Department of Physics and Astronomy Heidelberg University

Bachelor Thesis in Physics
submitted by

Lara Irida Belschner

born in Schwäbisch Hall (Germany)

2026

Validation of cluster parameters for the simulation of microphysics in superfluid helium-based detectors

This Bachelor Thesis has been carried out by Lara Irida Belschner at the
Kirchhoff Institute in Heidelberg
under the supervision of
Prof. Dr. Belina von Krosigk

Abstract

DELight is a direct detection experiment, currently still in the designing phase, searching for dark matter (DM) in the sub-GeV light dark matter (LDM) parameter space. It employs superfluid helium as target material, which thanks to its low nuclear mass allows to maximize the momentum transfer from the DM particle to the nucleus. This target material also enables the measurement of nuclear recoils, which provides a significant advantage, as most background sources predominantly contribute to the electronic recoil channel. The experiment is expected to reach sensitivities down to $\sim 10\,\mathrm{MeV/c^2}$, thereby probing regions of parameter space beyond the reach of many existing direct detection experiments.

In this work, the spatial and temporal clustering parameters used to reconstruct interaction events from Monte Carlo energy depositions are justified and validated. Clustering is required to group individual space-time energy depositions into physically meaningful events and to reproduce the expected signal formation in superfluid helium. Different background sources: gamma rays, neutrons, and the decay of ^{235}U, were simulated and analyzed to study the impact of clustering on the reconstructed energy distributions. It is shown that the spatial clustering parameter has a significant influence on the reconstructed spectra, while the temporal parameter plays a minor role within the relevant detector time scales.

Zusammenfassung

DELight ist ein Direktnachweisexperiment zur Suche nach Dunkler Materie (DM) im Sub-GeV-Bereich leichter Dunkler Materie (LDM) und befindet sich derzeit noch in der Planungs- und Entwicklungsphase. Als Targetmaterial wird suprafluides Helium verwendet, dessen geringe Kernmasse eine maximale Impulsübertragung vom DM-Teilchen auf den Atomkern ermöglicht. Dadurch wird die Messung von Kernrückstößen realisierbar, was einen entscheidenden Vorteil darstellt, da die meisten Hintergrundquellen überwiegend elektronische Rückstoßsignale erzeugen. Es wird erwartet, dass das Experiment Empfindlichkeiten bis zu $\sim 10\,\mathrm{MeV/c^2}$ erreicht und somit Parameterbereiche erschließt, die von vielen bestehenden Direktnachweisexperimenten bislang nicht zugänglich sind.

In dieser Arbeit werden die räumlichen und zeitlichen Clustering-Parameter, die zur Rekonstruktion von Wechselwirkungsereignissen aus Monte-Carlo-Energiedepositionen verwendet werden, begründet und validiert. Das Clustering ist erforderlich, um einzelne Raum-Zeit-Energiedepositionen zu physikalisch sinnvollen Ereignissen zusammenzufassen und die erwartete Signalentstehung in suprafluidem Helium realistisch nachzubilden. Hierzu wurden verschiedene Hintergrundquellen: Gammastrahlung, Neutronen sowie der Zerfall von ^{235}U simuliert und analysiert, um den Einfluss des Clusterings auf die rekonstruierten Energiespektren zu untersuchen. Es zeigt sich, dass insbesondere der räumliche Clustering-Parameter einen signifikanten Einfluss auf die rekonstruierten Spektren hat, während der zeitliche Parameter innerhalb der relevanten Detektor-Zeitskalen nur eine untergeordnete Rolle spielt.

Disclaimer

This thesis was prepared within the framework of the DELight collaboration. A significant contribution was made by Dr. Francesco Toschi, who developed the Monte Carlo simulation code employed in this study. All simulations presented in this work, as well as the analysis and interpretation of the results, were carried out independently by the author.

Use of Generative AI

The ChatGPT model (GPT-5.2), developed by OpenAI, was used as a supporting tool during the preparation of this thesis. In particular, it was employed to improve grammatical clarity and stylistic consistency of the written text, as well as to assist in the simplification and refinement of Python code used for data analysis and visualization. All scientific content, methodological decisions, implementation of the analysis, and interpretation of the results were developed and carried out independently by the author.

Contents

1 Dark Matter and Particle Interactions **5**
1.1 Dark Matter Evidence and Candidate 5
1.2 Direct detection of dark matter 6
1.3 Particle interaction in superfluid helium 8

2 DELight experiment **10**
2.1 Detection concept and apparatus 10
2.2 Why superfluid helium? 12
2.3 Signal formation 13
2.4 Motivation for energy deposition clustering 15

3 Monte Carlo Simulation Framework **17**
3.1 GEANT4 17
3.2 SPASS clustering algorithm 17
3.3 GEANT4 Simulation of Particle Interactions 19

4 Analysis of clustering behavior **22**
4.1 Sensitivity to temporal clustering parameter 22
4.2 Sensitivity to spatial clustering parameter 26

5 Event topologies limiting the spatial clustering parameters **33**
5.1 Lower limit to spatial clustering 33
5.2 Upper limit to spatial clustering 34
5.3 Secondary cut-off energy impact on clustering 36

6 Conclusion **39**

A Appendix **45**

B Appendix **46**

C Appendix **50**

1 Dark Matter and Particle Interactions

1.1 Dark Matter Evidence and Candidate

The modern paradigm of dark matter is hard to deny, even though this was not always the case. Fritz Zwicky's observations of the Coma Cluster in 1933 gave the first true hint of dark matter, but his work was not acknowledged immediately. The idea that unseen matter dominates most of galaxies' masses was hard to accept for many astronomers and physicists [1]. This began to change dramatically with Vera Rubin and Kent Ford. Their spectroscopic observations of the Andromeda Galaxy made it possible to create a full rotation curve, which represents the rotational velocity from the center to the outskirts of the visible galaxy. This rotational velocity can be estimated via Newtonian laws as Eq. 1.1:

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{1.1}$$

with G, the gravitational constant, $M(r)$ the mass within a certain distance, r the distance to the center and v the resulting rotational velocity. Therefore, a so-called Keplerian fall-off $v(r) \propto r^{-1/2}$ was expected [1]. But what was observed was a flat behavior even to the outskirts of the galaxy, as can be seen in Figure 1.1. The observation was unmistakable - most of Andromeda's mass is invisible. The visible disk's mass is not able to accelerate the outer regions to such high speeds [2]. The existence of a dark matter halo in which the galaxy is embedded is the theoretical solution to this missing mass problem. Astronomical observations lead to the assumption that 85% of the matter in our universe is dark matter [3]. Although galactic rotation curves are not the only evidence for the existence of dark matter, they clearly demonstrate that most galaxies' masses are dominated by non-luminous mass. If this consists of particles, several constraints have to be fulfilled. The most popular and also most searched for candidate is the Weakly Interacting Massive Particle (WIMP). That is because it naturally appears in many beyond Standard Model (BSM) theories and reproduces the observed relic abundance of dark matter when undergoing thermal freeze-out with an annihilation cross-section of the order of the electroweak scale. This is often referred to as the "WIMP miracle". In the early universe, temperatures were such that DM and standard model (SM) particles were created and were in thermal equilibrium - constantly created and annihilated. As the universe expanded and cooled down, there was not enough thermal energy to create WIMPs anymore but they could still annihilate, hence the number of DM particles dropped. Once the universe had expanded sufficiently, the interaction rate dropped below the Hubble expansion rate, and the DM abundance froze out. The co-moving

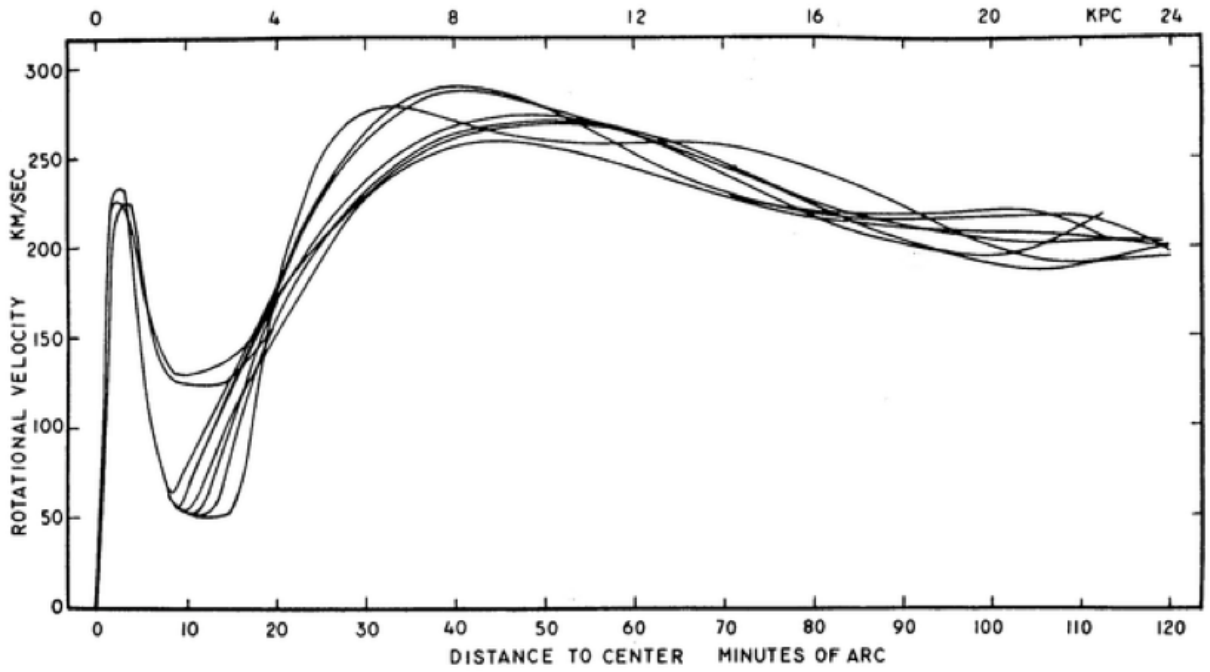


Figure 1.1: Rotational curves of the Andromeda galaxy, observed by Vera Rubin and Kent Ford, showing a flat behavior. Image adapted from [2]

number density of dark matter stayed constant and can be measured today [4].
As the name "WIMP" suggests, it is a massive particle, interacting with the SM particles only through gravity and the weak nuclear force. The mass of a standard WIMP is in the weak scale range $\sim 10\,\mathrm{GeV/c^2}$ to $10\,\mathrm{TeV/c^2}$[5]. In comparison the mass of an electron is only $0.511\,\mathrm{MeV/c^2}$. Even though the WIMP is a very promising dark matter candidate, it has remained undetected for several decades, which has lead to the idea that dark matter may also be in a lower mass range. This is called light dark matter (LDM), it is a WIMP-like particle including masses far below the $\mathrm{GeV/c^2}$, often referred to as sub-GeV [6]. It is estimated to be $m_\chi \gtrsim 10\,\mathrm{MeV}$. The Direct search Experiment for Light Dark Matter with superfluid helium (DELight) is aiming at exactly that range.

1.2 Direct detection of dark matter

DM particles may be detected through indirect searches, where two dark matter particles annihilate and produce SM particles (DM DM $\rightarrow$ SM SM). They can also be created at colliders, e.g. the LHC, where two SM particles annihilate and produce dark matter particles (SM SM $\rightarrow$ DM DM), and lastly WIMPs can be detected through direct interactions with SM particles (DM SM $\rightarrow$ DM SM), as has been summerized in Figure 1.2 [6].
As already stated, if WIMP-like LDM exists, it should be possible to detect it directly within

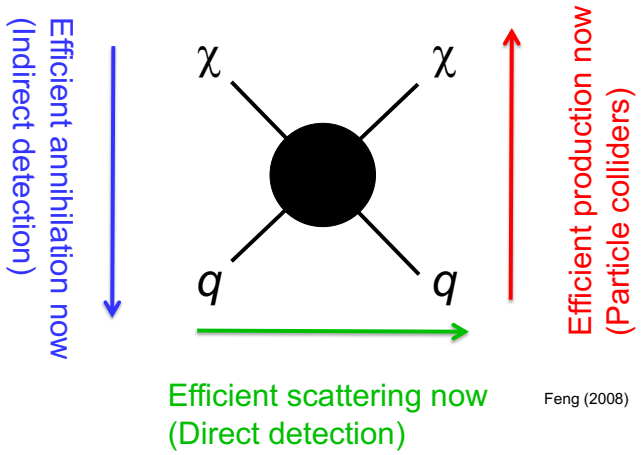


Figure 1.2: Different detection modes of DM particles. Here χ stands for DM and q for SM particle. Image adapted from [5]

a detector material. The DM density within the Milky Way has been determined to be $\rho_{\mathrm{dm}} = (0.21 - 0.55)\,\mathrm{GeV/cm^3}$ [3]. Direct detection experiments aim to measure the particles of the DM halo through scattering interactions with either the nucleus or the electron of the target material. In the most basic case, a DM particle crosses the detection material with a speed of $\sim 10^{-3}$ relative to the speed of light [7], interacts with a nucleus, and deposits a fraction of its kinetic energy in the recoiling nucleus. This process is equivalent to a classical elastic collision, the kinematics can therefore be expressed easily through the following (non-relativistic):

$$E_A = \frac{4 M_{\mathrm{A}} M_{\mathrm{DM}}}{(M_{\mathrm{A}} + M_{\mathrm{DM}})^2} E_{\mathrm{DM}}, \tag{1.2}$$

with E_{A} the recoiling energy of the target, E_{DM} the kinetic energy of the DM particle, and M_{i} the mass of the particles. This results in nuclear recoil energies below 10 keV for dark matter masses of the order of the weak scale, which are even smaller for LDM masses. Therefore, the energy threshold of these experiments needs to be correspondingly lower. Since the interaction strength and mass of LDM are at the weak scale, or even below typical neutrino cross sections ($10^{-40}\,\mathrm{cm^2}$), extremely low-background environments deep underground are required [4, 8, 9]. In particular, cosmic muons, due to their high energy and large mass compared to electrons and positrons, are unlikely to interact in the atmosphere and can easily reach the Earth's surface. Once entering denser material, they are more likely to interact. To reduce the impact of these particles, direct detection experiments must therefore be located deep underground. Typical background sources include cosmic radiation, gamma rays and radioactive decays from the material surrounding the target. These primarily cause electronic recoils; therefore, detectors that are capable of distinguishing whether a signal is produced by ER or NR have significant advantages [8]. Helium enables this identification, which will be discussed in Section 2.3. In the following, the focus

will lie on gamma rays and neutrons, as these are the main particles interacting in the target material. Therefore, the interaction properties and the response of helium must be studied thoroughly. Neutrons are of particular concern, as they are the only background source that can also produce NR responses mimicking the DM signal.

1.3 Particle interaction in superfluid helium

1.3.1 Gamma ray interaction mechanisms

Photoelectric absorption, Compton scattering, and pair production are the three most important interaction processes between gamma rays and matter. In all of these a fraction or total amount of the initial energy of the gamma is transferred to an electron. There exist more processes but they will not be discussed in the following as their contribution is negligible. In the case of photoelectric absorption, the photon interacts with an atom of the target material, the photon is entirely absorbed, and a so-called photoelectron is released from the atom from one of its bound shells. Not only a free electron is now appearing but also an ionized atom is left over. The energy of the now free photoelectron is:

$$E_{\mathrm{e}^-} = h\nu - E_{\mathrm{b}} \tag{1.3}$$

where $h\nu$ is the photon energy before the interaction, with h the Planck constant, ν the frequency, E_{b} the binding energy of the photoelectron in its former shell. This interaction mechanism is predominant for rather low energetic gamma rays and high atomic number Z of the absorber material. Since the probability of photoelectric absorption per atom is roughly given by $\tau \propto \frac{Z^n}{E_\gamma^{3.5}}$, where n lies between 4 and 5 for the gamma energies considered. As evident from this relation, high atomic number materials such as lead make good shields. Since helium has an atomic number of 2 the probability of a gamma ray interacting via the photoelectric effect is rather low. For example, at $E_\gamma \sim 100\,\mathrm{keV}$ one obtains $\tau_{\mathrm{Pb}}/\tau_{\mathrm{He}} \approx (82/2)^n = 41^n \sim 10^6$–$10^8$ (for $n = 4$–5), illustrating why photoelectric absorption is strongly suppressed in helium compared to high-Z materials such as lead.

Another important interaction mechanism is Compton scattering, where the photon interacts with the electron within the absorber material. But other than for the photoelectric effect the photon is not completely absorbed, but deflected at an angle θ. Only a fraction of the photon's initial energy is transferred to the electron, which is then called a recoil electron. The energy of the photon after the scattering is then given by

$$h\nu' = \frac{h\nu}{1 + \frac{h\nu}{m_0 c^2}(1 - \cos\theta)} \tag{1.4}$$

where m_0c^2 is the initial energy of the electron, which is assumed to be at rest. As evident from Eq. 1.4, for small scattering angles θ only a small fraction of the photon energy is transferred. The probability for a gamma ray to interact with the absorber material per atom via Compton scattering is dependent on the number of electrons available and therefore $\sigma \propto Z$.
Lastly, the production of an electron–positron pair, also called pair production, is worth mentioning, even though it is not a dominant effect for gamma energies below 10 MeV. For this process, the energy of the incident gamma ray must exceed a certain threshold, namely $E_\gamma > 2m_0c^2$, which corresponds to twice the rest-mass energy of an electron. If such a gamma ray interacts within the Coulomb field of a nucleus, it is completely absorbed and an electron–positron pair is created. The positron and electron then share the excess energy, i.e. the energy above 1.02 MeV. The positron will annihilate very quickly within the absorber material and produce two additional secondary photons. The exact probability is not easily expressed in a simple analytical form, but its relevance increases rapidly for gamma energies above approximately 1.5 MeV [10].

1.3.2 Interactions of neutrons

Neutrons, just as gamma rays, carry no charge, thus no Coulomb interaction can take place. If a neutron interacts with the target material, it will be with the nucleus, which is the crucial part: neutrons mimic the signal of DM through elastic scattering with the nucleus of the target material. The nuclear recoil resulting from the neutron interaction is therefore the most dangerous background for NR DM searches [8].
In a broad approximation, neutrons can be divided into two classes: "slow" and "fast neutrons", with a dividing line around 0.5 eV. Slow neutrons (often referred to as thermal neutrons) can in principle undergo neutron capture processes, which may lead to nuclear reactions and the emission of secondary radiation. In such cases, the Q-value of the reaction determines whether the process is energetically allowed. However, for ^{4}He this is not relevant, as the Q-value for neutron capture is negative and the neutron capture cross section is extremely small.
For higher neutron energies, the dominant interaction is elastic scattering with the helium nucleus. In this process, the neutron transfers part of its kinetic energy to the nucleus, resulting in a nuclear recoil and a change in the neutron's direction. These elastic scatterings are particularly relevant for direct dark matter searches, as they perfectly mimic the nuclear recoil signature expected from dark matter interactions.
If the neutron energy is sufficiently high, inelastic scattering processes may also occur, in which the nucleus is excited during the interaction and subsequently de-excites by emitting a gamma ray [10]. However, for the energies considered in this work, elastic scattering remains the dominant and most concerning background process.

2 DELight experiment

DELight is a direct detection experiment aiming to search for LDM including masses below $100\,\mathrm{MeV/c^2}$, so far barely probed by any other experiment. It uses superfluid helium as target material to realize low threshold measurements. The detector concept and the signal production will be explained in the following sections.

2.1 Detection concept and apparatus

The general detector layout is shown in Figure 2.1. The detector concept consists of a 1 L very high aspect ratio shape ("pancake"-like) filled with superfluid helium, which serves as the target medium.

When particles interact within the LHe, either with the nucleus or the electron of a helium atom, three primary signal channels are produced: excitation, ionization, and quasiparticle (qp) production. Excitation and ionization processes lead to the formation of scintillation light, while quasiparticle production results in phonon and roton excitations that propagate through the superfluid. These channels form the basis of the observable UV and phonon signals.

Above the liquid surface, a vacuum region is maintained. A large-area cryogenic micro-calorimeter (LAMCAL), based on magnetic micro-calorimeter (MMC) technology, is positioned in this vacuum to detect quasiparticle excitations generated in the helium. These excitations can lead to the quantum evaporation of ^{4}He atoms into the vacuum, where they are subsequently adsorbed and measured by the calorimeter. Along the bottom within the LHe, additional calorimeters are placed. However, they are not able to efficiently detect quasiparticles because of the high Kapitza resistance between LHe and solid materials [11].Only after multiple reflections within the helium volume the quasiparticles are able to leave through quantum evaporation. Film burners are used to avoid a thin layer of helium on the detectors in the vacuum, because superfluid helium has the property to creep up walls at temperatures below 2.7 K and would therefore cover all surfaces. This layer would then reduce the sensitivity of the detector.

As already stated, LAMCALs inside the helium volume will cover all walls of the cell, and are sensitive to photons and triplet excimers caused by ionization or excitation of an helium atom [12, 13, 11]. More insight into the signal production will be discussed in the next section.

The magnetic micro-calorimeters are the actual detectors, exploiting Curie's law $\delta M \propto \frac{\partial M}{\partial T}\delta T$, which shows the change of magnetization (M) of a paramagnet is proportional to the change of temperature (T). The setup of the DELight large area micro-calorimeters is shown in Figure 2.2. How it works is that energy depositions within the absorber material will lead to a temperature

change, which is then transmitted to a paramagnetic temperature sensor. The sensor is placed in a weak external magnetic field: any temperature change leads to a change in the magnetization (Curie's law), thus changing the magnetic flux within the paramagnetic sensor, since $\delta\phi \propto \delta M$. These variations are then converted into a change in voltage by a direct current superconducting quantum interference device (dc-SQUID).
MMCs have shown very high energy resolutions down to 1.25 eV, which is perfect for LDM search since energy resolutions of order $O(1\,\text{eV})$ are crucial [14].

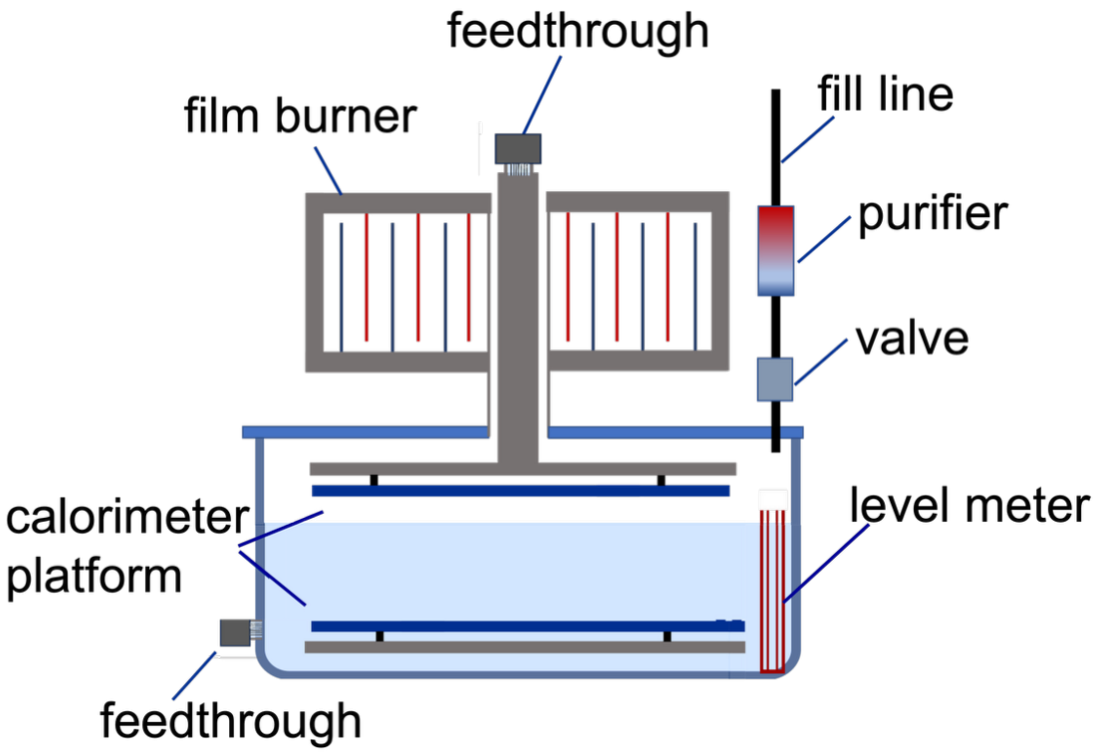


Figure 2.1: Schematic layout of DELight helium cell with film burner, heat flush purifier, level sensor and MMC. Image adapted from [12]

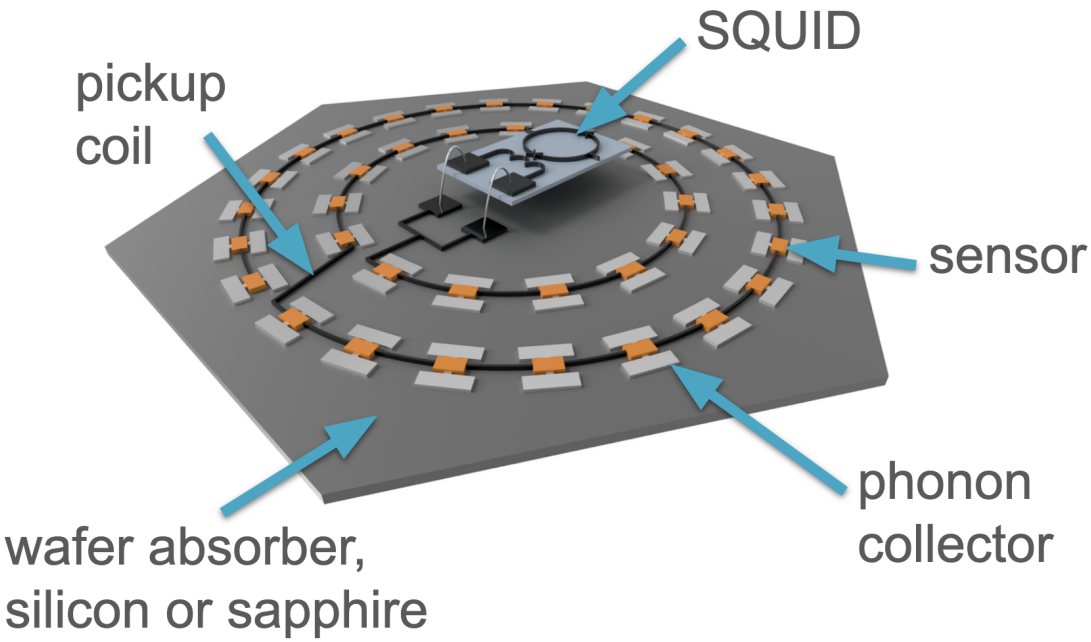


Figure 2.2: Schematic of large area MMC with absorber wafer (silicon or sapphire), the sensor, phonon collector and the SQUID device. Image adapted from [15]

2.2 Why superfluid helium?

Superfluid helium offers several advantages over commonly used detection materials such as xenon [16] or argon [17], particularly in the sub-GeV dark matter mass range. One key advantage is its low nuclear mass. For a fixed dark matter velocity, the kinetic energy transfer in elastic scattering is maximized when the target nucleus mass is comparable to the dark matter mass. Consequently, helium provides a larger recoil energy for LDM particles than heavier nuclei, resulting in an enhanced detectable energy signal in the sub-GeV regime.

In addition, helium is comparatively inexpensive and readily available among the noble elements. At temperatures of a few mK, superfluid ^{4}He supports the production and long-range propagation of quasiparticles, namely phonons and rotons. This heat channel is crucial for achieving energy thresholds below 100 eV.

The superfluid phase further enables a so-called self-cleaning mechanism: contaminants and impurities freeze out at low temperatures and do not remain dissolved in the liquid, leading to an extremely pure detection medium. Moreover, liquid helium contains no intrinsic long-lived radioactive isotopes, which significantly reduces internal radioactive backgrounds and makes it particularly well suited for rare-event searches.

Finally, signal formation in superfluid helium is particularly advantageous due to the presence of three independent and distinguishable signal channels: excitation, ionization, and quasiparticle production. A detailed description of the signal formation processes in superfluid helium is

provided in the next section. This is a crucial feature that makes it possible to discriminate between NR and ER, thereby strongly reducing the background.

2.3 Signal formation

As already mentioned in Section 1.2, direct detection experiments either measure nuclear recoil energies produced by LDM or neutrons, or electronic recoil energies caused by other background sources. An incident particle interacting with the helium target results in three different signal channels: electronic excitation, ionization, and quasiparticle excitation. This signal tree is visualized in Figure 2.3.

Excitation of ^{4}He leads directly to de-excitation via IR photon emission to either a singlet or triplet state of the first excited state of the helium atom, appearing as excited dimers (excimers). The singlet state ($A^1\Sigma_u^+$) is a very short-lived state, decaying via the emission of a UV photon within nanoseconds and producing an expected photon energy of $\sim 16\,\mathrm{eV}$, which is then directly detectable by the LAMCALs. The triplet excimer ($a^3\Sigma_u^+$) is long-lived, with a lifetime of $\tau = 13\,\mathrm{s}$. Given the detector dimensions and the propagation speed of $2 - 4\,\mathrm{m/s}$ of these excimers, the triplet state can be detected before decaying [14].

The ionized helium atoms recombine quickly into an excited state, which then follows the same mechanism as for excitation, de-exciting via IR photon emission and appearing either as a singlet or triplet excimer state.

Lastly, quasiparticle excitation occurs. Quasiparticle excitations in superfluid helium consist mainly of phonons and rotons, produced via non-radiative energy emission in excitation and ionization processes. Quasiparticles are also the only energy deposition mechanism in elastic scattering. Upon reaching the liquid surface, quasiparticles carrying an energy higher than the helium–helium binding energy can induce quantum evaporation of ^{4}He atoms into the vacuum. The evaporated atoms are subsequently adsorbed by the calorimeter and produce a measurable thermal signal.

The most important process for the clustering is Penning quenching. Charged particles interact in the target medium with the surrounding atoms and create a cloud of excited singlet and triplet states from excitation and ionization. These then interact with each other, and the outcome of these interactions can be observed, as the product of each individual interaction cannot be differentiated. The dashed lines in Figure 2.4 in the NR channel account for this effect, while the solid line does not. Clustering therefore helps to reconstruct the microphysics, which cannot be resolved by the detector [18].

These three signal channels have different time scales for arriving at the calorimeter. Whereas prompt scintillation occurs almost simultaneously with the recoil, the quantum evaporation oc-

curs on a timescale of order 100 µs[19]. In the case of nuclear recoils, the expected partitioning of the energy loss of ions within the target material, also called stopping power, can be expressed as follows 2.1:

$$E = \eta + \nu \tag{2.1}$$

where η denotes electronic stopping and ν nuclear stopping. Nuclear stopping mainly leads to quasiparticle excitations, while electronic stopping results in ionization and excitation.
The ratio of singlet to triplet outcomes for the ionization process has been determined by Ito and Seidel [20] and by Guo and McKinsey [21] to be 1:3. For electronic excitation, Seidel and Ito [20] estimated the ratio of singlet to triplet states as 0.86:0.14.
Since the excitation threshold of LHe is 19.77 eV, all incoming particles with lower deposited energies will result only in the quasiparticle channel [11].

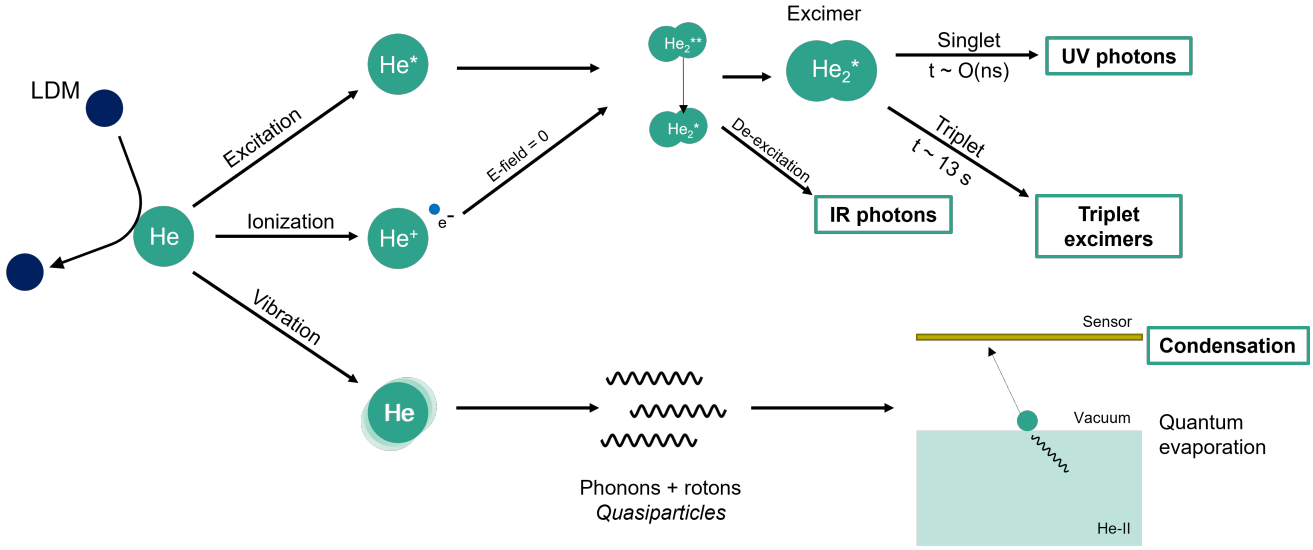


Figure 2.3: Schematic of signal formation [14]

Monte Carlo simulations lead to the same conclusion, as shown in Figure 2.4. In both NR and ER, the quasiparticle (QP) channel dominates at low energies, especially below ~ 20 eV. Below the excitation threshold, no singlet or triplet states can be produced, and therefore 100 % of the deposited energy goes into the QP channel.
For the NR case, the QP channel remains dominant even at higher energies, which distinguishes it from the ER case. In the ER scenario, after a peak at around 40 eV and 60 eV, the quasiparticle channel decreases and eventually falls below the UV channel.

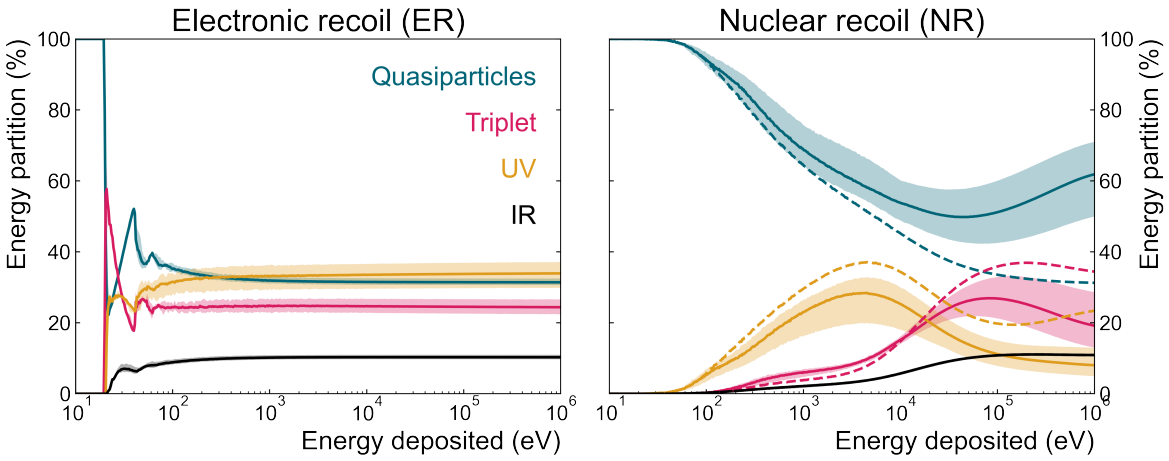


Figure 2.4: Mean energy partitioning of electronic (left) and nuclear (right) recoils into different signal channels. Image adapted from [18].

2.4 Motivation for energy deposition clustering

Clustering algorithms such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which forms the baseline of the DELight clustering algorithm Simulation Processing And cluStering for Superfluid Helium (SPASS), have become important analysis tools in particle and astrophysics. It is a topometric algorithm that groups energy depositions into space-time clusters. In the context of DELight, SPASS is used to reconstruct energy depositions in a way that closely resembles how signals would form and be observed in the detector. SPASS also performs signal partitioning of the simulations following the logic described in Section 2.3 and is based on what is described in [14].

Rather than serving primarily as an optimization tool, clustering in this work aims to reproduce the physical event topology expected in the detector. Background interactions (e.g. gammas) typically generate extended cascades with multiple secondary depositions, whereas a potential LDM interaction is expected to result in a localized single-site event. Clustering therefore provides a physically motivated way to model how multi-site background events and localized signal events would realistically appear in the detector, by requiring spatial and temporal coherence consistent with the detector response. An important aspect motivating the need for clustering is the microscopic nature of signal formation in superfluid helium. After an interaction, the deposited energy does not result in isolated, independent excitations. Instead, a spatially extended cloud of quasiparticles and excimers is created. These excitations can interact with each other before the final observable signal is formed. Processes such as Penning quenching modify the final light and quasiparticle yields depending on the local density of excitations. Therefore, the detector does not observe the simple sum of individual microscopic interactions, but rather

the collective outcome of this excitation cloud. Clustering provides a practical way to approximate this unresolved microscopic physics by grouping spatially and temporally correlated energy depositions into physically meaningful units.

3 Monte Carlo Simulation Framework

3.1 GEANT4

GEANT4 (GEometry ANd Tracking) is a toolkit developed by an international collaboration of physicists and software engineers to simulate the passage of particles through matter. As the name suggests, it provides functionalities such as particle tracking, geometry construction, physics process modeling, and hit recording [22]. GEANT4 relies on Monte Carlo techniques, sampling interaction probabilities based on implemented physics models to determine particle trajectories and secondary production. The interactions are simulated as discrete processes, meaning energy depositions at a given position and time. However, there are also continuous processes, such as described by the Bethe–Bloch equation, which introduce an additional level of discretization into the simulations.
In this work, simulation software developed for DELight is used to simulate different background sources within the planned detector geometry [23].
For each event, the spatial and temporal coordinates, the particle type, and the energy deposited in LHe and in the LAMCAL were recorded for all interactions for subsequent clustering analysis. Additionally, the primary spatial and temporal coordinates, track ID, parent ID, primary energy, parent particle type, sensitive volume, and the initial momentum were stored for each event.
The track ID is a number assigned to every particle within an event, starting with the primary particle assigned the number 1. Once this particle interacts and produces secondaries, these are assigned track IDs of 2, 3, and so on. The parent ID is similar, but instead gives the track ID of the particle responsible for the production of the given particle, the primary particle therefore has parent ID 0. These two properties are useful for identifying the track of a single particle and for understanding the physical properties of an event.
The sensitive volume is either 1 or 0, referring to whether the energy deposition occurred in the LHe or in the LAMCALs. This distinction is important, as only the energy depositions within the LHe are relevant for this study.

3.2 SPASS clustering algorithm

SPASS is a processing tool that uses the GEANT4 output as input. It implements a clustering procedure based on the DBSCAN algorithm. For a given $\epsilon > 0$, the algorithm searches within an ϵ-neighborhood in space around an energy deposition and determines whether additional energy

depositions are located within this region. In addition, it applies a density-based criterion by requiring a minimum number of energy depositions within the ϵ-neighborhood to form a cluster [24, 25]. There are three classes of points: core points, border points and outliers. Core points are energy depositions where both aspects of DBSCAN are fulfilled, border points are within the ϵ-neighborhood, but the density is below the threshold, and outliers are points where no criterion is fulfilled. The general working principle can be viewed in Figure 3.1. However, the

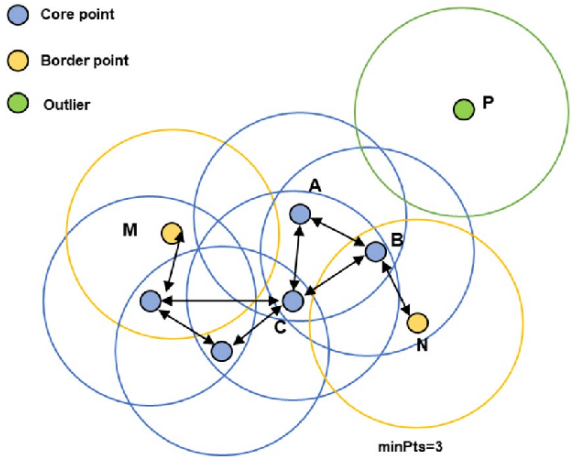


Figure 3.1: DBSCAN working principle, adapted from [25]

density criterion of the original DBSCAN algorithm is not applied, as a cluster is defined by at least one neighboring element. Therefore, no distinction between core points, border points, or outliers is made [26]. SPASS clusters not only in space, but also in time. The additional temporal aspect of the clustering is particularly important, since two events may occur spatially close to each other but at significantly different times and therefore do not physically affect each other and should not be clustered together. The main purpose of SPASS is to process the GEANT4 output and provide signal partitioning following the concepts described in Section 2.3.
The standard clustering parameters in this study were set to $dt = 50\,\mu s$ and the spatial parameter to $dd = 1\,mm$. The goal of this study is to validate the chosen parameters and estimate the impact of their choice on the observable spectrum.
SPASS provides the following output: the spatial coordinates of the primary particle, the spatial and temporal coordinates of the respective cluster, where each energy deposition contributes weighted by its deposited energy, the energy of the primary, the NR and ER energies with their respective fractions in the LHe, the partitioned phonon, triplet, UV, and IR energies, as well as the cluster ID, time ID, and combined ID. The cluster ID refers only to the spatial clustering, the time ID only to the temporal clustering, and the ID to the combination of both. Before starting the main analysis, the working principle of SPASS is illustrated in Fig. 3.2.
Here, the same event is processed with SPASS using different spatial clustering parameters. The original GEANT4 simulation consisted of 10^5 neutrons and was processed using the standard time

parameter of 50 µs, once with a spatial parameter of 1 mm (left) and once with 1 µm (right). The difference is clearly visible: the larger spatial parameter clusters all energy depositions into a single cluster, while the smaller parameter resolves the individual interaction track.
By calculating the spatial distances and time differences between the energy depositions, this behavior can be confirmed (Table 3.1). The energy depositions are not all contained within a region of 1 µm, and therefore should not be grouped together when using $dd = 1\,\mu\text{m}$.

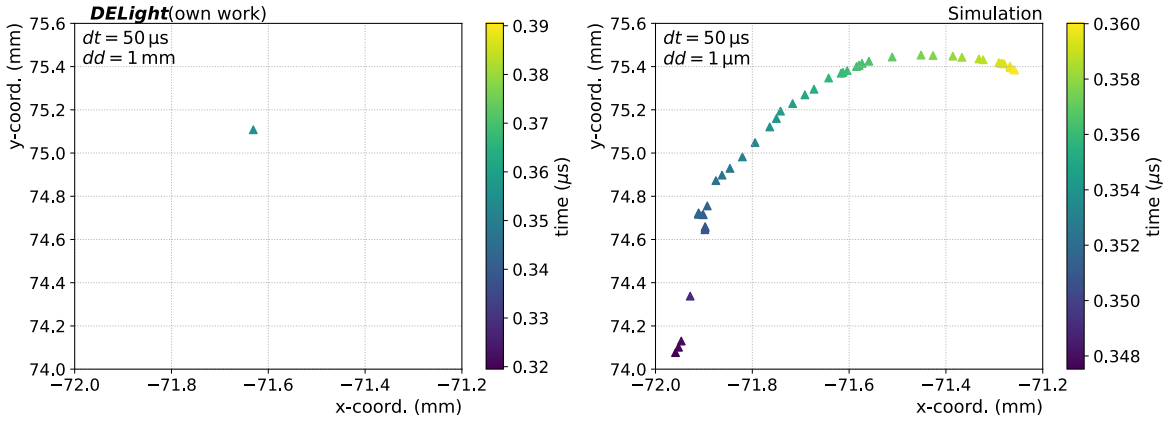


Figure 3.2: Energy deposition of different spatial clustering parameters, demonstrating the working principle

Table 3.1: Spatial distances and time differences for 1 µm clustering. The minimum spatial distance of the clustered data is still above 1 µm; therefore, the clustering behavior is reasonable. The time differences are within 50 µs.

Difference	Minimum	Maximum
Spatial distance (mm)	2.08×10^{-3}	0.328
Time difference (µs)	9.21×10^{-9}	1.80×10^{-3}

The same behavior can be found for time clustering as shown in Appendix A, Figure A.1, the SPASS algorithm can therefore be reliably applied in the following study.

3.3 GEANT4 Simulation of Particle Interactions

The purpose of these simulations is to generate realistic spatial and temporal energy deposition patterns in liquid helium. These patterns serve as input for the clustering algorithm and allow the validation of clustering parameters under physically motivated conditions, ensuring that reconstructed events reflect the expected macroscopic signal formation rather than artificial algorithmic effects.
As the clustering parameters are to be investigated and no detector is yet available, simulations

are required. Therefore, two particle types and one radioactive decay were chosen for the study. Gamma rays and neutrons are expected to be among the most relevant particles interacting within the LHe. ^{235}U is not the only relevant decay channel for background sources in direct detection experiments, but it has been selected for the simulations performed in this work, and therefore the focus will lie on it. In general, radioactive decays do not vary significantly in their interaction characteristics within the detector, such that ^{235}U serves as a representative example. There are two main types of radiation: α and $\beta^{\pm}$.

In the process of α radiation, two protons and two neutrons, i.e. a ^{4}He nucleus (also called an α particle), are emitted. The high (+2e) electric charge leads to rapid energy loss within the absorber material. As a result, α particles travel only short distances, typically on the order of centimeters. Other α emitters include thorium and radium.

β radiation emits either electrons or positrons. It originates from weak-interaction decays of neutrons or protons within a nucleus. As with α radiation, these charged particles interact with the target material ($\pm 1e$) and slow down due to continuous interactions with surrounding atoms [27, 10]. Through these simulations, the physical behavior of such particles within the detector can be reconstructed. Using the processing algorithm SPASS, the clustering and its impact on the energy distributions and signal formation can subsequently be analyzed. In the following, the simulations are described in more detail.
The detector geometry has already been described in Section 2.1. It consists of a 1 L cylindrical volume, 2 cm in height and with a diameter of approximately 20 cm, filled with liquid helium (LHe) and instrumented with LAMCALs at the top and bottom.
The origins of the neutron source and the ^{235}U decay were chosen to be outside a confined cylindrical region (cell) surrounding the detector, since helium itself contains no intrinsic radioisotopes at mK temperatures and therefore only external sources are relevant. Due to the detector dimensions and the low atomic number of helium, gammas are likely to traverse the LHe volume without interaction and instead interact with the absorber material of the LAMCALs, which has a higher atomic number. To increase the probability of interaction within the detector volume, the gamma source was therefore placed at the center of the detector. All sources were simulated isotropically. The origin positions are shown in Figure 3.3.
A summary of the simulation parameters is given in Table 3.2.

Table 3.2: Simulation properties of background sources

Background source	Gammas	Neutrons	^{235}U decay
Origin	center	confined cell	confined cell
Number of events	10^7	1.9×10^6	10^7
Primary energy	100 keV	(0,10) keV	–

The number of simulated events was chosen to ensure sufficient statistical precision for clustering calibration.

The purpose of these simulations is to characterize the spatial and energetic properties of relevant background sources for the calibration and validation of the clustering algorithm.

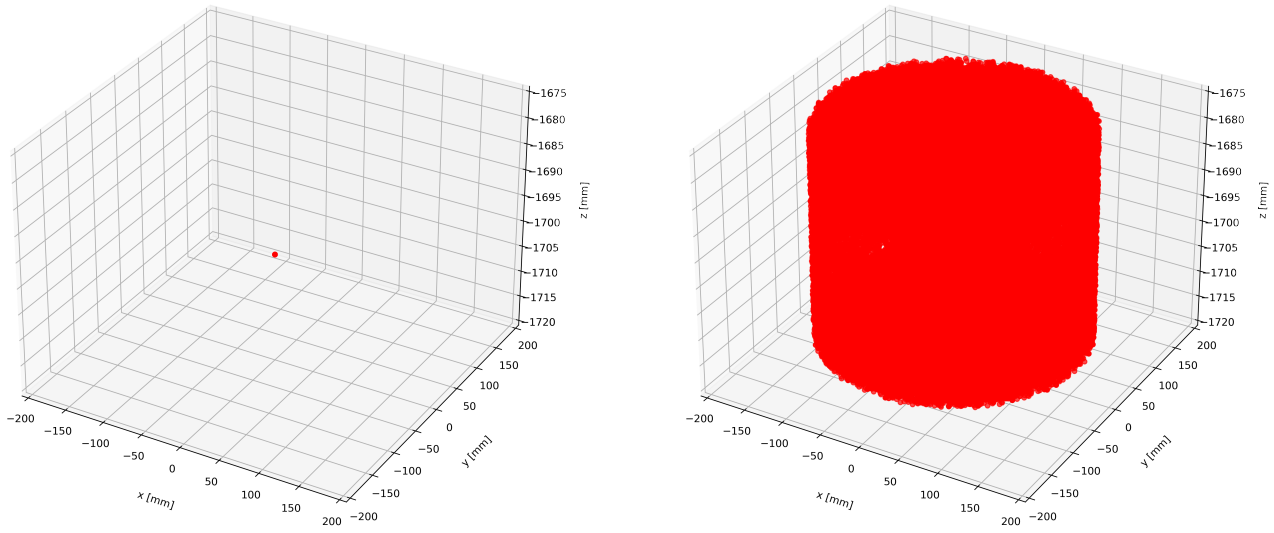


Figure 3.3: Coordinates/geometry of source origin. Gamma origin (left) and ^{235}U decay and neutron origin (right)

4 Analysis of clustering behavior

In this chapter, the impact of modifications of the clustering parameters in LHe is studied. Therefore, the three discussed background signals, gamma rays, neutrons, and the decay of ^{235}U, are considered.

4.1 Sensitivity to temporal clustering parameter

First, the behavior of the total energy deposition is studied by varying only the temporal parameter while keeping the spatial parameter constant ($dd = 1\,\mathrm{mm}$). To obtain a general overview of how significantly the time parameter affects the clustering, extreme values have been chosen. The results can be seen in Figures 4.1. On the x-axis, the total energy deposited within the LHe is shown, while the y-axis represents the probability density for a given deposited energy. The shown uncertainties are purely statistical and follow the Poissonian statistics.
The gamma distribution (Figure 4.1,top) exhibits a plateau-like structure with relatively sharp edges at approximately 15 eV and 25 keV, which is at around the Compton edge 28 keV. Since most of the energy is transferred to an electron via Compton scattering, this shape is expected. The electron subsequently loses its energy continuously through ionization.
The neutron distribution (Figure 4.1, middle), on the other hand, is flat up until approximately 10 meV. The distribution then decreases slowly up to $\sim$ 10 keV, where it drops more steeply. The ^{235}U simulations show a flat behavior from 20 eV to around 25 keV. From there it slowly decreases(Figure 4.1, bottom).
The more relevant observation for this study is that no visible change between the distributions can be observed when varying the temporal clustering parameter.
To quantify the difference between the distributions, the Kolmogorov–Smirnov (KS) and Jensen–Shannon (JS) distances have been computed. Both are statistical tools used to compare probability distributions without assuming any specific model. The KS test measures the maximum difference between the cumulative distributions and is therefore sensitive to the largest local deviation between two spectra. It returns a value between 0 and 1, where 1 corresponds to identical distributions [28]. The JS distance, on the other hand, compares the overall shape of two probability distributions. It is based on information theory and evaluates how similar the full distributions are, rather than focusing only on the largest deviation. It also returns a value between 0 and 1, with 1 again representing identical distributions [29].
Using both measures is helpful, since the KS distance highlights the largest deviation at a specific point, while the JS distance reflects the global similarity of the distributions.

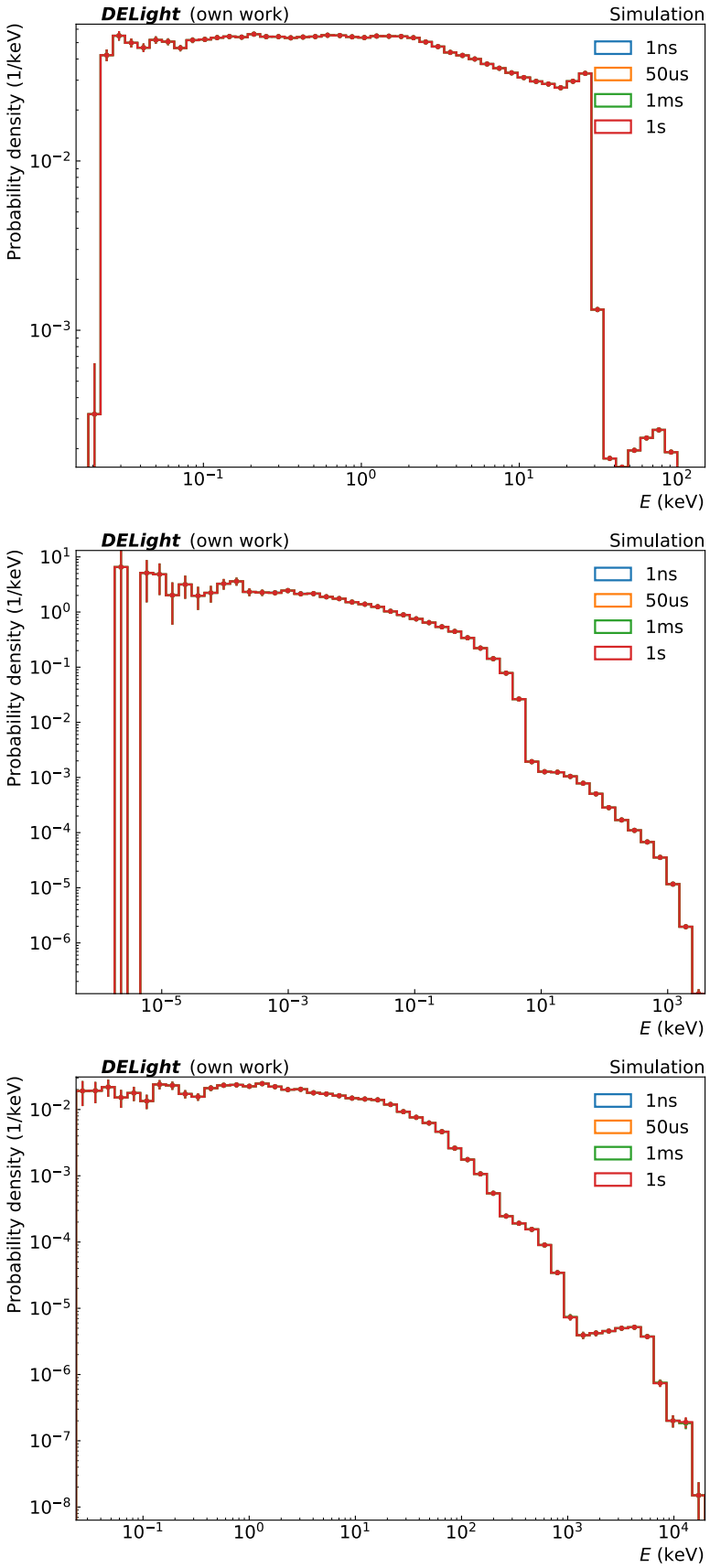


Figure 4.1: Total energy distribution for different time clustering parameters. Gamma simulation (top), neutron simulation (middle) and ^{235}U simulation (bottom).

Here the comparison distribution is the 'standard' ($dt = 50\,\mu s$, $dd = 1\,mm$), the results for each method can be seen in Table 4.1.

Table 4.1: Kolmogorov-Smirnov and Jensen-Shannon distance for all distributions

Simulation	KSD	JSD
Gamma ($dt = 1\,ns$)	1.000	1.000
Gamma ($dt = 1\,ms$)	1.000	1.000
Gamma ($dt = 1\,s$)	1.000	1.000
Neutron ($dt = 1\,ns$)	0.999	0.998
Neutron ($dt = 1\,ms$)	0.999	0.998
Neutron ($dt = 1\,s$)	1.000	1.000
^{235}U ($dt = 1\,ns$)	1.000	1.000
^{235}U ($dt = 1\,ms$)	1.000	1.000
^{235}U ($dt = 1\,s$)	1.000	0.998

For the gamma, ^{235}U, and neutron simulations, both the KS and JS distance show that there is essentially no difference between the different clusterings. Since, as explained in Section 2.3, the time scale of the signals lies in the µs regime, 1 ns is not a realistic parameter choice.
This leads to the conclusion that the temporal clustering parameter is not dominant in the clustering algorithm. An explanation for this can be obtained by examining the time differences between individual energy depositions. Taking the mean time difference Δt per event of the gamma simulation, $\Delta t \approx 1.6388 \times 10^{-5}\,\mu s$, this can be explained by the fact that gamma rays travel at the speed of light and therefore traverse the detection volume within nanoseconds. As described in Section 1.3.1, their interactions lead to the production of electrons, which lose their energy continuously through ionization. Both of these properties result in very small time differences between individual energy depositions. This is the reason why there is no change in the distribution: the mean of all Δt values is much smaller than 1 ns and would therefore be clustered anyway. A change would only be visible for parameters below $10^{-5}\,\mu s$. For the ^{235}U decay the mean difference Δt is either 0.0 or extremely high, $\sim 10^{18}\,\mu s$. Therefore, also here it makes sense that there is no change for the different temporal clustering parameters.
The signal partition has also been studied. Here the focus lies on the phonon and UV channel, since these are the important signals for the final analysis and are particularly useful for identifying NR events. As expected from the previous observation, no significant change can be observed, the distributions exhibit only statistical deviations. These fluctuations originate from the inherent statistical nature of the signal formation process, as it would occur in the real physical system, and are not induced by the clustering procedure. (Figures 4.2, 4.3, and 4.4). The phonon channel looks similar to the total energy distribution, while the UV channel shows, at very low energies, a quantization of energy, which is due to the quantization of light. As there is no effect on the cluster size, or distribution from choosing the temporal clustering parameter

obtained, the impact of spatial clustering is studied in Section 4.2

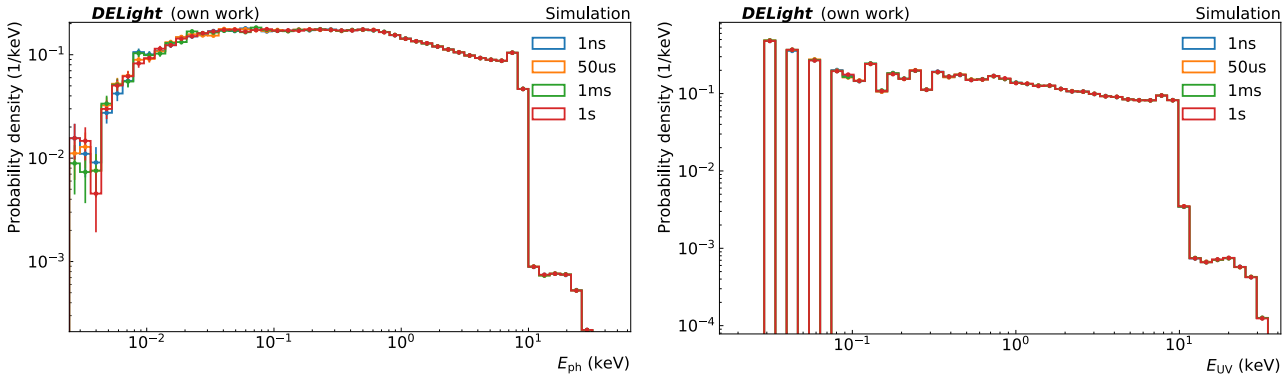


Figure 4.2: Gamma simulation: (left) phonon channel, (right) UV channel energy distribution

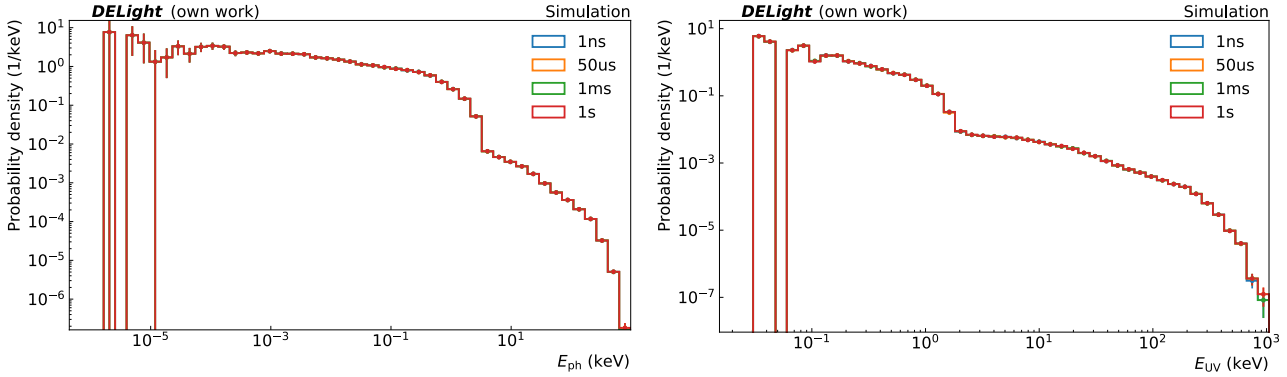


Figure 4.3: Neutron simulation: (left) phonon channel, (right) UV channel energy distribution

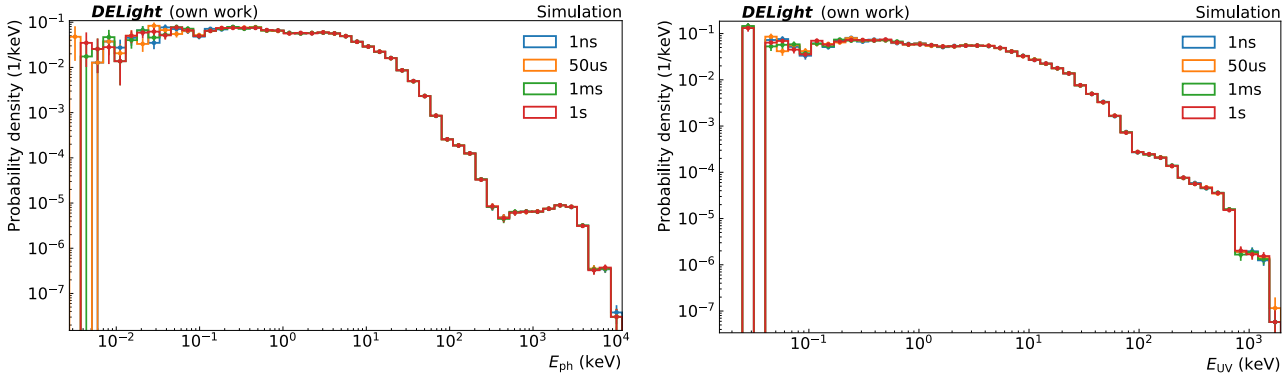


Figure 4.4: ^{235}U simulation: (left) phonon channel, (right) UV channel energy distribution

4.2 Sensitivity to spatial clustering parameter

As for the temporal clustering, the general response of the total energy deposition for the "extreme" spatial values is studied. Here, 1 fm, 1 nm, 1 µm, 100 µm, the standard 1 mm, and 1 cm have been chosen. In Figure 4.5, the results are presented. In contrast to the temporal clustering, significant changes are visible.

Starting with the gamma distribution (Figure 4.5 top), several features are directly visible. First, the 1 fm and 1 nm (blue and orange) parameters show the same distribution, indicating that there are no energy depositions within less than 1 nm. The 1 µm (green), as well as the smaller parameters (blue and orange), exhibit a peak at around 25 eV. This can be understood from the ionization energy of helium (24.6 eV), as small clustering parameters tend to isolate individual ionization events, each depositing approximately this characteristic energy. Lastly, the 100 µm (red), 1 mm (purple), and 1 cm (brown) parameters appear very similar. A plateau up to around 30 keV is visible, followed by a step down towards 100 keV. The transition between the regimes can be understood in terms of the typical spatial extent of secondary electron tracks in liquid helium. For clustering parameters below the characteristic separation of individual ionization steps, single interactions are resolved individually, leading to the pronounced peak at the ionization energy. As the clustering parameter approaches the typical track length of secondary electrons (micrometer to sub-millimeter regime for keV-scale energies [10]), spatially separated energy depositions are merged into single clusters, resulting in higher reconstructed cluster energies. A more detailed discussion of individual event topologies and their impact on the clustering behavior is provided later in this section.

Turning to the neutron simulation (Figure 4.5 middle), the general shape remains similar and also resembles the distribution obtained for temporal clustering (Figure 4.1 middle). Here as well, the 1 mm (purple) and 1 cm (brown) distributions lie very close together. The 100 µm (red), however, shows noticeable deviations, which will be explained using an example later.

The ^{235}U simulation again exhibits changes in shape. The 1 fm (blue) and 1 nm (orange) distributions lie directly on top of each other. As in the neutron case, the 1 mm (purple) and 1 cm (brown) distributions are very similar, while the 100 µm (red) shows slight deviations. This can be understood since the ^{235}U decay produces both α particles and secondary gamma radiation, combining features of both interaction types.

In general, for all three simulations, smaller spatial parameters lead to a higher probability in the low-energy region, while larger parameters allow for higher deposited energies. This is expected, since the cluster size increases with increasing spatial parameter, and larger clusters accumulate more deposited energy.

To quantify the changes for each spatial step, the KS and JS distance have again been determined, using the standard 1 mm clustering parameter as reference. To focus on the physically relevant

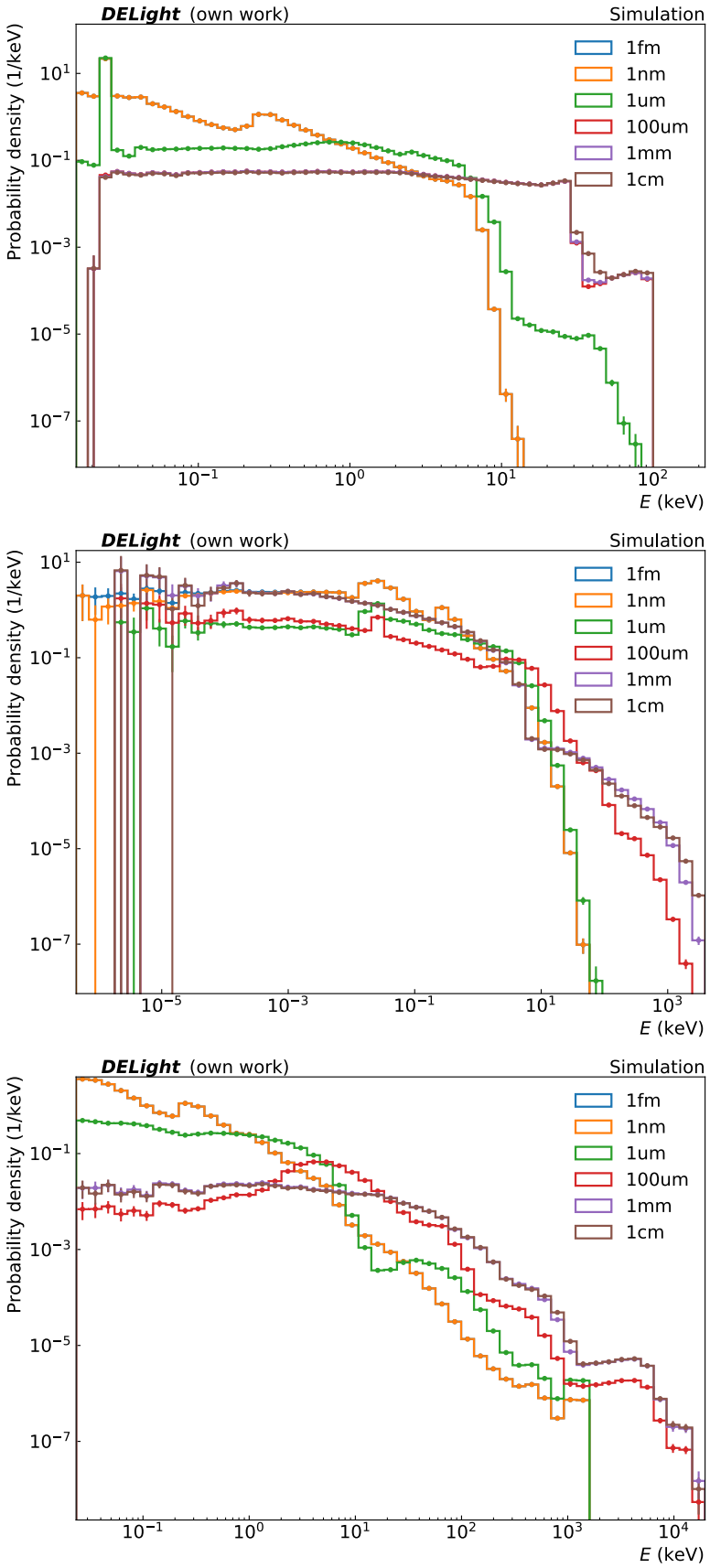


Figure 4.5: Total energy distribution for different spatial clustering parameters. Gamma simulation (top), neutron simulation (middle) and ^{235}U simulation (bottom).

parameters, only the results for 1 µm, 100 µm, and 1 cm are presented in Table 4.6, the remaining values can be found in Appendix B, Table B.1.

Figure 4.6: Kolmogorov-Smirnov and Jensen-Shannon distance for all distributions (spatial clustering)

Simulation	KSD	JSD
Gamma ($dd = 1\,\mu m$)	0.395	0.410
Gamma ($dd = 100\,\mu m$)	0.989	0.980
Gamma ($dd = 1\,cm$)	0.961	0.925
Neutron ($dd = 1\,\mu m$)	0.767	0.650
Neutron ($dd = 100\,\mu m$)	0.653	0.604
Neutron ($dd = 1\,cm$)	0.958	0.912
^{235}U ($dd = 1\,\mu m$)	0.331	0.370
^{235}U ($dd = 100\,\mu m$)	0.652	0.624
^{235}U ($dd = 1\,cm$)	0.961	0.943

As already suggested by Figure 4.5, the KS and JS distance confirm that significant differences between the distributions exist, particularly for parameters below 1 mm. The most interesting case is 100 µm, as the typical secondary electron track length lies on the order of this scale, and the deviation from the $dd = 1$ mm clustering is quite pronounced. Before focusing on this parameter, it is necessary to clarify which particles are responsible for these differences. In this work, ER is defined as any energy deposition originating from electrons, positrons, muons, antimuons, or photons, while NR refers to all energy depositions caused by particles other than these. For this purpose, the NR and ER distributions are plotted separately. As explained in Section 1.3, gamma interactions lead exclusively to ER events, therefore, only the neutron and decay modes are considered in this context.

In Figures 4.2 and 4.3, the NR and ER channel energy distributions within the LHe are presented. For the neutron simulation, the observed difference originates primarily from the ER channel. For the ^{235}U simulation, both channels appear to contribute, although this only applies to the 1 µm clustering parameter, the reason is discussed in Chapter 5. For the larger spatial parameters, the distributions remain similar and the changes are again predominantly driven by the ER channel. There is also a difference in the shape of the NR channel between the neutron and the ^{235}U simulations, which is most likely due to the significantly different energy scales of the resulting nuclear recoils: for neutrons the NR energy depositions extend from the meV to keV range, whereas the ^{235}U decay produces recoils reaching from keV up to MeV.

The number of events contributing to the NR channel (N_{NR}), defined as events for which the reconstructed cluster contains a non-zero nuclear recoil energy contribution ($E_{\mathrm{NR}} > 0$) and no electronic recoil contribution ($E_{\mathrm{ER}} = 0$), has been counted and compared to the total number of events in the entire simulation (N_{tot}). In addition, the energy deposited in NR events (E_{NR})

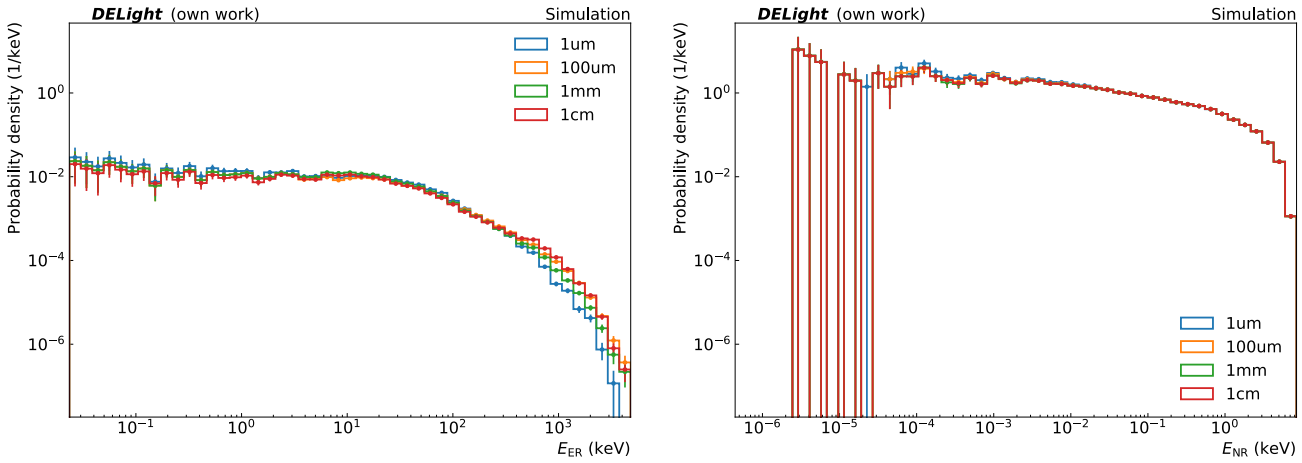


Table 4.2: Neutron simulation: (left) ER channel, (right) NR channel energy distribution

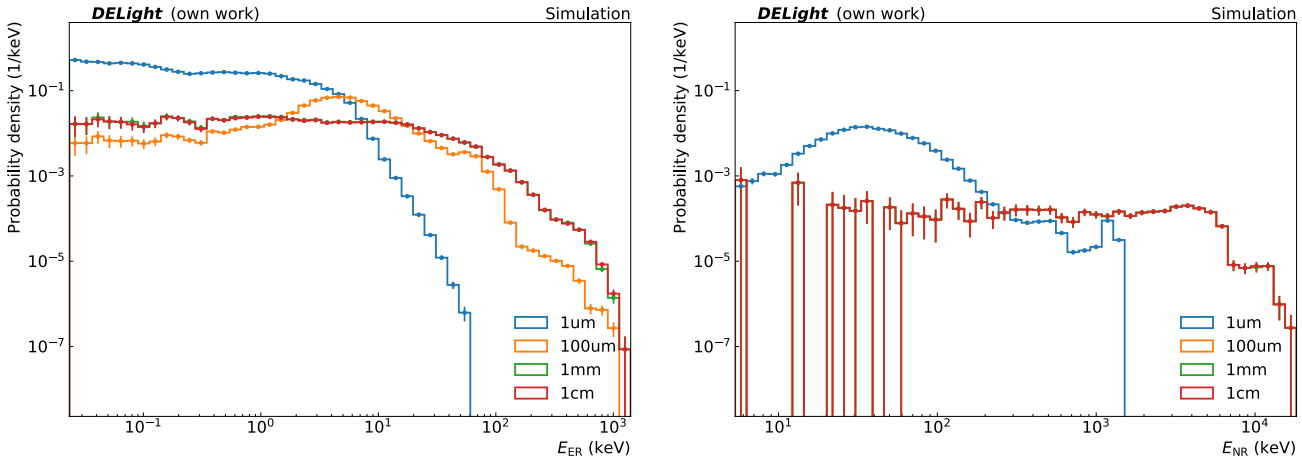


Table 4.3: ^{235}U simulation: (left) ER channel, (right) NR channel energy distribution

has been summed and compared to the total deposited energy in the entire simulation (E_{tot}). The fractions of the NR contributions are presented in Table 4.7. It is clearly visible that the fraction of NR counts increases with increasing spatial clustering parameter. This is due to the fact that, for larger clustering parameters, more ER events are merged together and therefore the total number of reconstructed events decreases. It is also evident why the NR contribution does not significantly affect the change of the total energy distribution, since it constitutes only $\sim 3\%$ of the total energy deposited within the LHe.

Consequently, even though neutrons represent a particularly dangerous background due to their NR signature, the clustering behavior of the total energy distribution is effectively governed by the ER component, as the NR contribution to the total deposited energy remains small. Nevertheless, from a background discrimination perspective, the NR component is of primary importance, since it mimics the expected dark matter signal. This observation may also be influenced by the choice of the secondary production cut, which could artificially enhance the

ER contribution. This will be briefly addressed in Chapter 5.3, primarily as an outlook for future studies.

Figure 4.7: Neutron simulation: Fraction of NR counts to total counts and fraction of NR energy to total energy

Clustering choice	$N_{\mathrm{NR}}/N_{\mathrm{tot}}$	$E_{\mathrm{NR}}/E_{\mathrm{tot}}$
$dd = 1\,\mu\mathrm{m}$	$(7.09 \pm 0.02)\,\%$	$(3.15 \pm 0.05)\,\%$
$dd = 100\,\mu\mathrm{m}$	$(22.47 \pm 0.06)\,\%$	$(3.15 \pm 0.01)\,\%$
$dd = 1\,\mathrm{mm}$	$(83.75 \pm 0.10)\,\%$	$(3.15 \pm 0.02)\,\%$
$dd = 1\,\mathrm{cm}$	$(85.50 \pm 0.09)\,\%$	$(3.15 \pm 0.05)\,\%$

Before proceeding, as in Section 4.1, to the signal formation, it is useful to discuss the actual events that lead to the differences between two clustering settings and their physical interpretation. Figure 4.8 shows one example event from the unprocessed neutron simulation contributing to the ER channel. The physical interpretation is the following: a neutron interacts outside

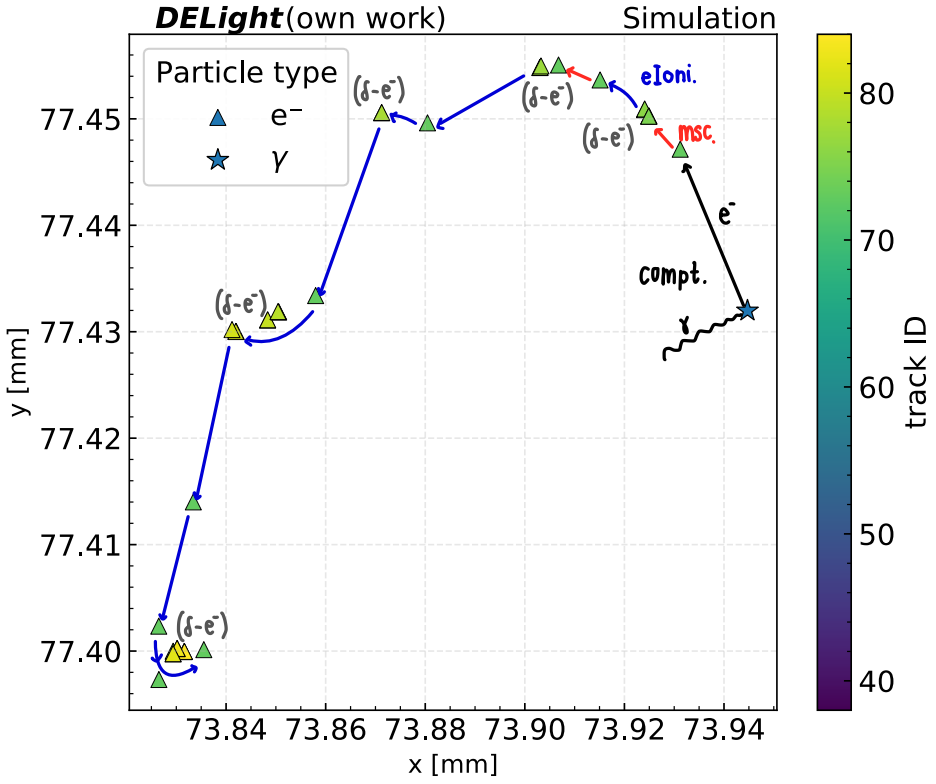


Figure 4.8: Energy deposition with neutron origin. A photon undergoes Compton scattering and releases an electron, which then propagates in the LHe and deposits energy along its trajectory through ionization and excitation processes.

the detector in such a way that a secondary photon is produced. This photon then enters the

detector volume and undergoes Compton scattering, releasing an electron from its atomic shell, which subsequently propagates through the LHe. As described previously, the electron loses its energy continuously via ionization and excitation processes.
The abbreviations "eIoni.", "msc.", and "compt." correspond to the respective interaction types: "compt." denotes Compton scattering, "msc." refers to multiple scattering, which in this context corresponds to excitation processes without full ionization, and "eIoni." represents electronic ionization. Along its trajectory, the "primary" secondary electron produces additional so-called δ-electrons with very low energies. Several other events with a similar structure have been identified and therefore appear to be responsible for the clustering differences. The color coding according to the track ID allows one to verify that it is indeed a single electron track responsible for the majority of the energy depositions. Since these depositions originate from a single electron track, they should physically be reconstructed as one cluster. Here, disadvantages of GEANT4 are also visible, since it simulates electron interactions in steps on the order of hundreds of micrometers. As discussed earlier, electrons deposit energy continuously along their path, therefore, these gaps would not physically exist, and all of these energy depositions should be clustered together.

The signal formation is also investigated for the spatial clustering, analogous to the temporal case. Again, only the UV and phonon channels are considered, since these represent the dominant observable signals. Here the observable corresponds to the signal that the detector would ultimately resolve. Since detectors cannot achieve time resolutions below approximately 50 µs, energy depositions occurring within that time window are naturally clustered. The observable takes this effect into account and displays the corresponding energy distribution.
In Figures 4.9 and 4.10, only the neutron simulation is shown, as the gamma and ^{235}U simulations exhibit the same qualitative behavior and are therefore presented in Appendix B. Figure 4.9 shows the phonon channel distribution, with the SPASS output (left) and the observable output (right). As expected, the SPASS output for both the UV and phonon channels exhibits differences between the spatial clustering parameters, similar to what was observed in the ER channel. The UV channel is affected more strongly, as the photon production is discrete. The observable output, on the other hand, shows a rather positive result: no significant changes are visible in the low-energy region and only slight deviations appear at higher energies. This indicates that although spatial clustering is important on a theoretical level, events occurring within very short time intervals cannot be distinguished by the detector in practice. This demonstrates that, while spatial clustering modifies the microscopic grouping of energy depositions, the finite detector time resolution effectively suppresses these differences at the observable level.

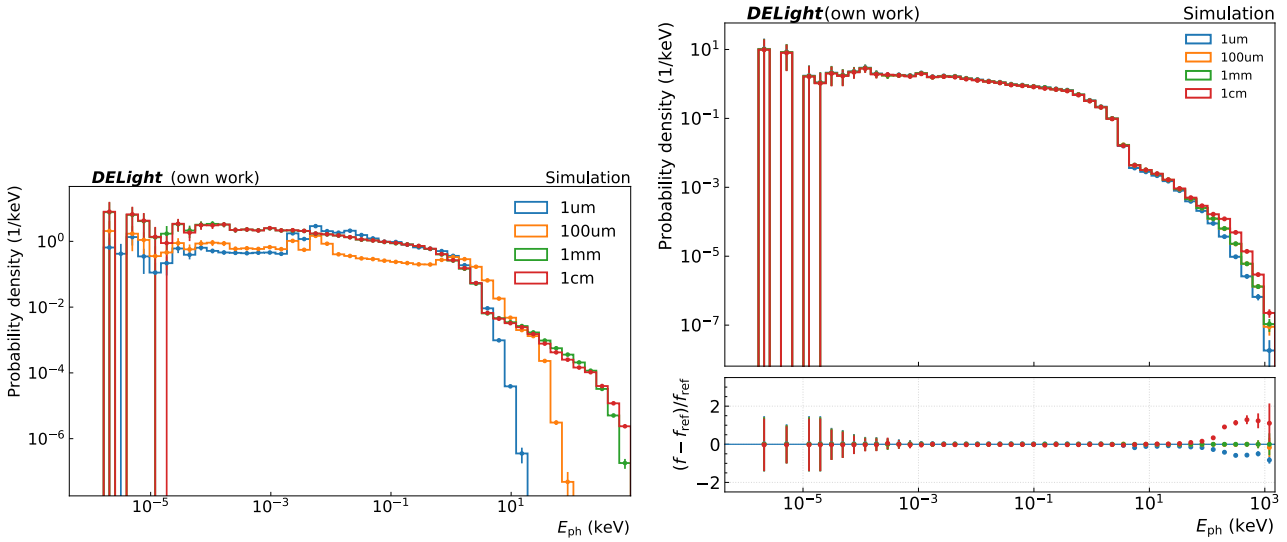


Figure 4.9: Neutron simulation: (left) phonon channel SPASS output, (right) phonon channel observable output, energy distribution

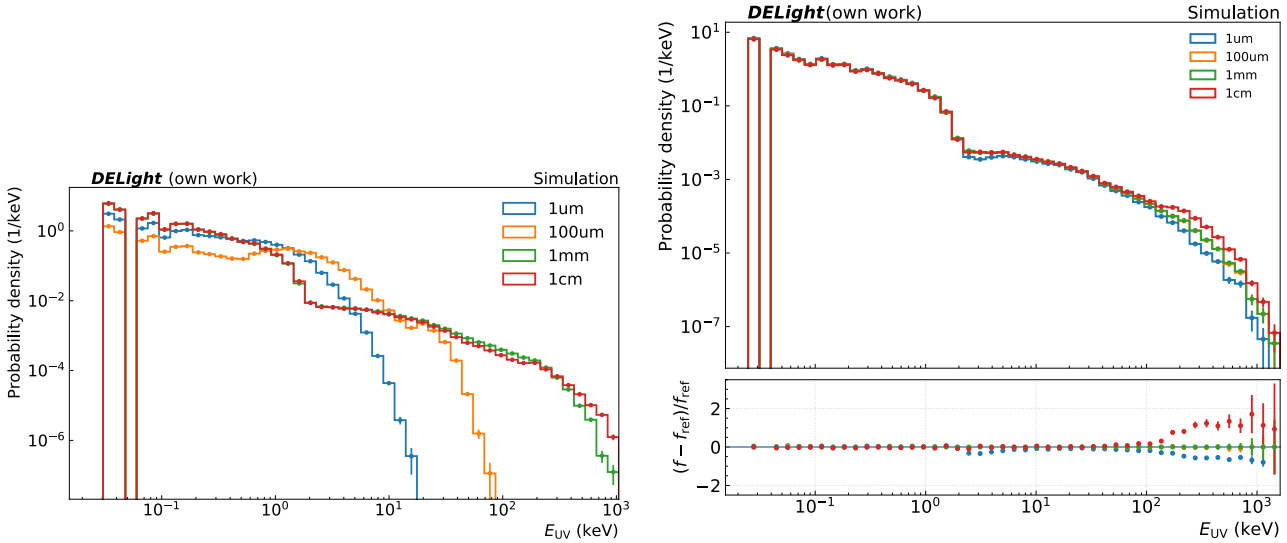


Figure 4.10: Neutron simulation: (left) UV channel SPASS output, (right) UV channel observable output, energy distribution

5 Event topologies limiting the spatial clustering parameters

In this chapter, a more physical and data-based reasoning for or against certain spatial clustering parameters is given. Since we have seen that most processes leading to differences are caused by ER, or more specifically by electron processes, this will be the focus. The mean free path λ of secondary electrons in LHe is on the order of $\sim 10^{-7}\,\mathrm{cm}$ for energies at around $\sim 20\,\mathrm{eV}$, which already rules out everything below $1\,\mu\mathrm{m}$ [30]. In the following sections, two example events are presented which rule out two explicit parameters and prove that $O(1\,\mathrm{mm})$ is actually reasonable.

5.1 Lower limit to spatial clustering

In the process of understanding the physics behind the energy depositions and the implementation of these physics in GEANT4, multiple events have been studied. Events were found that clearly exclude certain spatial clustering parameters. In Figure 5.1, a representative event is presented. The event originates from a decay of $^{235}\mathrm{U}$, which emits an α particle. Because of its high electric charge, it loses its energy very rapidly through ionization of neighboring helium atoms. Through this process, secondary electrons are produced, which then themselves lose their energy continuously along their trajectories via ionization and excitation. This process occurs twice, at different times ($\Delta t \sim 3.0 \times 10^{11}\,\mathrm{s} \approx 9500\,\mathrm{years}$), and is indicated by the two green arrows and the corresponding different track ID colors, yellow and green. Additionally, a secondary photon undergoes Compton scattering, releasing an electron which also loses its energy through ionization and emits multiple δ-electrons along its path.
The more important part to focus on here is the indicated box on the right. It shows a zoomed-in fragment of the electron track. The distance between those energy depositions is around $400\,\mu\mathrm{m}$, but these should definitely be clustered together. The electron loses its energy continuously through ionization and excitation, physically there would not be a gap, therefore GEANT4 is producing them. To reconstruct the actual physical properties these should be clustered. The respective processed energy deposition plots are shown in Appendix B, Figure B.5.
Consequently, the spatial parameter cannot be set below this threshold; otherwise, the concept of clustering regarding the GEANT4 simulations, is not fulfilled. This provides a first indication for a reasonable initial guess of a spatial parameter of $1\,\mathrm{mm}$. The value $100\,\mu\mathrm{m}$ is therefore excluded as a parameter candidate, as it separates physically continuous electron tracks.

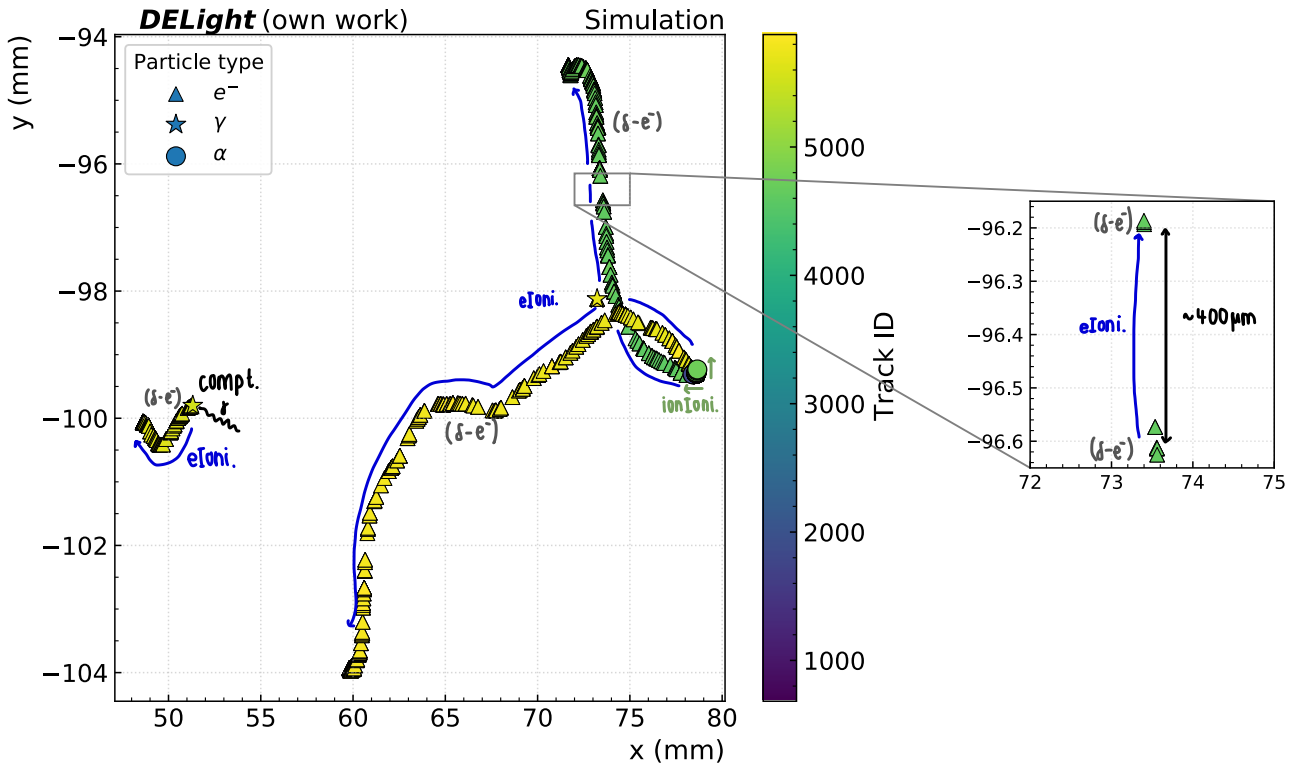


Figure 5.1: Energy deposition of one event from a ^{235}U simulation. ^{235}U decays into an α particle which, through ionization (ionIoni.), releases secondary electrons, which themselves have long tracks and ionize the helium atoms.

To further validate the choice of $dd = 1\,\text{mm}$, an upper limit has been identified. This is discussed in the next Section 5.2.

5.2 Upper limit to spatial clustering

As in the 100 µm case, there are several of events showing these properties, originating from ^{235}U decay. In Figure 5.2, a secondary gamma ray undergoes Compton scattering not once but twice within the liquid helium. Secondary electrons are produced but lose their energy rather rapidly via electronic ionization. Although it was previously stated that the same track ID corresponds to the same particle and should therefore be clustered together, in this case the distance between the two interactions is too large. Other than electrons, gammas do not lose their energy continuously but with discrete and separated processes (Compton, photo ionization and pair production), creating particles that do not affect each other.
Another reason is that gamma rays can produce interactions at large spatial separations while occurring within very short time intervals. Since gammas travel at the speed of light, independent of their energy, spatially distant interactions can still be temporally close. The mean free path of gammas in LHe is shown in Figure 5.3. The mean free path (λ) strongly depends on the energy,

and only for energies below $\sim 1\,\mathrm{keV}$ the free mean path λ is below 1 cm. This type of event is also rather rare. The difference between 1 mm and 1 cm is therefore only minimal. And the lower limit is rather important.
This indicates that $dd = 1\,\mathrm{cm}$ is too large of a clustering parameter, since these interactions should not be merged into a single cluster. The resulting clustered energy deposition plots are presented in Appendix B, Figure B.6.

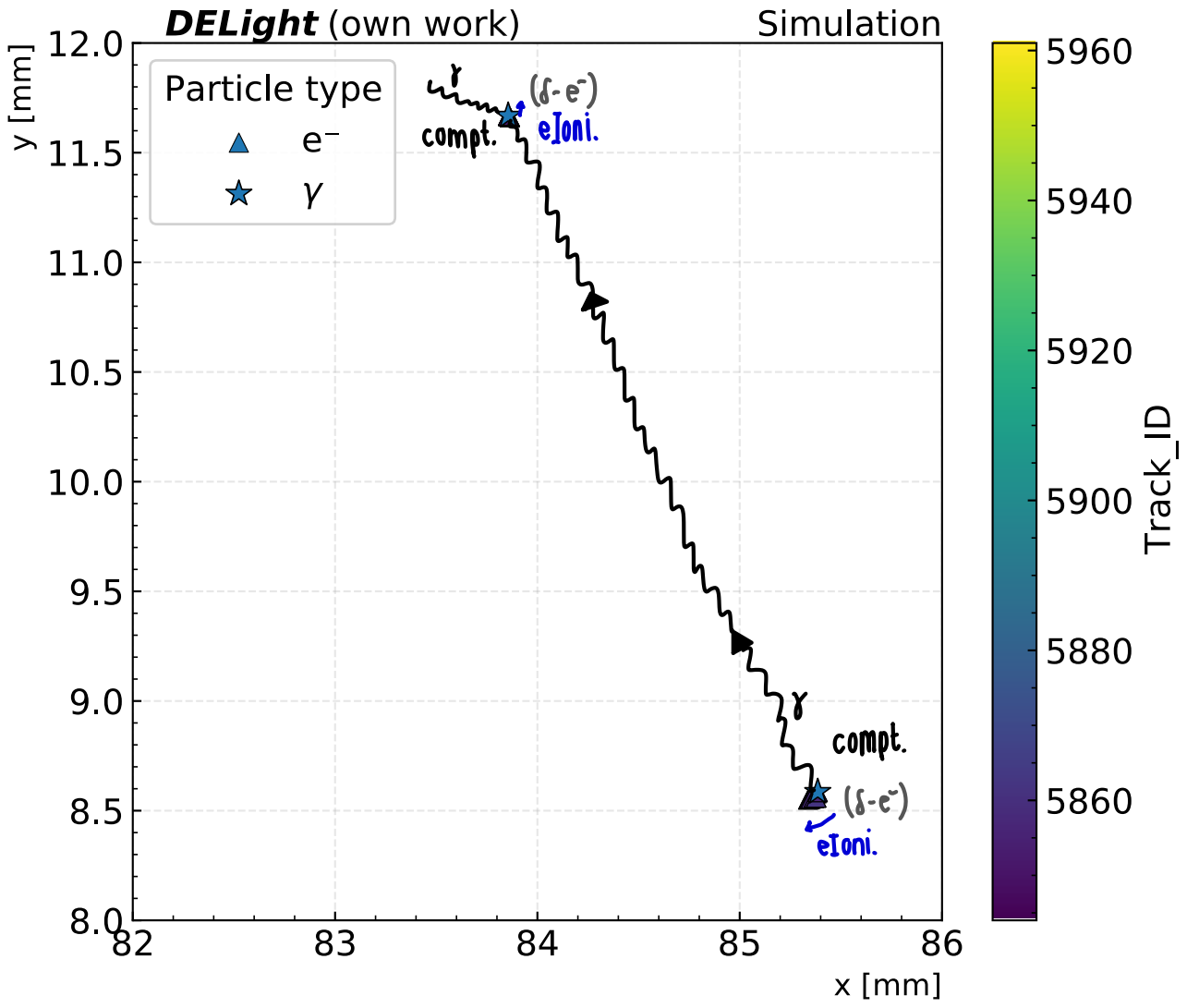


Figure 5.2: ^{235}U simulation: secondary photon interacting twice via Compton scattering, separated by more than 1 mm.

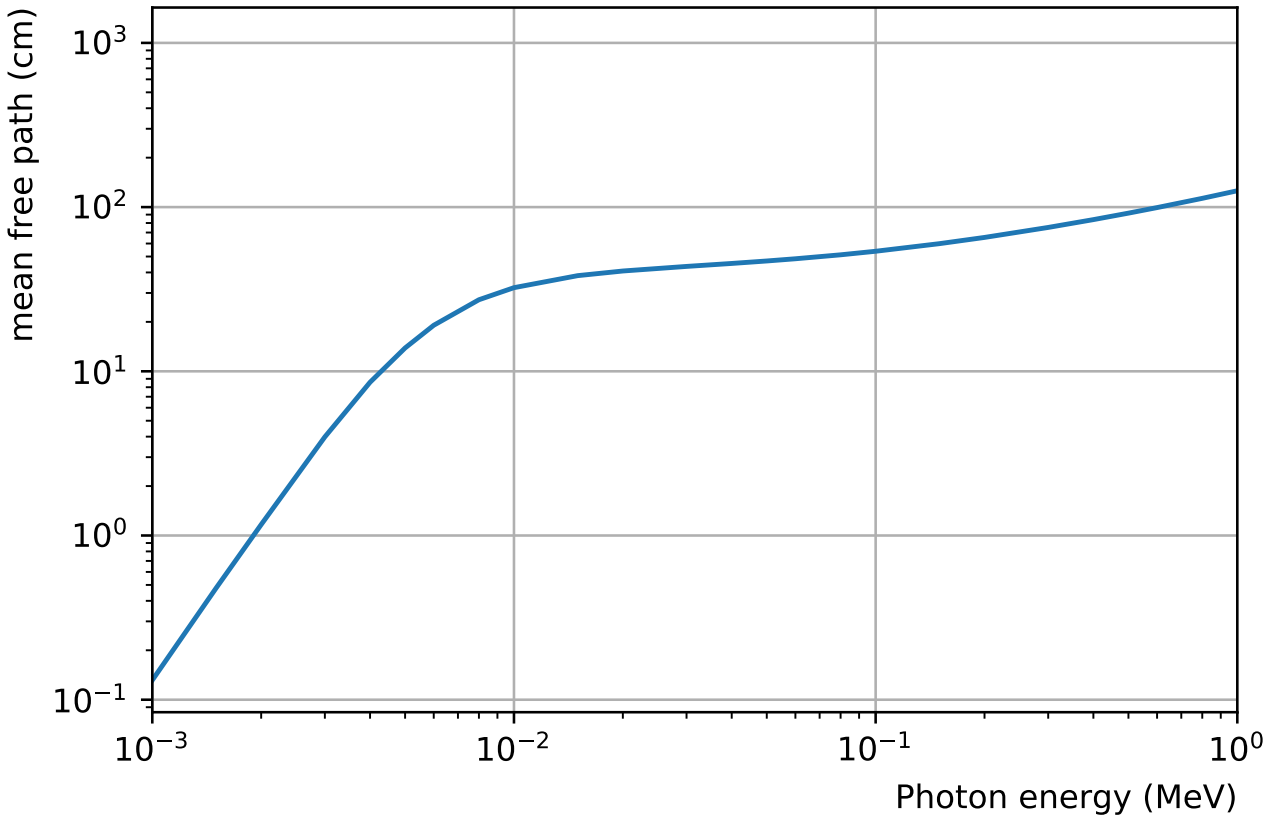


Figure 5.3: Mean free path of gamma rays in liquid helium [31].

5.3 Secondary cut-off energy impact on clustering

As already mentioned in Chapter 4, the low contribution of nuclear recoil effects to the change in the energy distribution might be a simulation effect rather than a physical one. The cut-off energy is a threshold, meaning that secondaries below this cut-off energy are not produced individually but are instead summed together. Here, it was set to $E_{\text{cutoff}} = 250\,\text{eV}$. To investigate whether it has an effect on the signal formation, further simulations have been carried out. In Chapters 4 and 5, the choice of the clustering parameters $dd = 1\,\text{mm}$ and $dt = 50\,\mu\text{s}$ has been justified. Therefore, only these values will be used in the following. The cut-off energy is set to 5 keV and 50 keV for comparison. The fractions of events contributing to NR ($E_{\text{NR}} > 0$) and ER (E_{ER}) at the same time have been considered first. The total number of these events for $E_{\text{cutoff}} = 250\,\text{eV}$ is 36 and goes down to 26 for the $E_{\text{cutoff}} = 5\,\text{keV}$ simulation, just to go back up to 36 events for the $E_{\text{cutoff}} = 50\,\text{keV}$ simulation. This is not expected and needs to be investigated further.

The fractions of the number of only NR events ($E_{\text{NR}} > 0$ and $E_{\text{ER}} = 0$) and NR energy are presented in Table 5.1. The fraction $E_{\text{NR}}/E_{\text{tot}}$ increases only slightly with increasing cut-off

energy, but the 5 keV cut-off energy simulation remains within the statistical errors.

Table 5.1: Fraction of NR counts and NR energy relative to the total counts and total energy for different cut-off energies

E_{cutoff} (keV)	$N_{\text{NR}}/N_{\text{tot}}$	$E_{\text{NR}}/E_{\text{tot}}$
0.250	$(83.61 \pm 1.35)\%$	$(3.16 \pm 0.06), \%$
5	$(69.71 \pm 1.52)\%$	$(3.22 \pm 0.06)\,\%$
50	$(70.70 \pm 1.52)\%$	$(3.26 \pm 0.06)\,\%$

To further investigate the impact, a signal space diagram has been produced for all cut-off energies. Figure 5.4 shows the UV energy versus the phonon energy.
There is clearly a change in how the energy is distributed. For $E_{\text{cutoff}} = 250\,\text{eV}$, the distribution extends up to 800 keV, whereas for the higher cut-off energies it extends only up to $\sim 300\,\text{keV}$. This is rather odd and indicates that it should be investigated more deeply in the future.
Another change is at especially low energy regions, below 100 keV where for the 250 eV shows a rather broad distribution and the other two a really fine distribution. Also the slope changes from 5 keV to 50 keV getting less steep. Further energy distributions for the different cut-off settings are presented in the Appendix C in Figures C.1, C.2, and C.3.
Overall, this suggests that the choice of secondary cut-off energy can influence the apparent NR fraction and should therefore be carefully considered in future, more systematic studies.

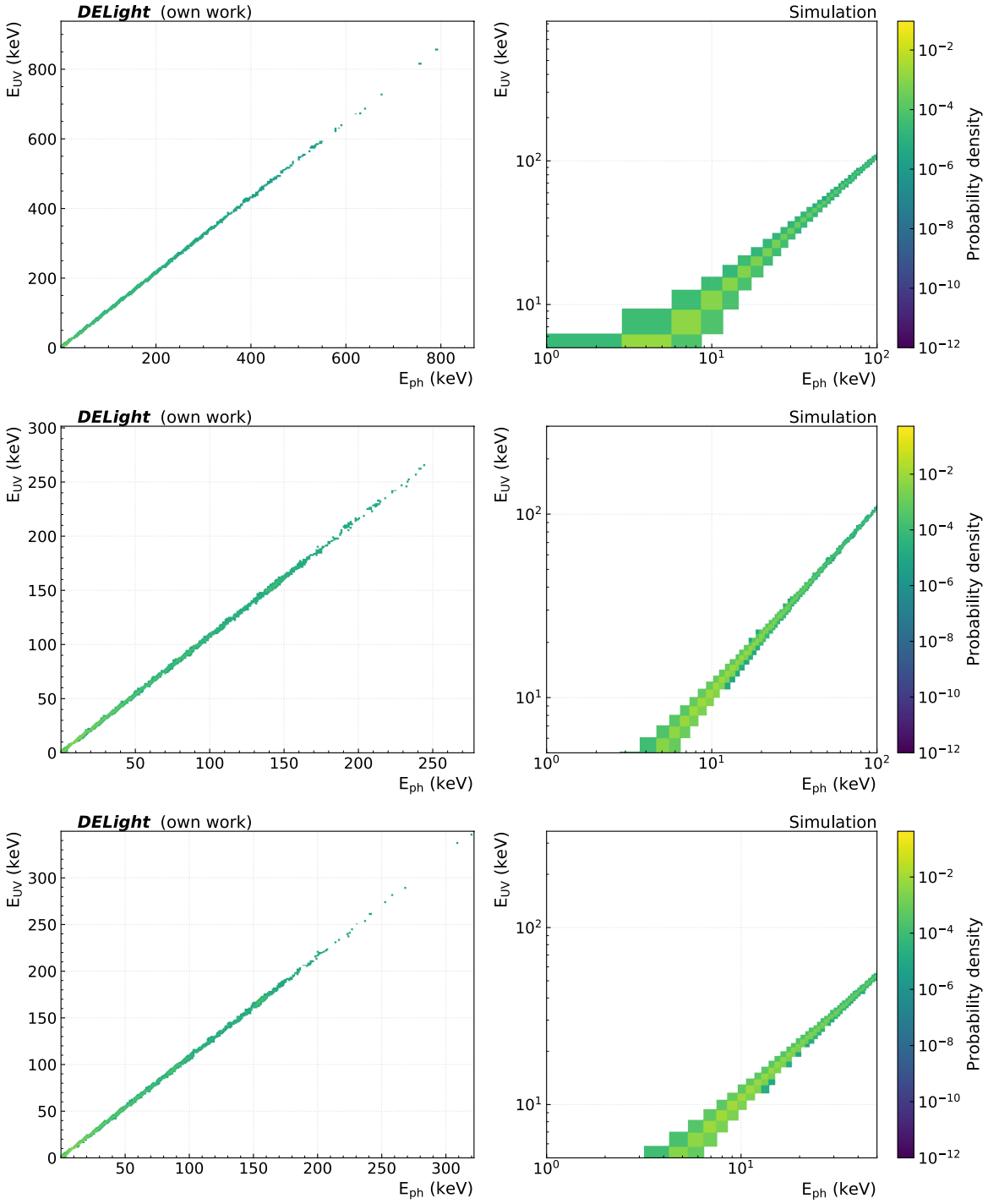


Figure 5.4: Observable signal space of neutron simulation. UV energy versus phonon energy with cut-off energy 250 eV (top), 5 keV (middle), and 50 keV (bottom)

6 Conclusion

The purpose of this work was to validate the spatial and temporal clustering parameters used in the DELight experiment by carrying out Monte Carlo simulations of relevant background sources: gamma rays, neutrons, and the decay of ^{235}U. In the simulation framework, individual energy depositions are produced at different positions and times. In order to reconstruct physically meaningful interaction events and reproduce the expected signal formation in superfluid helium, these depositions must be grouped using a clustering procedure. The clustering is defined by spatial and temporal parameters, which determine whether depositions are considered part of the same physical event.

First, the temporal clustering parameter was investigated. Detailed examinations of the reconstructed energy distributions have shown that, within the relevant range, the temporal parameter does not significantly affect the output. Variations of less than 1% compared to the standard choice support this conclusion.

In contrast, the sensitivity to different spatial clustering parameters revealed strong deviations in the total energy distributions. This was quantified using the Kolmogorov–Smirnov and Jensen–Shannon distances, which showed that clustering parameters below 1 mm lead to pronounced changes in the reconstructed spectra. A study of the origin of these deviations demonstrated that, for all three background sources, the ER channel exhibits the largest changes. Nevertheless, from the perspective of dark matter searches, the NR component remains the most relevant, as it mimics the expected signal.

During the analysis of individual event topologies, several spatially limiting cases were identified. One class of events indicates that clustering parameters below 100 µm artificially interrupt continuous electron tracks. This behavior originates from the stepwise treatment of energy deposition in GEANT4 and therefore reflects a simulation limitation rather than a physical effect. On the other hand, example events were found that disfavor clustering parameters above 1 mm, where spatially separated interactions, such as multiple Compton scatterings of a photon, would be merged into a single cluster.

It is important to distinguish between two different conceptual aspects: the detector time resolution and the clustering algorithm. The detector time resolution (assumed to be $\sim 50\,\mu s$) determines what can experimentally be resolved. The clustering algorithm, however, must ac-

count for the underlying microphysics of signal formation and the spatial extent of interaction processes. The comparison with the observable indicates that, within the assumed detector time resolution, spatial clustering does not drastically alter the experimentally accessible spectra.

Based on these studies, the choice of clustering parameters $dt = 50\,\mu\text{s}$ and $dd = 1\,\text{mm}$ can be justified. Nevertheless, further investigations of the microphysical modeling and simulation assumptions should be continued in future studies.

Bibliography

[1] G. Bertone and D. Hooper. History of dark matter. *Reviews of Modern Physics*, 90(4), October 2018.

[2] V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. , 159:379, February 1970.

[3] N. G. de Isídio, K. Menéndez-Delmestre, T. S. Gonçalves, M. Grossi, D. C. Rodrigues, N. Garavito-Camargo, A. Araujo-Carvalho, P. P. B. Beaklini, Y. Cavalcante-Coelho, A. Cortesi, L. H. Queiroga-Nuñez, and T. Randriamampandry. Dark matter distribution in milky way-analog galaxies, 2024.

[4] N. Bozorgnia, J. Bramante, J. M. Cline, D. Curtin, D. McKeen, D. E. Morrissey, A. Ritz, S. Viel, A. C. Vincent, and Y. Zhang. Dark matter candidates and searches. *Canadian Journal of Physics*, 103(8):671–703, August 2025.

[5] J.L. Feng. The wimp paradigm: Theme and variations. *SciPost Physics Lecture Notes*, June 2023.

[6] G. B Gelmini. Light weakly interacting massive particles. *Reports on Progress in Physics*, 80(8):082201, June 2017.

[7] T. Lin. Dark matter models and direct detection. *PoS*, TASI2018:009, 2019.

[8] J. Xu, P.S. Barbeau, and Z. Hong. Detection and calibration of low-energy nuclear recoils for dark matter and neutrino scattering experiments. *Annual Review of Nuclear and Particle Science*, 73(1):95–121, September 2023.

[9] M. Misiaszek and N. Rossi. Direct detection of dark matter: A critical review. *Symmetry*, 16(2):201, February 2024.

[10] G. F Knoll. *Radiation detection and measurement*. Wiley, New York, NY, 2010.

[11] S. A. Hertel, A. Biekert, J. Lin, V. Velan, and D. N. McKinsey. Direct detection of sub-GeV dark matter using a superfluid ^{4}He target. *Phys. Rev. D*, 100(9):092007, 2019.

[12] B. von Krosigk, K. Eitel, C. Enss, T. Ferber, L. Gastaldo, F. Kahlhoefer, S. Kempf, M. Klute, S. Lindemann, M. Schumann, F. Toschi, and K. Valerius. Delight: A direct search experiment for light dark matter with superfluid helium. *SciPost Physics Proceedings*, (12), July 2023.

[13] R. E. Lanou, H. J. Maris, and G. M. Seidel. Detection of solar neutrinos in superfluid helium. *Phys. Rev. Lett.*, 58:2498–2501, Jun 1987.

[14] F. Toschi, B. Maier, G. Heine, T. Ferber, S. Kempf, M. Klute, and B. von Krosigk. Optimum filter-based analysis for the characterization of a high-resolution magnetic microcalorimeter. *Physical Review D*, 109(4), February 2024.

[15] DELight collaboration. The delight experiment – magnetic microcalorimeters. `https://delight.kit.edu/experiment.html`, 2026. Accessed: 2026-02-15.

[16] E. Aprile. The xenonnt dark matter experiment. *The European Physical Journal C*, 84(8), August 2024.

[17] A. Rubbia. Ardm: a ton-scale liquid argon experiment for direct detection of dark matter in the universe. *Journal of Physics: Conference Series*, 39:129–132, May 2006.

[18] F. Toschi, A. Brunold, L. Burmeister, K. Eitel, C. Enss, E. Fascione, T. Ferber, R. Gabriel, L. Hauswald, F. Kahlhoefer, S. Kempf, M. Klute, B. von Krosigk, S. Lindemann, B. Maier, M. Schumann, M. Solmaz, K. Valerius, and F. C. Wagner. Signal partitioning in superfluid. *Physical Review D*, 111(3), February 2025.

[19] R. Anthony-Petersen, A. Biekert, C. L. Chang, Y. Chang, L. Chaplinsky, A. Dushkin, C. W. Fink, M. Garcia-Sciveres, W. Guo, S.A. Hertel, X. Li, J. Lin, R. Mahapatra, W. Matava, D.N. McKinsey, D.Z. Osterman, P.K. Patel, B. Penning, H.D. Pinckney, M. Platt, M. Pyle, Y. Qi, M. Reed, G.R.C. Rischbieter, R.K. Romani, A. Serafin, B. Serfass, R.J. Smith, P. Sorensen, B. Suerfu, A. Suzuki, V. Velan, G. Wang, Y. Wang, S.L. Watkins, M.R. Williams, and J.K. Wuko. Demonstration of the herald superfluid helium detector concept. *Physical Review D*, 110(7), October 2024.

[20] T. M. Ito and G. M. Seidel. Scintillation of liquid helium for low-energy nuclear recoils. *Phys. Rev. C*, 88:025805, Aug 2013.

[21] W. Guo and D. N. McKinsey. Concept for a dark matter detector using liquid helium-4. *Phys. Rev. D*, 87:115001, Jun 2013.

[22] S. Agostinelli and Allison et al. Geant4—a simulation toolkit, 2003.

[23] F. Toschi. Delight_mc, 2025. DELight internal note, Accessed 2026 Febuary 12.

[24] Ester-M. Kriegel HP. et al. Sander, J. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. pages 169–194, 1998. Data Mining and Knowledge Discovery 2.

[25] C. Zhang, W. Huang, T. Niu, Z. Liu, G. Li, and D. Cao. Review of clustering technology and its application in coordinating vehicle subsystems. *Automotive Innovation*, 01 2023.

[26] F. Toschi. Spass: Signal processing and clustering in superfluid helium, 2025. DELight internal note, Accessed 2026 Febuary 12.

[27] L. Cerrito. *Radiation and Detectors*. 01 2017.

[28] Z. W. Birnbaum. Numerical tabulation of the distribution of kolmogorov's statistic for finite sample size. *Journal of the American Statistical Association*, 47(259):425–441, 1952.

[29] B. Fuglede and F. Topsøe. Jensen-shannon divergence and hilbert space embedding. *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, pages 31–, 2004.

[30] G. M. Seidel, T. M. Ito, A. Ghosh, and B. Sethumadhavan. Charge distribution about an ionizing electron track in liquid helium. *Phys. Rev. C*, 89:025808, Feb 2014.

[31] NIST,In this database, it is possible to obtain photon cross section data for a single element, compound, or mixture (a combination of elements and compounds), accessed 2025 december 15th.

Acknowledgement

First, I would like to thank Prof. Belina von Krosigk for giving me the opportunity to write my Bachelor's thesis within her group and for making this work possible. I am especially grateful to Dr. Francesco Toschi for his constant support and patience to answer my many questions throughout the course of this project, as well as for carefully proof-reading this thesis. I would also like to thank Prof. Monica Dunford for kindly agreeing to serve as my second examiner. Finally, I am very thankful to the entire group for the welcoming atmosphere, the helpful discussions, and the support I received whenever questions arose.

Declaration

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.



Heidelberg, den 28.02.2026

A Appendix

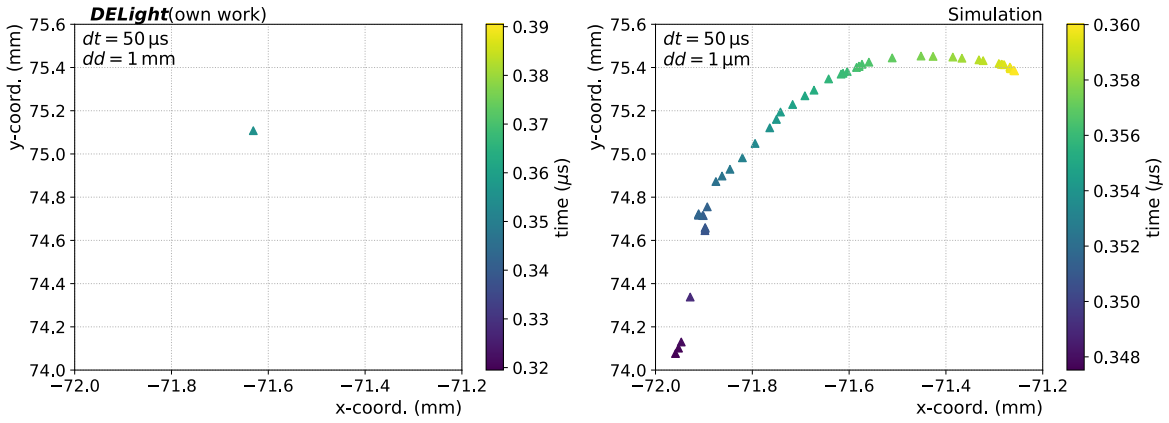


Figure A.1: Energy deposition of different temporal clustering parameters, demonstrating the working principle

B Appendix

Table B.1: Kolmogorov-Smirnov and Jensen-Shannon distance for all distributions (spatial clustering)

Simulation	KSD	JSD
Gamma ($dd = 1\,\mu m$)	0.395	0.410
Gamma ($dd = 100\,\mu m$)	0.989	0.980
Gamma ($dd = 1\,cm$)	0.961	0.925
Neutron ($dd = 1\,\mathrm{fm}$)	0.712	0.693
Neutron ($dd = 1\,\mathrm{nm}$)	0.712	0.691
^{235}U ($dd = 1\,\mathrm{fm}$)	0.281	0.290
^{235}U ($dd = 1\,\mathrm{nm}$)	0.282	0.293

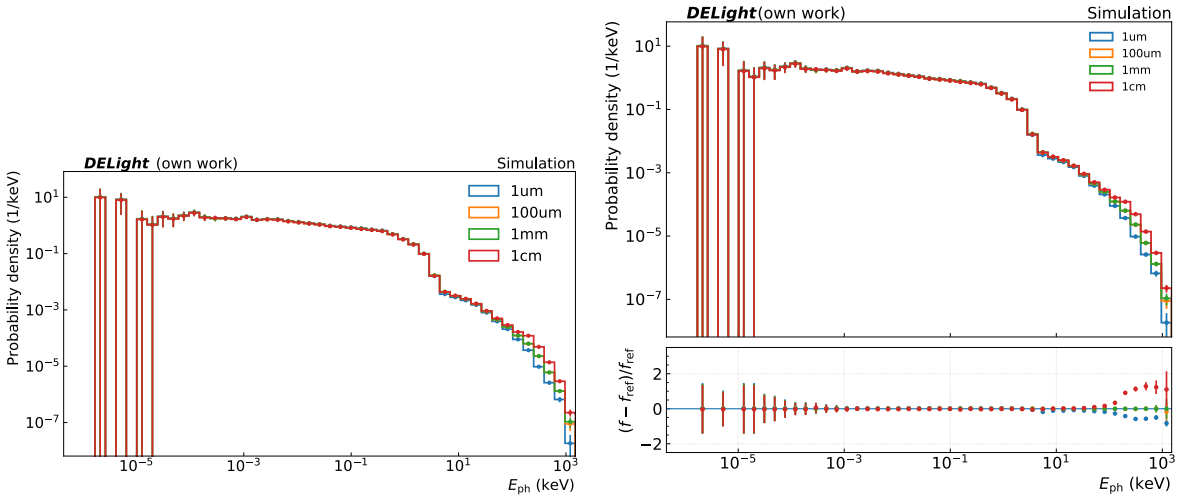


Figure B.1: Gamma simulation: (left) phonon channel SPASS output, (right) phonon channel observable output, energy distribution

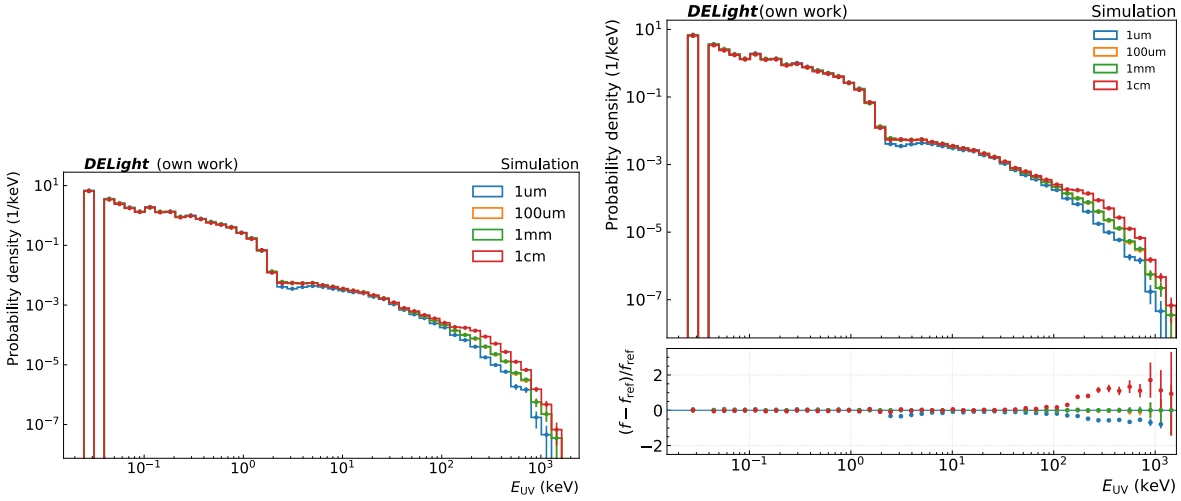


Figure B.2: Gamma simulation: (left) UV channel SPASS output, (right) UV channel observable output, energy distribution

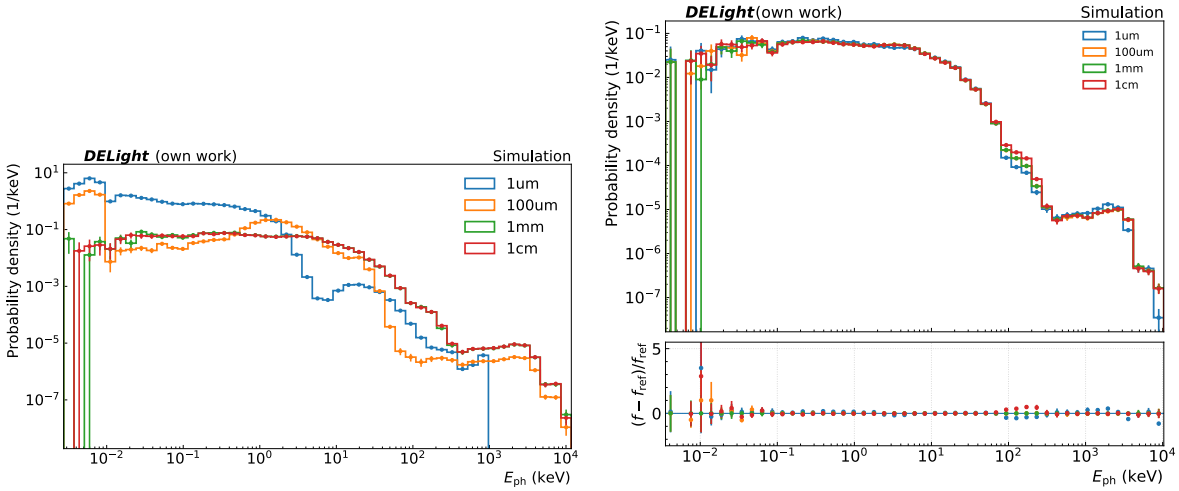


Figure B.3: ^{235}U simulation: (left) phonon channel SPASS output, (right) phonon channel observable output, energy distribution

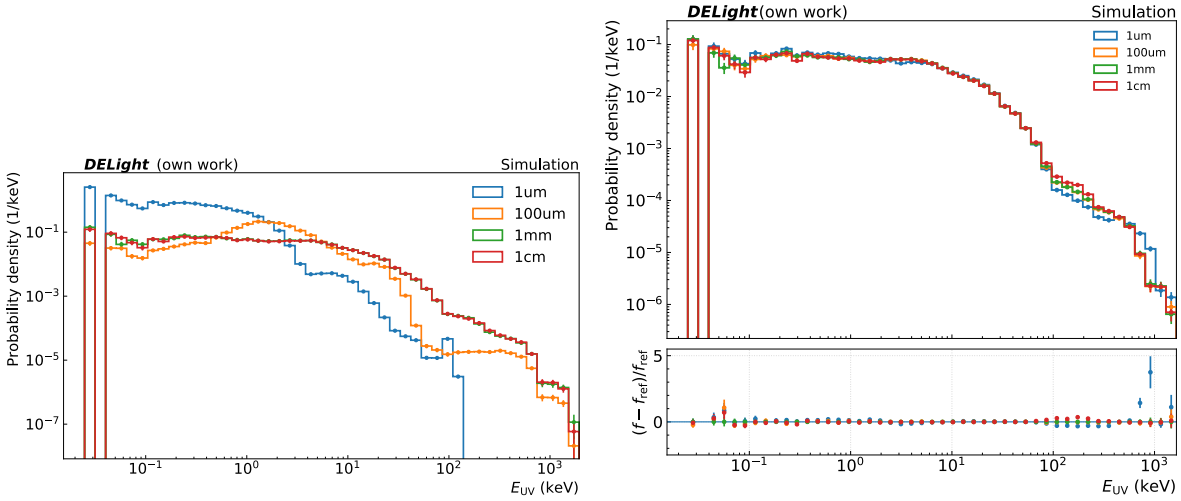


Figure B.4: ^{235}U simulation: (left) UV channel SPASS output, (right) UV channel observable output, energy distribution

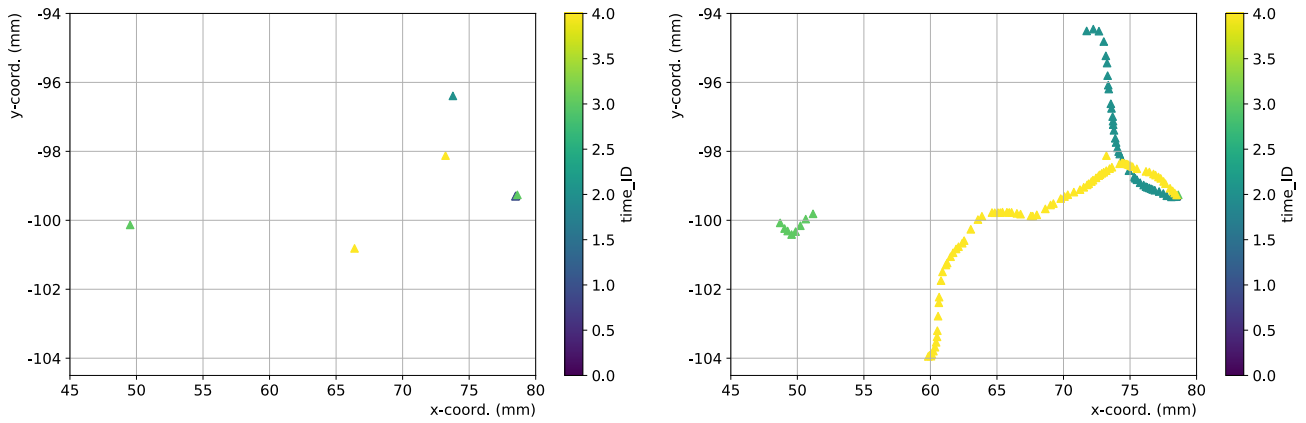


Figure B.5: Resulting clustered energy depositions, with clustering parameters $dd = 1\,\text{mm}$, $dt = 50\,\mu\text{s}$ (left) and clustering parameters $dd = 100\,\mu\text{m}$, $dt = 50\,\mu\text{s}$ (right)

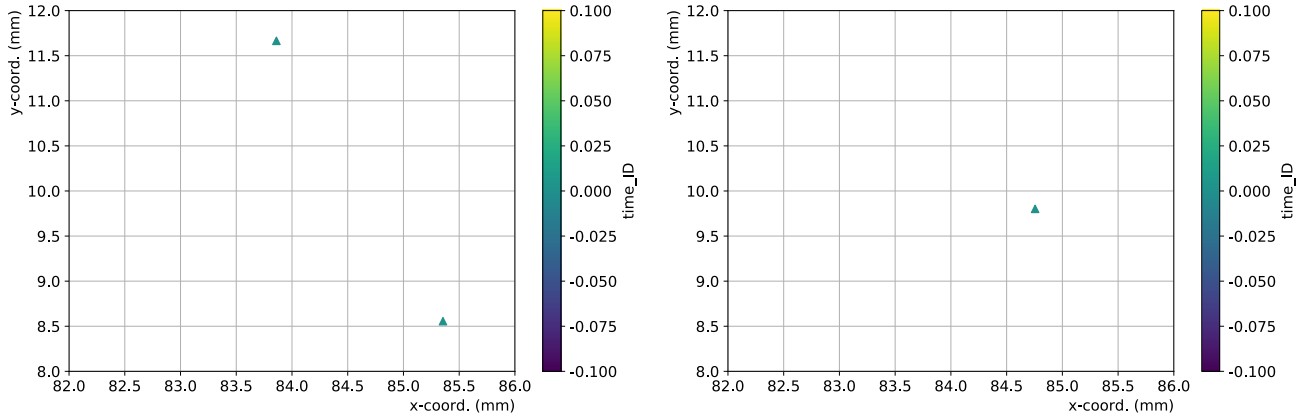


Figure B.6: Resulting clustered energy depositions, with clustering parameters $dd = 1\,\text{mm}$, $dt = 50\,\mu\text{s}$ (left) and clustering parameters $dd = 1\,\text{cm}$, $dt = 50\,\mu\text{s}$ (right)

C Appendix

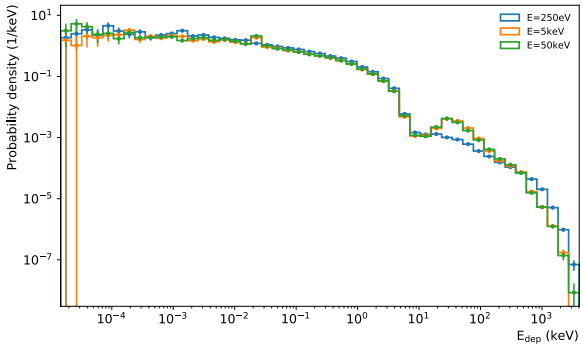


Figure C.1: neutron simulation: Energy deposition of different cut-off energies

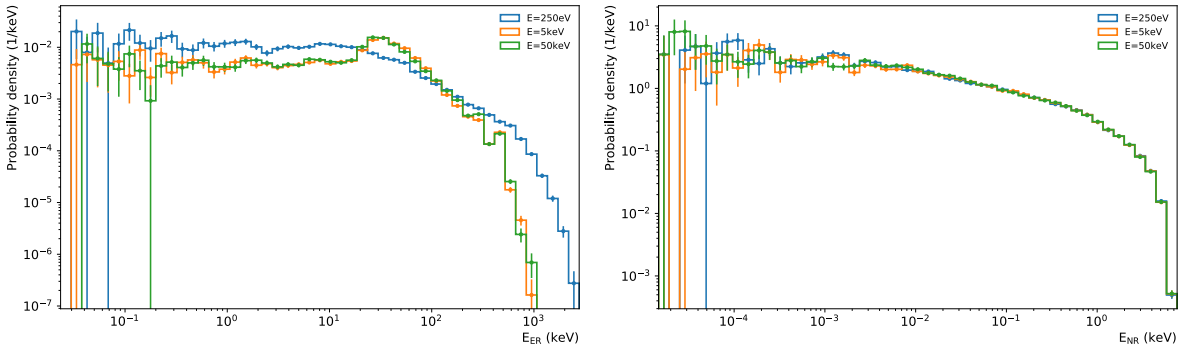


Figure C.2: Neutron simulation: Energy deposition of ER channel with different cut-off energies (left), NR channel (right)

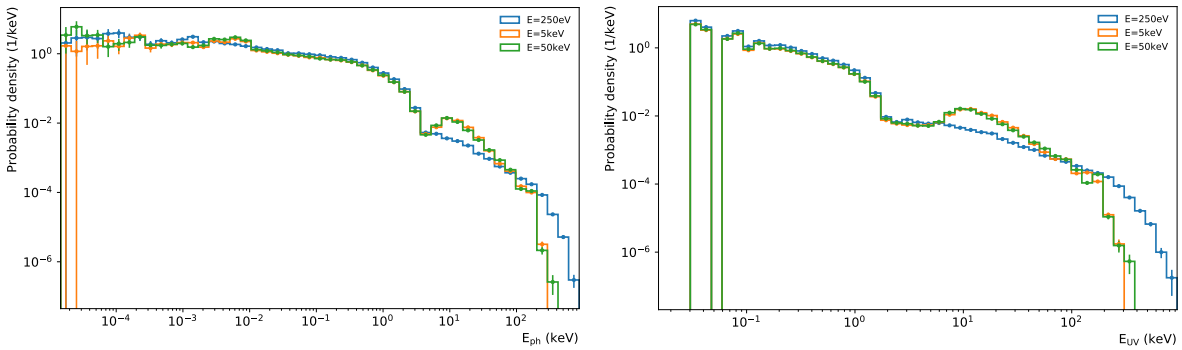


Figure C.3: Neutron simulation: Energy deposition of phonon channel with different cut-off energies (left), UV channel (right)