entered the fiber, its arrival time, and its propagation direction. A summary of the execution time of this whole process is presented in Table 4.2. It is important to note that this processing method offers only an approximation of how photons propagate through fibers to reach the SiPMs. A more detailed and sophisticated analysis is necessary to achieve a truly accurate detector response. However, this approach is valuable for preliminary studies on detection efficiency and fiber optimization. In particular, it can be used to explore strategies for maximizing photon

Particle	# Events	Energy [MeV]	Total Simulated Photons	Total Detected Photons	Execution Time [s]
e^-	2500	1	2.27×10^7	1.31×10^6	46.17
e^+	2500	1	4.16×10^7	2.40×10^6	201.25
γ	2500	1	1.83×10^7	1.06×10^6	89.85

Tab. 4.2.: Type of simulated events used in this work. All events were simulated at the center of the detector and fixed energy. Additionally, the total simulated and detected photons and the execution times for the fiber propagation and detection model.

collection efficiency, which is especially relevant in applications involving OWL fibers. By refining these models, one can optimize fiber properties to enhance detector performance in future experiments.

5

Cambridge-Aachen Algorithm for Opaque Scintillators

Originally developed for high-energy physics applications to identify particle jets in collider experiments, jet clustering algorithms are well-suited for the reconstruction challenges presented by opaque scintillators. These algorithms can group energy deposits based on proximity and kinematic properties, making them a natural choice for analyzing the spatially distributed light signals in an opaque detector. Additionally, their established theoretical foundation, ease of implementation, and lower dependence on large-scale training data make them a practical solution for this context.

The next section presents the background and definition of a family of jet clustering algorithms, focusing on the Cambridge-Aachen (C-A) algorithm and how this can be adapted and implemented as a robust and scalable reconstruction method for events in an opaque scintillator.

5.1 Jet Clustering Algorithms

Jet clustering algorithms are essential tools in high-energy physics. They are designed to identify jets: collimated bunches of particles produced by the hadronization of quarks and gluons in particle collisions. The concept of jet clustering dates back to the late 1970s when the need to analyze the complex particle data generated in experiments at colliders emerged [69].

The primary goal of these algorithms is to group particles into jets in a manner consistent with the underlying partonic process. This grouping allows for the reconstruction of the kinematic properties of the initial quarks and gluons, providing insights into the dynamics of the collision. Early approaches consisted of simple cone-based algorithms [69] where jets were defined as groups of particles within fixed angular regions. However, they were limited by the ambiguities in jet definition and sensitivity to soft radiation, which led to the development of more sophisticated approaches such as sequential recombination algorithms [70]. The formal development of these algorithms, also called "k_t-algorithms", was proposed in the 1990s [70, 72, 73]. These algorithms introduced a more systematic framework for jet reconstruction by grouping particles based on their kinematic and spatial properties. These algorithms became the standard for jet reconstruction until the last couple of years with the implementation of machine-learning methods [74]. Sequential recombination algorithms

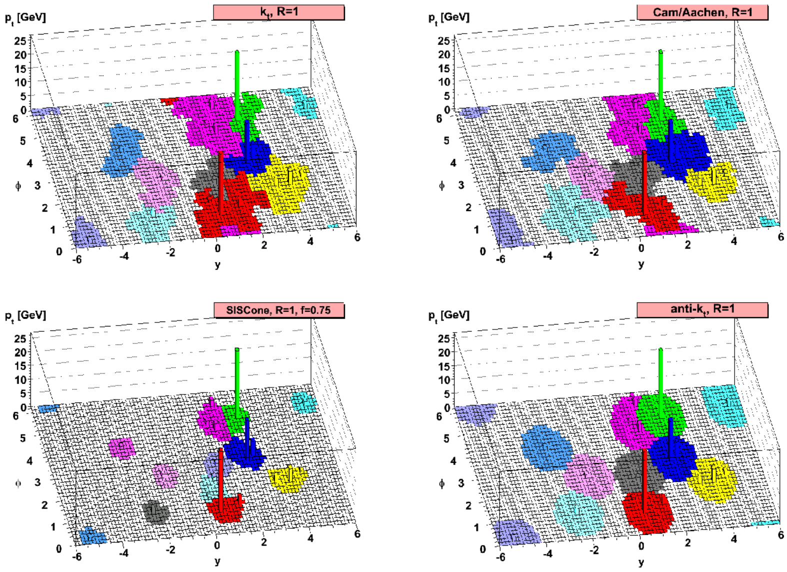


Fig. 5.1.: Example of the different $k_t t$ clustering algorithms [71]. It illustrates how each of the methods defines a way to define an area in which the particles form clusters.

work by calculating pairwise distances between particles and iteratively merging them according to specific distance measures, ultimately reconstructing the jets produced in the collision [71]. The distance measure in these algorithms is defined as:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta R_{ij}^2}{R^2} \tag{5.1}$$

where d_{ij} is the distance between particles i and j, k_t represents the transverse momentum, ΔR_{ij}^2 is the squared angular distance (usually in terms of the pseudorapidity and azimuthal angle), R is a resolution parameter that controls the angular size of the jets [71], and p is the *jet algorithm exponent* or *power parameter*. Additionally, a beam distance is defined for each particle as:

$$d_{iB} = k_{ti}^{2p} \tag{5.2}$$

where d_{iB} is the beam distance measure for particle i, k_{ti} is its transverse momentum and it determines how a final jet is constructed. The value of the parameter p results on different versions of the algorithm [71]:

- k_t algorithm ($p = 1$) tends to cluster soft particles first and is less sensitive to detector noise.
- Cambridge-Aachen algorithm ($p = 0$) uses only angular distance for clustering, making it simpler and particularly effective for jet substructure studies, where resolving the internal splitting patterns is important [75].
- anti-k_t algorithm ($p = -1$) preferentially forms geometrically regular, cone-like jets, making it particularly robust in high-pileup environments.

Despite their advantages, early sequential recombination algorithms were computationally intensive, scaling as $O(n^3)$, where n represents the number of particles [70]. This inefficiency restricted their use in environments with high multiplicity. FastJet [76] overcomes this challenge by exploiting the rapidity-azimuth plane's geometry. It uses spatial partitioning and optimized data structures for efficient nearest neighbor searches, reducing computational complexity to $O(n \ln n)$ [77]. Its efficiency enabled the application of jet clustering algorithms to large datasets in real-time analyses, transforming them into a practical tool in high-energy physics experiments.

The next section delves into the Cambridge-Aachen algorithm [78, 79] and its implementation within an opaque scintillator detector.

5.2 Implementation in Opaque Scintillator Detectors

5.2.1 Adaptation of the Clustering Algorithm

The Cambridge-Aachen version of the k_t-algorithms is suggested for event reconstruction in an opaque scintillator because it clusters signals based solely on the angular (geometrical) separation rather than their kinematics. This geometrical approach results useful given the spatial pattern of light depositions observed in these kind of detectors. The algorithm corresponds to the case where $p = 0$ in Equations 5.1 and 5.2, defining then the distance between particles i and j as the squared distance ΔR_{ij}^2 over the resolution parameter R and a beam distance of one:

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \tag{5.3}$$

Adapting the algorithm in this context requires modifications to account for the optical properties and responses of the detector. First, the input data differs from the original formulation: instead of particle information, the algorithm processes optical photons, and rather than forming jets, it clusters these photons into blobs. Next, the distance measure d_{ij} must be tailored to the specific experimental conditions. While in the traditional context, the resolution parameter R is assigned a value reflecting the underlying jets physics (typically between 0.8 and 1), there is not a direct translation into the experimental context of an opaque scintillator detector. Consequently, a value of $R = 1$ cm is adopted. The squared distance ΔR_{ij}^2 can be defined in any parameter space depending on the experimental goals and available input variables. For example, in the cases discussed in Section 4, where light is collected at both ends of the detector, one could perform the clustering algorithm first in the (x, y) plane. In this scenario, a two-dimensional Euclidean distance is employed:

$$\Delta R_{ij}^2 = \Delta x_{ij}^2 + \Delta y_{ij}^2 \tag{5.4}$$

where x and y denote the positions of the fibers and associated SiPMs in the detector. Another important consideration is the beam distance (d_{iB}) introduced in Eq. 5.2. This parameter serves as a threshold in determining when a particle belongs to a final jet. In the context of an opaque scintillator detector, this can be translated as the threshold $(\mathbf{d_{th}})$ to determine when a photon belongs to a final blob. This threshold is defined as a constant related to the fiber pitch and the average blob size and is one of the parameters that must be optimized (see Subsection 5.2.3).

5.2.2 Algorithm's Implementation

The algorithm implementation for an opaque liquid scintillator detector follows a structured approach to grouping first the SiPM hits based only on their spatial proximity in the transversal (x, y) plane. Each SiPM hit has its own set of properties like spatial coordinates (where x and y are the position of the SiPM in the grid and z identifies if it is a bottom or top SiPM), arrival time and an assigned energy of 1. Initially, each hit is treated as an individual cluster. The clustering can be broken down into three main steps:

1. **Distance Matrix Calculation:** A symmetric distance matrix D is constructed, where each entry represents the distance between two clusters based on the given distance measure d_{ij}.

2. **Iterative Merging:** The algorithm systematically identifies the closest pair of clusters and merges them if their separation is below a predefined threshold d_{th}. The center of the newly formed cluster is determined as a center of mass weighted by its energy (number of hits in the cluster); its time is updated as the mean time. Using this methodology, the clustering process is solely influenced by the spatial separation of SiPMs hits.

3. **Updating Clusters:** After merging, the cluster list and distance matrix are updated by substituting the old clusters with the newly merged one and calculating the pairwise distances again.

This process continues until no clusters remain within the merging threshold. The algorithm is implemented as a C++ and ROOT application (See Appendix A.1). It uses as input the optical photons detected following the procedure described in Section 4.2.2. Each cluster is characterized by the coordinates of its center of mass (X, Y, Z), its time information, and energy. While only the clustering on the transversal plane has been discussed here, a second step of the clustering for the longitudinal plane (z) is discussed in detail in Subsection 6.1.1.

5.2.3 Distance Threshold Optimization

The distance threshold is a critical parameter in the clustering algorithm, directly determining how light depositions are grouped into clusters. A properly chosen threshold ensures that contiguous light signals are correctly identified as distinct clusters, without merging separate features or fragmenting a single deposition into multiple clusters. If the threshold is too large, distinct light originating from the same particle interaction in the scintillator may be merged, obscuring fine structural details. Conversely, if the threshold is too small, a single deposition may be artificially split into several clusters, degrading the reconstruction quality.

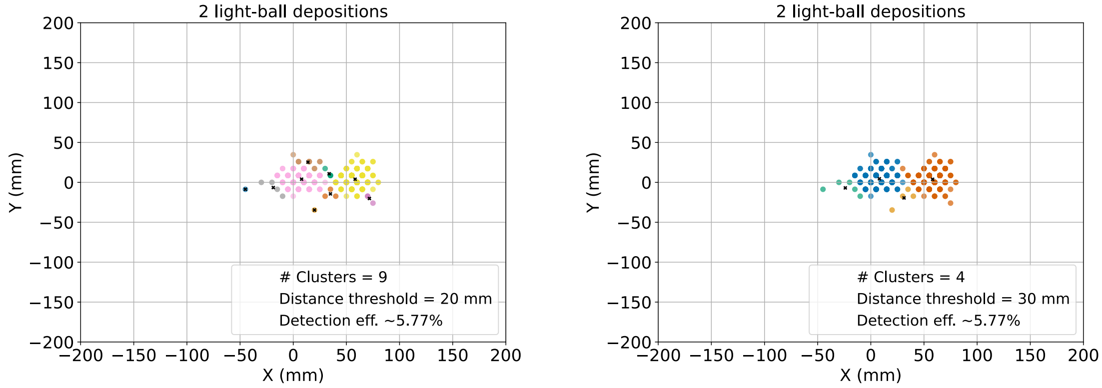
To define a reasonable threshold range, the blob size and fiber pitch must be considered. For an energy deposition of 1 MeV, a fiber pitch of 1 cm and a scattering length of 2 mm, a blob's characteristic diameter is approximately 5 cm, where 90% of the light is confined. This reflects the effective spatial extent of the signal. Conversely, the fiber pitch indicates the detector's intrinsic granularity, establishing a lower bound on the spatial resolution. Therefore, these parameters provide a scale for the clustering analysis, suggesting that the distance threshold should be selected within a range sensitive to both. To refine the selection process, the clustering algorithm was tested using controlled light-ball depositions (photon bomb[1] events). Two test cases were examined:

1. **Single photon bomb:** In this case, the ideal outcome is the formation of a single, centralized cluster, reflecting an undivided light deposition.
2. **Two photon bombs:** Here, they were placed in the center of the detector, 5 cm relative to each other. The selection of this separation is primarily based on the diameter of a typical light deposition and it is used to simulate the critical distance at which two overlapping clusters are just barely distinguishable. This configuration, allows adjusting the threshold so that the two depositions are resolved as separate clusters without fragmenting either individual blob.

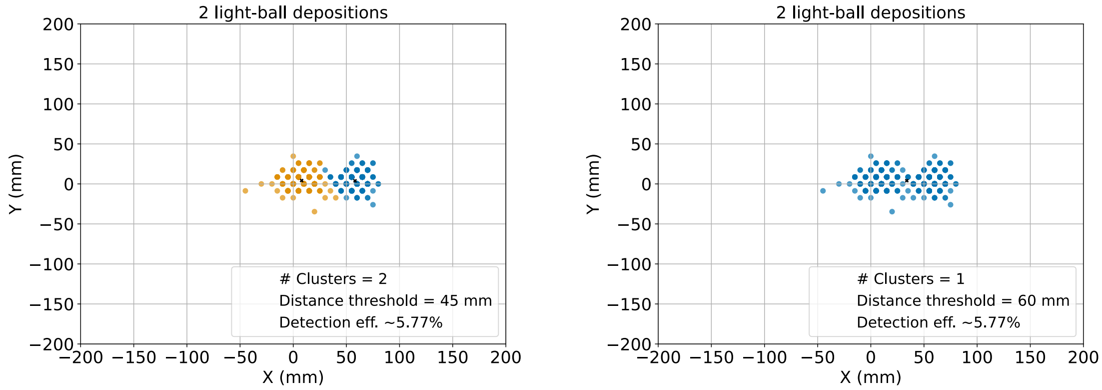
The clustering algorithm was run for these tests with values of the distance threshold (d_{th}) ranging from 20 mm to 60 mm in 5 mm increments. An event visualization for the two photon bombs is displayed in Figure 5.2. Additionally, a quantitative analysis of the average number of clusters versus the threshold (Figure 5.3) indicates that at approximately 45 mm, the algorithm can distinguish two contiguous light depositions into separate clusters while maintaining the integrity of a single photon bomb. Based on this, a 45-mm distance threshold is selected for subsequent clustering

---

[1] spherical photon source with a fixed number of photons used to mimic a concentrated energy deposition



(a) Clustering with $d_{th} = 20$ mm **(b)** Clustering with $d_{th} = 30$ mm



(c) Clustering with $d_{th} = 45$ mm **(d)** Clustering with $d_{th} = 60$ mm

Fig. 5.2.: Event visualization in the (x, y) plane of the clustering with different distance threshold values for the case of two contiguous light-balls. Each cluster is represented by a distinct color and the 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2. The distance threshold used in the algorithm has a direct impact on how the depositions are grouped. If this value is too small, it results in the main deposition being split into multiple clusters. If a value exceeding the average size of a deposition is chosen all the light gets clustered together losing the spatial structure of the event.

analyses. This value provides a justified balance between resolving adjacent blobs and conserving the integrity of a single cluster. Future work should incorporate additional considerations, such as complex event topologies with multiple overlapping depositions or variations in detector response and light intensity, to further refine the distance threshold selection.

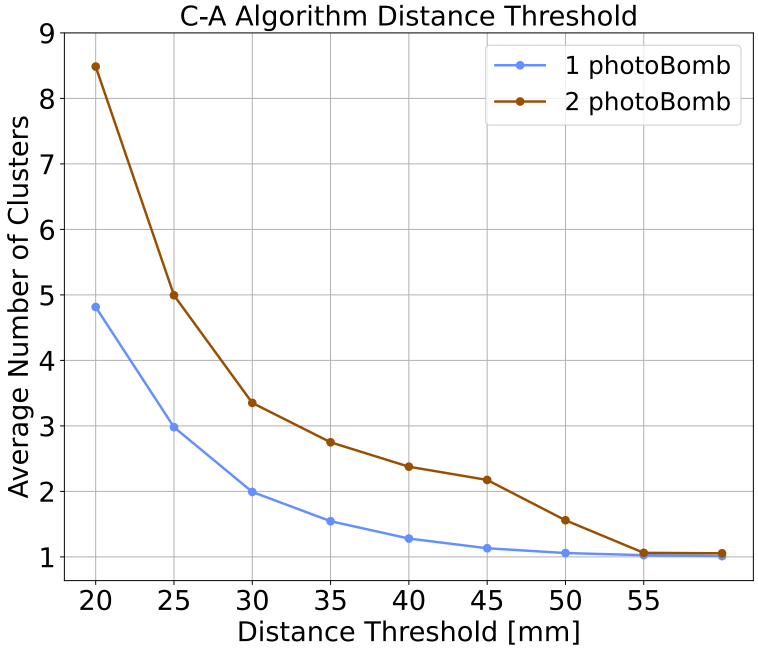


Fig. 5.3.: Optimization of the distance threshold for the two cases: a single photon bomb and a double photon bomb. For each case 1000 events were simulated.

5.2.4 Cluster Separation

As briefly mentioned before in Subsection 5.2.2, after clustering in the (x, y) plane on the top and bottom SiPM arrays and determining an appropriate distance threshold for the clustering (Figure 5.4 shows these clustering results for a 1 MeV e^+ event), a matching strategy is required to determine whether clusters from opposite sides correspond to the same light deposition event. After identifying if a SiPM hit was recorded in the top or bottom of the detector, all hits are merged into a single dataset. The clustering algorithm is then applied to this set, resulting in single collection of clusters that reflect the overall spatial structure of the light deposition. By subsequently separating the light contributions of each cluster, the signals at the top and bottom belonging to a single cluster can be separated. This procedure also facilitates the reconstruction of the Z position of each cluster as it will

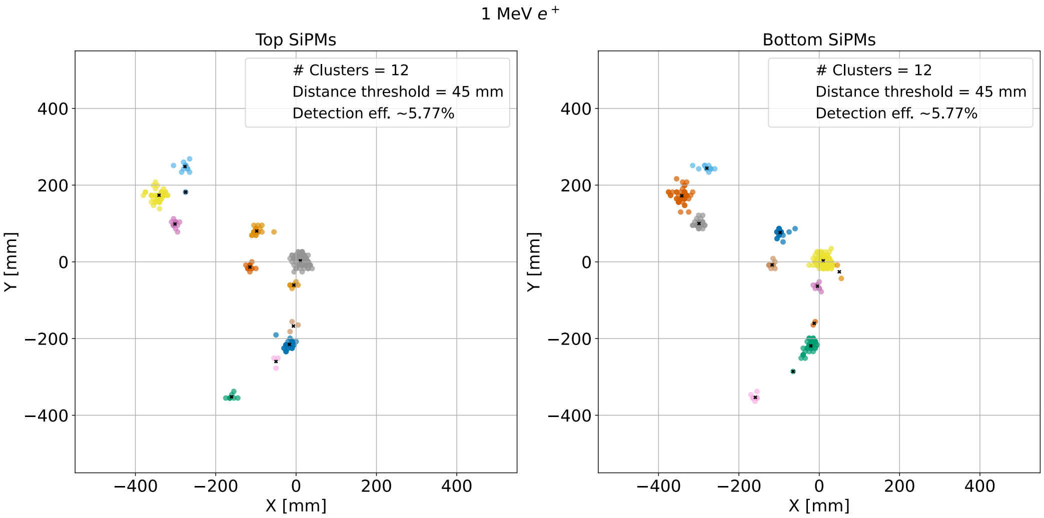


Fig. 5.4.: Clustering results for a 1 MeV e^+ event before the matching procedure. The figure illustrates the event topology, where light depositions are spatially grouped into distinct clusters. For a given detector side, each cluster is represented by a different color. The left (right) plot represents the clusters for the top (bottom) SiPMs. The "x" marker indicates the center of mass (COM) position in the (x, y) plane for each cluster. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2

be detailed in Subsection 6.1.1. Figure 5.5, shows the corresponding light depositions on each side of the detector by using the same color of the cluster they belong to.

It should be noted that this process may result in an unequal number of clusters on each side; for example, a predominantly top-originating light signal may leave some bottom clusters unmatched. Although these unmatched clusters still contain valuable event information, their optimal interpretation requires further investigation. Moreover, while the current matching approach is effective for detectors with parallel fiber arrangements, more complex configurations, such as those employing stereo geometries, may benefit from advanced techniques, including image pattern recognition or graph-based methods.

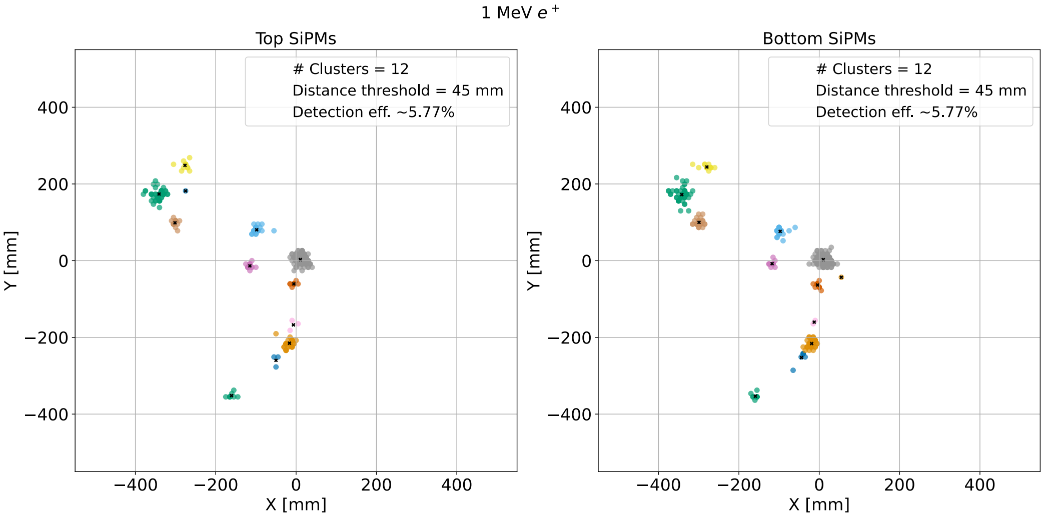


Fig. 5.5.: Matching results for the clusters of a 1 MeV e^+ event. The figure illustrates the event topology, in this case of multiple light depositions across the (x, y) plane of the detector. The left (right) plot represents the clusters for the top (bottom) SiPMs, with the same color meaning they belong to the same cluster. The "x" marker indicates the center of mass (COM) position. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2.

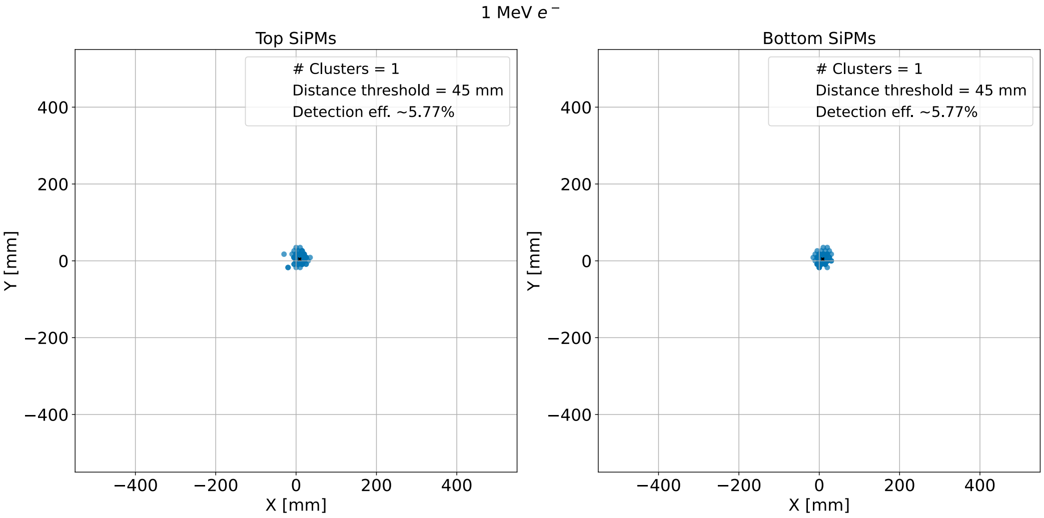


Fig. 5.6.: Matching results for the clusters of a 1 MeV e^- event. The figure illustrates the event topology, in this case a single light deposition. The left (right) plot represents the clusters for the top (bottom) SiPMs, with the same color meaning they belong to the same cluster. The "x" marker indicates the center of mass (COM) position in the (x, y) plane. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2.

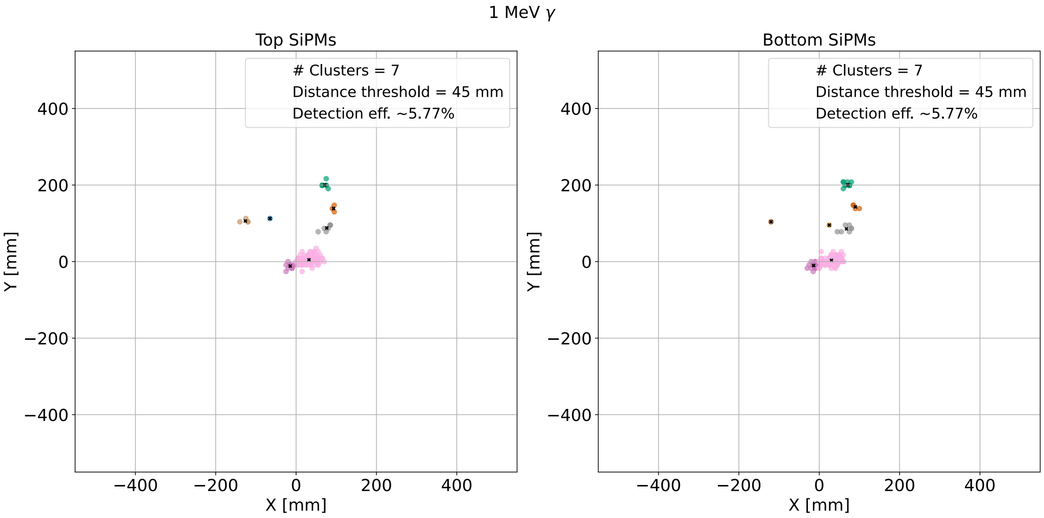


Fig. 5.7.: Matching results for the clusters of a 1 MeV γ event. The figure illustrates the event topology, in this case of multiple light depositions across the (x, y) plane of the detector. The left (right) plot represents the clusters for the top (bottom) SiPMs, with the same color meaning they belong to the same cluster. The "x" marker indicates the center of mass (COM) position. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2.

5.2.5 Computational Performance

The primary computational drawback of the C-A algorithm is its inherent $O(n^3)$ complexity: it calculates all distances between all particles (a $O(n^2)$ step), before recombining them ($O(n)$ step), where n denotes the number of hits. Although detection efficiency criteria, such as light loss in the fibers and SiPMs, significantly reduce the size of the input passed onto the algorithm, it is important to understand how its runtime scales with the number of hits. Specifically, if it is eventually to be used in scenarios with high event rates, high-energy depositions, or complex interactions (e.g., from positrons or gammas), this could result in a sufficiently large n that creates a bottleneck. To explore this behavior, both independent transversal plane clustering on each side of the detector and the matching clusters approach were executed for different numbers of SiPM hits. Figure 5.8 shows the number of hits with the total clustering time. To estimate the

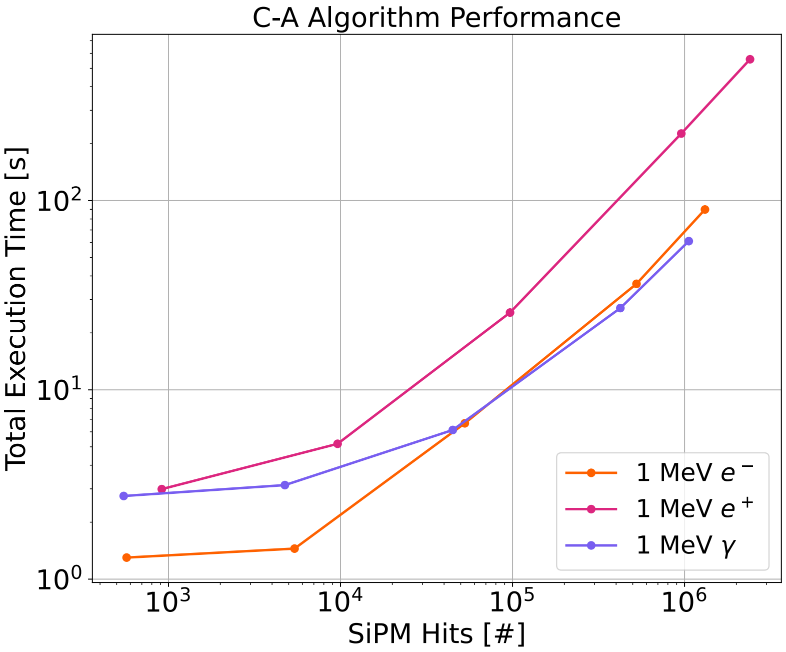


Fig. 5.8.: Number of detected photons versus the total clustering time for the three types of particle events analyzed.

complexity, after plotting the logarithm of both the execution time (T) and number of SiPM hits (n) we can check if the slope (p) or effective exponent follows a power law:

$$T(n) = n^p \tag{5.5}$$

then taking the logarithm:

$$\log(T) = p\log(n) \tag{5.6}$$

to estimate p a linear regression on $\log(n)$ and $\log(T)$ was performed the value found was of approximately ≈ 0.72 for the three type of simulated particles. The consistent slope across three execution time measurements suggests that the algorithm exhibits a sub-linear complexity close to $O(n^{2/3})$. This reduction from the worst case complexity of $O(n^3)$ to the observed $O(n^{2/3})$ might be due to the data reduction from detection efficiency considerations, the early termination when no more clusters can be merged (meaning the worst-case scenario does not always occur), or applying the clustering separately to the top and bottom SiPM hits. However, this preliminary finding should be validated with further profiling by isolating the processing time of the distance calculations in the clustering, a detailed examination of the optimizations such as the data processing, and an assessment of hardware and memory effects.

However, the underlying computational challenge is expected to be the brute-force calculation of the distance matrix, which becomes increasingly demanding as the number of photons grows. Optimizing this step is essential to making the algorithm feasible for opaque scintillator detectors. A potential improvement is to instead of calculating all the pairwise distances in a *brute-force* way, incorporate geometric considerations to determine when the distance between two hits should be calculated, such as spatial partitioning[2] or nearest-neighbor approximations[3]. Additionally, parallelizing the matrix calculation itself by distributing pairwise calculations

---

[2] A geometric technique that divides space into smaller regions (e.g., grids) to efficiently organize and process spatial data.

[3] Form of proximity search to identify the closest point to a given point without exhaustively computing all pairwise distance [80].

across multiple threads could reduce the execution time. Alternatively, adapting existing software frameworks such as FastJet, which already implemented some of these optimizations as discussed in Subsection 5.1, could provide an effective solution for reducing the computational workload and significantly improving the performance in more realistic scenarios.

6

Event Reconstruction

This chapter presents an initial reconstruction framework using the Cambridge-Aachen clustering algorithm for events in the NuDoubt^{++} detector, which employs a slow opaque scintillator. The focus is on identifying individual light depositions, determining their position and energy, and ultimately identifying the particle.

6.1 Position and Energy Reconstruction

Position and energy reconstruction are the key components in the development of any reconstruction framework. While the X and Y center of mass coordinates of the clusters are obtained from the fiber and associated SiPM positions, the cluster Z coordinate calculated through the clustering procedure is derived from the true z interaction point of the photon with the fiber. In practice, this information is unavailable. Additionally, our primary focus is to reconstruct the cluster position rather than the position of each photon. Therefore, an alternative approach is necessary to estimate the cluster Z position without relying on the true z information. A method utilizing the cluster timing information is proposed to address this issue, as detailed in the following section.

6.1.1 Estimating the Cluster Z-Position using Timing Information

For detectors with fibers arranged parallel to the z-axis, the time information on both sides of the detector offers a practical way to estimate the cluster Z-position:

$$Z = \frac{(t_{\text{top}} - t_{\text{bottom}})(1/v_{\text{fiber}})}{2} \tag{6.1}$$

where t_{top} and t_{bottom} denote the time assigned to the cluster at the top and bottom of the detector, and $\frac{1}{v_{\text{fiber}}}$ represents the average speed of light in the fiber due to different photon paths.

However, the definition of cluster time is not unique. Various methods may be employed, such as computing the mean photon arrival time within a cluster, selecting the earliest photon time, or adopting a quantile-based measure (e.g., using the first quantile or a random selection among the earliest few photons). Although these alternatives could account for effects such as electronic noise or signal degradation, further investigation is required to identify the optimal strategy. Figure 6.1 illustrates the cluster time distribution using the mean photon time and earliest photon time for the case of 1 MeV e^-, e^+, and γ. As described in 4, the slow scintillator broadens the photon arrival time distribution (see Figure 4.5b). This broadening can skew the mean arrival time, as late photons from the scintillation tail may dilute the direct correlation with the event's initial interaction. In contrast, the earliest photon time better approximates the actual start of the process by capturing the very first detectable light signal, thus being more sensitive to the initial interaction and less affected by delayed photon contributions. Consequently, the earliest photon time is adopted as the preferred variable for cluster time.

Comparing the cluster Z_{true} position calculated as the weighted average using the true fiber hit z information from the Geant4 simulation to the reconstructed cluster Z_{reco} position from the timing approach quantifies the resolution of this method. In the top-left plot of Figure 6.4, the distribu-

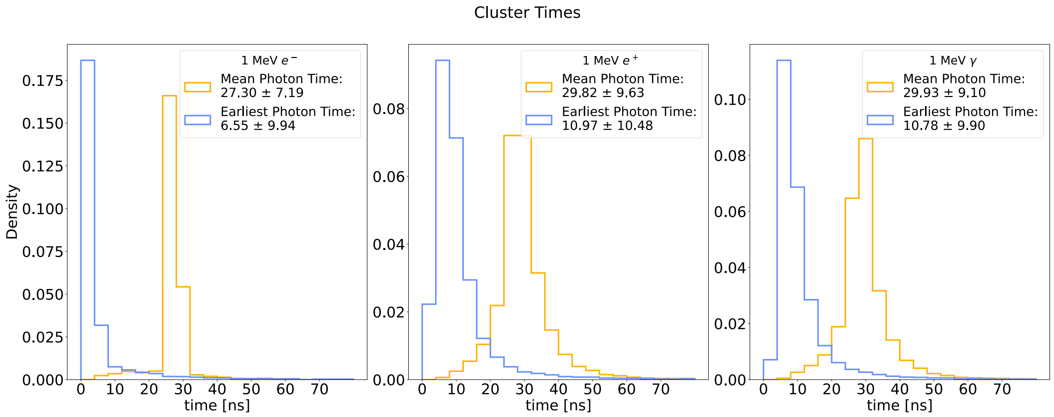
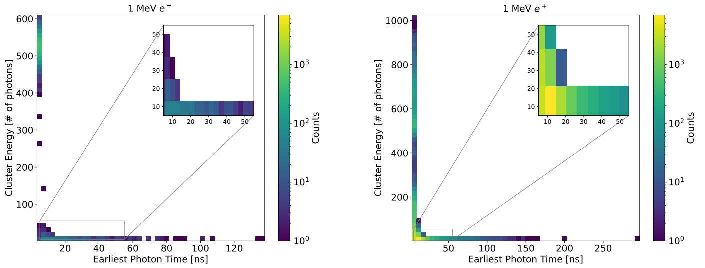


Fig. 6.1.: Cluster times distribution for 1 MeV e^-, e^+, and γ events. The figure compares the timing characteristics of the clusters. The blue histogram represents the earliest photon time arrival time for each cluster, providing insight into the first detected light signals. The yellow histogram corresponds to the mean photon arrival time for each cluster, reflecting the overall time distribution of detected photons. These distributions help characterize the temporal response of the scintillator to different particle interactions.

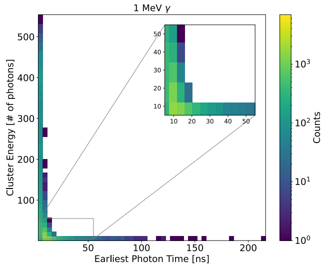
tion of $Z_{reco} - Z_{true}$ is shown for clusters associated with the three particle types used in this analysis. The results reveal large distribution tails, with values extending beyond the physical dimension of the detector, leading to poor and non-physical resolution. This can be attributed primarily to two factors:

- **Late-Arriving Clusters:** Clusters that occur later in the event often lead to large time differences.
- **Small-Energy Clusters:** There is an apparent correlation between clusters with a low number of hits (i.e., small energy depositions) and later recorded times.

To investigate these effects, the relationship between cluster time and its energy for 1 MeV e^-, e^+, and γ events was examined (Figure 6.2). The analysis indicates that although clusters span the full-time range, low-energy clusters are the ones often being recorded at later times. This observation motivates the need for additional quality selections. First, the reconstruc-



(a) 1 MeV e^- **(b)** 1 MeV e^+



(c) 1 MeV γ

Fig. 6.2.: 2D histograms showing the relationship between cluster time and energy for 1 MeV electrons, positrons, and gammas. The x-axis represents the cluster time, while the y-axis corresponds to the cluster energy. While small clusters are present throughout the entire time range, only small clusters tend to appear at later times. This suggests a time-dependent evolution in cluster formation, where late-emerging clusters are primarily low-energy. The distribution provides insight into the timing characteristics of light collection and potential scintillator response effects.

tion was done only on the clusters containing more than 5 SiPM hits (top left plot in Figure 6.4), this already had a visible impact on the overall Z resolution ((σ_Z)) and noticeably altered the mean cluster time distribution (Figure 6.3). An anomalous bump previously observed in the electron distribution within the first 20 nanoseconds, was eliminated which indicates that was associated with small clusters. Showing that the removal of these clusters improves the consistency of the timing information by reducing the influence of late depositions. Then, a spatial selection was imposed. This selection consists of only reconstructing clusters whose estimated Z position falls within the physical limits of the detector (bottom left plot in Figure 6.4). Finally, the reconstruction was done on both selections: only

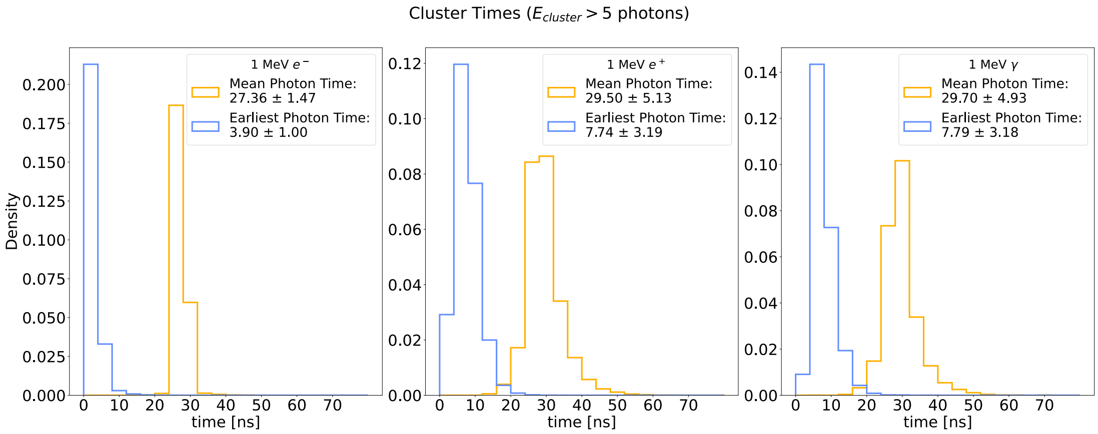


Fig. 6.3.: Cluster times distribution for 1 MeV e^-, e^+, and γ events after applying a photon selection of > 5 photons per cluster. This selection significantly impacts the mean cluster time distribution. It removes an anomalous bump previously observed in the electron distribution within the first 20 nanoseconds. This was primarily associated with small clusters, indicating that low photon clusters contributed disproportionately to early-time fluctuations.

clusters above the energy threshold and within the detector limits (bottom right plot in Figure 6.4). The results of these cuts showed an improvement in the resolution from $\sigma_Z \approx 28.3$ cm with no selection to $\sigma_Z \approx 5.4$ cm for electrons with the spatial and energy selection.

Additionally, to further assess the effect of imposing an energy threshold on the reconstruction of the Z position and establishing that a selection of 5 hits represents a good compromise between precision and information retention, the resolution for various threshold values (5, 10, 15, 20, and 100 SiPM hits) was studied. Figure 6.5 shows how σ_Z evolves as a function of the energy threshold. The results are summarized in Table 6.1. While a higher threshold (e.g., 10 hits) might clearly improve the resolution, it also risks discarding clusters that contain essential topological information about the event. The results presented here demonstrates that applying a 5-hit selection, in combination with the spatial detector selection, yields results comparable to those obtained with a 10-hit threshold. Moreover, resolution performance varies with particle type: electrons tend to produce a localized blob, which results in better resolution, whereas positrons and

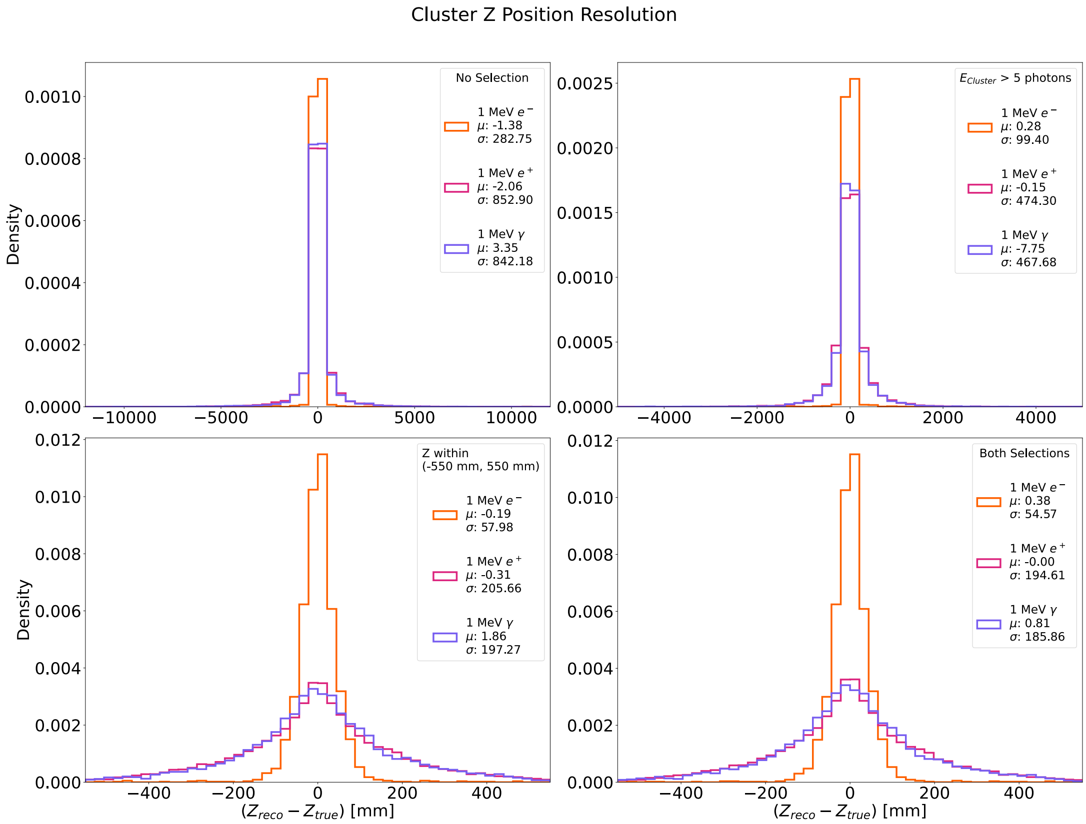


Fig. 6.4.: Cluster Z-position resolution for 1 MeV e^-, e^+, and γ events, where the x-axis represents the difference $Z_{\text{reco}} - Z_{\text{true}}$. With Z_{true} the cluster Z position calculated as the weighted average using the true fiber hit z information from the Geant4 simulation. The histograms are normalized to form a probability density, ensuring the total area sums to one. **Top left:** Resolution before applying any selection criteria, serving as a baseline comparison. **Top right:** Resolution after applying an energy selection of > 5 detected photons. This selection improves resolution by mitigating the impact of small clusters. **Bottom left:** Resolution after imposing the detector limit selection ($Z \in (-550, 550)$), ensuring only clusters having a Z-position reconstructed within the detector volume. **Bottom right:** Resolution after applying both selections, demonstrating the combined improvement in the cluster Z-position reconstruction.

gammas typically generate extended trails of depositions, which naturally yield a poorer resolution.

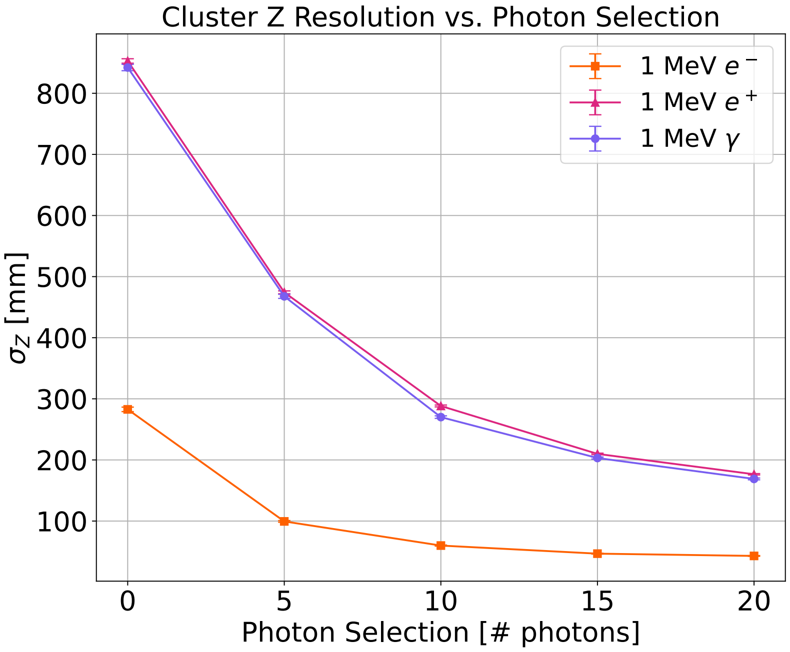


Fig. 6.5.: Evolution of the cluster Z-position resolution (σ_Z) as a function of the energy threshold. The x-axis represents the different threshold values (5, 10, 15, 20, and 100), while the y-axis shows the corresponding σ_Z values. The figure highlights the impact of selecting an appropriate hit threshold.

Selection	σ_Z for e^- [mm]	σ_Z for e^+ [mm]	σ_Z for γ [mm]
No Selection	282.75 ± 3.68	852.90 ± 3.63	842.18 ± 5.21
>5 hits	99.40 ± 1.38	474.30 ± 2.31	467.68 ± 3.28
>10 hits	59.87 ± 0.84	288.31 ± 1.54	270.09 ± 2.05
>15 hits	46.64 ± 0.65	209.90 ± 1.21	203.15 ± 1.65
>20 hits	42.92 ± 0.60	176.52 ± 1.09	168.91 ± 1.46
>100 hits	38.90 ± 0.55	75.99 ± 0.75	79.15 ± 1.00
Detector limit	57.98 ± 0.81	205.66 ± 1.05	197.27 ± 1.44
Both selections:			
Detector limit & 5 hits	54.57 ± 0.76	194.61 ± 1.02	185.86 ± 1.40

Tab. 6.1.: Values for σ_Z for the different studied selections for 1 MeV e^-, e^+, and γ events.

Another consideration resulting from this "quality" selection analysis is that clusters that fail to pass one the selections are categorized as "noise" but are retained for potential further analysis since they still represent valuable information. In other words, only a proportion of the total clusters produced in an event will be kept as "valid" clusters and used for subsequent analysis. The proportion of the clusters being kept after each selection is given in Table 4 6.2.

Selection	Valid number of clusters for e^- [%]	Valid number of clusters for e^+ [%]	Valid number of clusters for γ [%]
No Selection	100 ± 0.0	100 ± 0.0	100 ± 0.0
>5 photons	92.8 ± 1.7	78.2 ± 1.3	81.3 ± 1.6
>10 photons	92.6 ± 1.8	67.0 ± 1.4	71.9 ± 1.9
>15 photons	92.4 ± 1.8	58.7 ± 1.5	65.5 ± 2.1
>20 photons	92.3 ± 1.8	52.3 ± 1.6	59.9 ± 2.2
>100 photons	92.5 ± 1.9	22.5 ± 0.9	32.8 ± 2.0
Detector limit	92.8 ± 1.7	72.2 ± 1.4	76.5 ± 1.7
Both selection	92.7 ± 1.8	68.5 ± 1.4	73.2 ± 1.9

Tab. 6.2.: Percentages of clusters kept as valid after different selections for 1 MeV e^-, e^+, and γ events.

6.1.2 Reconstruction of the Interaction Position

After estimating the cluster Z position with the timing approach and establishing a selection of valid clusters, the subsequent step is to evaluate the accuracy of the three-dimensional (X, Y, Z) reconstruction of individual clusters. The assessment is performed by comparing the true interaction positions recorded in the Geant4 simulation (see Figures 6.7, 6.8, and 6.9) with the reconstructed cluster positions. Because the number of reconstructed clusters does not necessarily match the number of simulated interactions, a nearest-neighbor pairing strategy is adopted. For each cluster, the three-dimensional Euclidean distance to every true interaction point is computed as

$$d = \sqrt{(X_{\text{reco}} - X_{\text{true}})^2 + (Y_{\text{reco}} - Y_{\text{true}})^2 + (Z_{\text{reco}} - Z_{\text{true}})^2}. \tag{6.2}$$

The interaction that minimizes d is paired with the corresponding cluster, and the difference reco $-$ true is calculated. Importantly, true interactions are not removed from the list after pairing; this ensures that each cluster is independently compared against the complete set of interactions, thereby providing a fair and unbiased assessment of the reconstruction performance. Figure 6.6 displays the spatial resolution for individual clusters corresponding to electrons (e^-), positrons (e^+), and gammas (γ) events, respectively. The results indicate a significant disparity between the transverse resolutions ($\sigma_X(\sigma_Y) \approx 0.6$ cm for electrons) and the longitudinal resolutions ($\sigma_Z \approx 5$ cm for electrons). The much finer resolution in X and Y can be attributed to the well-defined fiber and SiPM positions in the detector, which provide precise spatial constraints. In contrast, the poorer Z resolution can be attributed to the reliance on timing information, which is intrinsically more uncertain. This uncertainty is further exacerbated by the slow scintillator, whose extended light emission profile diminishes the precision of time-based measurements and, consequently, the accuracy of the Z position reconstruction. Additionally, positrons and gammas exhibit

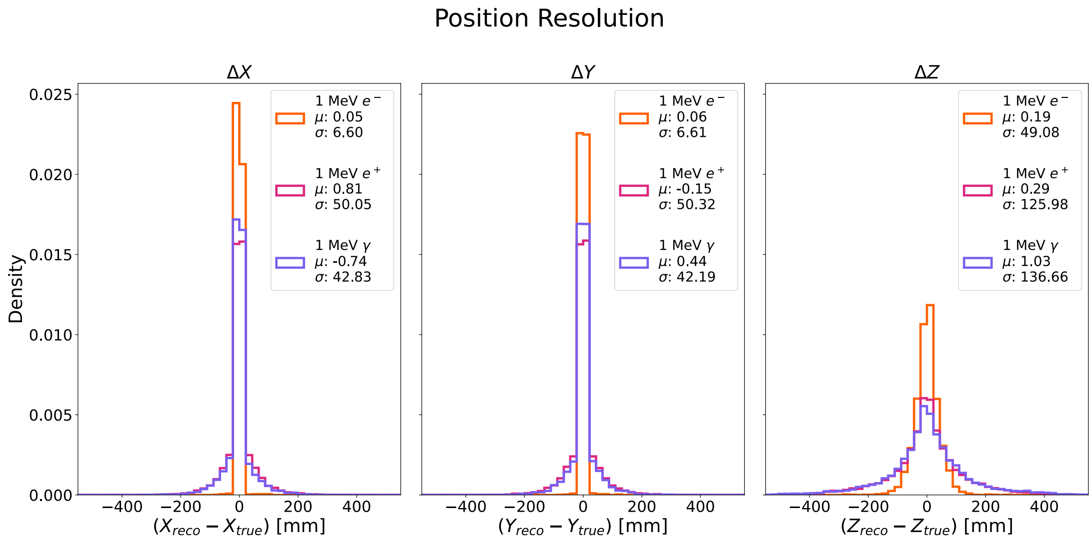


Fig. 6.6.: Position resolution for X, Y, and Z for the three kind of simulated particles. The histograms are normalized to form a probability density, ensuring the total area sums to one

larger discrepancies in all dimensions due to their more complex interaction patterns, leading to greater uncertainties in cluster localization.

While this provides an initial approximation to the position reconstruction, using the interactions recorded in the Geant4 simulation might not be the most reliable way of assessing vertex reconstruction (Figures 6.7, 6.8, and 6.9). It might introduce uncertainties stemming from the way physical processes are modeled and recorded within the simulation framework. A more robust approach to determining the true interaction vertex is necessary to improve the evaluation reconstruction performance.

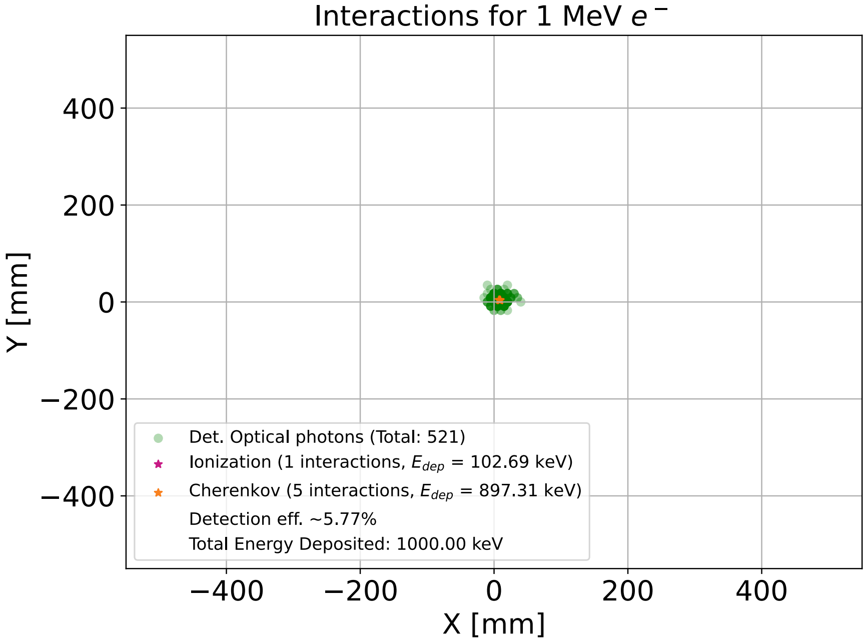


Fig. 6.7.: Example of the detected photons for a 1 MeV e^- event, indicating the position of the recorded interactions.

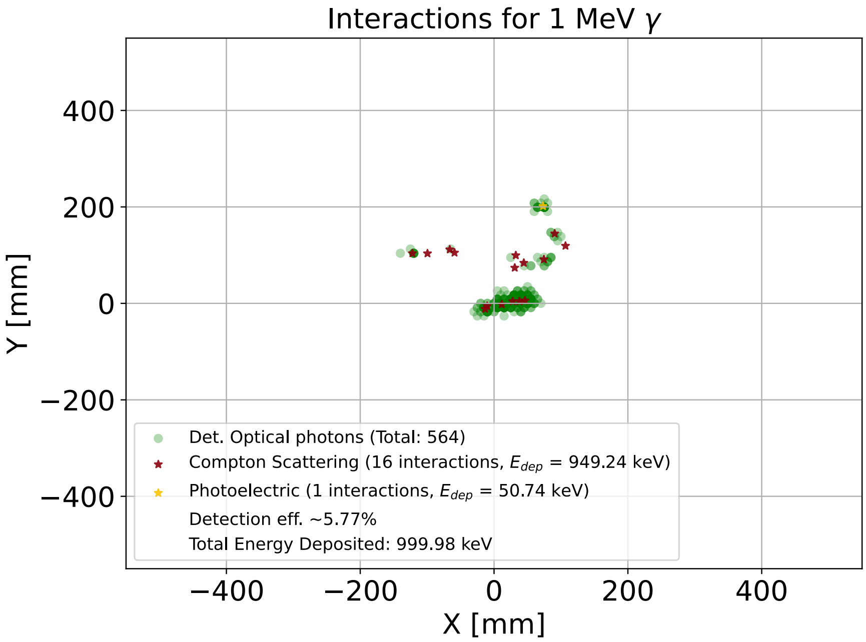


Fig. 6.8.: Example of the detected photons for a 1 MeV γ event, indicating the position of the recorded interactions.

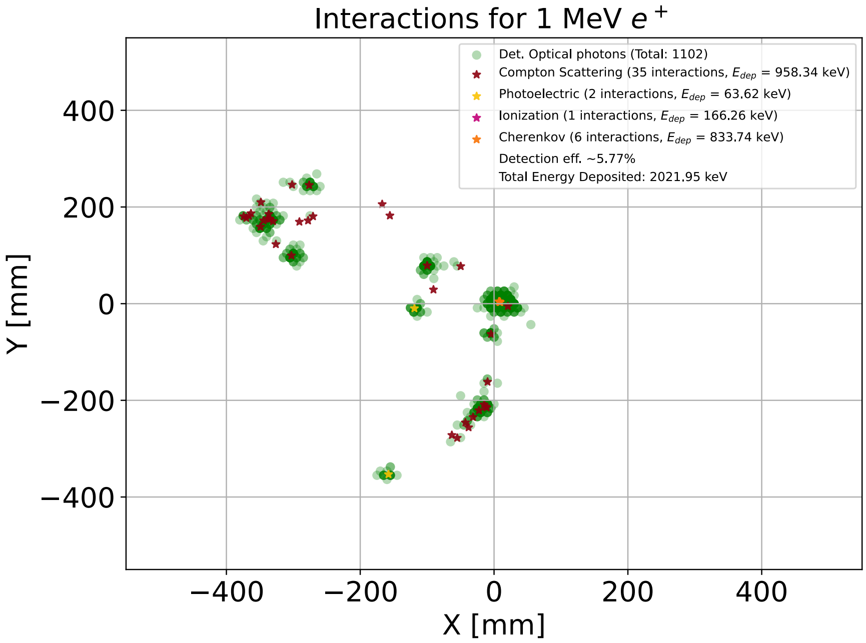


Fig. 6.9.: Example of the detected photons for a 1 MeV e^+ event, indicating the position of the recorded interactions.

6.1.3 Energy Reconstruction

Another important quantity is the cluster energy, defined as the total number of optical photons in a cluster. Since the scintillation process produces photons directly proportional to the energy deposited in the detector, this photon count serves as a proxy for that energy. The cluster energy distribution presented in Figure 6.10 provides insight into the energy deposition characteristics of different particle types. For instance, the distribution for 1 MeV electrons exhibits a relatively narrow peak of around 500 photons. This reflects the localized nature of electron energy depositions, which primarily occur via ionization and excitation within the scintillator, producing a consistent number of optical photons per interaction. The energy spectrum of 1 MeV gammas follows a continuous distribution rather than a well-defined peak. This is characteristic of Compton scattering, where gammas transfer only a fraction of their energy to electrons in each inter-

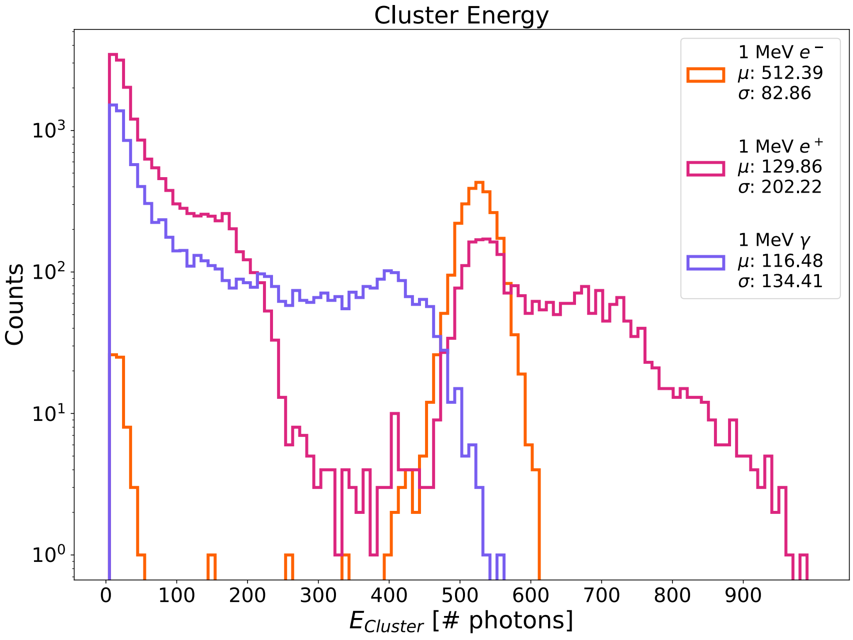


Fig. 6.10.: Cluster energy distributions for 1 MeV e^-, e^+, and γ. The patterns of light deposition can be observed.

action. The resulting energy deposits span over a broad range, producing clusters with widely varying photon counts. In contrast, the energy distribution for 1 MeV positrons shows a broader peak of around 500 photons and a significantly low-energy continuum. This behavior can be attributed to the fact that positrons undergo annihilation upon interaction with an electron in the medium, producing two 511 keV gammas. The broad peak corresponds to the primary energy deposition from the positron before annihilation, while the low-energy tail can be associated with clusters originating from the secondary annihilation gammas. These annihilation photons may Compton scatter before being absorbed, leading to lower energy clusters within the detector. The distinction between particle types highlights the utility of cluster energy in particle identification, which will be discussed in more detail in the next section.

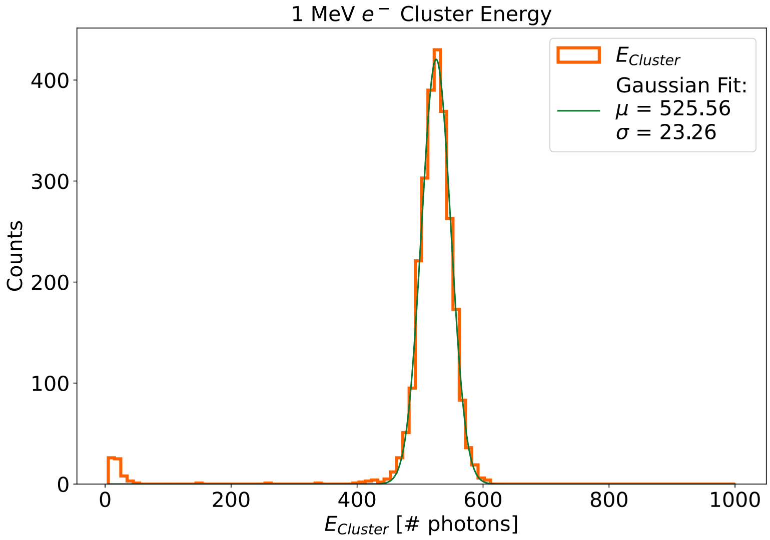


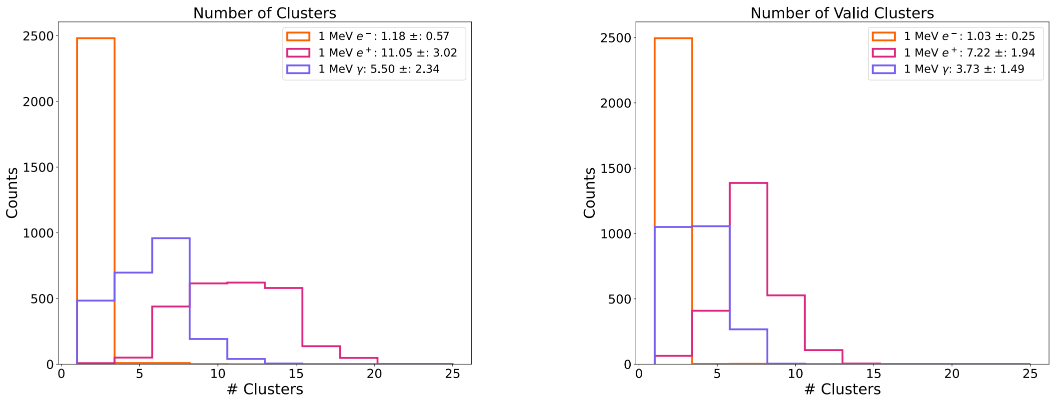
Fig. 6.11.: Preliminary fit of the electron cluster energy distribution to a Gaussian model.

A preliminary approach to determine a calibration constant, the electron cluster energy distribution, was fitted to a Gaussian model. Since the electron energy deposition is dominated by statistical fluctuations, its photon yield is well represented by a normal distribution. The parameters obtained from this fit provide both a calibration point (μ) and an energy resolution (σ). Dividing any measured cluster's energy by μ will proportionate an estimate of the deposited energy, establishing a first direct conversion between the optical signal and the energy scale. The fit was performed using the SciPy library [81], and the results are presented in Figure 6.11. Although the fit captures the overall trend, further refinement and optimization are required for a better accuracy. In particular, a small peak in the distribution was excluded from the fitting process because it corresponds to low-energy clusters, that despite passing the selection criteria, likely represent noise. A similar approach could be used to fit the gamma depositions to a Compton continuum model. Subsequent analyses, such as energy resolution studies, require a precise calibration constant. In addition, non-linearities in the detector response and other systematic effects must be accounted for to ensure a robust energy reconstruction across the detector's full range, likely necessitating a more sophisticated model than the one used here.

6.2 Particle Identification

Particle identification is a crucial component of the overall event reconstruction framework. Accurately classifying the type of particle interacting with the detector is essential for effective signal detection and background discrimination, especially in neutrino events experiments.

The opaque scintillator's ability to preserve the topological information of an event ensures that the distinct patterns created by different particles are conserved. Moreover, the use of a slow scintillator enables particle identification through analysis of the ratio between Cerenkov and scintillation light. Although this method presents a promising PID approach [58], it is not further explored in this thesis. A set of event variables is defined to characterize each event. One straightforward variable is the total number of clusters per event (Figure 6.12a), which directly corresponds to the number of light depositions in the detector. For instance, an electron typically produces a single energy deposition (and hence a single



(a) Number of total Clusters for 1 MeV e^-, e^+, and γ events

(b) Number of clusters categorized as "valid" for 1 MeV e^-, e^+, and γ events.

Fig. 6.12.: The number of clusters represents a direct variable for the number of light depositions in the detector, and thanks to the topological information it can prove to offer valuable information for particle identification.

cluster), whereas gammas and positrons generate multiple depositions, resulting in multiple clusters. Therefore, the cluster count provides an initial indication of the particle type. Even after applying the selection criteria described in Section 6.1.1, which reduce the overall number of clusters (Figure 6.12b), this variable continues to offer valuable information for event characterization.

Another event variable is the total cluster energy, as discussed in the previous section, the cluster energy provides insight into the particle type, as different particles are expected to deposit varying amounts of light. Accordingly, the total event energy is defined as the sum of all cluster energies (Figure 6.13). The final event variable considered in this work is the event extension. This extension quantifies the spatial spread of the clusters and requires careful treatment. A basic approach is developed using only the cluster X, Y, and Z positions. For each event, the maximum and minimum values of these coordinates are determined, yielding the range (ΔX_{event}, ΔY_{event}, and ΔZ_{event}), which defines the spatial extent along each axis. These values are used to construct a bounding box, which is the smallest rectangular box that contains the entire set of clusters in three-dimensional space.

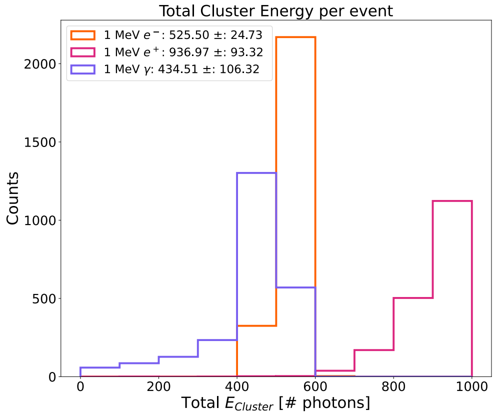


Fig. 6.13.: Distribution of the total event cluster energy for 1 MeV e^-, e^+, and γ events. The total event energy is defined as the sum of the energies of all individual clusters within the event. This highlights the differences in the amount of light deposited between the three particle types

Figure 6.14 shows the box built for a 1 MeV e^+ event in a first representation of this method. Then, the Euclidean norm (Equation 6.3) of these ranges is calculated to provide a single measure of the maximum diagonal distance across the event's bounding box.

$$d = \sqrt{(\Delta X)^2 + (\Delta Y)^2 + (\Delta Z)^2} \tag{6.3}$$

The selection of this approach is based in the fact that avoids dependencies on time or energy sorting, providing a pure geometry measure of spatial extent, where the sign of the coordinates does not affect the spread. This may be further refined by more sophisticated methods and including the RMS of the clusters or calculating the weighted centroid of the event. Figure 6.15 shows the diagonal spread for the three different types of events. Electrons typically produce a single cluster, leading to a significantly lower spread compared to positrons or gammas, which generate multiple clusters and exhibit a larger spatial extent. Once the event variables were determined, two classification approaches were implemented: a basic binary approach using only the number of clusters as a classifier and a more comprehensive

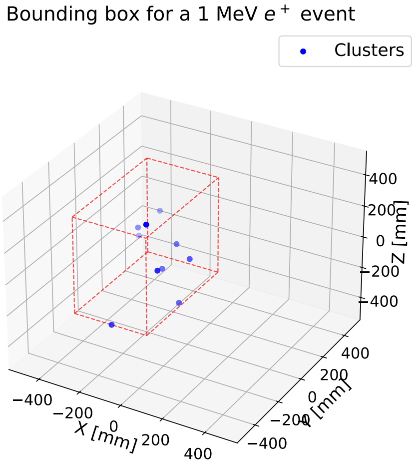


Fig. 6.14.: Bounding box for a 1 MeV e^+ event. The blue dots represent the COM positions of the clusters, the box is represented in dashed red lines and it is constructed with the dimensions $(X_{\text{min}}, X_{\text{max}})$, $(Y_{\text{min}}, Y_{\text{max}})$, and $(Z_{\text{min}}, Z_{\text{max}})$.

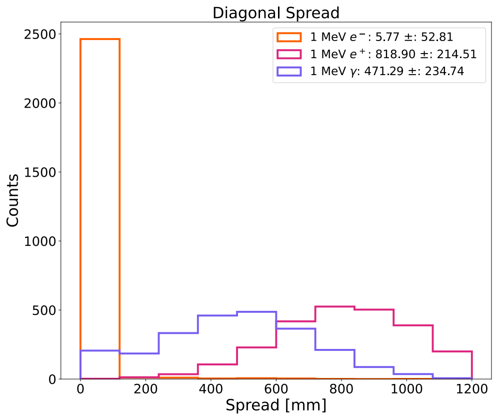


Fig. 6.15.: Diagonal spread of the clusters for 1 MeV e^-, e^+, and γ events. The spread is defined as the maximum diagonal distance across the bounding box that contains all clusters of the event.

method, employing a random forest classifier [82], that included the extra event variables.

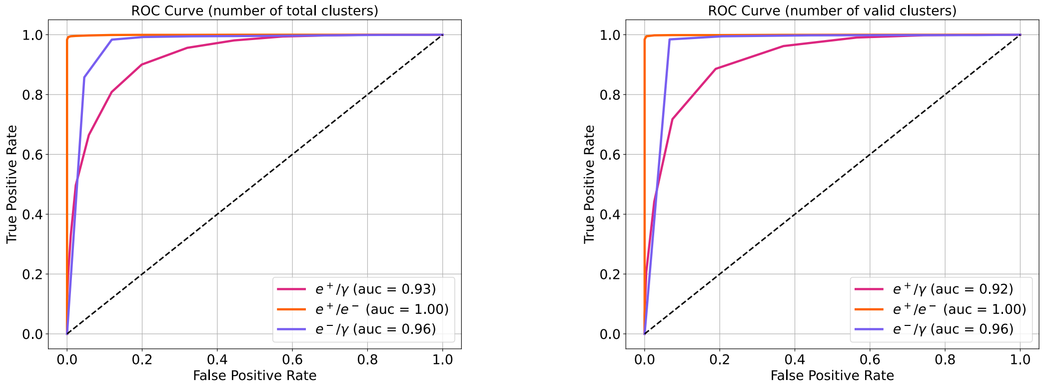
6.2.1 Binary classification

This analysis examines the particle discrimination capabilities of the clustering algorithm by using a binary classification procedure based solely on the number of clusters detected in each event. Three comparison cases are considered: e^-/e^+, e^+/γ, and γ/e^-. A classification function is defined to distinguish particle types based on the number of reconstructed clusters. A threshold is applied: if the cluster meets or exceeds this threshold, the event is assigned a specific label (typically 1). To determine the optimal threshold, the range of observed cluster counts across all events is identified, and each possible threshold value is tested systematically. By iterating over these values, the impact of different threshold choices on classification performance can be evaluated. For each threshold, the True Positive Rate (TPR) and False Positive Rate (FPR) are computed using the standard formulas:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6.4}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{6.5}$$

where TP, FP, TN, and FN represent the numbers of True Positives, False Positives, True Negatives, and False Negatives, respectively. The Receiver Operating Characteristic (ROC) curve is generated from these values, and the Area Under the Curve (AUC) is calculated. The AUC serves as a quantitative measure of the effectiveness of the classifier. To check if the reduction on the number of clusters after the quality selections had an impact on the classification, the TPR and FPR values were calculated using the total number of clusters and the number of valid clusters, Figure 6.16 presents the ROC curves obtained from these values. The excellent performance in both cases demonstrate that reducing the number of clusters does not affect overall classification performance.



(a) ROC curve for the different binary classification scenarios using the total number of clusters as classifier.

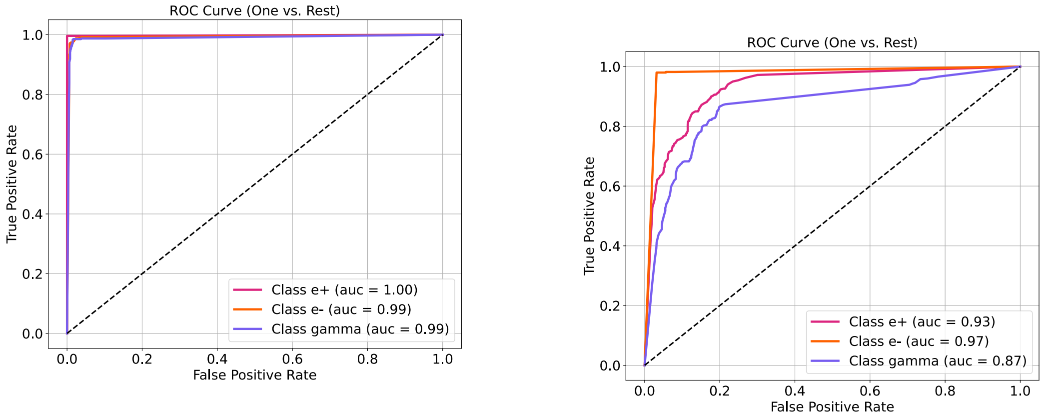
(b) ROC curve for the different binary classification scenarios using the number of valid clusters as classifier.

Fig. 6.16.: Combined ROC curves for binary classification cases.

6.2.2 Random Forest Classifier

A multi-class classification approach based on a Random Forest algorithm provided by scikit-learn [83] was implemented to integrate the rest of the event variables. The implementation of the RF classifier offers a streamlined and flexible framework for building such models.

A Random Forest (RF) classifier is an ensemble method built upon Decision Trees that aggregates the outputs of multiple trees constructed on different sub-samples of the data. Decision Trees are non-parametric, supervised learning models that partition the feature space into regions defined by a sequence of binary decisions. They predict a target variable by learning simple decision rules from the input features. Then, the ensemble approach improves predictive accuracy and robustness by reducing over-fitting and capturing complex patterns. In this study, a RF classifier was implemented to distinguish among particle types using as features the total event energy, the number of valid clusters, and the event extension. The RF model was built with 100 decision tress (estimators) using default bootstrap sampling. The evaluation focused on how well the classifier could distinguish each particle type by analyzing the ROC curve with a one-versus-rest strategy



(a) ROC curve obtained from the Random Forest classification using the number of valid clusters, the total event energy, and the diagonal event spread as features.

(b) ROC curve obtained from the Random Forest classification using only the number of valid clusters and the diagonal event spread as features.

Fig. 6.17.: Combined ROC curves for the two Random Forest classification methods. The AUC represents how well that class is classified versus the two other classes.

(Figure 6.17a). An additional case was considered by removing the total cluster energy as a feature to check the impact of this feature; the results are shown in Figure 6.17b. The overall performance of the classifier is excellent, as indicated by an accuracy of approximately 98.04% using all the features and 81.84% for the model excluding the total cluster energy. Below (Table 6.3) is the detailed classification report, which provides insight into the performance per class:

Class	Model (All Features)				Model (without total cluster energy)			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
e^-	1.00	1.00	1.00	500	0.78	0.79	0.78	500
e^+	0.97	0.97	0.97	500	0.92	0.98	0.95	500
γ	0.97	0.97	0.97	476	0.74	0.68	0.71	476
Overall Accuracy	0.98				0.82			

Tab. 6.3.: Comparison of classification performance with all features and without total cluster energy as a feature.

Where *precision* is defined as the measure of accuracy of the positive predictions. For example, a precision of 1 for positrons means that every event classified as positron was indeed a positron event, indicating no false positives for that class. The *recall* values represent the ability of the classifier to find all relevant instances of a given class. A recall of 1 indicates that all actual events of that class were correctly identified by the model. The *F1-score* is the harmonic mean of precision and recall, providing a single metric that balances both. An F1-score close to 1 signifies that the classifier maintains both high precision and high recall. The *support* refers to the number of actual occurrences of that class in the dataset. Finally, the *overall accuracy* is the proportion of total correctly classified events over all events.

The near-perfect performance observed with both classification methods is partly due to the specific event types considered. A decline in performance is anticipated when applying this procedure to more realistic scenarios, as detailed in Chapter 7. Overall, the binary approach based solely on the event topology demonstrates the potential of opaque scintillators for particle identification. The current analysis also provides evidence that the RF classifier can be implemented for event classification with enough room for improvement through detailed feature importance analysis, applying cross-validation, and testing on larger, more diverse datasets.

7

Conclusion and Outlook

This thesis established an initial event reconstruction framework for opaque liquid scintillator detectors, with a particular focus on the NuDoubt^{++} detector. This detector employs a hybrid-slow scintillator coupled with OWL fibers to collect and direct the produced light to the SiPMs. A Geant4-based simulation, developed under the assumption of constant optical properties, generated optical photons from particle interactions and provided ground truth for reconstruction. Although the simulation was based on idealized events (photon bombs, 1 MeV electrons, positrons, and gammas) to offer a controlled benchmark, future work should incorporate detailed physics models and a broader range of event energies and positions to more accurately represent detector performance.

A simplified analytical model for fiber propagation and SiPM detection was implemented, yielding a detection efficiency of 5.7% and assigning arrival times to the detected photons. While this model is valuable for preliminary studies, further refinement is required to accurately simulate photon trajectories, account for detailed electronic responses, and optimize triggering strategies. The reconstruction framework employs the Cambridge-Aachen clustering algorithm, which uses a two-dimensional Euclidean distance in the (x, y) plane to group light depositions. An optimal distance threshold of 45 mm was determined to preserve the spatial structure without merging distinct signals or fragmenting individual depositions. A secondary clustering step then matches clusters from different detector sides; although this process sometimes results in unmatched clusters, these still contain valuable event information, requiring additional analysis to determine whether they represent noise or are part of a significant signal. Their treatment might involve further tailoring or refining the clustering algorithm. Furthermore, while the current matching approach is practical for parallel fiber arrangements, more complex detector geometries (such as stereo fiber layouts) may benefit from advanced techniques like image pattern recognition or graph-based methods. In terms of computational performance, although an $O(n^3)$ complexity was initially anticipated, an empirical complexity of $O(n^{2/3})$ was observed. This result should be validated through further profiling. The clustering results are encouraging, with clear visual evidence of grouped light depositions. This approach is notably flexible and shows promise for tuning to specific geometries like CLOUD; however, further work is needed to extend its application to more complex events and address potential detector edge effects.

After clustering, a method to estimate the cluster's Z position based on timing information was developed. The analysis showed that given the slow scintillator's temporal profile, using the time of the earliest photon yields the most reliable cluster time variable. This approach revealed a clear relationship between cluster time, energy, and Z position; notably, smaller clusters tend to appear later, which can produce unrealistically high reconstructed cluster Z values. Two quality cuts were implemented: a energy threshold to discard low-energy clusters and a spatial selection to restrict reconstruction to clusters whose reconstructed Z is in the detector's physical limits. These adjustments improved the Z position resolution from $\sigma_Z = 282.75 \pm 3.68$ mm uncertainty to $\sigma_Z = 54.57 \pm 0.76$ mm for electrons. Future work should explore additional selection strategies, incorporate more complex timing models and reconstruction techniques to improve the resolution of the cluster Z position.

For full position reconstruction, the interaction data from the Geant4 simulation was used to assign the nearest true interaction vertex to each cluster, providing an initial position resolution estimate. The transverse resolution is finer due to the precise fiber positions, while the longitudinal resolution suffers from timing uncertainties. Additionally, based on cluster energy distributions, the preliminary energy reconstruction shows promise for establishing a calibration constant. Future work should refine vertex determination methods to minimize uncertainties and enhance overall reconstruction accuracy.

Finally, the particle identification capabilities were assessed using event-level variables such as the number of clusters, total cluster energy, and spatial spread. A simple binary classifier based on the number of clusters yielded near-perfect performance, highlighting the topological advantages of opaque scintillators, while a random forest classifier using all three variables produced similarly high accuracy under controlled conditions. However, in more realistic and challenging scenarios, for example differentiating between a 1.78 MeV positron and two 0.38 MeV positrons depositions, the performance is expected to decline. Future work will extend the classification analysis to more complex event topologies and integrate advanced classification techniques into broader signal/background discrimination frameworks. Additionally, gathering more data and exploring alternative event variables, including reintegrating previously removed noise clusters, will be crucial for evaluating classification performance in more realistic conditions.

In summary, the results demonstrate the potential of the proposed reconstruction framework for event reconstruction and effective particle classification in opaque scintillator detectors. The work lays a solid foundation for further refinements, including improved simulation and more realistic detector response, to fully exploit opaque scintillator technology in future neutrino and rare event searches.

Bibliography

[1] J. Chadwick. “The intensity distribution in the magnetic spectrum of β particles from radium (B + C)”. In: *Verh. Phys. Gesell.* 16 (1914), pp. 383–391 (cit. on p. 3).

[2] W. Pauli. “Dear radioactive ladies and gentlemen”. In: *Phys. Today* 31N9 (1978), p. 27 (cit. on p. 3).

[3] E. Fermi. “An attempt of a theory of beta radiation. 1.” In: *Z. Phys.* 88 (1934), pp. 161–177 (cit. on p. 4).

[4] H. BETHE and R. PEIERLS. “The “Neutrino””. In: *Nature* 133.3362 (1934), pp. 532–532 (cit. on p. 4).

[5] Luis W Alvarez. *A Proposed Experimental Test of the Neutrino Theory*. Tech. rep. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), Apr. 1949 (cit. on p. 4).

[6] F. Reines and C. L. Cowan. “A Proposed Experiment to Detect the Free Neutrino”. In: *Phys. Rev.* 90 (3 1953), pp. 492–493 (cit. on p. 4).

[7] F. Reines and C. L. Cowan. “Detection of the free neutrino”. In: *Phys. Rev.* 92 (1953), pp. 830–831 (cit. on p. 4).

[8] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. “Detection of the free neutrino: A Confirmation”. In: *Science* 124 (1956), pp. 103–104 (cit. on p. 4).

[9] D. Decamp et al. “A Precise Determination of the Number of Families With Light Neutrinos and of the Z Boson Partial Widths”. In: *Phys. Lett. B* 235 (1990), pp. 399–411 (cit. on p. 4).

[10] P. Aarnio et al. “Measurement of the Mass and Width of the Z^0 Particle from Multi - Hadronic Final States Produced in e^+e^- Annihilations”. In: *Phys. Lett. B* 231 (1989), pp. 539–547 (cit. on p. 4).

[11] B. Adeva et al. “A Determination of the Properties of the Neutral Intermediate Vector Boson Z^0”. In: *Phys. Lett. B* 231 (1989), p. 509 (cit. on p. 4).

[12]M. Z. Akrawy et al. “Measurement of the Z^0 Mass and Width with the OPAL Detector at LEP”. In: *Phys. Lett. B* 231 (1989), pp. 530–538 (cit. on p. 4).

[13]Samuel J. Ling, Jeff Sanny, and William Moebs. *Introduction to Particle Physics*. Accessed: 2025-02-25. 2021 (cit. on p. 5).

[14]Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266 (cit. on p. 5).

[15]Y. Fukuda et al. “Evidence for oscillation of atmospheric neutrinos”. In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. arXiv: `hep-ex/9807003` (cit. on pp. 5, 6).

[16]B. Pontecorvo. “Mesonium and anti-mesonium”. In: *Sov. Phys. JETP* 6 (1957), p. 429 (cit. on p. 6).

[17]Ziro Maki, Masami Nakagawa, and Shoichi Sakata. “Remarks on the unified model of elementary particles”. In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880 (cit. on p. 6).

[18]Hiroshi Nunokawa, Stephen J. Parke, and Jose W. F. Valle. “CP Violation and Neutrino Oscillations”. In: *Prog. Part. Nucl. Phys.* 60 (2008), pp. 338–402. arXiv: `0710.0554 [hep-ph]` (cit. on p. 6).

[19]Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. “Search for Neutrinos from the Sun”. In: *Phys. Rev. Lett.* 20 (21 1968), pp. 1205–1209 (cit. on p. 6).

[20]K. S. Hirata et al. “Observation of B-8 Solar Neutrinos in the Kamiokande-II Detector”. In: *Phys. Rev. Lett.* 63 (1989), p. 16 (cit. on p. 6).

[21]J. N. Abdurashitov et al. “Measurement of the solar neutrino capture rate by SAGE and implications for neutrino oscillations in vacuum”. In: *Phys. Rev. Lett.* 83 (1999), pp. 4686–4689. arXiv: `astro-ph/9907131` (cit. on p. 6).

[22]S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001 (cit. on p. 7).

[23]STEVEN R. ELLIOTT and Jonathan Engel. “Double-beta decay”. In: *Journal of Physics G: Nuclear and Particle Physics* 30.9 (2004), R183–R215 (cit. on p. 7).

[24]J. D. Vergados, H. Ejiri, and F. Šimkovic. “Neutrinoless double beta decay and neutrino mass”. In: *Int. J. Mod. Phys. E* 25.11 (2016), p. 1630007. arXiv: `1612.02924 [hep-ph]` (cit. on p. 7).

[25]Katermickie (https://physics.stackexchange.com/users/195443/katermickie). *The neutrinoless double beta decay can only occur if....* Physics Stack Exchange. URL:https://physics.stackexchange.com/q/646916 (version: 2021-06-21). eprint: `https://physics.stackexchange.com/q/646916` (cit. on p. 8).

[26]Alexander Barabash. "Precise Half-Life Values for Two-Neutrino Double-β Decay: 2020 Review". In: *Universe* 6.10 (2020), p. 159. arXiv: `2009.14451 [nucl-ex]` (cit. on pp. 8, 9).

[27]M. Goeppert-Mayer. "Double Beta-Disintegration". In: *Phys. Rev.* 48 (6 1935), pp. 512–516 (cit. on p. 8).

[28]N. Takaoka and K. Ogata. "The half-life of Te-130 double beta-decay". In: *Z. Naturforsch. A* 21 (1966), pp. 84–90 (cit. on p. 9).

[29]T. Kirsten, O. A. Schaeffer, E. Norton, and R. W. Stoenner. "Experimental evidence for the double-beta decay of Te-130". In: *Phys. Rev. Lett.* 20 (1968), pp. 1300–1303 (cit. on p. 9).

[30]S. R. Elliott, A. A. Hahn, and M. K. Moe. "Direct Evidence for Two Neutrino Double Beta Decay in ^{82}Se". In: *Phys. Rev. Lett.* 59 (1987), pp. 2020–2023 (cit. on p. 9).

[31]Stefano Dell'Oro, Simone Marcocci, Matteo Viel, and Francesco Vissani. "Neutrinoless Double Beta Decay: 2015 Review". In: *Advances in High Energy Physics* 2016.1 (2016), p. 2162659 (cit. on p. 9).

[32]J. B. Albert et al. "Search for Majorana neutrinos with the first two years of EXO-200 data". In: *Nature* 510 (2014), pp. 229–234. arXiv: `1402.6956 [nucl-ex]` (cit. on p. 9).

[33]Michelle J. Dolinski, Alan W. P. Poon, and Werner Rodejohann. "Neutrinoless Double-Beta Decay: Status and Prospects". In: *Ann. Rev. Nucl. Part. Sci.* 69 (2019), pp. 219–251. arXiv: `1902.04097 [nucl-ex]` (cit. on p. 10).

[34]Izumi Ogawa. "Study of Ca-48 double beta decay with CANDLES". In: *J. Phys. Conf. Ser.* 136 (2008). Ed. by Jenni Adams, Francis Halzen, and Stephen Parke, p. 042053 (cit. on p. 10).

[35]Stefan Schonert et al. "The GERmanium Detector Array (GERDA) for the search of neutrinoless beta beta decays of Ge-76 at LNGS". In: *Nucl. Phys. B Proc. Suppl.* 145 (2005). Ed. by P. Bernardini, G. Fogli, and E. Lisi, pp. 242–245 (cit. on p. 10).

[36]Richard Gaitskell et al. “The Majorana zero neutrino double beta decay experiment”. In: (Nov. 2003). arXiv: `nucl-ex/0311013` (cit. on p. 10).

[37]Irina Nasteva. “The SuperNEMO double beta decay experiment”. In: *10th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications*. Oct. 2007. arXiv: `0710.4279 [physics.ins-det]` (cit. on p. 10).

[38]V. Alenkov et al. “Technical Design Report for the AMoRE $0\nu\beta\beta$ Decay Search Experiment”. In: (Dec. 2015). arXiv: `1512.05957 [physics.ins-det]` (cit. on p. 10).

[39]D. Muenstermann. “Current status and recent results of the COBRA experiment”. In: *PoS* HEP2005 (2006). Ed. by Gaspar Barreira, p. 170 (cit. on p. 10).

[40]I. C. Bandac. “Search for neutrinoless double beta decay with the CUORE detector”. In: *J. Phys. Conf. Ser.* 110 (2008). Ed. by Roger Barlow, p. 082001 (cit. on p. 10).

[41]C. Hall. “EXO: A Next generation double beta decay experiment”. In: *4th Workshop on Neutrino Oscillations and their Origin (NOON2003)*. Feb. 2003, pp. 332–337 (cit. on p. 10).

[42]Justo Martin-Albo. “NEXT, a HPXe TPC for neutrinoless double beta decay searches”. In: *PoS* NUFACT08 (2008). Ed. by Amos Breskin, Marc Henneaux, Viatcheslav Mukhanov, and Hector Rubinstein, p. 130 (cit. on p. 10).

[43]Mark C. Chen. “The SNO+ Experiment”. In: *34th International Conference on High Energy Physics*. Oct. 2008. arXiv: `0810.3694 [hep-ex]` (cit. on p. 10).

[44]Bernhard Schwingenheuer. “Status and prospects of searches for neutrinoless double beta decay”. In: *Annalen Phys.* 525 (2013), pp. 269–280. arXiv: `1210.7432 [hep-ex]` (cit. on p. 10).

[45]J. D. Vergados, H. Ejiri, and F. Simkovic. “Theory of Neutrinoless Double Beta Decay”. In: *Rept. Prog. Phys.* 75 (2012), p. 106301. arXiv: `1205.0649 [hep-ph]` (cit. on p. 10).

[46]M. I. Krivoruchenko, Fedor Simkovic, Dieter Frekers, and Amand Faessler. “Resonance enhancement of neutrinoless double electron capture”. In: *Nucl. Phys. A* 859 (2011), pp. 140–171. arXiv: `1012.1204 [hep-ph]` (cit. on p. 11).

[47]Glenn F. Knoll. *Radiation Detection and Measurement*. 4th. John Wiley & Sons, 2010 (cit. on p. 12).

[48]T. Marrodan Undagoitia, F. von Feilitzsch, L. Oberauer, et al. “Fluorescence decay-time constants in organic liquid scintillators”. In: *Rev. Sci. Instrum.* 80 (2009), p. 043301. arXiv: `0904.4602 [physics.ins-det]` (cit. on p. 13).

[49]J.B. BIRKS. “CHAPTER 8 - ORGANIC LIQUID SCINTILLATORS”. In: *The Theory and Practice of Scintillation Counting*. Ed. by J.B. BIRKS. International Series of Monographs in Electronics and Instrumentation. Pergamon, 1964, pp. 269–320 (cit. on pp. 13, 14).

[50]Alex M Mooney. *FRET Jablonski Diagram*. CC BY-SA 3.0, via Wikimedia Commons. 2012 (cit. on p. 13).

[51]Th. Förster. “Zwischenmolekulare Energiewanderung und Fluoreszenz”. In: *Annalen Phys.* 437.1-2 (1948), pp. 55–75 (cit. on p. 14).

[52]Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. 3rd. Springer, 2006 (cit. on p. 14).

[53]Christian Buck, Benjamin Gramlich, and Stefan Schoppmann. “Novel Opaque Scintillator for Neutrino Detection”. In: *JINST* 14.11 (2019), P11007. arXiv: `1908.03334 [physics.ins-det]` (cit. on pp. 14, 16, 17).

[54]Stefan Schoppmann. “Review of Novel Approaches to Organic Liquid Scintillators in Neutrino Physics”. In: *Symmetry* 15.1 (2023) (cit. on pp. 14, 15).

[55]P. A. Cherenkov. “Visible luminescence of pure liquids under the influence of γ-radiation”. In: *Dokl. Akad. Nauk SSSR* 2.8 (1934), pp. 451–454 (cit. on p. 15).

[56]Árpád Horváth. *Cherenkov Radiation Diagram*. CC BY-SA 2.5, via Wikimedia Commons. 2005 (cit. on p. 15).

[57]M. Yeh, S. Hans, W. Beriguete, et al. “A new water-based liquid scintillator and potential applications”. In: *Nucl. Instrum. Meth. A* 660 (2011), pp. 51–56 (cit. on p. 15).

[58]Kyra Mossel. “Reconstruction and Background Discrimination in a Hybrid and Opaque Scintillator Detector”. Master’s thesis. Johannes Gutenberg University Mainz, 2024 (cit. on pp. 16, 21, 58).

[59]A. Cabrera et al. “Neutrino Physics with an Opaque Detector”. In: *Commun. Phys.* 4 (2021), p. 273. arXiv: `1908.02859 [physics.ins-det]` (cit. on pp. 16, 17).

[60]Manuel Böhles et al. “Combining Hybrid and Opaque Scintillator Techniques in the Search for Double Beta Plus Decays”. In: (July 2024). arXiv: `2407.05999 [hep-ex]` (cit. on pp. 17, 21).

[61]H. de Kerret et al. “The CHOOZ experiment: Proposal to search for neutrino vacuum oscillations to $\Delta m^2 = 10^{-3} eV^2$ using a 1 km baseline reactor neutrino experiment”. In: (1993) (cit. on p. 19).

[62]D. Motta. “The Double Chooz experiment”. In: *Acta Phys. Polon. B* 37 (2006). Ed. by A. Zalewska, K. Golec-Biernat, and M. Skrzypek, pp. 2027–2037 (cit. on p. 19).

[63]Diana Navas Nicolas. “CLOUD: the First Reactor Antineutrino Experiment using the Novel LiquidO Detection Technology”. In: *PoS* ICHEP2024 (2024), p. 157 (cit. on p. 20).

[64]Christian Buck and Minfang Yeh. “Metal-loaded organic scintillators for neutrino physics”. In: *J. Phys. G* 43.9 (2016), p. 093001. arXiv: `1608.04897 [physics.ins-det]` (cit. on p. 20).

[65]Bastian Keßler. “Development of Wavelength-shifting Fibers with high Photon Capture-Rate”. Master’s thesis. Johannes Gutenberg University Mainz, 2024 (cit. on p. 22).

[66]Benjamin Bastian-Querner et al. “The Wavelength-Shifting Optical Module”. In: *Sensors* 22.4 (2022), p. 1385. arXiv: `2112.12258 [astro-ph.IM]` (cit. on p. 22).

[67]John Rack-Helleis et al. “The Wavelength-shifting Optical Module (WOM) for the IceCube Upgrade”. In: *PoS* ICRC2021 (2021), p. 1038. arXiv: `2107.10194 [astro-ph.HE]` (cit. on p. 22).

[68]S. Agostinelli et al. “GEANT4–a simulation toolkit”. In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303 (cit. on p. 23).

[69]George F. Sterman and Steven Weinberg. “Jets from Quantum Chromodynamics”. In: *Phys. Rev. Lett.* 39 (1977), p. 1436 (cit. on p. 30).

[70]S. Catani, Yu.L. Dokshitzer, M.H. Seymour, and B.R. Webber. “Longitudinally-invariant k_t-clustering algorithms for hadron-hadron collisions”. In: *Nuclear Physics B* 406.1 (1993), pp. 187–224 (cit. on pp. 30, 31).

[71]Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti-k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. arXiv: `0802.1189 [hep-ph]` (cit. on pp. 30, 31).

[72]Stephen D. Ellis and Davison E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. arXiv: `hep-ph/9305266` (cit. on p. 30).

[73]Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. “Better jet clustering algorithms”. In: *JHEP* 08 (1997), p. 001. arXiv: `hep-ph/9707323` (cit. on p. 30).

[74]Andrew J. Larkoski, Ian Moult, and Benjamin Nachman. “Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning”. In: *Physics Reports* 841 (2020). Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning, pp. 1–63 (cit. on p. 30).

[75]Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. “Jet substructure as a new Higgs search channel at the LHC”. In: *Phys. Rev. Lett.* 100 (2008), p. 242001. arXiv: `0802.2470 [hep-ph]` (cit. on p. 31).

[76]Matteo Cacciari. “FastJet: A Code for fast k_t clustering, and more”. In: *41st Rencontres de Moriond: QCD and Hadronic Interactions*. July 2006, pp. 487–490. arXiv: `hep-ph/0607071` (cit. on p. 31).

[77]Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “FastJet User Manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. arXiv: `1111.6097 [hep-ph]` (cit. on p. 32).

[78]M. Wobisch and T. Wengler. “Hadronization corrections to jet cross-sections in deep inelastic scattering”. In: *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*. Apr. 1998, pp. 270–279. arXiv: `hep-ph/9907280` (cit. on p. 32).

[79]Stan Bentvelsen and Irmtraud Meyer. “The Cambridge jet algorithm: Features and applications”. In: *Eur. Phys. J. C* 4 (1998), pp. 623–629. arXiv: `hep-ph/9803322` (cit. on p. 32).

[80]Wikipedia contributors. *Nearest neighbor search — Wikipedia, The Free Encyclopedia*. [Online; accessed 6-March-2025]. 2025 (cit. on p. 43).

[81]SciPy community. *SciPy: Open source scientific tools for Python*. Accessed: 2025-03-02. 2020 (cit. on p. 58).

[82]Leo Breiman. “Random Forests”. In: *Mach. Learn.* 45.1 (Oct. 2001), 5–32 (cit. on p. 61).

[83]F. Pedregosa, G. Varoquaux, A. Gramfort, et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 63).

List of Figures

2.1. The Standard Model of Particle Physics. Neutrinos belong to the lepton family and interact only by the weak force through the Z and W bosons [13]. 5

2.2. Two possible mass ordering of neutrinos. Normal ordering and inverted ordering. 7

2.3. Feynman diagrams illustrating two double beta decay processes. The diagram on the left depicts the two-neutrino double beta decay $(2\nu\beta\beta)$, where two electrons and two corresponding antineutrinos are emitted, reflecting the behavior of Dirac fermions. The diagram on the right shows the neutrinoless double beta decay $(0\nu\beta\beta)$, in which the two neutrinos annihilate, leaving only two electrons in the final state, consistent with neutrinos being Majorana particles [25]. 8

2.4. Schematic view of the $2\nu\beta\beta$ and the $0\nu\beta\beta$ energy deposition. For the latter, a monochromatic peak at the Q-value is expected [31]. 9

3.1. Jablonski diagram illustrating the Fluorescence Resonance Energy Transfer (FRET) mechanism in organic scintillators. The diagram illustrates the electronic transitions between donor and acceptor molecules, along with their typical timescales. Credits: A. Mooney [50]. 13

3.2. Schematic of the light cone emitted by Cherenkov radiation. Credit: A. Horváth [56]. 15

3.3. Example of a 1 MeV e^+ simulated event in an opaque scintillator (left) and in a transparent scintillator (right). Credit: LiquidO Consortium [59]. 16

3.4. Example of different energy and particle depositions in an opaque scintillator. It is observed that the topology of the event is conserved which allows for particle discrimination [60]. 17

4.1. First proposed design of the CLOUD experiment. The inner volume (1) will contain 5-10 tons of opaque liquid scintillator instrumented with approximately 10,000 wavelength-shifting fibers running through the volume connected to SiPMS on both extremities. A veto volume (2) consisting of transparent liquid scintillator and PMTs. Surrounding this, a water tank, a concrete bunker (3) and a steel layer (4) are included for shielding. . . . 19

4.2. Proposed detector design of NuDoubt^{++}. It consists of a target volume containing an isotope-loaded opaque scintillator placed inside the inner volume. This is filled with a hybrid-opaque scintillator and instrumented with OWL fibers along the detector's vertical axis. Dedicated SiPMs and electronics will be at both ends of each fiber. Finally, a veto detector vessel equipped with a transparent scintillator and PMTs. 21

4.3. Top view (x-y) of the inner detector volume illustrating the positions of OWL-fibers in the simulation. The fibers are arranged in a parallel hexagonal distribution along the z-axis (HexZ) with a 1 cm pitch, providing coverage for optimal light collection. . . 23

4.4. Geant4 visualization of 1 MeV positron, gamma and electron depositing energy in the detector, and optical photons propagation. The particles are generated at the center of the detector. 25

4.5. Time distributions before and after the fiber propagation and SiPM model. 27

5.1. Example of the different $k_t t$ clustering algorithms [71]. It illustrates how each of the methods defines a way to define an area in which the particles form clusters. 30

5.2. Event visualization in the (x, y) plane of the clustering with different distance threshold values for the case of two contiguous light-balls. Each cluster is represented by a distinct color and the 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2. The distance threshold used in the algorithm has a direct impact on how the depositions are grouped. If this value is too small, it results in the main deposition being split into multiple clusters. If a value exceeding the average size of a deposition is chosen all the light gets clustered together losing the spatial structure of the event. 36

5.3. Optimization of the distance threshold for the two cases: a single photon bomb and a double photon bomb. For each case 1000 events were simulated. 37

5.4. Clustering results for a 1 MeV e^+ event before the matching procedure. The figure illustrates the event topology, where light depositions are spatially grouped into distinct clusters. For a given detector side, each cluster is represented by a different color. The left (right) plot represents the clusters for the top (bottom) SiPMs. The "x" marker indicates the center of mass (COM) position in the (x, y) plane for each cluster. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2 . . . 38

5.5. Matching results for the clusters of a 1 MeV e^+ event. The figure illustrates the event topology, in this case of multiple light depositions across the (x, y) plane of the detector. The left (right) plot represents the clusters for the top (bottom) SiPMs, with the same color meaning they belong to the same cluster. The "x" marker indicates the center of mass (COM) position. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2. . . . 39

5.6. Matching results for the clusters of a 1 MeV e^- event. The figure illustrates the event topology, in this case a single light deposition. The left (right) plot represents the clusters for the top (bottom) SiPMs, with the same color meaning they belong to the same cluster. The "x" marker indicates the center of mass (COM) position in the (x, y) plane. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2. 40

5.7. Matching results for the clusters of a 1 MeV γ event. The figure illustrates the event topology, in this case of multiple light depositions across the (x, y) plane of the detector. The left (right) plot represents the clusters for the top (bottom) SiPMs, with the same color meaning they belong to the same cluster. The "x" marker indicates the center of mass (COM) position. The 5.7% detection efficiency corresponds to the total amount of detected light based on the model implemented in Subsection 4.2.2. . . . 41

5.8. Number of detected photons versus the total clustering time for the three types of particle events analyzed. 42

6.1. Cluster times distribution for 1 MeV e^-, e^+, and γ events. The figure compares the timing characteristics of the clusters. The blue histogram represents the earliest photon time arrival time for each cluster, providing insight into the first detected light signals. The yellow histogram corresponds to the mean photon arrival time for each cluster, reflecting the overall time distribution of detected photons. These distributions help characterize the temporal response of the scintillator to different particle interactions. 47

6.2. 2D histograms showing the relationship between cluster time and energy for 1 MeV electrons, positrons, and gammas. The x-axis represents the cluster time, while the y-axis corresponds to the cluster energy. While small clusters are present throughout the entire time range, only small clusters tend to appear at later times. This suggests a time-dependent evolution in cluster formation, where late-emerging clusters are primarily low-energy. The distribution provides insight into the timing characteristics of light collection and potential scintillator response effects. . . 48

6.3. Cluster times distribution for 1 MeV e^-, e^+, and γ events after applying a photon selection of > 5 photons per cluster. This selection significantly impacts the mean cluster time distribution. It removes an anomalous bump previously observed in the electron distribution within the first 20 nanoseconds. This was primarily associated with small clusters, indicating that low photon clusters contributed disproportionately to early-time fluctuations. 49

6.4. Cluster Z-position resolution for 1 MeV e^-, e^+, and γ events, where the x-axis represents the difference $Z_{\text{reco}} - Z_{\text{true}}$. With Z_{true} the cluster Z position calculated as the weighted average using the true fiber hit z information from the Geant4 simulation. The histograms are normalized to form a probability density, ensuring the total area sums to one. **Top left:** Resolution before applying any selection criteria, serving as a baseline comparison. **Top right:** Resolution after applying an energy selection of > 5 detected photons. This selection improves resolution by mitigating the impact of small clusters. **Bottom left:** Resolution after imposing the detector limit selection ($Z \in (-550, 550)$), ensuring only clusters having a Z-position reconstructed within the detector volume. **Bottom right:** Resolution after applying both selections, demonstrating the combined improvement in the cluster Z-position reconstruction. 50

6.5. Evolution of the cluster Z-position resolution (σ_Z) as a function of the energy threshold. The x-axis represents the different threshold values (5, 10, 15, 20, and 100), while the y-axis shows the corresponding σ_Z values. The figure highlights the impact of selecting an appropriate hit threshold. 51
6.6. Position resolution for X, Y, and Z for the three kind of simulated particles. The histograms are normalized to form a probability density, ensuring the total area sums to one 53
6.7. Example of the detected photons for a 1 MeV e^- event, indicating the position of the recorded interactions. 54
6.8. Example of the detected photons for a 1 MeV γ event, indicating the position of the recorded interactions. 55
6.9. Example of the detected photons for a 1 MeV e^+ event, indicating the position of the recorded interactions. 55
6.10. Cluster energy distributions for 1 MeV e^-, e^+, and γ. The patterns of light deposition can be observed. 56
6.11. Preliminary fit of the electron cluster energy distribution to a Gaussian model. 57
6.12. The number of clusters represents a direct variable for the number of light depositions in the detector, and thanks to the topological information it can prove to offer valuable information for particle identification. 59
6.13. Distribution of the total event cluster energy for 1 MeV e^-, e^+, and γ events. The total event energy is defined as the sum of the energies of all individual clusters within the event. This highlights the differences in the amount of light deposited between the three particle types 60
6.14. Bounding box for a 1 MeV e^+ event. The blue dots represent the COM positions of the clusters, the box is represented in dashed red lines and it is constructed with the dimensions ($X_{\text{min}}, X_{\text{max}}$), ($Y_{\text{min}}, Y_{\text{max}}$), and ($Z_{\text{min}}, Z_{\text{max}}$). 61
6.15. Diagonal spread of the clusters for 1 MeV e^-, e^+, and γ events. The spread is defined as the maximum diagonal distance across the bounding box that contains all clusters of the event. 61

6.16. Combined ROC curves for binary classification cases. 63
6.17. Combined ROC curves for the two Random Forest classification methods. The AUC represents how well that class is classified versus the two other classes. 64

List of Tables

2.1. Q-values and natural abundance for isotopes that can go through $\beta\beta$ decays. Half-life of $2\nu\beta\beta$ process for those isotopes and some of the experiments performing $2\nu\beta\beta/0\nu\beta\beta$ searches with them [44]. 10

4.1. Scintillator and fiber properties used in the detector simulation and readout model. L.Y. is the light yield in photoelectrons per MeV, τ_1 and τ_2 are the fast and slow decay times of the scintillator, the scattering and absorption lengths characterize light propagation, and R.I. is the refractive index. For the fibers, the composition (Polystyrene) is not explicitly simulated, τ_{fiber} is the fiber decay time, ϵ_{fiber} is the efficiency (accounting for trapping, quantum efficiency, and optical coupling), t_{fiber} average time for a photon to travel 1 m of fiber, and l_{att} is the attenuation length (m). 24
4.2. Type of simulated events used in this work. All events were simulated at the center of the detector and fixed energy. Additionally, the total simulated and detected photons and the execution times for the fiber propagation and detection model. 27

6.1. Values for σ_Z for the different studied selections for 1 MeV e^-, e^+, and γ events. 51
6.2. Percentages of clusters kept as valid after different selections for 1 MeV e^-, e^+, and γ events. 52
6.3. Comparison of classification performance with all features and without total cluster energy as a feature. 64

A

Appendix

A.1 Implementation of the C-A algorithm

The C-A algorithm is generally implemented as follows:

```
procedure CAClustering(PhotonHits, DistanceMeasure, Threshold)
    Input: A list of optical photons, PhotonHits, each with spatial coordinates $(x, y, z)$, time $t$ and direction (-1 or 1).
    Output: A list of clusters with weighted centers of mass, time, energy proxy, and direction.
    Initialize an empty list Clusters.
    for all $p \in$ PhotonHits do
        Create a new cluster $C$ containing only $p$.
        Set the cluster center:
            $C.\text{center} \leftarrow (x,\ y,\ z,\ t)$
        Assign energy proxy:
            $C.\text{energy} \leftarrow 1 \quad \text{(each photon carries unit energy)}$
        Append $C$ to Clusters.
    end for

    // Compute Initial Distance Matrix
    Let $D$ be a symmetric matrix where each element $D(i, j)$ represents the distance between clusters $C_i$ and $C_j$ (using the given distance measure).
```

```
    for $i = 1$ to $|\texttt{Clusters}|$ do
        for $j = i + 1$ to $|\texttt{Clusters}|$ do
            Compute
                $$D(i, j) \leftarrow d(C_i, C_j)$$
            Set
                $$D(j, i) \leftarrow D(i, j)$$
        end for
    end for

    while the minimum element in $D$ is less than the distance threshold Threshold do
        Identify indices $(i, j)$ such that
            $$D(i, j) = \min\{D(k, l) \mid k \neq l\}$$
        if $D(i, j) \geq$ Threshold then
            break from loop.
        end if

        // Merge Clusters:
        Let $C_{\text{new}} \leftarrow C_i \cup C_j$.
        Compute the weighted center of mass for $C_{\text{new}}$:
            $$C_{\text{new}}.\text{center} \leftarrow \frac{C_i.\text{center} \times C_i.\text{energy} + C_j.\text{center} \times C_j.\text{energy}}{C_i.\text{energy} + C_j.\text{energy}}$$
        Update the energy proxy:
            $$C_{\text{new}}.\text{energy} \leftarrow C_i.\text{energy} + C_j.\text{energy}$$

        // Update Cluster List and Distance Matrix:
        Remove clusters $C_i$ and $C_j$ from Clusters.
        Append $C_{\text{new}}$ to Clusters.
```

```
        Remove the corresponding rows and columns $i$ and $j$ from the distance matrix $D$.
        for all each cluster $C_k$ in Clusters (excluding $C_{\text{new}}$) do
            Compute the distance:
                $d(C_{\text{new}}, C_k)$
            Update the matrix entries $D(\text{new}, k)$ and $D(k, \text{new})$ accordingly.
        end for
    end while
    return Clusters.
end procedure
```

Declaration

I hereby declare that I have written the submitted thesis independently using only the resources and sources listed (this includes AI-based applications or tools*). All direct or indirect quotations and paraphrased texts have been marked and cited appropriately. I certify that I did not use any tools whose use was explicitly prohibited by the examiner. I documented the AI tools I used in the appendix "Use of AI Tools." By submitting this work, I take responsibility for the entirety of its content. I thereby also take responsibility for any and all AI-generated content included in my work. I have checked the veracity of the included (AI-generated) content to the best of my knowledge. I have not submitted this work in an identical or similar form to earn credit in any other context. I am aware that a violation of the above specifications has consequences in terms of examination law and in particular can lead to the coursework/examination being marked as "failed." Enrollment can be revoked for up to two years if a student has cheated in coursework or examinations two or more times (section 69 subsections 4 and 5 University Act, HochSchG).

Mainz, March 26, 2025

Kitzia Michelle
Hernandez Curiel