



SAPIENZA
UNIVERSITÀ DI ROMA

Studies of the performance of the CYGNO experiment in low-energy nuclear recoils detection

Facoltà di Scienze Matematiche Fisiche e Naturali
Corso di Laurea Magistrale in Fisica

Luca Zappaterra
ID number 1803126

Advisor
Dr. Davide Pinci

Academic Year 2023/2024



Thesis defended on 28 October 2024
in front of a Board of Examiners composed by:

Prof. Roberto Di Leonardo (chairman)

Prof. Claudio Conti

Prof. Giovanni Organtini

Prof. Francesco Zamponi

Prof. Marco Drago

Prof. Ugo Giuseppe Aglietti

Prof. Fabio Columbro

Studies of the performance of the CYGNO experiment in low-energy nuclear recoils detection

Sapienza University of Rome

© 2024 Luca Zappaterra. All rights reserved

This thesis has been typeset by L^AT_EX and the Sapthesis class.

Author's email: luca.zappaterra@me.com

Abstract

The majority of the Universe remains unseen, with over 95% of its composition attributed to Dark Matter and Dark Energy. The Λ CDM model, which includes the cosmological constant (Λ) and Cold Dark Matter (CDM), is the prevailing framework explaining this unseen mass. Within this model, Weakly Interactive Massive Particles (WIMPs) are a leading candidate for Dark Matter, largely due to the so-called WIMP miracle, which aligns observed relic densities with those predicted by thermal freeze-out scenarios.

This thesis contributes to the ongoing search for Dark Matter by analyzing data from the LIME prototype detector, a Time Projection Chamber (TPC) with an optical readout, developed under the CYGNO experiment. This detector is designed to overcome limitations posed by the neutrino floor through directional detection, exploiting the potential anisotropy of WIMP interactions that could manifest as an apparent "WIMP wind" from the Cygnus constellation.

The analysis focuses on low-energy nuclear recoil events induced by an Americium-Beryllium neutron source. It includes data pre-processing techniques such as normalization, gain correction, and light yield calibration. Additionally, the study explores directionality measurements and Monte Carlo simulations to validate the detection technique. The thesis concludes with insights into 3D range reconstruction of nuclear recoils, providing early evaluations of the Lime detector's performance in identifying potential WIMP signatures.

Contents

Introduction	vii
1 Dark Matter: from evidences to direct detection	1
1.1 Visible matter	1
1.2 Astrophysical evidences of Dark matter	2
1.2.1 Galaxy clusters	2
1.2.2 Gravitational lensing and Bullet cluster	5
1.2.3 Galactic rotational curves	5
1.3 Cosmological evidences of Dark matter	8
1.3.1 Cosmic Microwave Background	8
1.3.2 Large scale structures	9
1.4 Dark matter candidates	11
1.4.1 Generalities	11
1.4.2 The WIMP miracle	13
1.5 Direct detection of dark matter	16
1.5.1 Standard Halo Model	16
1.5.2 WIMP - Nucleus scattering cross section	18
1.5.3 Dark matter event rate	18
1.5.4 Annual modulation and directional approach	19
1.5.5 Current experimental status	23
2 The CYGNO Project	25
2.1 Time Projection Chambers	25
2.1.1 Working principles	25
2.1.2 Directional TPCs	26
2.2 CYGNO approach and technology	28
2.2.1 Optical readout	28
2.2.2 GEM amplification stage	30
2.2.3 Gas mixture	31
2.3 The LIME prototype	34
2.3.1 Vessel and field cage	36
2.3.2 Optical readout	37
2.3.3 Voltage supply and gas flow	38
2.3.4 DAQ and triggering	39
3 Data reconstruction and pre-processing	43
3.1 Image reconstruction algorithm	43
3.1.1 Noise reduction	44
3.1.2 Vignetting correction	45
3.1.3 iDBSCAN seeding	45

3.1.4	Directional super-clustering step	46
3.1.5	Super-clusters relevant features	48
3.2	PMT data integration	49
3.3	AmBe data analysis and pre-processing techniques	51
3.3.1	Data normalisation	54
3.3.2	Gain non-uniformity correction	54
3.3.3	Light Yield stability and calibration	58
3.4	Pre-processed energy spectra	62
4	Low energy nuclear recoils analysis and directionality	67
4.1	AmBe data taking reference frame	67
4.2	Nuclear recoils selection strategy	68
4.2.1	Mowed clusters removal	70
4.3	Nuclear recoils directionality	72
4.3.1	Clusters direction estimation and distribution	72
4.3.2	Monte Carlo validation	75
4.4	First attempt to nuclear recoils 3D reconstruction	78
4.5	Final remarks	87
	Conclusions	91
A	Validating the gain non-uniformity correction map	93
A.1	Strategy	93
A.2	Daily calibration tests	93
B	Elastic scattering kinematics from the angular point-of-view	97
	Bibliography	113

Introduction

When we look at the sky, pretending to see the Universe, we are actually looking at a really small fraction of it.

In fact, there are a number of evidences which lead the scientific community to the conclusion that more than 95% of the Universe is *not visible*. This introduces the concepts of Dark Matter and Dark Energy, which constitute the 25.9% and 69.1% of the Universe, respectively. These numbers are commonly associated to the Λ CDM cosmological model, which is the simplest model in agreement with the observations.

In particular, Λ recalls the *cosmological constant*, which regulates the theory behind the Universe accelerated expansion, while CDM stands for *Cold Dark Matter*, which is the model adopted to describe the dark matter component of the Universe.

This is one of the most outstanding facts of the last century and the search for the best Dark Matter candidate is still one of the most interesting research field in particle physics.

Note that the temperature concept applied to particles' density distributions refers to their mean kinetic energy. In fact, in CDM, *cold* refers to the idea that dark matter is composed by particles with very small kinetic energy, which means almost stationary particles, distributed today following a certain *relic density* originated from the dilution of a hotter density, which described the dark matter distribution just after the Big Bang, by means of Universe expansion (typically known as *thermal freeze-out*).

Between the plethora of possible candidates, nowadays searched in laboratories all around the world, the Weakly Interactive Massive Particle (WIMP) hypothesis, proposed near the '80s, is one of the most studied. To use technical terminology: its cross-section vs energy phase-space (hereafter only *phase-space*) is the most constrained one.

One of the most important motivation behind this incredible interest is the so-called *WIMP miracle*, i.e. the coincidence between the observed relic density of dark matter and the one calculated on the hypothesis of the thermal freeze-out of a weakly interactive particle with mass near the *weak scale*, $\mathcal{O}(1 \text{ GeV} - 10 \text{ TeV})$.

These kind of elusive candidates are typically studied by means of very large detectors, to maximise the exposed volume (hence the interactions), generally positioned underground, to minimise the cosmic rays' background.

In this context, a very popular kind of detector is the Time Projection Chamber, where charged particles or recoiling nuclei, which travel across the gas inside the detector, are able to produce a cloud of electrons that are drifted towards the anode thanks to an almost constant electric field provided by a potential difference applied to a field cage.

WIMPs are supposed to only interact via elastic scattering, producing nuclear recoils in the keV range when interacting in the gas inside the TPCs.

Up to now, the leading TPCs are able to capture only the energy deposit released by the tracks produced inside the detector and this is good enough to explore

the phase-space, at least until the so-called *neutrino floor* is reached. In fact, the neutrino floor (or neutrino fog) is a region of the phase-space where the nuclear recoils produced from WIMPs or neutrinos are indistinguishable.

To overcome this issue, a new kind of detector has been proposed in the recent years, based on the concept of *directional detection*, i.e. with the idea to exploit the direction of collected tracks in order to identify the astrophysical source which produced the interaction, if any. A possible technique is the implementation of an optical readout featuring an Active Pixel Sensor CMOS camera to obtain 2D information about a certain track from photos and a set of photomultiplier tubes (PMTs) to obtain timing information and close the 3D reconstruction, after the track is associated to the correct set of waveforms.

The reason behind this idea resides in the possibility that an apparent WIMP wind coming from the Cygnus constellation could be experienced from our detectors on Earth, since our solar system follows the Cygnus constellation while revolving around the galactic center, moving inside a cold WIMP dark matter halo.

In this thesis, I report my contribution to the analysis of data acquired from the LIME prototype detector of the CYGNO experiment, whose purpose is exactly the one described in the last paragraphs, with the aim of providing conclusions on the performance in low-energy nuclear recoils detection.

In chapter 1, I describe the main cosmological knowledge required to understand the theoretical context of the thesis, starting from the concept of visible matter and then reviewing the astrophysical and cosmological evidences of dark matter, proceeding with a focus on the WIMP candidates and concluding with an overview of the main concepts of experimental techniques in dark matter direct detection.

In chapter 2, after a short preamble on the time projection chambers' working principles, I describe the CYGNO project approach and technology, together with the specifications of the LIME prototype, a gaseous TPC with optical readout.

In chapter 3, I first describe the algorithm for data reconstruction and feature extraction for the reconstructed clusters. Then, the focus is shifted on the first part of my analysis, oriented to the energy data pre-processing for a dataset taken with an Americium-Beryllium neutron source, treating in particular data normalisation, gain non-uniformity correction and light yield calibration.

Finally, in chapter 4, I discuss the second part of my analysis, focused on a selected sample of low-energy nuclear recoils induced from the neutron source. First of all, I describe the sample selection strategy, then I provide a first approach to directionality measurements, together with a Monte Carlo simulation to validate the technique. In the end, I discuss how to reconstruct the 3D length using *lonely* nuclear recoils from the sample, which are easily associated with the corresponding data from PMTs, providing the first insights on possible results and problems.

Chapter 1

Dark Matter: from evidences to direct detection

As already stated in the introduction, 95% of the Universe is made of Dark Matter and Dark Energy [1]. The characterisation of the particle nature of dark matter is one of the most outstanding mysteries of the whole physics field.

In the following, our interest will be focused on the dark matter component of the Universe, in particular on WIMP **Cold¹ Dark Matter** (CDM) candidates.

It is hence useful to review some key aspects and concept of this research field, starting from the astrophysical and cosmological evidences of dark matter, moving to the description of the so-called Weakly Interactive Massive Particles and giving an overview of the experimental status in direct dark matter detection, with a special focus on the expected dark matter event rate and its annual modulation.

1.1 Visible matter

Trying to catch relevant information about our Universe by looking at the sky is a very complicated task, but there are a lot of friendly objects which helped us in this mission from the beginning of the astronomical research: someone named them “*stars*”.

Stars emit light in a really broad range of wavelength, mostly in the infrared, visible and ultraviolet range of the electromagnetic spectrum.

We can define the concept of **visible matter** and, simultaneously, of **visible mass** as the matter/mass that can be inferred by looking at the Universe using a V-band filter, which allows the telescope to detect photons in the wavelength range $500 \text{ nm} < \lambda < 590 \text{ nm}$. In this context, the “V” stands for “visual”.

As a reference, when looking at the Sun with a V-band filter, we get a luminosity of roughly 12% of the total Sun luminosity, namely $L_{\odot,V} = 0.12L_{\odot}$. With this in mind, if we survey the galaxies in the local Universe (up to $d \sim 0.1c/H_0$ where $H_0 \approx 70 \text{ km/s} \cdot \text{Mpc}$ is the Hubble constant) we can manage to extract the luminosity density in the V-band [2]:

$$\Psi_V = 1.1 \times 10^8 L_{\odot,V} \text{ Mpc}^{-3}$$

From this, knowing the mean *mass-to-light* ratio $\langle M/L_V \rangle$ of the local Universe, we can compute the mass density. In particular, since the local Universe is a mixture

¹we will elaborate more on the meaning of “cold” later in this chapter.

of quiescent and star-forming region, we can safely take the mass-to-light ratio to approximately be $\langle M/L_V \rangle \approx 4 \times M_\odot/L_{\odot,V}$.

The resulting mass density of stars emitting in the V-band in the local Universe today (where “today” is captured by the “0” subscript) is then:

$$\rho_{*,0} = \langle M/L_V \rangle \Psi_V \approx 4 \times 10^8 \text{ Mpc}^{-3}$$

Remembering the concept of critical density (labeled with “c”), namely the matter density of a spatially flat Universe, together with the concept of density parameter, usually indicated by $\Omega = \rho/\rho_c$, one can define the density parameter of stars today:

$$\Omega_{*,0} = \frac{\rho_{*,0}}{\rho_{c,0}} \approx 0.003$$

This number is telling us that *the stars in the visual band are making up only the 0.3% of the matter density required to flatten the Universe!*

If we now compare this number with the BBN estimate of the baryonic matter density parameter from [3], namely $\Omega_{B,0} \approx 0.04$, we deduce that the stars’ contribution to baryonic matter is approximately the 8%!

Of course, we are making a big approximation.

Nonetheless, even broadening our definition of “stars” including objects like neutron stars, black holes, stellar remnants, white dwarfs etc., one ends up with $\Omega_{*,0} < 0.005$.

Most of the baryonic matter inside galaxies is, in fact, in the form of hot interstellar gas which emits an extremely low amount of (visible) light.

An example of this can be seen from the comparison between figure 1.1 and figure 1.2, from which we can infer the inter-cluster gas distribution in the Coma Cluster from its X-ray emission.

We demonstrated only a part of the opening statement of this subsection, the reality is far more intriguing: not only most of the matter inside the Universe is faintly emitting in the visible band, *a large part of it is not even baryonic!*



Figure 1.1. The Coma Cluster in the visible range [SkyCenter/University of Arizona].



Figure 1.2. The Coma Cluster in the x-ray range [ESA].

1.2 Astrophysical evidences of Dark matter

1.2.1 Galaxy clusters

It was 1933 when Fritz Zwicky observed a very large radial velocity dispersion in the Coma Cluster. It was an incredibly accurate measurement confirmed after almost 50 years from an independent work [4].

This dispersion, of $\mathcal{O}(10^3)$ km/s [5], was too high to be explained by means of the visible mass only.

This was the first evidence of, to cite Zwicky himself, *Dunkle Materie*, but it was not taken too seriously until almost 40 years later, with the measurement of M31's rotational curves from Vera Rubin, but we will come to that in section 1.2.3.

If we once again look at figure 1.2 we can clearly see the confinement of the inter-cluster gas in the inner regions of the Coma Cluster, this is consistent with the hypothesis of a distribution of invisible matter which modifies the gravitational balance in this region.

To say this in other words: *if there were no dark matter to produce the gas confinement, it would have expanded (due to the Universe expansion) beyond the cluster limits on time scales much shorter than the Hubble time!*

Cluster dynamics and virial theorem results

Clusters of galaxies are gravitationally bound objects, meaning that they do not expand with the same rate as the Universe around them.

Hence, treating the cluster as a rigid body, one can study the system with a purely Newtonian approach and apply the virial theorem to deduce the mass of the cluster.

Consider a cluster composed by N galaxies, each of which approximated as a point mass with mass m_i and $i \in [1, N]$.

Let $\vec{x}_i$ and $\dot{\vec{x}}_i$ be, respectively, the i -th galaxy's coordinate and velocity, from purely Newtonian arguments we can derive the potential energy W of the system as:

$$W = -\frac{G}{2} \sum_{j \neq i} \frac{m_i m_j}{|\vec{x}_j - \vec{x}_i|} \quad (1.1)$$

namely, the energy required to pull the N galaxies up to an infinite distance from each other.

It is useful to rewrite equation 1.1 in the following way:

$$W = -\alpha \frac{GM^2}{r_h} \quad (1.2)$$

where we defined $M = \sum_i m_i$ as the total mass of the galaxies in the cluster, r_h as the radius containing half the mass of the cluster and where α is an $\mathcal{O}(1)$ parameter which depends on the cluster's density profile.

In particular, $\alpha \approx 0.45$ gives a good fit to the potential energy for observed clusters.

Defining $\langle v^2 \rangle \equiv \frac{1}{M} \sum_i m_i |\dot{\vec{x}}_i|^2$ as the mean square velocity of all the galaxies in the cluster, we can write the kinetic energy K as follows:

$$K = \frac{1}{2} M \langle v^2 \rangle \quad (1.3)$$

We can now invoke the virial theorem, which in its most general form reads as:

$$\ddot{I} = 2W + 4K \quad (1.4)$$

and remembering our initial assumptions we can remove the extra complication introduced by the moment of inertia, since a not expanding nor contracting system has $I = \text{const.}$

Hence, we are left with a simplified version of the virial theorem which reads as:

$$K = -\frac{W}{2} \quad (1.5)$$

Combining now the virial theorem with equations 1.2 and 1.3, with some little algebraic steps we can compute the mass of our system, i.e.:

$$M = \frac{\langle v^2 \rangle r_h}{\alpha G} \quad (1.6)$$

At this point we can follow Zwicky's reasoning applying equation 1.6 to the Coma cluster. To do so we need an estimate of the mean squared velocity and half-mass radius:

- for the mean squared velocity, we can take first of all the 1D velocity dispersion, along the Earth line of sight, to be $\sigma_r = 880 \text{ km/s}$ [6]. Assuming this dispersion as isotropic, we can compute the mean squared velocity simply as:

$$\langle v^2 \rangle = 3\sigma_r^2 \approx 2.32 \times 10^{12} \frac{\text{m}^2}{\text{s}^2}$$

- for the half-mass radius we can rely on [2], which reports:

$$r_h \approx 1.5 \text{ Mpc} \approx 4.6 \times 10^{22} \text{ m}$$

Since $G = 6.67 \times 10^{-11} \text{ m}^3 \text{ s}^{-2} \text{ kg}^{-1}$ and, as already stated, $\alpha \approx 0.45$ is a good fit to observations, we end up with:

$$M_{\text{Coma}} \approx 2 \times 10^{15} M_{\odot} \quad (1.7)$$

where $M_{\odot}$ is the Solar mass.

It is now important to compare this results with the estimates of visible mass inside the Coma cluster. In particular:

- for the stars it is found $M_{\text{Coma},\star} \approx 2 \times 10^{13} M_{\odot}$, which is only 1% of the computed virial mass;
- for the hot inter-cluster gas we have instead $M_{\text{Coma,gas}} \approx 2 \times 10^{14} M_{\odot}$, roughly 10% of the total.

The numbers are crystal clear: *almost 90% of the Coma cluster virial mass is invisible to current detection techniques.*

An example of independent confirmation of the Coma virial mass is reported in [7], which computed a similar value studying the weak lensing effect produced from the cluster.

Moreover, one could apply the same logic to a large number of clusters of galaxies to provide an estimate of their virial mass. This in turn provides an estimate of their density parameter, namely $\Omega_{\text{clus},0} \approx 0.2$ [2], which can be seen as a lower limit to the matter density of the Universe.

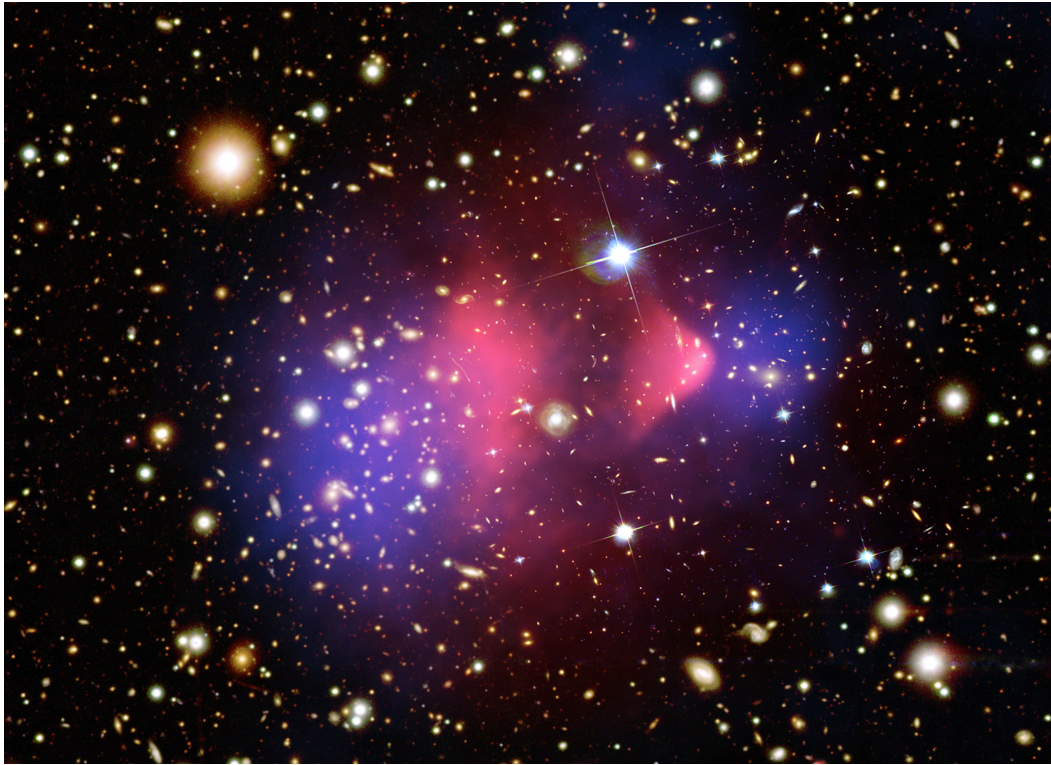


Figure 1.3. A composite image of the Bullet Cluster, **in pink** the X-Ray emission from the inter-cluster hot gas (slowed down from the collision), **in blue** the computed matter distribution required to produce the observed lensing effect. [ESA].

1.2.2 Gravitational lensing and Bullet cluster

Since we just cited the gravitational lensing effect, it is interesting to note the importance of the lensing produced by the so-called “Bullet cluster” (which can be admired in figure 1.3) in the context of dark matter evidences.

The Bullet cluster was produced by the collision between two clusters of galaxies and it is evident, from the image, that the galaxies did not collide with each other.

On the contrary, the hot inter-cluster gas of one cluster did collide with the other, slowing down and emitting X-Rays (whose distribution is highlighted in pink).

When looking at the Bullet cluster it is possible to measure a certain lensing effect and one can compute the mass distribution required in order for it to be possible. This mass distribution is highlighted in blue in the image.

The point is that the inter-cluster gas contribution to the cluster total mass is found to be far larger than the one relative to galaxies, as seen in the Coma cluster, hence one should expect the mass distribution to follow the gas. Nonetheless, the contrary is observed! This is considered a strong evidence of the existence of dark matter [8].

1.2.3 Galactic rotational curves

Let’s start from a simple, reasonable assumption: *starlight is a good tracer for matter*. It follows that where we observe more light, e.g. in galaxies’ cores, we expect more matter.

Take now a spiral galaxy as a case of study. Near its core there will of course be brighter regions, hence more matter (which in turn means a stronger gravitational potential), with respect to the outer regions of the galaxy.

Thus, it is pretty straightforward to assume that, since spiral galaxies spins around their core, stars in the outer regions should move slower with respect to the inner ones.

Nonetheless, as Vera Rubin's work highlighted back in 1970 [9], Nature did not agree on this.

To be more formal, Rubin's idea was to measure the speed of the stars inside the Andromeda galaxy. To perform such a measure, one can exploit the HI regions' emission, where the famous "21 cm line" is largely produced from the spin-flip of electrically neutral Hydrogen atoms, to extract the stars' velocity using the observed Doppler shift (an example can be seen in figure 1.5, where this concept is applied to M33).

These measurements provide what is known as *galactic rotational curves*.

In principle, from Newtonian dynamics, we expect that

$$v(R) = \sqrt{\frac{GM(R)}{R}}$$

where G is the gravitational constant, R is the distance from the galactic core and $M(R)$ is the (visible) mass enclosed inside a given R .

Hence, when R is very large, we expect $v(R) \propto 1/\sqrt{R}$, since $M(R)$ will essentially be constant.

Observations showed that this was not the case. In fact, what Rubin [9] observed was a flattening of the rotational curve in M31 for $R \gg 1$.

For the sake of aesthetics, it is interesting to report a very nice plot (figure 1.6) from another work which did the same reasoning for the NGC 3198 galaxy [10]. From this we can see the superimposition of the observed data with the predicted behaviour one should expect when considering the visible mass only (labeled as "disk").

Moreover, the plot provides the computed rotational curve contribution from the so-called "dark matter halo", i.e. an hypothetical halo of gravitational-only interacting non-baryonic matter which is supposed to be the responsible of the rotational curve flattening.

In particular, this halo should have $M_{\text{halo}} \propto R$, which in turn requires a mass density $\rho_{\text{halo}} \propto R^{-2}$, since in general $M(R) = \int_R \rho(r) dr$.



Figure 1.4. Vera Rubin operating the Kitt Peak National Observatory's telescope [NOAO/AURA/NSF].

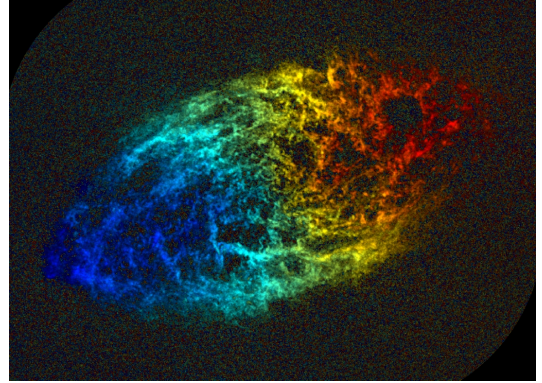


Figure 1.5. M33 Doppler shift from the 21 cm Hydrogen line, which is telling us the rotation direction of the galaxy.

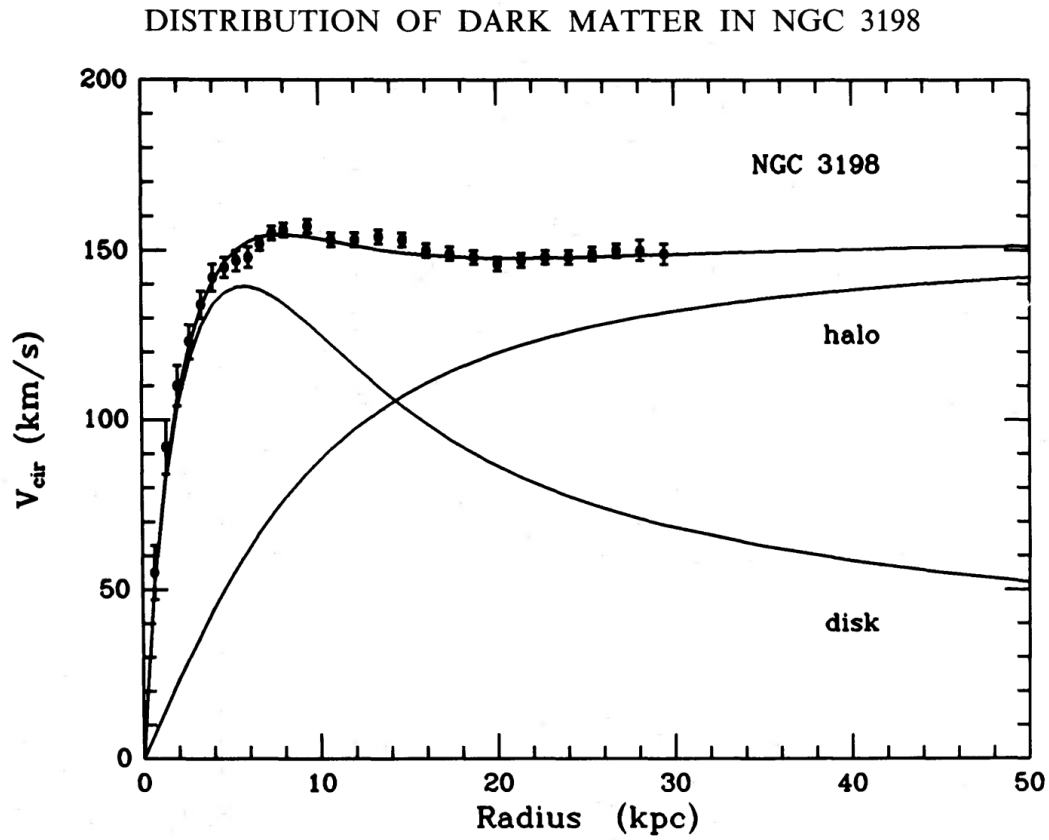


Figure 1.6. Superimposition of rotational curves for the NGC 3198 galaxy from [10]. The points are observed data, “disk” is the rotational curve expected from the visible mass alone, “halo” is the rotational curve expected from the dark matter halo alone, while the solid curve consistent with the data is the combination disk + halo.

1.3 Cosmological evidences of Dark matter

1.3.1 Cosmic Microwave Background

Moving the discussion from astrophysics to cosmology, the first evidence which should be reported is the one that can be extracted from the Cosmic Microwave Background (**CMB**).

First of all, the CMB was produced at the so-called “*time of decoupling*” or “*time of last scattering*”, which corresponds to the time when the photons decoupled from the photo-baryonic plasma existing in the early stages of the Universe, according to the Big Bang theory. In redshift terms, it is estimated to be at $z_{\text{LS}} = 1090$.

After decoupling the Universe became transparent to radiation and photons (which were thermalized, i.e. resembling a black body spectrum) became free to propagate in the surrounding expanding Universe, without any Thomson scattering to constraint them inside the plasma. The cosmic microwave background critical to understand the early Universe processes, since it is a picture of the Universe at the time of last scattering.

The Universe expansion diluted the CMB density, which in turn diluted its temperature. Today, we observe an almost isotropic radiation, up to a part every 10^5 , with a mean temperature of $T_{\text{CMB}}^0 \approx 2.73 \text{ K}$ [11].

Following [2], we can model the so-called “*temperature anisotropies*” of the CMB by expanding them in spherical harmonics, since the fluctuations are distributed over a spherical surface, i.e.:

$$\frac{\delta T}{T}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} Y_{lm}(\theta, \phi) \quad (1.8)$$

From this, we can compute the two-point correlation function to explore the spatial scale of the anisotropies, namely the average product of temperature anisotropies in two different points separated by an angle θ over the celestial sphere:

$$C(\theta) = \left\langle \frac{\delta T}{T}(\hat{n}_1), \frac{\delta T}{T}(\hat{n}_2) \right\rangle_{\hat{n}_1 \cdot \hat{n}_2 = \cos \theta} \quad (1.9)$$

which in terms of spherical harmonics becomes:

$$C(\theta) = \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) C_l P_l(\cos \theta) \quad (1.10)$$

where C_l are measures of the fluctuations over an angular scale $\theta \sim \pi/l$ and P_l are the Legendre polynomials.

We can, hence, equivalently use the multipoles and the angular scale.

Looking at the measured power spectrum produced from the Planck collaboration [12] in figure 1.7, we can have an idea of its shape. It is interesting to note that the different multipoles provide different information; in particular:

- $l = 1$ term is related to the dipole anisotropy, which is generated from the Doppler effect.
- $l \geq 2$ terms are related to the temperature anisotropies at the time of last scattering.

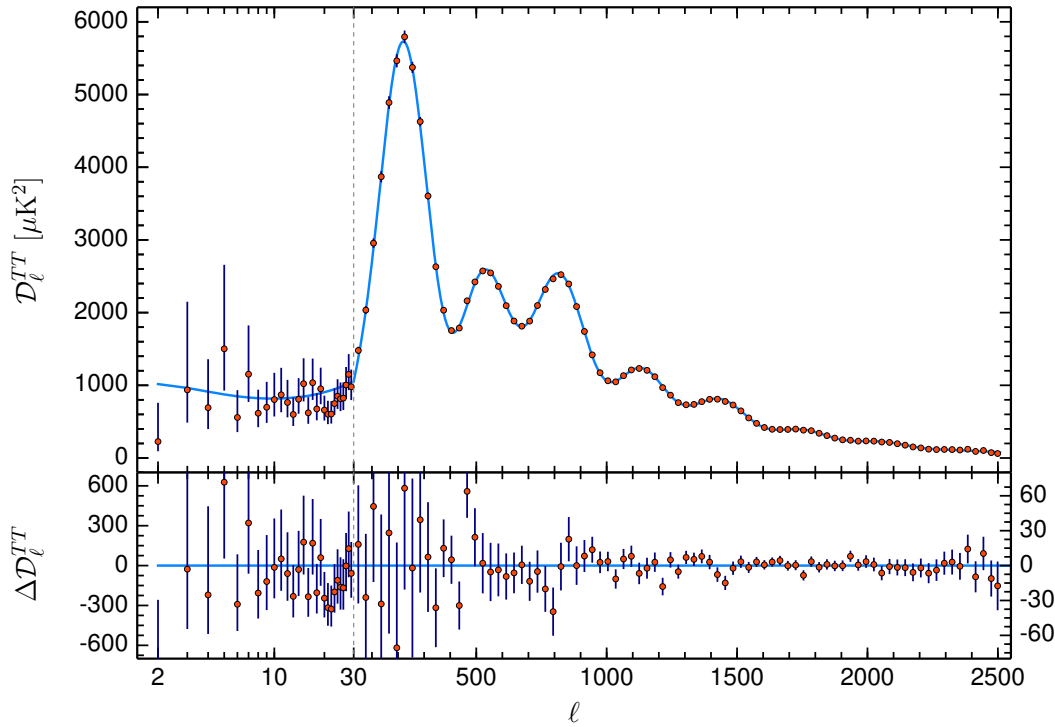


Figure 1.7. upper panel: CMB power spectrum produced from Planck2018 with data superimposed to the best prediction from the Λ CDM model. lower panel: residuals. [12]

Moreover, it is important to know that the first, larger peak we observe close to $\ell = 30$ is connected to the curvature of the Universe. In fact, it is expected to be at an angular scale $\theta \approx 0.7^\circ$ and every shift from this value provide an indication of a curved Universe.

At the moment, the observed first peak position is consistent with the expected value, and the best estimate for the total density parameter from Planck2018 [12], well reproduced and organized also in chapter 25 of PDG [13], points to $\Omega_{tot} = 0.9993 \pm 0.0019$, which is consistent with the hypothesis of a flat Universe ($\Omega_{tot} = 1$).

Another important result that we can extract from the CMB, reported in the same sources [12, 13], is the estimate for the cold dark matter (CDM) density parameter $\Omega_{CDM} \approx 0.25$.

In the context of this thesis, this is the most important result to cite, since it is telling us that, as already stated, more than 80% of the matter contribution to the energy density ($\Omega_m \approx 0.3$) is given by non-baryonic ($\Omega_b \approx 0.05$) cold dark matter.

1.3.2 Large scale structures

There is another worth-mentioning evidence on the existence of dark matter, in particular of the cold one.

This evidence come from the theory behind the mechanism of structures formation and we can restrict the discussion to the so-called “*large scale structures*”, i.e. structures like galaxies, clusters or even superclusters of galaxies.

Unfortunately, structure formation is a very large and complex topic and a whole

course could be required to deeply understand it. In the following we will see just the basic idea which is required to understand the crucial role that cold dark matter plays in this context following [2], where the discussion is taken with a more rigorous approach.

The key point is to define what is called a “*density contrast*”, δ , able to model the density of a given region of the Universe with respect to the surroundings.

By studying the equation of motion for a point on the surface of an expanding spherical region, exploiting mass conservation inside the sphere and remembering Friedmann’s acceleration equation, we end up with the equation governing the growth of small amplitude perturbations:

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\bar{\rho}\delta \quad (1.11)$$

where $H \equiv \dot{a}/a$ is the Hubble parameter, $2H\dot{\delta}$ is called *Hubble friction term*, which exists only for an expanding Universe slowing down its expansion, and $\bar{\rho}$ is the considered region’s matter density.

Equation 1.11 can be generalised deriving it with a fully relativistic calculation and rewriting the right-hand-side in terms of the matter density parameter $\Omega_m = \varepsilon_m/\varepsilon_c$:

$$\ddot{\delta} + 2H\dot{\delta} = \frac{3}{2}\Omega_m H^2 \delta \quad (1.12)$$

This equation is telling us *three* things:

- During the radiation-dominated epoch we can forget about the RHS and take $H \sim 1/t$. Solving the equation, we find that density fluctuations in this period grew only at a logarithmic rate:

$$\delta \propto \ln t$$

- On the other hand, considering a Λ -dominated epoch, Ω_m will again be negligible and we can take the Hubble parameter to be constant $H = H_\Lambda$. The solution will hence lead to a constant fractional amplitude for the matter density fluctuations:

$$\delta \approx C_1 + C_2 e^{-2H_\Lambda t}$$

- Lastly, in the matter-dominated regime (assuming $\Omega_m = 1$ and $H = \frac{2}{3t}$), the solution is a power law in t with two terms, one $\propto t^{-1}$, the other $\propto t^{2/3}$. The first one is clearly subdominant after a certain time, which leads to:

$$\delta \propto t^{2/3} \propto a(t) \propto \frac{1}{1+z}$$

as long as $|\delta| \ll 1$.

If we assume the baryonic matter as the dominant form of non-relativistic matter in the Universe, we could expect the density fluctuation growth to be started at the decoupling time ($z_{LS} = 1090$), leading to an amplitude today only ~ 1090 times greater than the starting amplitude.

Actually, since the dominant form of non-relativistic matter is cold dark matter, matter density fluctuation were able to start growing at an earlier stage of the Universe, namely the *radiation-matter equivalence* time ($z_{eq} = 3440$), when the non-baryonic dark matter decoupled from the photo-baryon plasma. Then, at the

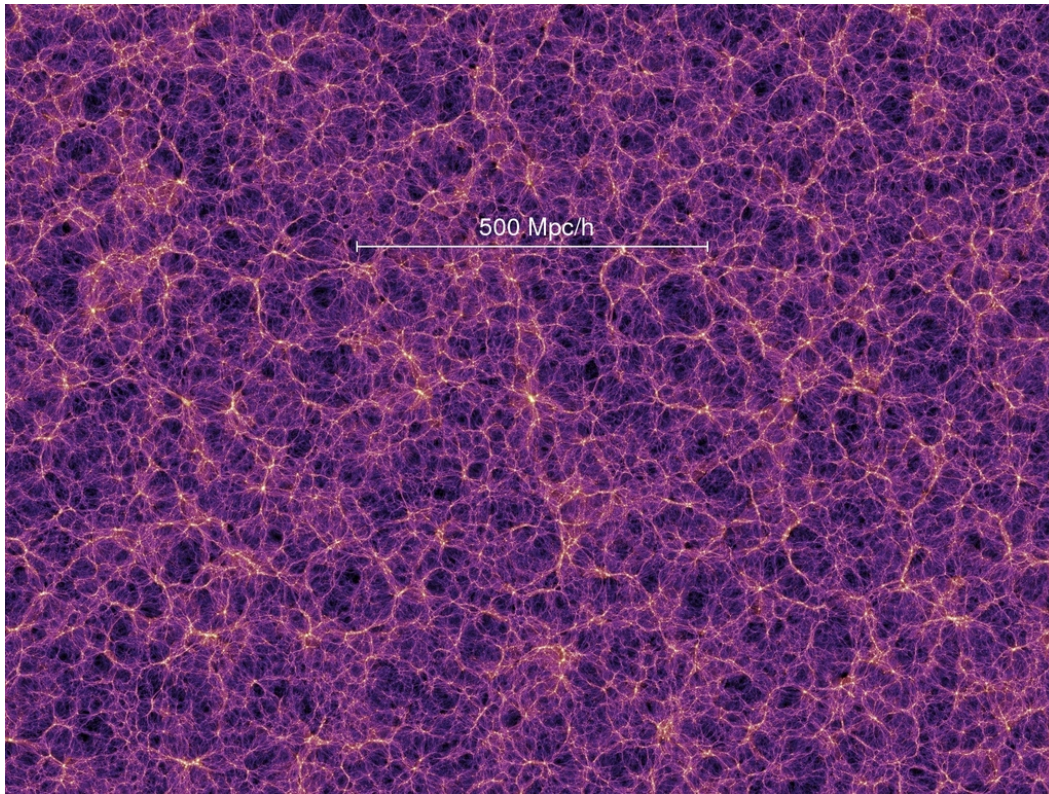


Figure 1.8. Large Scale Structures as simulated from the Millennium Project [14], also known as “*cosmic web*”.

time of decoupling, baryons fell into the gravitational wells created by cold dark matter, eventually collapsing to form structures!

The point in considering only cold dark matter as possible candidate for structure formations resides in the main characteristics of this type of dark matter, which will be formally defined in the following section.

1.4 Dark matter candidates

1.4.1 Generalities

At this point we should be convinced about the fact that the majority of the Universe is composed by an invisible constituent. It is natural to ask ourselves the question: “*what is the sub-atomic nature of this elusive matter?*”

In summary, the perfect dark matter candidate should be:

- **Non-baryonic**, since the best fit for the baryonic energy density parameter is $\Omega_b \approx 0.05$, only a small fraction of the total matter density parameter $\Omega_m \approx 0.3$.
- **At most weakly interacting** with Standard Model particles, in the sense that it does not emit, absorb or reflect light, otherwise we would have seen it.
- **With negligible self-interactions**, such that it is compatible with the constraints on structure formation theory.

ZOO OF DARK MATTER CANDIDATES

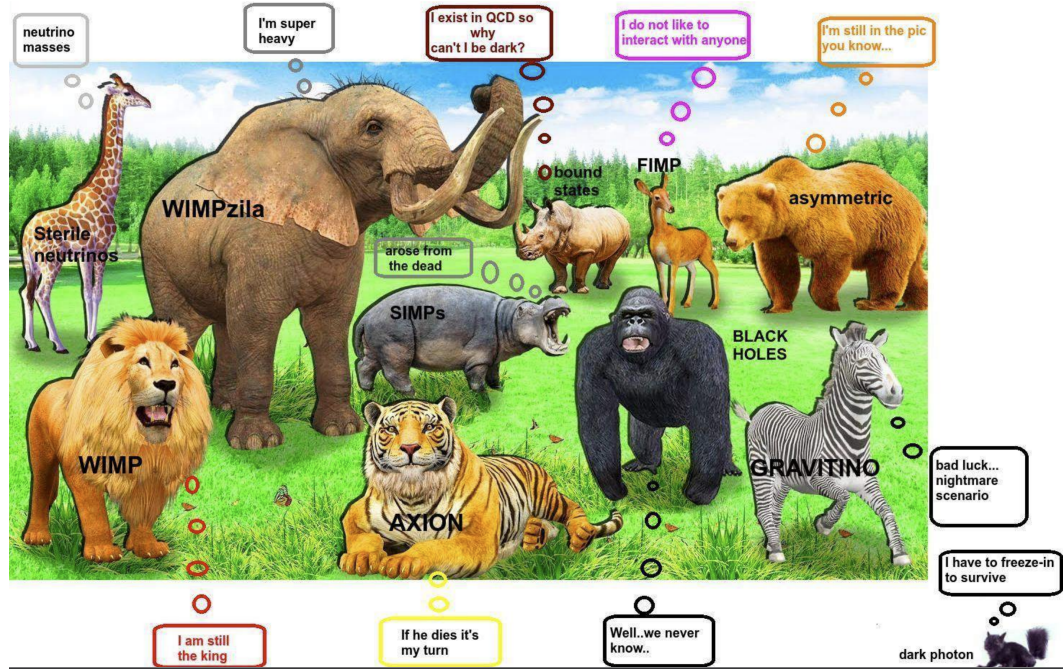


Figure 1.9. A not-in-scale overview of the most important proposed dark matter candidates [Source [Farinaldo Queiroz IIP-Natal](#)].

- **Stable**, since if not it would have already decayed.
- **Cold**, where the concept of temperature of a particle is related to its relativistic nature (cold is non-relativistic, hot is relativistic), in light of the cruciality of this argument in explaining the formation of large scale structures, as we elaborated in section 1.3.2

Note that the last point is an assumption corroborated by the fact that in a hot dark matter-dominated Universe, structures formation theory would require superclusters to form way earlier than galaxies themselves [15], and this is in contrast with astrophysical observations, which led to an immediate lost of interest in hot dark matter models.

Probably, the most rational scenario would be a mixed one. The hot dark matter hypothesis is in fact not completely rejected and a possible HDM candidate is the well-known neutrino, from which one can compute a lower limit on the HDM density parameter [2].

A lot of possible dark matter candidates have been proposed in the past years with completely different masses, interactions and temperature leading to an actual “dark matter zoo”. To have an idea about how different these candidates are, figure 1.9 shows a general picture of the main ones (not in scale), while figure 1.10 shows a more quantitative overview of the ranges for candidates’ masses, anomalies and search techniques in the dark matter field.

Anyway, since the focus of this thesis will be on a cold dark matter scenario, it is useful to expand the discussion on the characteristics of the so-called “*Weakly Interactive Massive Particles*” (**WIMPS**).

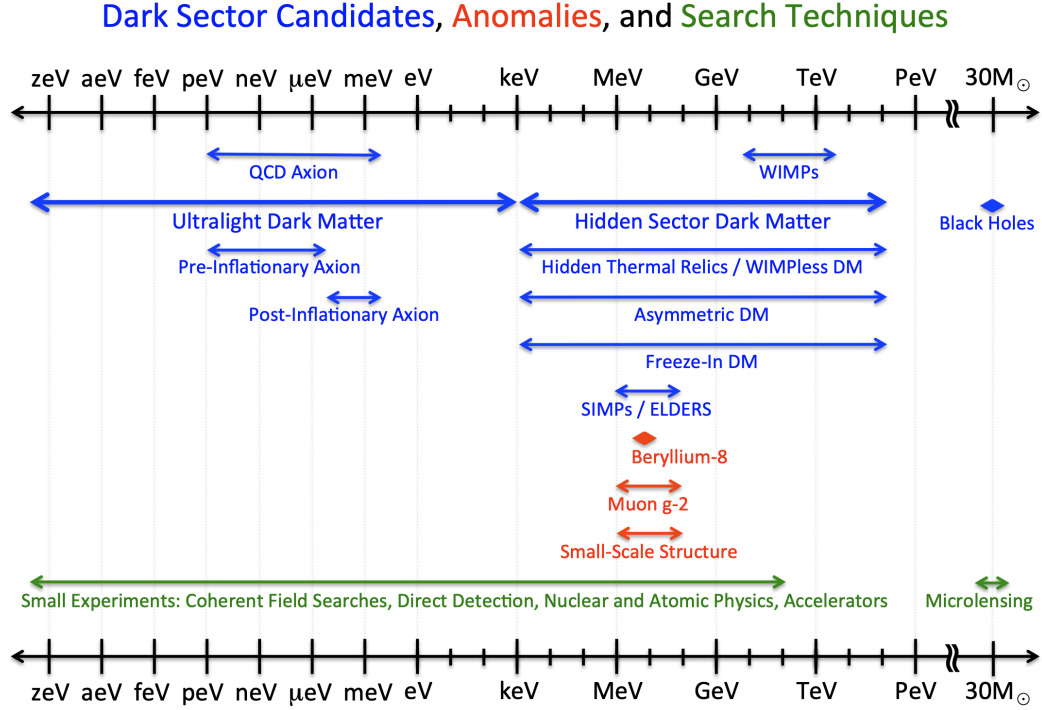


Figure 1.10. A quantitative overview of ranges for dark matter candidates masses, anomalies, and search techniques [16].

1.4.2 The WIMP miracle

Weakly Interactive Massive Particles are largely motivated by the “*WIMP miracle*” which, in a nutshell, consists in the coincidence between the observed relic density of dark matter and the one calculated on the hypothesis of the thermal freeze-out of a weakly² interactive particle with mass near the weak scale, which means $\mathcal{O}(1 \text{ GeV} - 10 \text{ TeV})$.

In order to describe more in detail the WIMP miracle, we can elaborate on the freeze-out concept.

In principle, the Universe was hot and dense, and both species from standard model (**SM**) and dark matter (**DM**) were in thermal equilibrium, with dark matter production balanced from its annihilation. The production and annihilation processes can be modeled as:

$$\chi\chi \longleftrightarrow \text{SM} \overline{\text{SM}}$$

assuming that the dark matter particle is a Majorana particle ($\chi = \bar{\chi}$) and that SM is a generic standard model particle.

Moreover, the condition of thermal equilibrium is assured by the assumption of a DM-SM interaction which enables energy exchange, e.g.:

$$\chi \text{ SM} \longleftrightarrow \text{SM} \chi$$

The Universe eventually cooled down, up to a temperature $T_f < m_\chi$ where the DM production processes became Boltzmann-suppressed, exponentially dropping

²Note that here “weakly” is describing an extremely low interaction rate with ordinary matter and it is not related anymore to the weak force as it was thought in the past.

with a factor $\propto e^{-m_\chi/T_f}$. This fact is essentially telling us that the dark matter is no longer able to be produced and can only exchange energy interacting with SM particles or annihilate interacting with DM particles.

At the same time, the universe was also expanding, up to the point where the annihilation rate became smaller than the Universe expansion rate, namely:

$$\Gamma_{\text{ann}} \leq H \quad (1.13)$$

At this point, both annihilation and production processes became inefficient, which set the number of dark matter particles to a constant value! We refer to this moment as *chemical decoupling* or *freeze-out* and T_f is known as the freeze-out temperature³.

At a later time, also energy exchange processes became inefficient, we refer to this moment as the *kinetic decoupling*.

The process that we just introduced is quantitatively described by the Boltzmann equation:

$$\frac{dn_\chi}{dt} = -3Hn_\chi - \langle\sigma_{\text{ann}}v\rangle(n_\chi^2 - n_{\chi,\text{eq}}^2) \quad (1.14)$$

where n_χ is the DM number density, H is the Hubble parameter, $\langle\sigma_{\text{ann}}v\rangle$ is the thermally averaged cross-section for the annihilation process and $n_{\chi,\text{eq}}$ is the DM number density when thermal equilibrium is verified.

It is clear that in equation 1.14 the first term accounts for the Universe expansion, the n_χ^2 term is related to the DM annihilation process, while the $n_{\chi,\text{eq}}^2$ term is related to the DM production process.

In this context, it is customary to speak about dark matter co-moving number density or dark matter yield, namely:

$$Y_\chi \equiv \frac{n_\chi}{s} \quad (1.15)$$

where $s \propto T^3$ is the entropy density.

Considering a WIMP with $m_\chi = 100$ GeV, its co-moving number density can be visualized as a function of temperature in figure 1.11, from which we recognise the initial exponential drop, eventually leading to a constant value when equation 1.13 is satisfied.

Following [17], we can produce a rough estimate of the dark matter density parameter.

First of all we can define the thermal freeze-out as the equivalence between the first and second term in equation 1.14, i.e.:

$$n_\chi^f \langle\sigma_{\text{ann}}v\rangle \approx H \quad (1.16)$$

where the superscript f denotes that we are considering the quantity at freeze-out.

We can now define the density parameter for dark matter as:

$$\Omega_{\text{CDM}} h^2 \equiv \frac{\rho_\chi^0}{\rho_c} h^2 = \left(\frac{s^0 h^2}{\rho_c} \right) m_\chi Y_\chi^0 \quad (1.17)$$

where ρ_c is the critical density and in the last equality we used the fact that $\rho_\chi^0 = m_\chi n_\chi^0$ combined with equation 1.15, indicating with the superscript 0 the quantities calculated at present time ($a = 1$).

³Note that the distinction between hot and cold dark matter is based on the fact that the dark matter candidate was relativistic or non relativistic at the freeze-out temperature

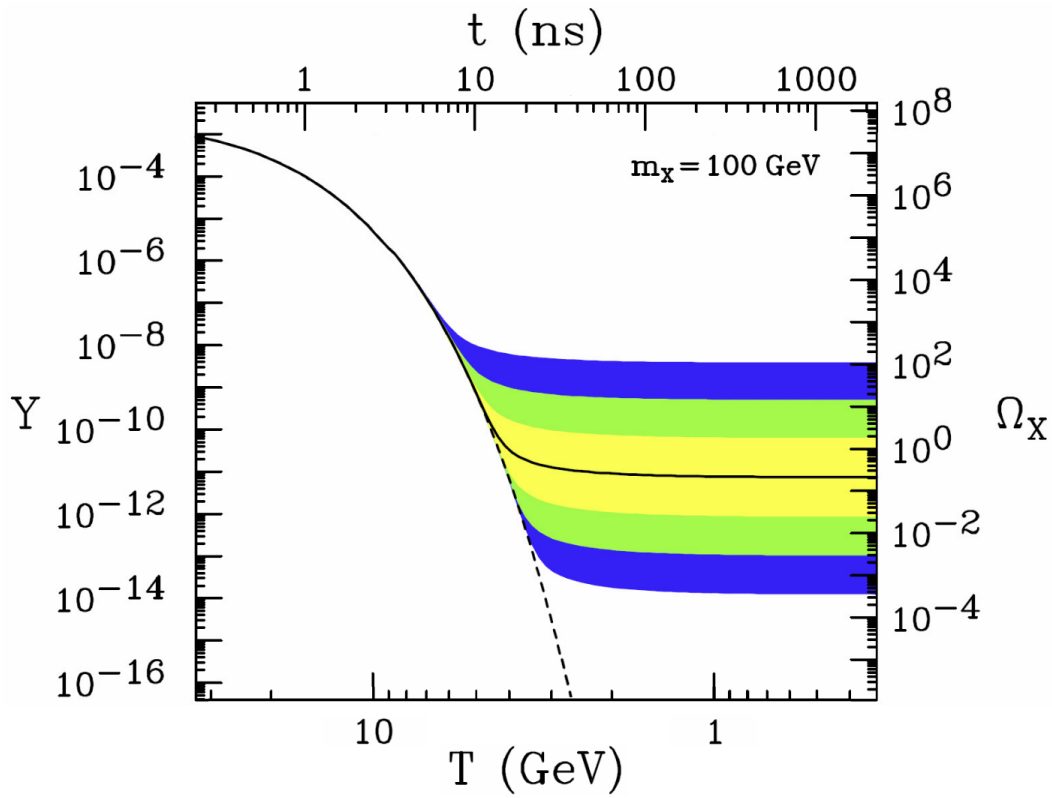


Figure 1.11. Dark matter yield or density parameter vs. Universe temperature or time for a 100 GeV WIMP. Shaded regions account for different cross-sections, the solid line is produced from the cross-section which yields the correct density parameter, while the dashed line is produced considering a particle which stays in thermal equilibrium. Figure from [17].

Now we can make use of the freeze-out process, which set $Y_\chi^0 = Y_\chi^f = n_\chi^f/s^f$, using the condition 1.16 and thus rewriting equation 1.17 as follows:

$$\Omega_{\text{CDM}} h^2 = \left(\frac{s^0 h^2}{\rho_c} \right) m_\chi \frac{n_\chi^f}{s^f} \stackrel{(1.16)}{\approx} \left(\frac{s^0 h^2}{\rho_c} \right) m_\chi \frac{H}{\langle \sigma_{\text{ann}} v \rangle s^f} \quad (1.18)$$

It is evident that $\Omega_{\text{CDM}} \propto \langle \sigma_{\text{ann}} v \rangle^{-1}$ and, considering m_χ as the only relevant scale (which is a totally reasonable assumption for many theories), we can approximate the cross section as:

$$\sigma \propto \frac{g^4}{m_\chi^2} \quad (1.19)$$

where we called g a generic coupling constant which describes the WIMP annihilation processes.

Here comes the “*miracle*”: plugging equation 1.19 into equation 1.18 assuming $g \approx g_{\text{weak}}$ and $v \approx 0.1c$ and we try to match $\Omega_{\text{CDM}} \approx 0.25$ [12, 13] on the LHS; what is found is $m_\chi \sim \mathcal{O}(100 \text{ GeV} - 10 \text{ TeV})$, the order of magnitude state at the beginning of this section!

Of course, **the WIMP miracle is not a proof of their existence**. Nevertheless, it is an interesting fact which pointed most of the last thirty years experiments towards trying to detect this promising candidates.

In the following, we will focus on the treatment of one of the most popular approaches in DM experimental research: **direct detection**.

1.5 Direct detection of dark matter

Direct detection experiments have their roots in the assumption that dark matter interacts with standard matter with processes which exceeds from the gravitational regime. In particular, they rely on the idea that dark matter particles with velocity v can interact with nuclei via elastic scattering causing a nuclear recoil with energy E_R .

The aim of these experiments is to create a setting where this recoil energy can be collected and measured to check if it is compatible with theoretical predictions for a certain dark matter candidate; this is typically done by going underground such that the detector is shielded from the cosmic rays.

Figure 1.12 shows a visual description of what just stated, both on macroscopic and microscopic level.

1.5.1 Standard Halo Model

The most used theoretical framework for the interpretation of direct detection experiments is known as the Standard Halo Model (**SHM**). In this model, the dark matter particles are characterized by the following key *assumptions*:

- Isotropic distribution in an isothermal sphere, i.e.

$$\rho(r) = \frac{\sigma^2}{2\pi G r^2} \quad (1.20)$$

where $\sigma(R) = \sqrt{3/2} v_c(R)$ is the velocity dispersion, related to the average circular velocity $v_c(R)$ calculated at the orbiting distance from the galactic

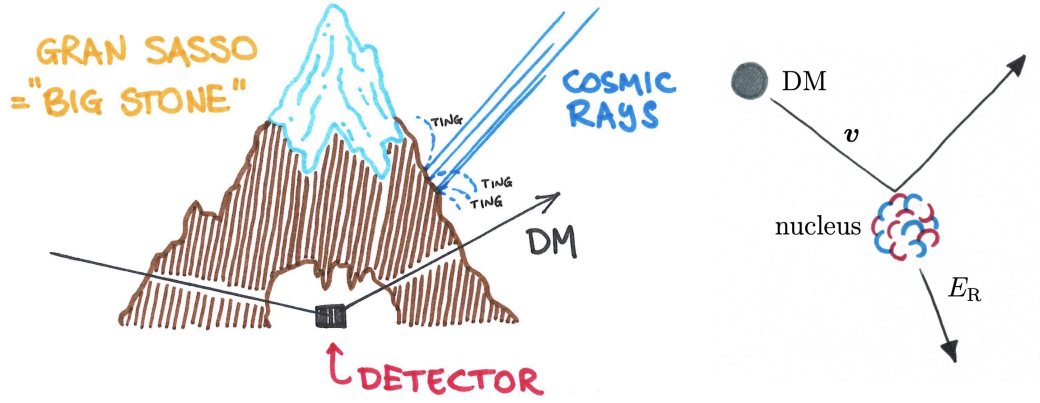


Figure 1.12. Macroscopic ($\sim \text{km}$ scale) and microscopic ($\sim \text{fm}$ scale) description of the main concepts of dark matter direct detection, from [18].

center R . Note that the scaling with r^{-2} is the one required to reproduce the flat galactic rotational curves introduced as a dark matter evidence in section 1.2.3.

- Maxwellian velocity distribution which is a consequence of the solution of the Boltzmann equation for collisionless particles, namely:

$$f(\vec{v}) = N \exp\left(-\frac{3|\vec{v}|^2}{2\sigma^2}\right), \quad (1.21)$$

where N is a normalization constant.

Moreover, it is important to summarize the three important values which are adopted from this model, all derived at the Earth position relative to the Milky Way galactic center $R_0 = 8.0 \text{ kpc}$:

- The **local dark matter density**, derived from the measured rotational curve for the Milky Way [19], is generally assumed to be:

$$\rho(R_0) \equiv \rho_0 \simeq 0.3 \text{ GeV}/\text{cm}^3.$$

- The **local circular velocity** can be measured from many independent observables, as e.g. luminous red giants using supervised machine learning models [20]. All measures are compatible and are averaged to the value:

$$v_c(R_0) \simeq 220 \text{ km/s}.$$

- The **local escape velocity**, namely the speed at which the WIMPs are no longer bound to the gravitational potential of the Milky Way, obtained from sample of high velocity stars from galactic surveys, e.g. [21], can be taken to be:

$$v_{\text{esc}}(R_0) \simeq 544 \text{ km/s}.$$

1.5.2 WIMP - Nucleus scattering cross section

First of all, from simple scattering kinematics, one can retrieve the recoil energy for a scattered nucleus in the non-relativistic regime, as a function of the center-of-mass (**CoM**) angle θ :

$$E_R = \frac{\mu^2 v^2}{m_N} (1 - \cos \theta), \quad (1.22)$$

where m_N is the mass of the nucleus and $\mu = m_N m_\chi / (m_N + m_\chi)$ is the reduced mass of the system.

At the same time, the squared momentum transfer is defined as:

$$q^2 = 2m_N E_R \quad (1.23)$$

The physics of a given process is encoded inside its differential cross section. In the case of the WIMP - Nucleus (elastic) scattering, this cross section can be seen as the coherent sum of two main components: a scalar, or spin-independent (**SI**), one and a spin-dependent (**SD**) one.

With [18] as reference for a more detailed derivation, we can write it in general as:

$$\frac{d\sigma}{dE_R} = \frac{m_N}{2\mu^2 v^2} \left[\sigma_0^{SI} F_{SI}^2(E_R) + \sigma_0^{SD} F_{SD}^2(E_R) \right]. \quad (1.24)$$

where, focusing on the quantities inside the square brackets, we have the cross sections at zero momentum-transfer, σ_0 , and the nuclear form factors $F(E_R)$, both in the scalar and spin-dependent cases.

If we limit the discussion to the spin-independent case, since the WIMP coupling to protons and neutrons is almost identical (*isosinglet approximation*), we can write the scalar cross section at zero momentum transfer as simply:

$$\sigma_0^{SI} = \sigma_p \frac{\mu_T^2}{\mu_p^2} A^2 \quad (1.25)$$

where A is the atomic mass number of the target element and the subscript in μ makes the distinction between the WIMP - Target and the WIMP - proton system.

1.5.3 Dark matter event rate

It is now interesting to define and describe the expected event rate for a WIMP - nucleus interaction, following in particular the approach from [22] and [23].

The differential rate with respect to the recoil energy, commonly reported in *differential rate units* (**dru**) $\equiv$ counts/kg/day/keV, can be expressed as follows, for a WIMP with mass m_χ [23]:

$$\frac{dR}{dE_R} = \frac{\rho_0}{m_N m_\chi} \int_{v_{min}}^{v_{esc}} v f(v) \frac{d\sigma}{dE_R} dv, \quad (1.26)$$

where:

- v_{min} is defined as the minimum WIMP velocity required to produce a nuclear recoil and can be computed from equation 1.22, in the case $\theta = \pi$:

$$v_{min} = \sqrt{\frac{E_R m_N}{2\mu^2}}.$$

- The upper integration extreme, v_{esc} , can be virtually infinite. Nonetheless, from the considerations in section 1.5.1, we know that there is a cap velocity after which the WIMPs are no longer bound to the galaxy's potential.
- $d\sigma/dE_R$ is the differential cross section of the process.

We can now define the total number of observed events, in a detector with efficiency $\epsilon(E_R)$, after a certain exposure time ΔT as:

$$N = \Delta T \int_{E_{thr}}^{E_{max}} \epsilon(E_R) \frac{dR}{dE_R} dE_R. \quad (1.27)$$

Here, $E_{max} = \mu^2 v_{esc}^2 / m_N$ is again given by the kinematics, while the threshold energy E_{thr} is the smallest energy that the detector is capable to resolve.

Differential rate in the isosinglet approximation

Relying on the isosinglet approximation and neglecting the spin-dependent contribution to the differential cross section, we can simplify the structure of equation 1.26 to be:

$$\frac{dR}{dE_R} \propto \int_{v_{min}}^{v_{esc}} \frac{1}{v} f(v) dv. \quad (1.28)$$

If we take $v_{esc} \rightarrow \infty$ we can integrate equation 1.28, ending up with [22]:

$$\frac{dR}{dE_R} = \frac{R_0}{E_0 r} e^{-\frac{E_R}{E_0 r}}, \quad (1.29)$$

where we defined the total event rate (R_0), the most probable kinetic energy for the dark matter particle (E_0) and the dimensionless efficiency in the momentum transfer r , proportional to the reduced mass of the system.

Note that $r = 1$ when $m_N = m_\chi$, i.e.: the efficiency is maximal when the target is composed by atoms with the same mass as the WIMPs. This of course is crucial to optimise the detector sensitivity.

$$R_0 = \int_0^\infty \frac{dR}{dE_R} dE_R, \quad E_0 = \frac{1}{2} m_\chi v_c^2, \quad r = \frac{4m_N m_\chi}{(m_N + m_\chi)^2}.$$

This is essentially telling us that the WIMP event rate suffers of a featureless exponential decrease, at least in the most simple case of a detector not moving around the galaxy.

This is confirmed (or worsen) from the complete calculation reported in figure 1.13 from [18], from which we can also see the very low event rate, in dru.

This is why direct detection of dark matter is pursued by means of large detector (to maximize the exposed mass) put under huge mountains or inside mines (in order to reduce the amount of background contaminating the data taking).

1.5.4 Annual modulation and directional approach

If we now relax the assumption of a static detector, we need to take into account the Earth motion through the WIMP halo. This is typically modeled as [1]:

$$v_E = v_\odot + v_\oplus \cos \theta \cos[\omega(t - t_0)] \quad (1.30)$$

where:

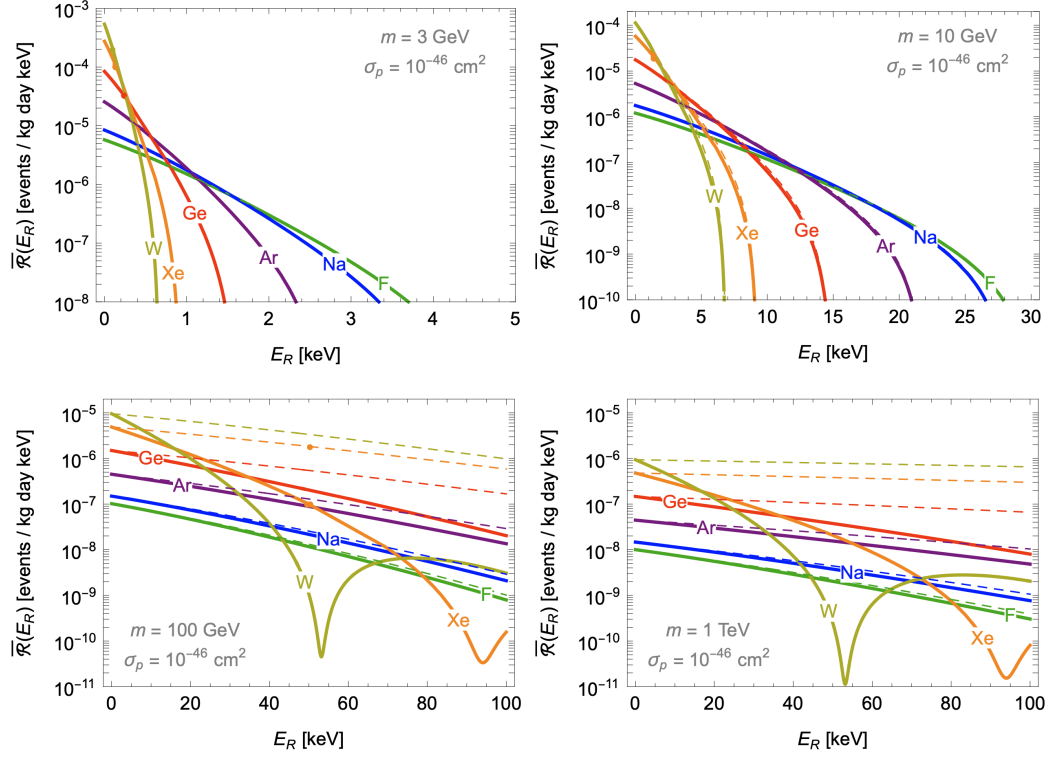


Figure 1.13. Annual average rate spectra for different configurations of WIMP - target masses in the spin-independent case, assuming isosinglet approximation and fixing $\sigma_p = 10^{-46} \text{ cm}^2$. Note the dips visible in the W and Xe case, related to the Helm form factor vanish at specific values of momentum transfer. Dashed lines are computed neglecting the form factor contribution, while solid lines takes it into account. Plot from [18].

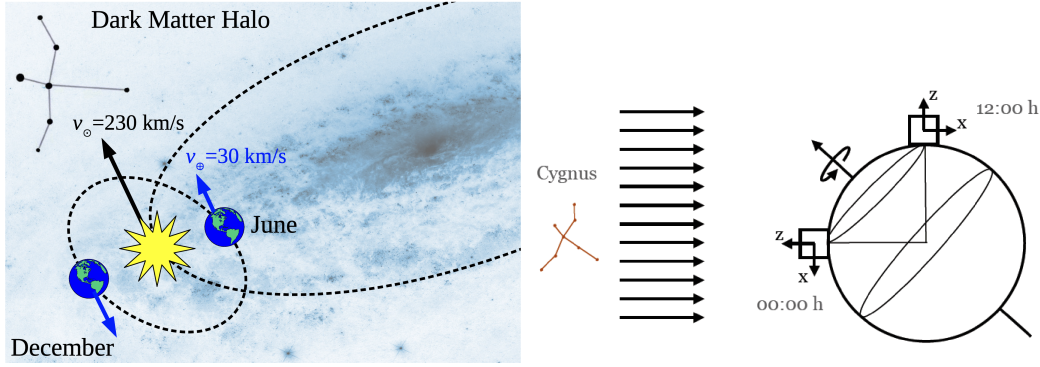


Figure 1.14. A schematic representation of the solar system orbiting around the galactic center always behind the Cygnus constellation (**left**) [1] and of the Earth inclination with respect to the galactic plane and Cygnus constellation (**right**) [24].

- $v_{\odot} = v_c + 12 \text{ km/s}$ is the Sun speed with respect to v_c ;
- $v_{\oplus} = 30 \text{ km/s}$ is the Earth velocity while orbiting around the Sun;
- $\theta \approx 60^\circ$ describes the inclination angle of the Earth axis with respect to the galactic plane;
- $\omega = 2\pi/T$ is the angular frequency, defined by the time period $T = 1 \text{ yr}$ and with phase fixed to $t_0 = \text{June } 2^{\text{nd}}$, which represents the day where $v_{\odot}$ and $v_{\oplus}$ are parallel.

On a general ground, assuming a certain detector threshold, we can describe the signal modulation over the year as follows:

$$S(t) = Bkg + S_0 + S_m \cos[\omega(t - t_0)], \quad (1.31)$$

where background and unmodulated signal S_0 are typically much larger than the modulated part and a detection which only relies on the modulation measure would require $\mathcal{O}(10^4)$ events to have a statistical meaning.

To disentangle background events from the actual signal it is useful to adopt the so-called “*directional approach*” to direct detection. This approach has its roots in the fact that Earth is always moving towards the Cygnus constellation, a schematic view of this concept can be seen in figure 1.14, while figure 1.16 displays a more *realistic* picture of the situation.

If the halo model is correct, this motion should create an apparent WIMP *wind* coming from the Cygnus constellation galactic coordinates.

In the galaxy frame, the doubly differential recoil rate with respect to nuclear recoil energy and the angle between the WIMP and recoil directions (ϕ) can hence be expressed as [1]:

$$\frac{d^2 R}{dE_R d\phi} \propto \exp \left[\frac{2(v_{\odot} \cos \phi - v_{\min})^2}{3v_{\odot}^2} \right] \quad (1.32)$$

This rate is expected to be peaked around the Sun’s direction of motion, and in figure 1.15 we can clearly see how, from the simulation, in the forward direction we have a rate which is 1 order of magnitude bigger than in the backward direction.

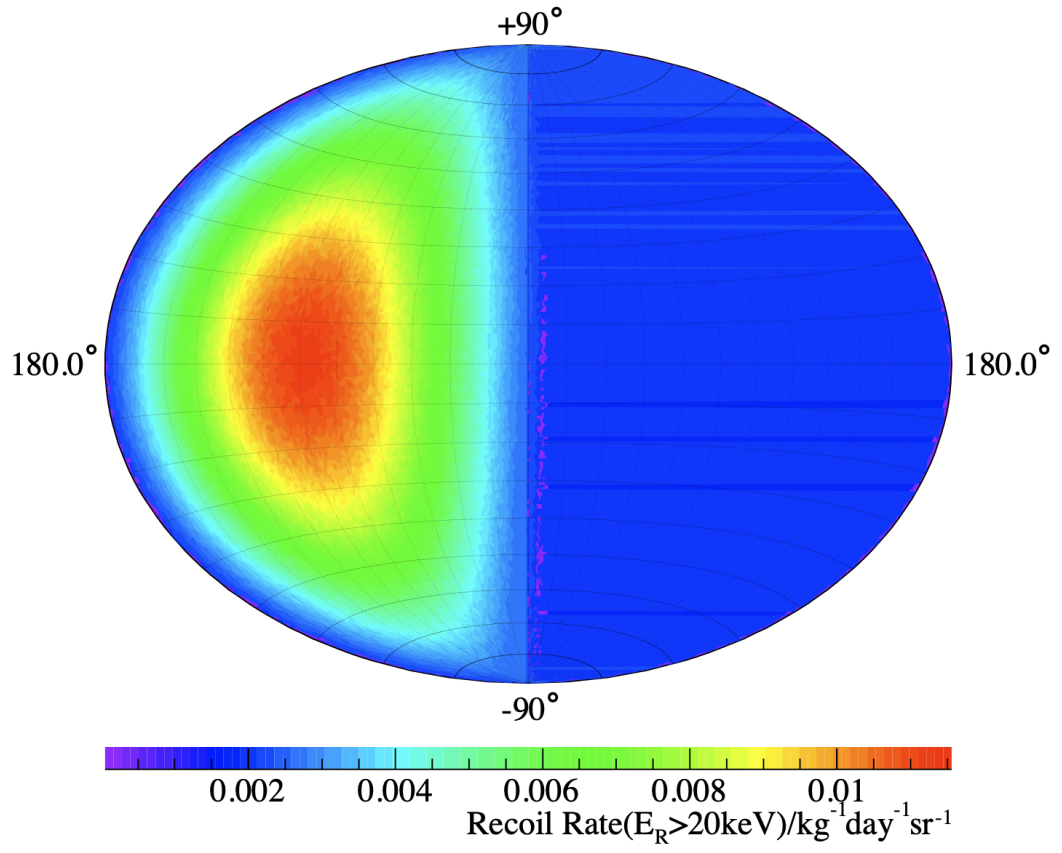


Figure 1.15. Expected differential recoil rate galactic map for a 100 GeV WIMPs population and for $E_R > 20\text{ keV}$ [23].

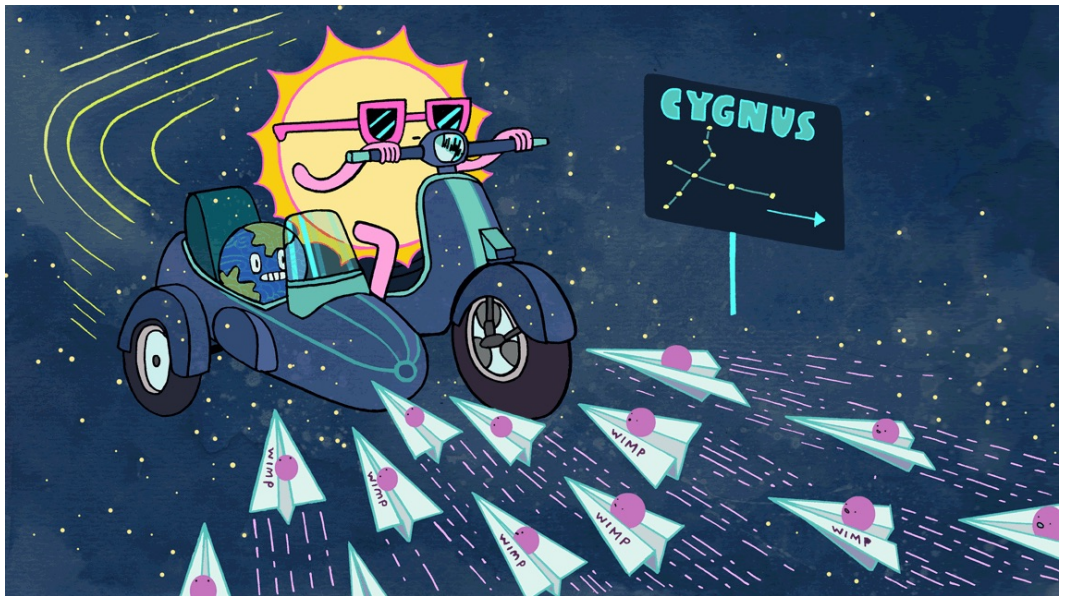


Figure 1.16. Very realistic representation of the apparent wind of WIMPs generated by the Solar System motion following the Cygnus constellation.

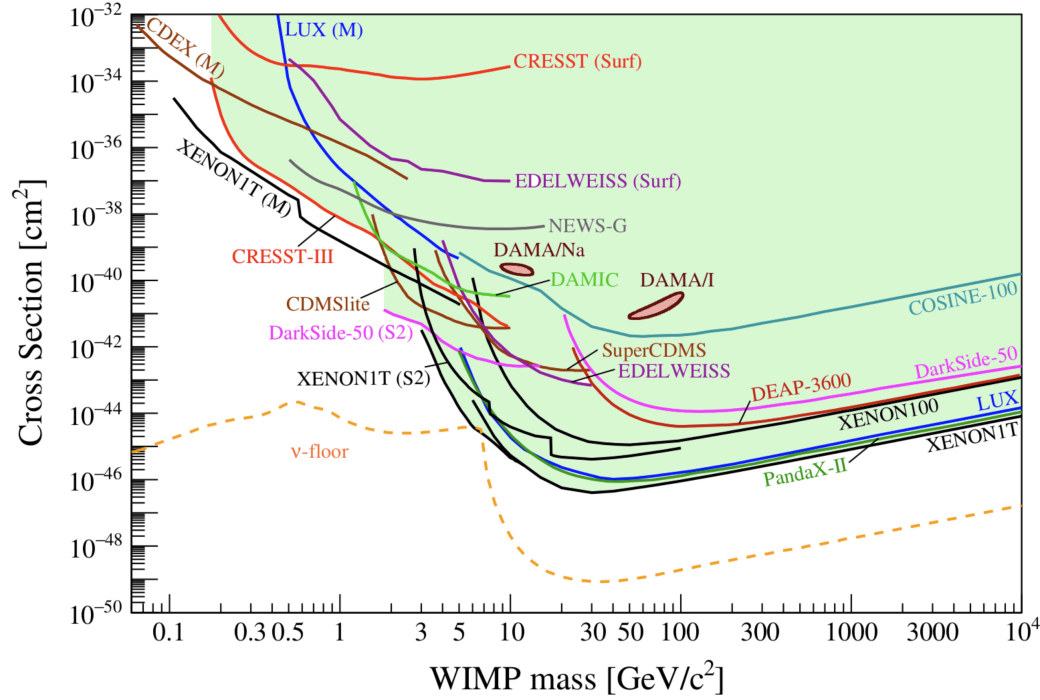


Figure 1.17. Current exclusion limits for spin-independent interactions from state-of-the-art experiments in the WIMP dark matter direct detection field. The excluded region (at 90% C.L.) is shaded in green, the individual contribution from the different experiments is reported in solid lines, the neutrino-floor region is delimited by the dashed orange line. Plot from [25].

1.5.5 Current experimental status

Limiting the discussion to the spin-independent interactions, the current experimental status in dark matter direct detection experiments is typically reported in a σ vs. m_χ plot (where σ is the total cross-section in cm^2 or barn).

The status as of 2022 is reported in figure 1.17 [25] and it is interesting to talk about some of its key aspects.

First of all, the green area denotes the parameter space region which is excluded with 90% confidence level from the various experiments altogether, whose individual contribution is highlighted by the differently coloured solid lines.

Note that there are two small elliptical regions [26], which correspond to the limits extracted from the results provided by the DAMA/Libra collaboration during the years [27–29]. These are particularly controversial results: they could be interpreted as an actual signal of dark matter, when considering the WIMP model, but they are also excluded from independent experiments such as XENON100 [30] and LUX [31] at 5.7σ and 9.2σ , respectively.

The bottom part of the plot (below the *neutrino-floor*, identified by the orange dashed line) is called the *neutrino-fog region*. In this region, coherent elastic neutrino-nucleus scattering (CE ν NS) and WIMP elastic scattering on nuclei would produce the same indistinguishable signature.

From right to left, we can see how the unexplored region of the parameter space increases. In particular:

- For $m_\chi > 3$ GeV, XENON1T is the leading experiment [32], followed by LUX and PandaX-II, all adopting the LXe TPC approach.
- For $2 < m_\chi < 3$ GeV, is DarkSide-50 the one with the best constraints with the LAr TPC approach [33].
- For $m_\chi < 2$ GeV, instead, the leading technique is the one which involves cryogenic detectors with ionization/phonon readout due to their extremely low thresholds, with the best constraints produced by CRESST-III [34]. It seems that by means of the Migdal effect also experiments like XENON1T could extend their range of sensitivity in this mass region, but this is still under study.

To increase the experiment sensitivity it is sufficient to increase the exposure time and/or the mass of the exposed target and, with the high-mass region almost entirely explored, the focus is shifting to the low-mass and neutrino-fog regions of the parameter space.

Eventually, the sensitivity is going to overlap with the neutrino-floor and, if no dark matter detection will be achieved by that time, going further is going to be a really big problem, at least for detectors which are able to reconstruct only the charge released by the nuclear recoil in the detector (such as XENON).

One possible solution to this problem is to adopt the directional approach which was introduced in section 1.5.4. In particular, one project moving in this direction is called CYGNO and it will be the focus for the rest of this thesis.

Chapter 2

The CYGNO Project

This chapter will review the experimental approach of the CYGNO collaboration regarding the directional detection of WIMP dark matter.

First of all, the working principles of the detector known as time projection chamber will be presented.

The second part of the chapter will be focused on the technological concepts on which the experiment is based, together with the description of its goal.

Finally, the focus will shift to an actual working prototype, LIME, which is currently taking data underground at Laboratori Nazionali del Gran Sasso (LNGS) as of the composition of this thesis.

2.1 Time Projection Chambers

2.1.1 Working principles

The first proposal of a Time Projection Chambers was related to the PEP4 experiment at the Stanford Linear Accelerator Center (SLAC) [35] with the aim of enabling full 3D reconstruction and particle identification in a single detector.

Nowadays, we can find them involved in a large variety of physics research fields, from high energy physics (e.g. ALICE [36]) to extremely rare events searches (e.g. Xenon [37]).

A generic TPC (schematically pictured in figure 2.2) is composed by the following key components:

- A **sensitive volume** inside an external structure filled with gas or, eventually, with gas and liquid (dual phase TPC) as in the case of Xenon.
- A **cathode**, negatively charged and positioned at the center of the TPC.
- Two **anodes**, positively charged, such that an electric field is generated between anodes and cathode, directed towards the latter.
- **Multiplication regions** in proximity of the anodes which enable signal amplification by means of avalanche processes.
- A **readout plane** after the multiplication stage, whose aim is to detect the position and time of arrival of the electrons to reconstruct the particle's track in 3D.

When a charged particle crosses the TPC active volume, it ionizes the gas and the electrons are drifted towards the anodes by the electric field, where they are amplified by avalanche processes, inducing the signal which is then detected by the electronics.

Of particular interest for applications in high energy physics like collider experiments, the time projection chamber can also be equipped with a magnetic field inside the sensitive volume, which permits the measurement of the particle's momentum, that in turn enables better particle identification (PID) by means of the energy deposition density per unit length (indicated with dE/dx).

It is well known that dE/dx follows the Bethe-Bloch formula, reported in its “universal” form (normalised to the target mass density ρ) in equation 2.1, where I is the **average ionization potential** of the medium, while $\delta(\beta\gamma)$ represents the **density correction**, which arises from the polarization of the medium and corrects for the logarithm explosion at high values of γ (the dependence of δ on $\beta\gamma$ is very weak).

$$-\frac{1}{\rho} \frac{dE}{dx} = C \frac{Z}{A} \frac{z^2}{\beta^2} \left[\ln \frac{2m_e c^2 \beta^2 \gamma^2}{I} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \quad (2.1)$$

$$C = 4\pi r_e^2 m_e c^2 N_A \simeq 0.307 \frac{\text{MeV cm}^2}{g}$$

Measuring the particles dE/dx inside the TPC, one can discriminate between different species, as it is done for example in ALICE [38], reported from the plot in figure 2.1, where also the corresponding Bethe-Bloch theoretical curve is superimposed.

When a particle's momentum is near the Bethe-Bloch minimum, it is said to be “*Minimum Ionizing Particle*” (**MIP**) and, after the minimum, the correction is so small that the particle can still be considered a MIP. Typical MIPs, as e.g. energetic muons or electrons, produce long and thin tracks.

On the contrary, for small momenta, dE/dx is really steep, particles ionize a lot and produce dense tracks (short and wide).

Note that it is completely possible to build a TPC taking only one half of what is described above, namely from the central cathode to one end of what can be seen in figure 2.3, the working principles are the same.

Another important concept to mention is the one that the *time* in “time projection chamber” refers to. In fact, different triggering systems are implemented in the common TPCs, in order to provide a *time of arrival* measurement of the drifted electrons on the anode. From the time of arrival, it is possible to evaluate the track length, while an external trigger (if implemented) can provide information about the track distance from the cathode.

Moreover, it is a known concept in literature, in the field of drift chambers, that diffusion of ionization electrons scales with the square root of the distance [40] in absence of magnetic field. Large experiments like ALICE make use of large magnetic fields in order to make ionisation electrons rotate around the field lines, which results in a reduction of their diffusion.

This will be particularly useful to know for a better understanding of the last part in chapter 4.

2.1.2 Directional TPCs

Since this thesis is devoted to dark matter directional research, in the following we will focus on a novel TPC design, whose aim is to measure the direction of nuclear

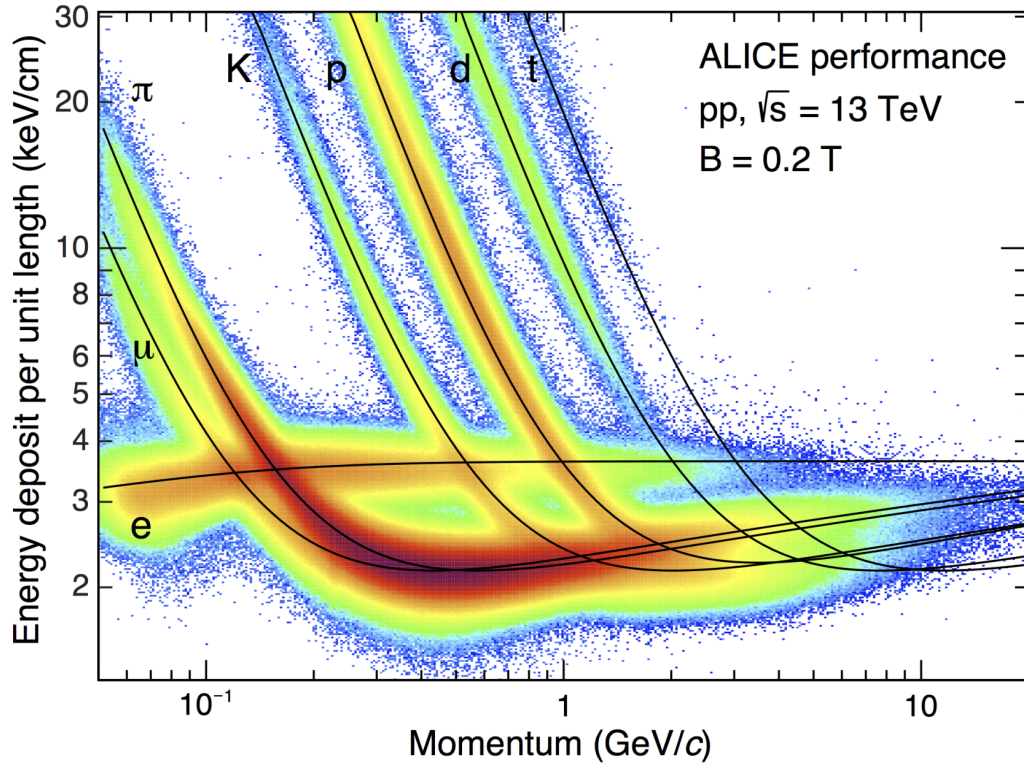


Figure 2.1. ALICE results in particle identification based on their energy deposition density dE/dx vs momentum with superimposed the corresponding Bethe-Bloch theoretical curve. Plot from [38]

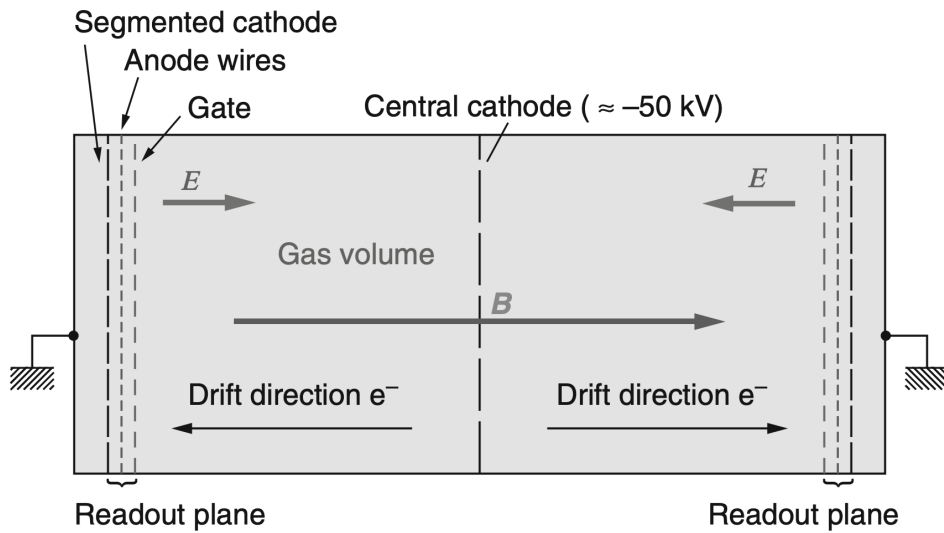


Figure 2.2. Time projection chamber geometry from [39].

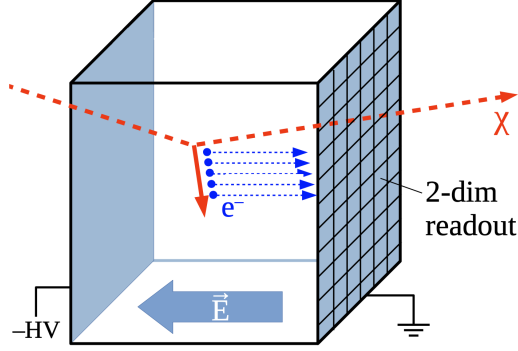


Figure 2.3. Directional time projection chamber geometry from [1], where with χ we indicate a WIMP DM particle.

recoils, in order to detect nuclear recoils induced from the apparent WIMP wind coming from the Cygnus constellation, as depicted in section 1.5.4. This is pursued by employing TPCs with **low density gas mixtures**, since the track length is correlated to the target density and longer tracks are easier to reconstruct, and **optical readout**, featuring Active Pixel Sensor cameras to obtain images of the tracks and photo-multiplier tubes to extract information about the time of arrival of the light.

A schematic view of this concept can be observed in figure 2.3.

2.2 CYGNO approach and technology

The CYGNO experiment detector concept involves a TPC filled with a gas mixture of helium (He) and carbon tetra-fluoride (CF_4) operated at atmospheric pressure. The electrons generated by the gas ionization are multiplied by the amplification stage and optically read out, joining information from both a scientific camera, for the bi-dimensional position, and a certain number of photomultiplier tubes (which is able to detect single photons), depending on the prototype, to exploit the time of arrival and hence the third dimension of the track.

To better understand the last introduced concepts, we can start the discussion from the optical readout.

2.2.1 Optical readout

The idea of the optical readout system is to detect the light emitted from the gas molecules during electron avalanches inside Gas Electron Multipliers (hereafter **GEMs**, with a more detailed characterization in section 2.2.2), which are employed in the amplification stage. Typical numbers ranges between $\mathcal{O}(10^{-2})$ and $\mathcal{O}(10^{-1})$ photons emitted per avalanche electron [41], and these number clearly changes with the type of gas mixture and with the electric field intensity.

The production mechanism behind this concept is the gas luminescence, which happens when the gas molecules or atoms are excited by the charged particles; concurrently with the de-excitation process, photons are emitted.

This approach is a good alternative to the charge readout for a number of reasons. For example, adopting an optical readout:

- permits to locate the sensors outside the active volume, which reduces both the radioactive background and the gas contamination;

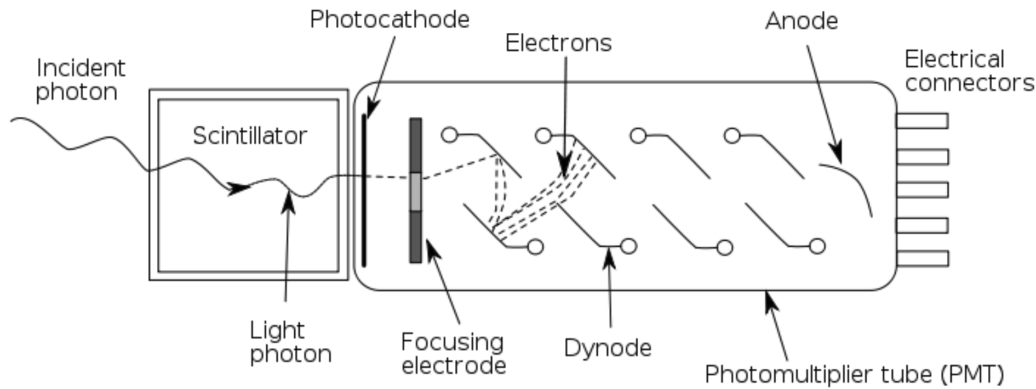


Figure 2.4. Schematic representation of a photomultiplier tube.

- the granularity is very high, around 2000×2000 pixels in a $1 \times 1 \text{ cm}^2$ sensor, in general. Imagine to have a $33 \times 33 \text{ cm}^2$ area to be imaged, this will result in a spatial granularity of almost 4000 pixels/cm^2 .

At the same time, given that the sensors are far from the active volume, one could expect to lose a large amount of photons. In the CYGNO experiment, this is mitigated by the triple GEM stack in the amplification stage, which assure very high gain [42].

As already introduced, CYGNO adopts an optical readout which is divided in two main kind of optical sensors whose joined work should enable a full 3D track reconstruction: an active pixel sensor camera and a certain number of photomultiplier tubes.

Active pixel sensor CMOS camera

Inside the Active Pixel Sensor (**APS**) family, scientific CMOS (sCMOS) cameras are particularly popular. They are based on complementary metal-oxide semiconductor pinned photodiode technology (CMOS PPD) and they are special photon detectors characterised by very low noise and high sensitivity / granularity. In the following, APS and sCMOS will be referred to with the same meaning, if not otherwise specified.

The underlying technology has a broad range of applications on the generalist market, in fact it can be found even inside smartphones camera modules, and this has been the key for its recent glow up in scientific applications. The competitive market for this kind of technology provided a large boost to its R&D, allowing the APS to surpass even the CCD performances, which were leading the scientific sector up to some years ago.

An area of up to 1 m^2 can be acquired with a single sensor if equipped with proper lenses achieving at the same time an optimal spatial resolution of $\mathcal{O}(100 \text{ }\mu\text{m})$ which is very close the pixel effective size.

Photomultiplier tubes

The addition of the photomultiplier tubes (PMTs) is fundamental to achieve full 3D tracking as demonstrated in [43], the APS camera alone is not sufficient given the too low frame rate.

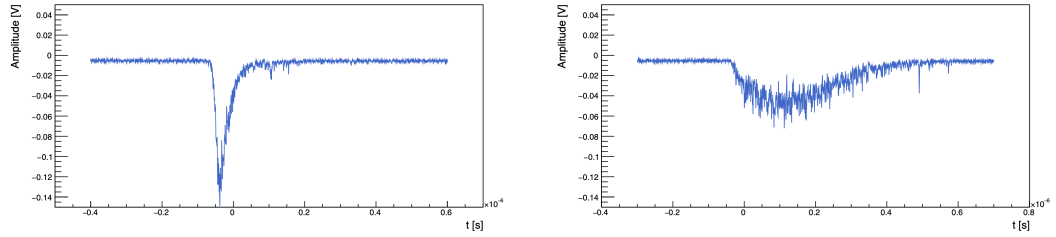


Figure 2.5. PMT waveforms for a track parallel (**left**) or tilted (**right**) with respect to the GEM plane.

A PMT is a single photon detector whose working principles are based over the photoelectric effect. It is essentially a vacuum tube with a photocathode on one side, placed just after an input window. Inside the tube there is a focusing electrode and a series of amplifiers (dynodes), separated by voltage dividers, which multiply the electron produced by the incident photon until reaching the anode placed at the other hand of the tube where the electrons are read out. A detailed sketch of a PMT can be observed in figure 2.4.

In general three or four electron are produced for each electron impinging on a given dynode, and a typical PMT gain can be as high as 10^6 , subject to optimization in case of different materials and voltage provided to the dynodes.

The total charge collected by the anode is proportional to the photons which were converted, which means that it corresponds to a charge measurement.

Moreover, one can easily infer if a track is tilted or parallel to the readout plane from the amplitude and the width of the collected signal and examples of parallel and tilted tracks' signals as seen from a PMT are reported in figure 2.5. The width of the signal, in particular, is a measure of the electrons' time of arrival, introduced at the end of section 2.1.1, and from this one can retrieve information about the length of the track in the drift direction when the drift velocity is known. This concept will be recalled and quantified in section 2.3.2, during the technical description of the current working detector prototype.

If a set of PMTs is employed in the readout one can use the relative amplitude of the waveforms seen by each PMT to reconstruct the planar position of the signal to correctly associate events seen by the PMTs with events seen by the camera.

2.2.2 GEM amplification stage

In 1997, the Gas Electron Multiplier (**GEM**) has been invented by Sauli and an extensive summary of the state-of-the-art for this detector can be found in [44].

GEMs fall within the family of Micro Pattern Gas Detectors (MPGDs) and largely employed in wire chamber-based experiments since they provide really good efficiency and granularity.

In a nutshell, GEMs are composed by:

- a thin insulating foil, generally made of KaptonTM or fiberglass;
- a conducting metallic coating around the insulator, generally made of copper.

The whole GEM foil is then pierced via chemical digging (when the foil is sufficiently thin) or via a double conical etching, in order to have an high density of holes in the foil. The holes diameter usually ranges from $50\text{ }\mu\text{m}$ to 1 mm with a pitch of $140\text{ }\mu\text{m}$.

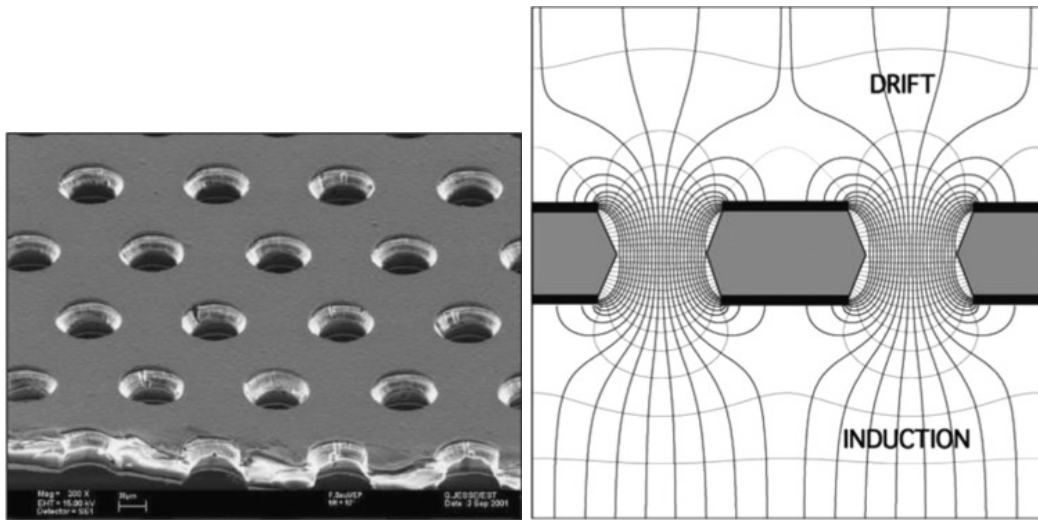


Figure 2.6. 50 μm thick GEM section pictured with electron microscope showing holes with 70 μm diameter and 140 μm pitch (**left**) and sketch of the electric field configuration inside GEM holes after the application of a large difference of potential (**right**).

A large difference is applied between the top and the bottom metallic foils and this produces strong electric field of $\mathcal{O}(10)$ kV/cm inside the holes. Such electric fields are strong enough to initiate electron avalanches.

In fact, electrons produced during the gas mixture ionization are drifted towards the GEM plane by the electric field inside the gas itself. When entering the GEM's holes they start to multiply thanks to the strong electric field in the holes.

To increment the overall gain, a stack of GEMs can be employed resulting in the end in a total gain ranging from $\mathcal{O}(10^4)$ to $\mathcal{O}(10^6)$, as depicted by several studies with different mixtures [44, 45].

2.2.3 Gas mixture

The choice of the components of the gas mixture was influenced by three key requirements:

- a **low-mass target**, which provides a better sensitivity towards low-mass dark matter nuclear recoils since the energy transfer in elastic scattering is known to be maximal when the two particles have the same mass, as illustrated in section 1.5.3;
- a **low density target**, which enables a longer travel distance for the tracks produced;
- a gas with **good scintillation properties** in the visible range, since the adopted camera has its maximum in quantum efficiency in the visible range, as will be better quantified in section 2.2.1.

For the first two points, the best candidate should be Hydrogen, which is a light enough element to enable tests in the $\mathcal{O}(\text{GeV})$ WIMP mass region. Nonetheless, Hydrogen is also very dangerous to treat when operated in its gaseous phase, and this forces the choice to be on Helium, which with $A = 4$.

It is also worth mentioning that the CYGNO collaboration is also testing the possibility of introducing small percentages of hydrocarbons (such as isobutane)

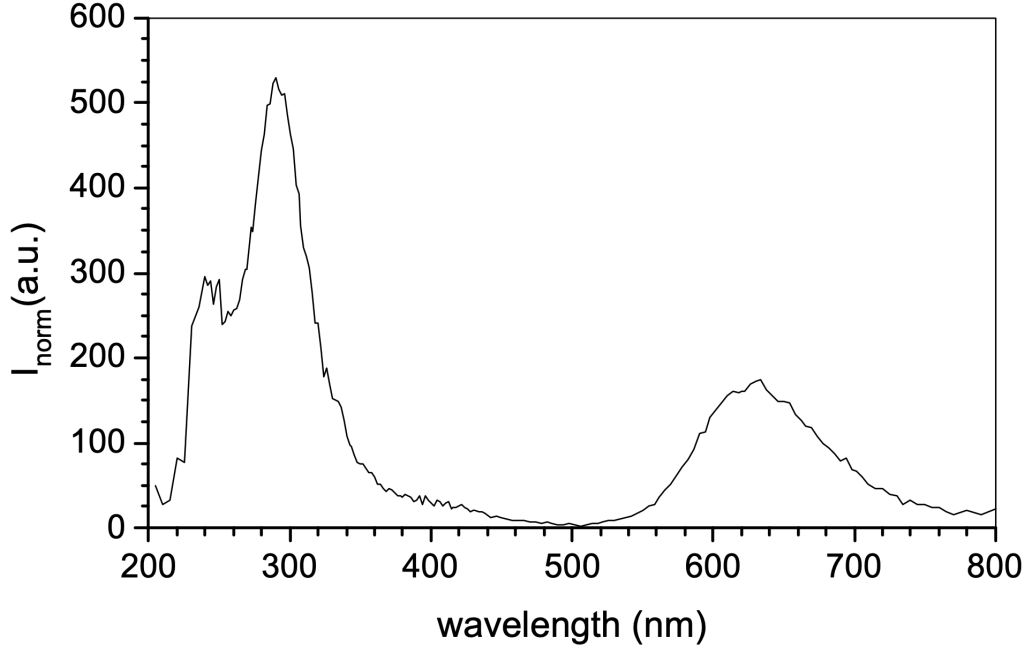


Figure 2.7. Light intensity vs. wavelength for a He:CF₄ (60:40) mixture. Plot from [41]

inside the gas mixture in order to test their impact on the overall performance of the detector [46].

For the scintillation, CF₄ is a perfect match. It has well known scintillating properties with a clear peak in the visible region also in combination with noble gases, as it is appreciable in figure 2.7 [41]. At the same time, the presence of Fluorine also opens to a possible sensitivity to spin-dependent couplings with low-mass WIMPs.

In order to optimize both light yield and gas mixture ratios, different configuration have been explored and tested [42, 47], converging in the end to the He:CF₄ 60:40 ratio, which maximizes gain and light yield, while minimizing the diffusion and the GEMs' discharge probability, as it can be seen from figure 2.8.

As a reference wrap-up, figure 2.9 reports the Garfield¹-simulated diffusion coefficients and drift velocity vs. drift field for a 60 to 40 mixture. Moreover, we can see that, for a drift field between 0.5 – 1 kV/cm, the simulations predict diffusion coefficients less than 130 $\mu\text{m}/\sqrt{\text{cm}}$ and a drift velocity of approximately 5 cm/ μs .

Lastly, it is interesting to report the GEANT4 and SRIM simulations of electron/He recoils 3D ranges (travelled distance between production and absorption) with respect to the initial kinetic energy, whose plot can be seen in figure 2.10, together with some important comments:

- electron / He recoils lengths are comparable if the initial kinetic energy is of the order of 1 keV_{nr};
- with increasing energy, recoils lengths start to separate but He recoils are always below the mm scale even for energies of $\mathcal{O}(100 \text{ keV}_{nr})$, while electron recoils which started with such energy produce tracks with lengths two order of magnitude bigger.

¹Informations about Garfield tool can be found here: [Garfield](#)

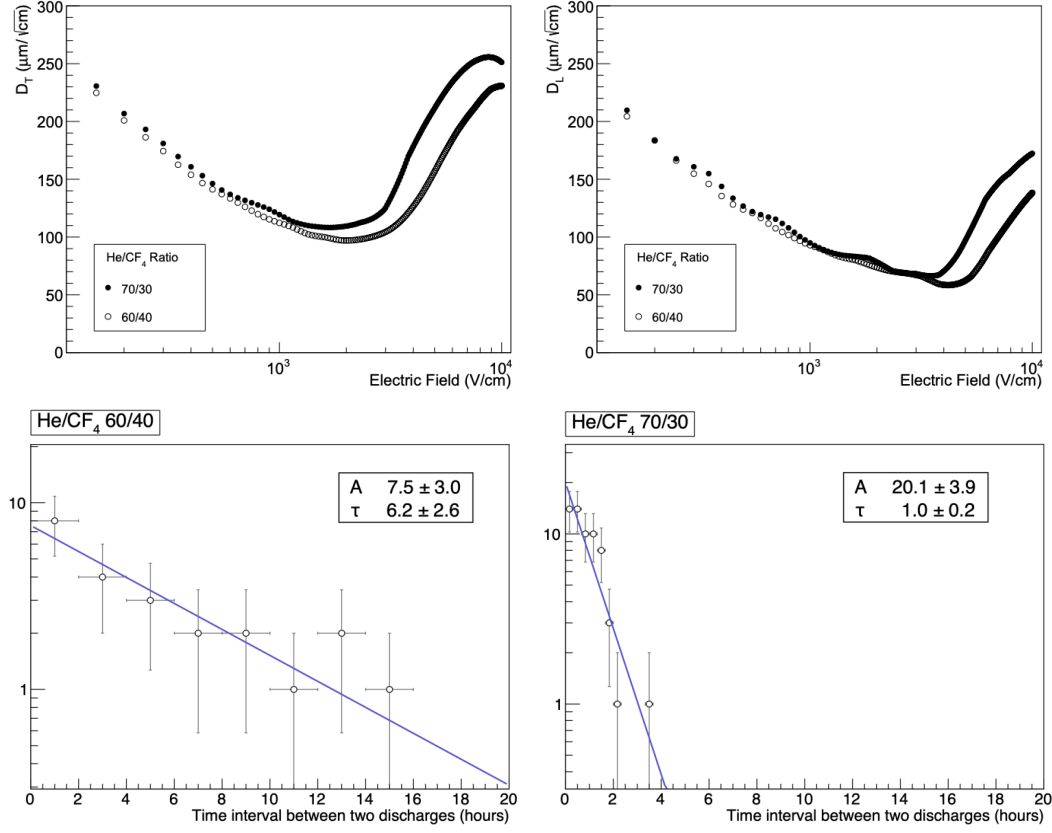


Figure 2.8. Diffusion parameters (upper panel) and discharge probability (upper panel) for a 60:40 and 70:30 He:CF₄ gas mixture. Plots from [47].

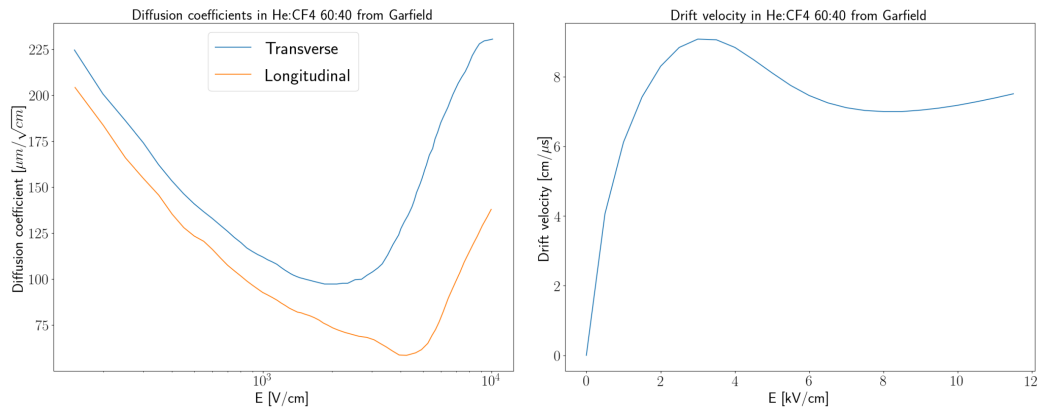


Figure 2.9. Garfield simulation for transversal and longitudinal diffusion coefficients per unit length (left) and drift velocity (right) vs. drift field for He:CF₄ (60:40). Plots from [48].

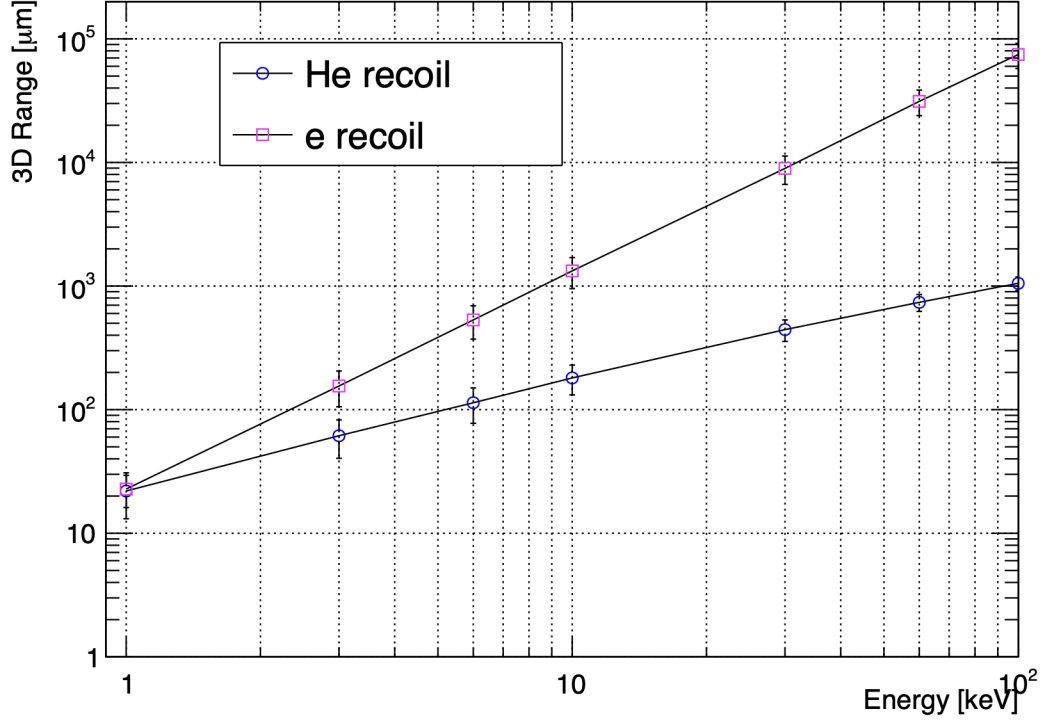


Figure 2.10. Average 3D range simulation for both He and electron recoils in a He:CF₄ (60:40) gas mixture.

It is fundamental to remark that **the mixture is operated at atmospheric pressure**, which provides the possibility to reduce the internal background eventually induced by an external vessel required to pressurize the detector. At the same time, this clearly reduces the costs, while also providing a good volume to target mass ratio and not compromising the track lengths, since also at atmospheric pressure the gas has a sufficiently low density.

2.3 The LIME prototype

The Long Imaging ModuleE (LIME) comes after a series of exploratory studies performed over small prototypes, listed in figure 2.11 (for more detailed descriptions, see e.g. [51, 52]), and is the first step towards the $\mathcal{O}(1\text{ m}^3)$ demonstrator which the CYGNO collaboration is trying to build. It is a time projection chamber with a 50 litres sensitive gas volume operated at atmospheric pressure and a 50 cm long electric field region closed by a $33 \times 33\text{ cm}^2$ triple-GEM plane for the amplification.

In this case, the optical readout is composed by a scientific CMOS camera and a set of four PMTs, arranged in the corners of a square around the camera.

The initial characterization of the detector was conducted overground at the INFN Frascati National Laboratories (LNF) [50], while it is currently installed and taking data underground at the INFN Gran Sasso National Laboratories (LNGS) since 2022. Figure 2.12 shows a schematic view of the detector as seen from above, while figure 2.13 presents a picture of the prototype from a diagonal perspective.

The aim of this prototype is first of all to test whether the technological approach is viable and/or sustainable, since the first demonstrator will be a modular detector

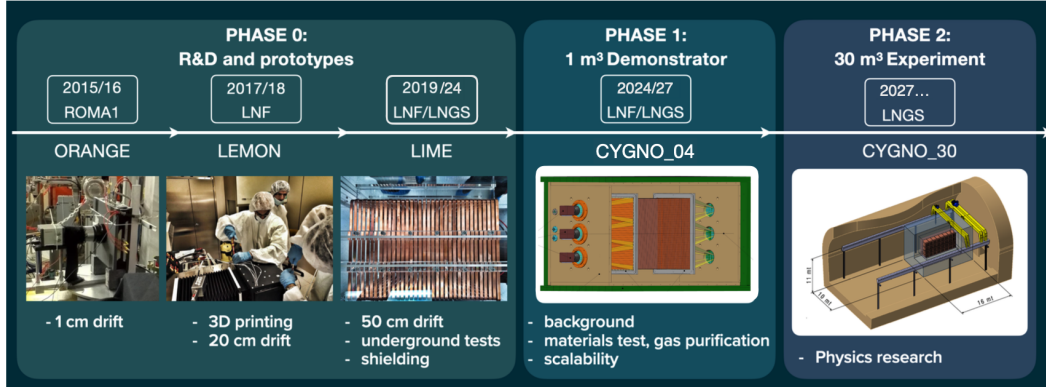


Figure 2.11. CYGNO timeline with all the steps completed and foreseen. As of writing this thesis, we are at the end of PHASE 0, with the CYGNO_04 demonstrator already commissioned [49] and its final installation expected in 2025.

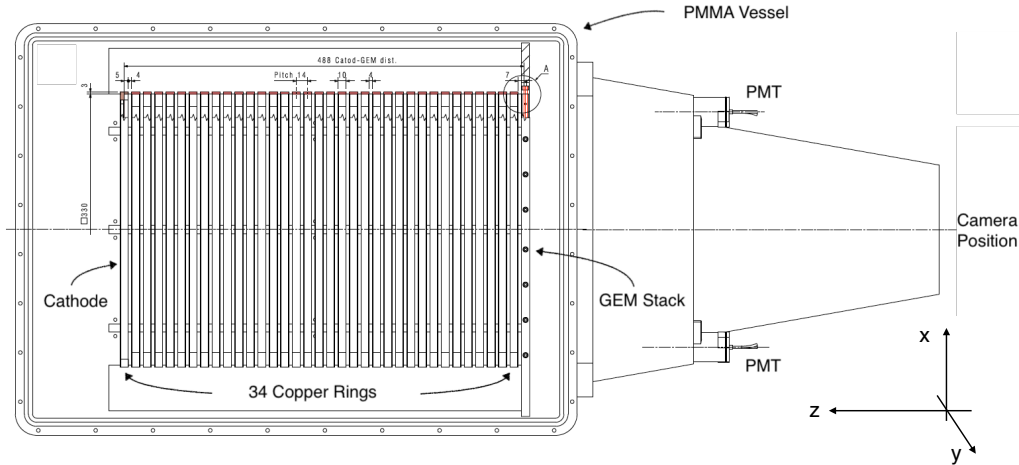


Figure 2.12. LIME schematics as seen from above, with notes indicating each part of the detector. Picture from [50].

composed by a series of LIME modules, with the idea of building larger detectors following this modular concept.

In the following, the main technical characteristics of the LIME components will be described.

First of all, it is important to note that the whole detector is enclosed in an external box, which differs in material between different data taking configurations. The external box as a dual function: it works as a Faraday cage and it also avoids that light or cosmic rays which survived the Gran Sasso natural shielding can reach the detector. Of course, feed-through connections are installed in order to allow the gas flow and the high voltages for the optical readout to enter inside the box.

2.3.1 Vessel and field cage

The gas vessel is a PMMA box with 10 mm thickness (except for the camera side where the thickness is reduced to 3 mm in order to ensure a more efficient light transmission). The box total volume is around 100 litres, but only half of it contains the heart of the detector.

In fact, a $330 \times 330 \times 488 \text{ mm}^3$ field cage ($\sim 53 \text{ L}$ of active volume) is installed inside the vessel (figure 2.13), devoted to the production of a uniform electric field which will be responsible for the drift of the electrons produced in the gas primary ionization towards the GEM plane. The gas mixture inside is generally some millibars above the atmospheric pressure, which translates the active volume in $\sim 87 \text{ g}$ of active mass when filled with He:CF_4 (60:40).

Moreover, the field cage can be described in more detail by dividing it into three main components:

- The **34 squared copper coils**, 10 mm wide and separated from each other by 4 mm (edge-to-edge), while the effective pitch (between the coils' center) is 14 mm. Each coil is connected to the closest coils by 100 M Ω resistors, such that an almost uniform electric field of $\mathcal{O}(1 \text{ kV/cm})$ is produced inside the cage.
- A **cathode**, also made of copper, 0.5 mm thin, which perfectly matches the frame of the field cage coils.
- The **amplification stage**, consisting of three superimposed GEMs, 50 μm of internal diameter for the holes and 140 μm pitch, separated from each other by 2 mm and the with first one 7 mm apart from the coils.

For calibration purposes, the upper side of the vessel features a 50 cm long and 5 cm wide window, sealed with a 150 μm thick ETFE (ethylene-tetrafluoroethylene) layer. This window acts as a track where a trolley with a X-ray-emitting radioactive source (such as ^{55}Fe) inside can be moved enabling controlled calibration along the

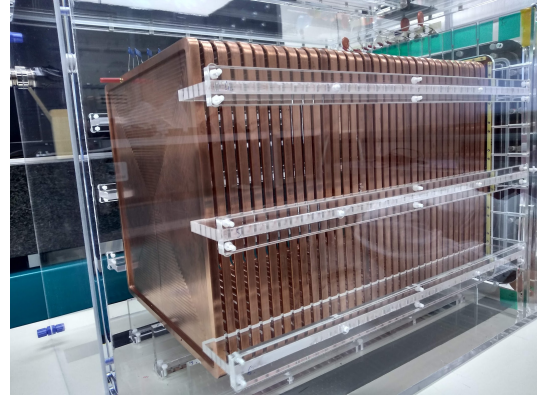


Figure 2.13. Picture of the LIME prototype from a diagonal perspective. From this view, one can see the cathode plane and the copper rings which ensure the uniform drift field, all enclosed in the transparent acrylic vessel. The GEM plane and the optical readout are covered by the rest of the detector, on the right [50].

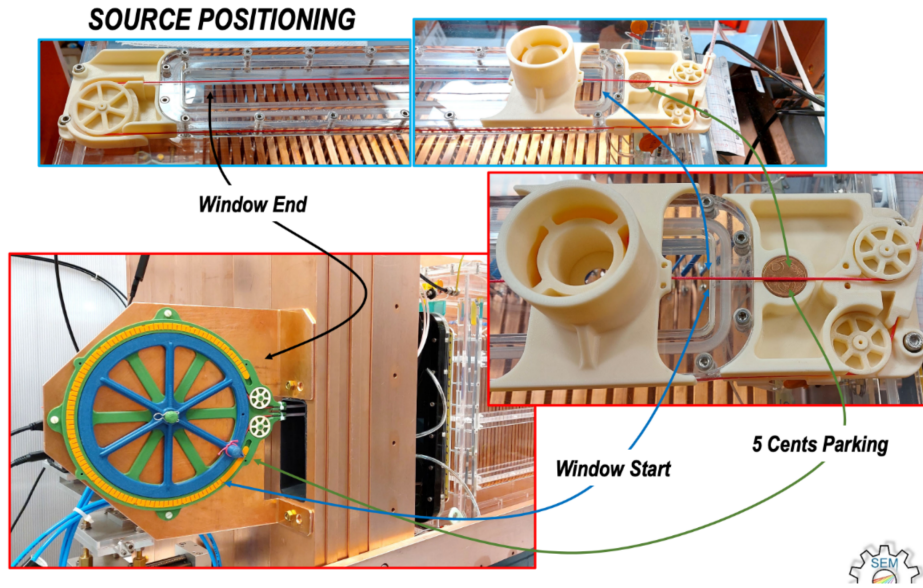


Figure 2.14. Detailed view of the automated mechanism which is used to move the trolley with the Iron source to calibrate the detector. The wheel on the left can be rotated with a small electric engine managed from an Arduino board in order to move the source from the start to the end of the upper window by means of a series of pulleys. When not required, the source can be put in a “parking” position, over a 5 cent coin.

z axis. In this configuration, the source is known to be 18 cm above the sensitive volume and a hole (5 mm in diameter) under the source allows the photon beam collimation.

More in detail, the trolley can be remotely controlled with an electro-actuated small engine, which moves a 3D-printed apparatus that is able to make the trolley slide between 5 cm and 45 cm far from the GEM plane (detailed view in figure 2.14). Everything is managed by an Arduino board and nicely integrated in the daily calibration procedure. Moreover, it will be useful to know, in the following, that the calibration procedure consists in 5 different *steps* in which the sliding range is divided, with step 3 being almost in the middle of the detector.

2.3.2 Optical readout

The optical readout is positioned outside the vessel, in front of the GEM plane, and is supported by a black PMMA conical structure, as it is visible in figure 2.12. Precisely, the optical readout is composed by:

- an Orca Fusion² sCMOS camera with 5.3 Megapixels (2304×2304 px), whose active area is $6.5 \times 6.5 \mu\text{m}^2$ for each pixel. Furthermore, a Schneider Xenon 0.95/25-0037³ lens with 25.6 mm focal length and a 0.95 aperture is installed at a distance of 623 mm from the GEM.
- 4 Hamamatsu R7378⁴ photomultipliers with a diameter of 22 mm.

²for technical details: [Hamamatsu \(Orca Fusion\)](#)

³for technical details: [Schneider Xenon 0.95/25-0037](#)

⁴for technical details: [Hamamatsu \(R7378 PMT\)](#)

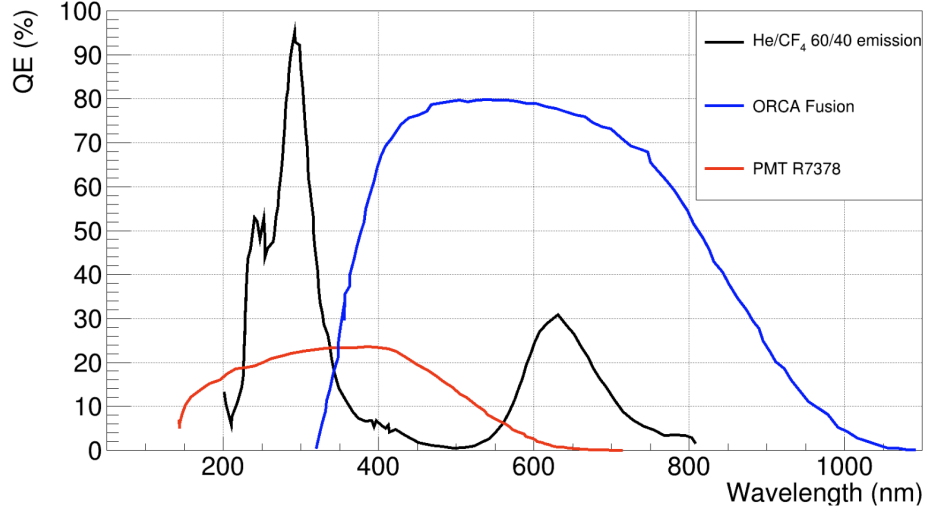


Figure 2.15. Quantum efficiency for both Orca Fusion and PMT R7378 (taken from Hamamatsu), with the He:CF₄ (60:40) scintillation spectrum superimposed [41]. The camera QE maximum precisely matches the He:CF₄ peak in the visible range.

As anticipated, one of the He:CF₄ electro-luminescence peaks is in the visible region, which is a nice match with the Orca Fusion best working point. In fact, the camera quantum efficiency (QE) is greater than 70% between 500 and 700 nm.

Moreover, in this configuration the camera acquires a surface of $35 \times 35 \text{ cm}^2$ which implies an area of $152 \times 152 \text{ }\mu\text{m}^2$ imaged by each pixel. The geometrical acceptance is instead $\epsilon_{\Omega} = 1.2 \times 10^{-4}$.

Another important technical detail regards the maximum time that the electrons can take to drift towards the anode plane: since the detector length in the drift direction is $L_{z,max} = 50 \text{ cm}$ and the drift velocity can be taken to be $v_{\text{drift}} = 5.47 \text{ cm}/\mu\text{s}$ [53], the maximum time of arrival will be $\Delta T_{max} = L_{z,max}/v_{\text{drift}} = 9.14 \text{ }\mu\text{s}$.

2.3.3 Voltage supply and gas flow

Two high voltage generators are employed to supply the detector components: ISEG HPn 500 and CAEN A1515TG. They permit the generation of a constant electric field inside the field cage ($E_{\text{drift}} \sim 0.9 \text{ kV/cm}$) and between the GEMs' transfer gaps ($E_{\text{transf}} \sim 2.5 \text{ kV/cm}$).

Moreover, the voltage difference between the GEM faces for each GEM layer is set to be $V_{\text{GEM}} = 440 \text{ V}$.

The gas mixture is continuously regenerated by flushing new gas inside the TPC to prevent its deterioration and to keep a stable light yield. The new gas is stored in cylinders and its flow, which can vary between 1 L/h and 20 L/h, is controlled by an independent gas system.

Since the start of the underground data taking, an additional recirculation system is employed and it can vary between 0 and 50 L/h, although its maximum value was never reached.

When exhausted, the gas is stored in another cylinder which is then disposed of by a designated company.

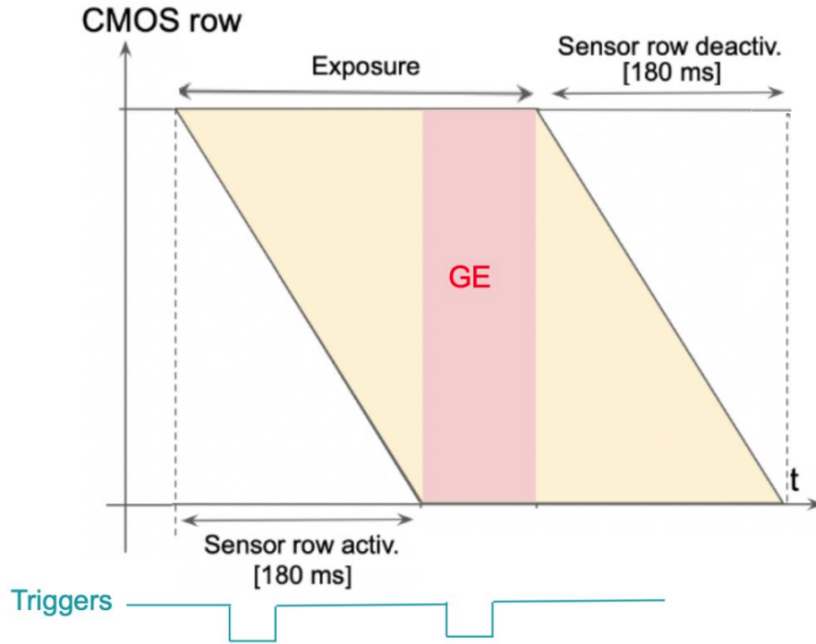


Figure 2.16. Visual representation of the camera data acquisition which is described in the text. The red region is called Global Exposure and it corresponds to the time window in which all the rows of the sensor are simultaneously exposed. In the lower part, the blue trigger line simulates two hypothetical events triggered and acquired from the PMTs, with the first one possibly not linked to any track in the corresponding image.

2.3.4 DAQ and triggering

LIME's data acquisition (DAQ) is designed to acquire both images from the camera and waveforms from the 4 PMTs.

This operation is performed by means of a triggered DAQ system developed within the Modular Integrated Distributed Analysis System (MIDAS) open source framework [54]: PMT signals are sent to a discriminator and data is acquired only when a coincidence is registered from the logic module in at least two PMTs during the so-called *Global Exposure* of the camera.

In fact, the camera data acquisition is peculiar and it is graphically described from figure 2.16:

- Before the start of the DAQ, a certain time of exposure is set, for the events taken underground this is typically 300 ms.
- Camera pixels need to be activated row by row starting from the top. Since each row needs 80 μ s to activate and the sensor has 2304 rows, this procedure requires about 180 ms in total. The same time is required to deactivate them at the end of the DAQ.
- Given the fixed exposure and the time of activation/deactivation, there will be a certain time frame where all the pixels are exposed: this is the *Global Exposure* (GE).

If the exposure is set to be 300 ms, the global exposure will be 120 ms.

To have an idea of the eV equivalent triggering threshold, a study was done evaluating the differential rate with an ^{55}Fe source with respect to the imposed

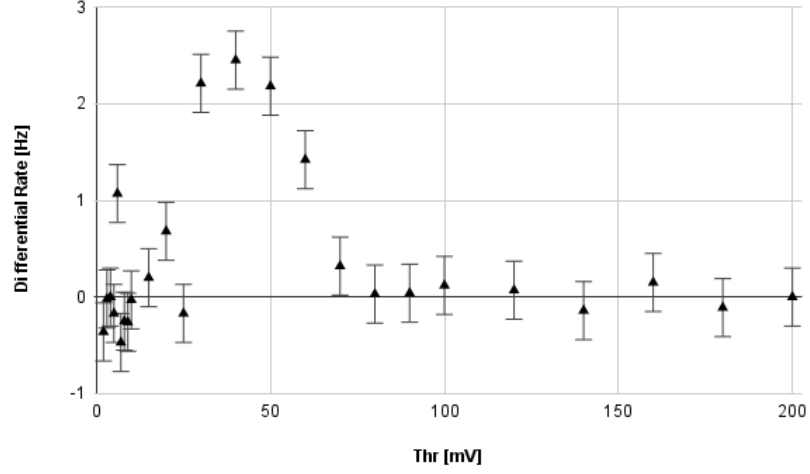


Figure 2.17. ^{55}Fe differential rate with respect to the triggering threshold. Plot from [55].

triggering threshold (where the differential rate is the number of triggered events at a certain threshold minus the number of triggered events at the next relevant threshold). After subtracting the background events, we see in figure 2.17 that an almost Gaussian shape appears with a mean value of 40 mV.

Given the fact that the ^{55}Fe emits X-Rays at 5.9 keV, a 1 mV increase in the triggering threshold correspond to 150 eV. In our case, the threshold is set to 2 mV for at least 2 PMTs in a sufficiently small time window, which means that we are triggering at 300 eV.

It is important to note that, as hinted from the “Triggers” blue line at the bottom of figure 2.16, it is possible that the PMTs are triggered outside the global exposure region. Since all the waveforms are acquired if a trigger happen during the GE, this can produce an image with a discrepancy between number of tracks and PMT events, if the ionization happens in a region of the chamber where the corresponding camera pixels are not activated, or with a partially acquired track, if it crosses only a part of the deactivated region.

Moreover, two different digitizers are employed to save the waveforms and they will be referred to as *fast* and *slow* digitizer:

- The **fast** digitizer is a CAEN V1742⁵. It is a 12 bit switched capacitor digitizer which is operated at a sampling rate of 750 MHz and acquires 1024 samples, with each sample corresponding to 1.33 ns.
- The **slow** digitizer is a CAEN V1720⁶. It a 12 bit ADC operated at a 250 MHz sampling rate, it acquires 4000 samples and each sample corresponds to 4 ns.

Having two digitizers acquiring waveforms with different time resolutions is crucial, since in the case of very tilted events it is possible to have waveforms which are too extended in time and could be cut when acquired with the fast digitizer only.

Hence, an *event* is in general defined by an image and $4n$ waveforms, where n is precisely the number of tracks in the image, if all the PMT triggers happened

⁵for technical details: [CAEN V1742](#)

⁶for technical details: [CAEN V1720](#)

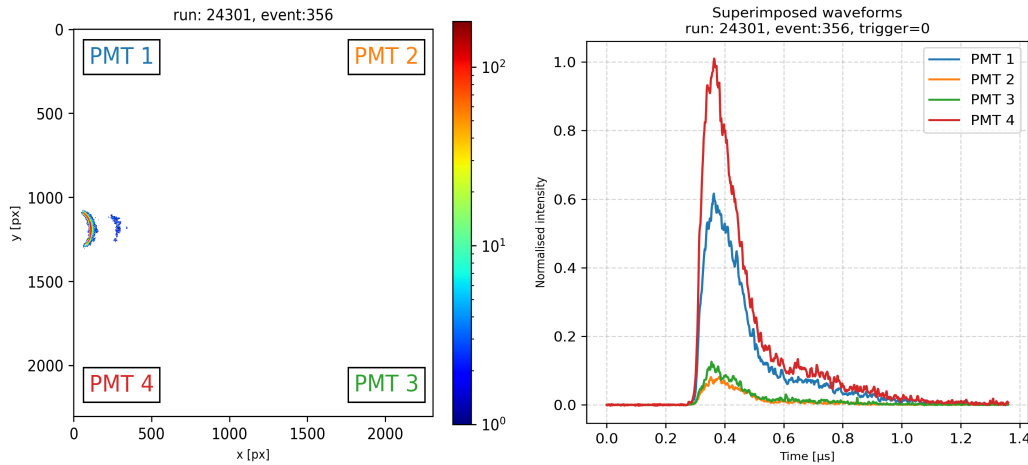


Figure 2.18. Example of a full event with only one track and four waveforms. The PMTs positions are labeled in the image reference frame, keeping coherence with the colors in the waveforms plot. From the waveforms relative intensity, it is easy to understand the position of the associated track by eye.

during the global exposure. An example event (after the reconstruction that will be described in detail in the next chapter) can be seen in figure 2.18, where the PMTs positions are reported with respect to the image reference frame. Note how the relative intensity of the waveforms hints the track position.

The DAQ is divided in *runs*, with each run composed by an average number of 400 events.

At the same time, different data taking campaigns are labeled with the prefix **RUN** followed by a number, which identifies a different configuration for the shielding. At the moment the following campaigns have been performed at LNGS:

- **RUN 1**, no shielding;
- **RUN 2**, 4 cm of copper;
- **RUN 3**, 10 cm of copper;
- **RUN 4**, water shielding + 10 cm of copper;

and RUN 5 is ongoing right now with a low gain campaign, to attempt the measurement of the environmental neutrons underground.

Different kind of data taking configuration are implemented and labeled with different names. Some of them are highly specific and related to a particular radioactive source present during the acquisition. Others, instead, are more frequently repeated and it is worth spending some lines to describe them, in order to better understand the terminology that will be used later on:

- **Standard / Background:** as the name suggests, this is the most common configuration and it permits to acquire background data related to the shielding configuration, based on the campaign.
- **Stability:** in this case, an ^{55}Fe source is positioned midway in drift direction, such that the detector stability can be evaluated.

- **Daily Calibration:** this is a multi-step procedure which involves the source motor described in section 2.3.1. As already described in the same section, the ^{55}Fe is moved between 5 different positions in the drift direction z , in order to evaluate the detector performance at different z .

In particular, a large part of my work is based on RUN 3 data, and a first look at the analysis will be given in the next chapter, together with the description of the reconstruction algorithm, which is a core building block of the experiment and it includes the reconstruction and organization of both camera and PMTs data.

Chapter 3

Data reconstruction and pre-processing

As illustrated in Chapter 2, the CYGNO experiment and, in particular, the LIME prototype data taking relies on an optical readout which consist of a sCMOS camera and 4 PMTs, collecting light produced in the amplification stage.

From the camera, one can extract information about the imaged tracks in the GEM plane ($x - y$), while from the PMTs it is possible to evaluate the evolution of the tracks in the drift direction (z).

Since this is an innovative readout approach in the time projection chambers field, a completely original algorithm has been developed to extract and organize information from both sCMOS and PMTs. In the following, this will be referred to as “*the reconstruction code/algorithm*”.

This chapter will be divided in two main blocks:

- First, it will be provided a description of the most important concepts behind the reconstruction algorithm.
- Then, the focus will shift to the first part of my contribution to the CYGNO experiment. In particular, this is about the characterization of the detector stability underground, during a particular RUN 3 sub-campaign involving an Americium-Beryllium source, with the aim to provide a reliable and general pipeline to perform data pre-processing oriented towards energy calibration.

Moreover, the results of this technique will be presented at the end of the chapter by means of energy spectrum comparisons and subtraction.

3.1 Image reconstruction algorithm

Very different shaped trajectories can be imaged within the LIME active volume. For examples, alpha particle will be straight, dense and shorter than MIPs (like e.g. muons), which will be way longer and thinner. At the same time, Compton electrons will produce thin and curly tracks with a dense Bragg peak at the end.

Some labeled visual examples of tracks pictured from the LIME prototype can be seen in figure 3.1.

Given the very different topology of the tracks, the reconstruction algorithm has been developed in such a way to be flexible enough to identify pattern in a meaningful way. To do so, a four-step procedure is adopted, denominated **iDDBSCAN**¹ [56],

¹the meaning will be made explicit in the following

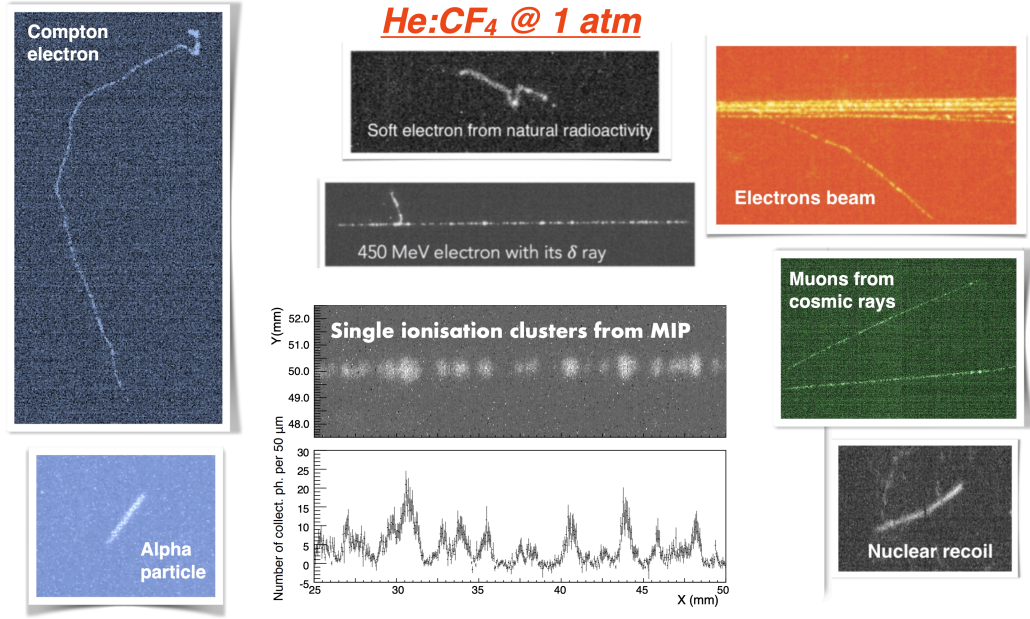


Figure 3.1. Examples of tracks pictured from the LIME prototype.

with the final output being a collection of pixels related to each identified track, together with their position in the image reference frame and a set of features which describe the tracks in detail.

3.1.1 Noise reduction

The sCMOS camera outputs a 2304×2304 matrix, filled with the raw counts of the number of photons hitting every single pixel. These counts are obviously affected by a number of possible noise sources, such as electronic noise or thermal energy.

The noise reduction step is divided in multiple substeps:

- First of all, noise is evaluated at the start of each data taking by acquiring a set of *pedestal* runs, namely $\mathcal{O}(100)$ images with GEMs off, producing a pedestal map which stores the mean noises μ_i and their standard deviations σ_i , evaluated for each pixel with a gaussian model.
- At this point a pixel-by-pixel subtraction is performed for each acquired image, ending up with a clean image.
- Moreover, pixels with intensity lower than the mean value plus a certain threshold zs_{thr} (*zero-suppression threshold*) are set to zero, suppressing the channels which are retained to be empty, in order to reduce the amount of information carried on. This threshold is typically related to the pedestal standard deviation, i.e. $zs_{thr} = 1.3\sigma$
- At the same time, the upper and lower regions of the image are removed from the reconstruction (see figure 3.2, left), since they are far noisier than the rest of the sensor (they constitute the tail of the noise rms distribution in the right panel of figure 3.2).

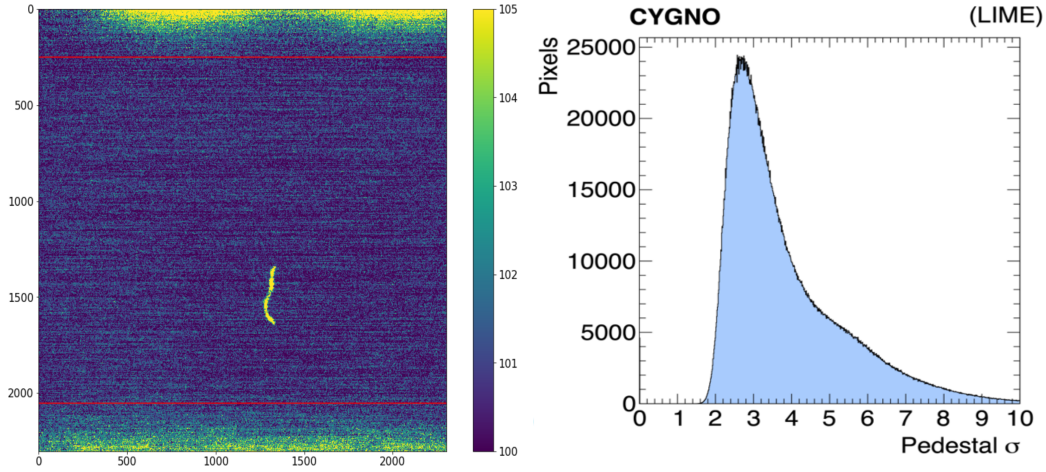


Figure 3.2. sCMOS image with the excluded regions delimited from the red lines (**left**) and noise rms distribution extracted from a set of pedestal runs (**right**) [50].

- To further reduce electronic noise fluctuation and to limit the computing power requirements, the image is re-binned by averaging counts in 4×4 matrices and by applying a median filter [57], ending up with a 576×576 matrix of pixels.

As a side note, it is important to talk about the *interactions on the sensor*. The name is self-explanatory: it is possible for particles to interact directly on the P-N junctions inside the camera sensor, producing meaningless information which cannot be compared with the rest of the data.

3.1.2 Vignetting correction

When an image is acquired with a generic camera, the intrinsic geometry of its lens will produce a distortion effect known as *vignetting* [58]. This is in particular due to the lens inclination with respect to the light cone, which in the end reduces the solid angle covered by the lens itself as a function of the distance from the optical center.

To overcome this problem, a vignetting map has been produced by imaging a white piece of paper, uniformly illuminated, and it is applied to the re-binned pixel matrix produced at the end of the previous step.

The vignetting measurement of the Schneider Xenon 0.95/25-0037 lens adopted installed in the LIME setup is reported in figure 3.3, together with an example of the vignetting map.

Note that the vignetting correction needs to be performed after the noise reduction, in order to avoid noise enhancement.

3.1.3 iDBSCAN seeding

After the noise reduction and the vignetting correction, the resulting pixel matrix is passed to the **clustering algorithm**, which is a variation of the “density-based spatial clustering of applications with noise” (**DBSCAN**) algorithm [59].

DBSCAN is a very popular technique in a variety of research field, it is able to group data points which are close in space into clusters. This algorithm features two main parameters: a radius ϵ , and the minimum number of data points into an hyper-sphere of radius ϵ required to form a cluster, N_{min} .

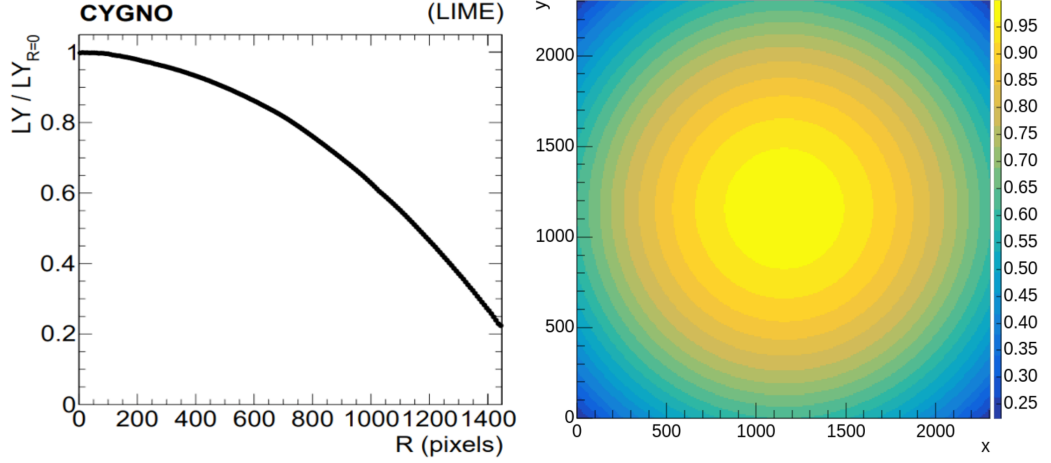


Figure 3.3. Measure of the vignetting effect, obtained by imaging a uniformly illuminated white sheet of paper (**left**) [50], together with the vignetting map (**right**) [48].

The adaptation developed for the CYGNO experiment [60], is named intensity-based DBSCAN (**iDBSCAN**), because of the fact that the pixels' intensities are used as weights, with N_{min} taking the role of the minimum total intensity to create a cluster. This approach helps in the discrimination of real clusters with respect to noise-induced clusters which for some reason did survive to the noise reduction steps.

In this context, the algorithm works as follows:

- When the sum of the intensities in an hyper-sphere of radius ϵ reaches the N_{min} value the center pixel is *activated* as a *seed* and the cluster is formed.
- The same is done for the neighbor pixels, adding them to the cluster until the condition is no longer met and the pixels of the cluster are removed from the image.
- This procedure is repeated over all the pixels until no cluster can be identified.

At the same time, a possible directionality between the identified clusters is evaluated with a linear RANSAC procedure [61]. First, half of the clusters are extracted from the pool and they are fitted with a line. Then, the ratio between the number of data points sufficiently close to the fitted line and the total number of clusters in the sample is computed. When this ratio is bigger than a certain threshold, the clusters are marked as *directional*, from which the algorithm takes the second "D" in iDDBSCAN.

If no directional cluster is produced, the algorithm ends here and the collection of clusters is returned. Otherwise, a last step concludes the procedure.

3.1.4 Directional super-clustering step

In this step the iDBSCAN algorithm is re-applied, this time starting from the tagged coordinates extracted from the previous step, employing a third order polynomial and three additional parameters: the width ω , a radius ϵ_{dir} and a distance $\rho_{isolation}$.

The idea is that for every pre-identified cluster a fit is computed. Every additional cluster which is inside a window of width ω around the fitted curve and inside a

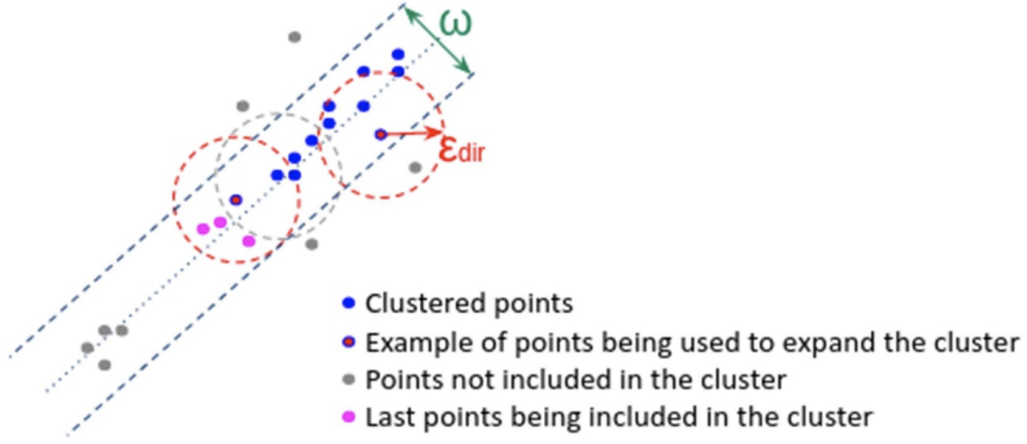


Figure 3.4. Visual description of the directional super-clustering procedure, taken from [56].

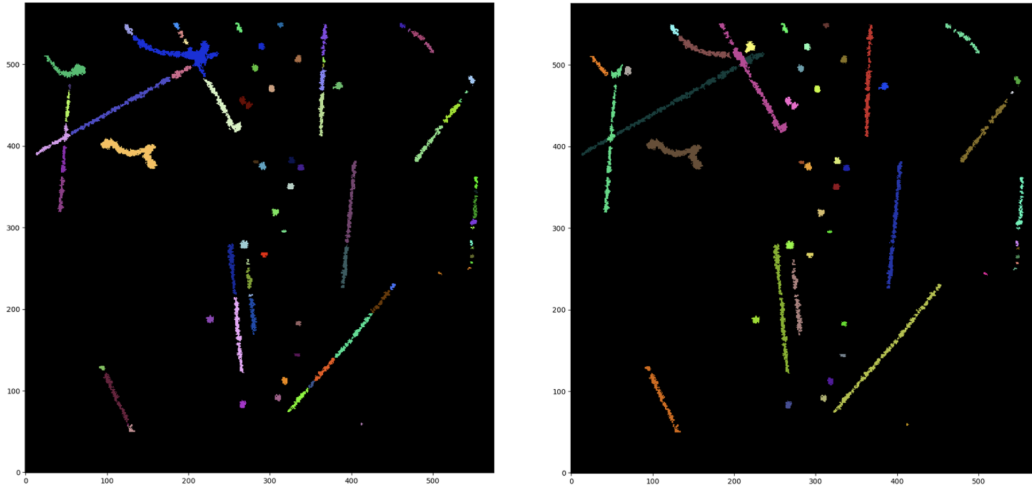


Figure 3.5. Examples of results produced by the iDBSCAN (left) and iDBSCAN (right). It is clear how the additional directional super-clustering step helps in connecting correlated clusters while also disentangling eventual superpositions. Pictures from [56].

circle with radius ε_{dir} , centered in one of the clustered points, is added to the cluster. At the same time, every pixel at a distance smaller than $\rho_{\text{isolation}}$ from the directional clusters is classified as noise and removed from the image (this is called *isolation seeding*), together with all the super-cluster's pixels.

The procedure is repeated for every directional cluster provided at the beginning of the step until all the super-clusters are formed.

In the end, the iDBSCAN is performed one last time on the remaining pixels, in order to identify clusters which survived the super-clustering and isolation seeding, hence small clusters which were not cut from the first iteration or which were obscured by closer longer tracks. In figure 3.5, one can have a look at the difference between clusters produced by the iDBSCAN only (**left**), where most of the long tracks are split up, or the ones produced after the iDDBSCAN (**right**), where clusters which are part of a single long track are joined together and overlapping tracks are disentangled.

3.1.5 Super-clusters relevant features

A plethora of features are computed and stored, all based on the iDDBSCAN output. Understanding their meaning will be fundamental to follow the analysis procedure and results.

While for a full description of every variable one can rely on the [GitHub Wiki page](#), here I would like to describe only the more relevant features:

- **sc_integral** is the sum of all the intensities inside the pixels which compose a super-cluster. Since the pixel intensity is proportional to the energy deposit, its value can be calibrated to reproduce the energy of the cluster.
- **sc_length** and **sc_width** are, obviously, the length and width of the super-cluster. They are computed by means of a singular value decomposition (SVD), which produces two eigenvectors (and relative eigenvalues) corresponding to the major (length) and minor (width) axis of the ellipsis which encloses the super-cluster.
- **sc_nhits** and **sc_size** are, respectively, the number of pixels with intensity bigger than the zero-suppression threshold and the total pixels in the super-cluster.
- **sc_xmean** and **sc_ymean** are the super-cluster's mean coordinates. With the same logic also maximum and minimum coordinates exists as features.
- **sc_tgausssigma** is the cluster's Gaussian transverse profile standard deviation and it can be interpreted as a measure of the dispersion in the transverse direction, which in turn gives an idea of the cluster position in the drift direction. This idea will become relevant in the next chapter.

By exploiting the uni- or multi-variate distributions of such features or their combination, one is able to select only the clusters which are relevant for a given analysis, as it will be seen in the following.

A fundamental remark is that all the spatial features are related to the bi-dimensional projection of a (in general) three-dimensional track. The super-cluster length, for example, is strongly dependent on the real track inclination with respect to the GEM plane. As already stressed before, PMTs data are the key to correct these approximation.

3.2 PMT data integration

At the moment of writing this thesis, the PMT reconstruction is in its first stable integration with the camera reconstruction [62]. The algorithm is far simpler with respect to the camera one, since there is no clustering involved and the complex part is the organization of the huge amount of data which is acquired which I will not detail here.

To have an idea of the volume of data: there are 4 acquisition channel, one for every PMT, with two digitizations, as discussed in section 2.3.4, to avoid the possibility of cut signal in the case that a track light arrives at the PMTs in more than $\Delta t_{\text{fast}} \simeq 1.5 \mu\text{s}$ (which is the time window of the fast digitizer), which result in 8 waveforms per event (4 with 1024 samples and 4 with 4000 samples).

Storing every waveform is not meaningful, since their relevant information can be enclosed in features one can extract during the reconstruction. I will now give a brief description of the most important ones:

- **pmt_wf_tot_integral** is the total area under the waveform and can be converted in units of charge, effectively providing a measure of the energy released from the particle in the detector.
- **pmt_TOT_[time/area]** are the so-called time and area over threshold, namely the time over which the waveform is extended over a certain threshold and the area under the waveform and over the same threshold. These features are useful because of the fact that they can provide, respectively, a measure of the electrons' drift time and a refined measure of the charge collected (with respect to **pmt_wf_tot_integral**).
- **pmt_wf_insideGE** is a flag to monitor whether a given triggered event is inside the camera global exposure (see section 2.2.1).
- **pmt_wf_TTT** is instead the trigger time tag, from which one can determine if an event will be inside the picture.
- **pmt_nPeaks** is the number of peaks in the waveforms and can be crucial to select events from ^{55}Fe used for calibration purposes excluding background data like cosmic rays, since iron events' waveforms have a single sharp peak, while cosmic rays can produce waveforms with several peaks.

These are the building blocks to achieve the previously mentioned 3D reconstruction, but there is still one major problem to solve: the **track-waveform association**.

In fact, while the PMTs' data can be uniquely associated to the corresponding image by means of the trigger time tag, when more than one cluster is present inside the same image there is no way to determine, a priori, which set of waveforms is connected to which cluster. This has to be done a posteriori, for example making use of the relative charge seen from each PMT with respect to the others when the clusters are sufficiently localized. For long and curly tracks the task is more difficult, since they can be closer to different PMTs in different time periods, but in principle this is also not so important given the fact that they will be for sure classifiable as background.

To better understand the geometrical meaning of this reasoning, an example is provided in figure 3.6, where the 4th PMT is the brightest and in fact the track is closer to the left-bottom corner (the PMTs' positioning is reflected in the waveforms' arrangement in the figure).

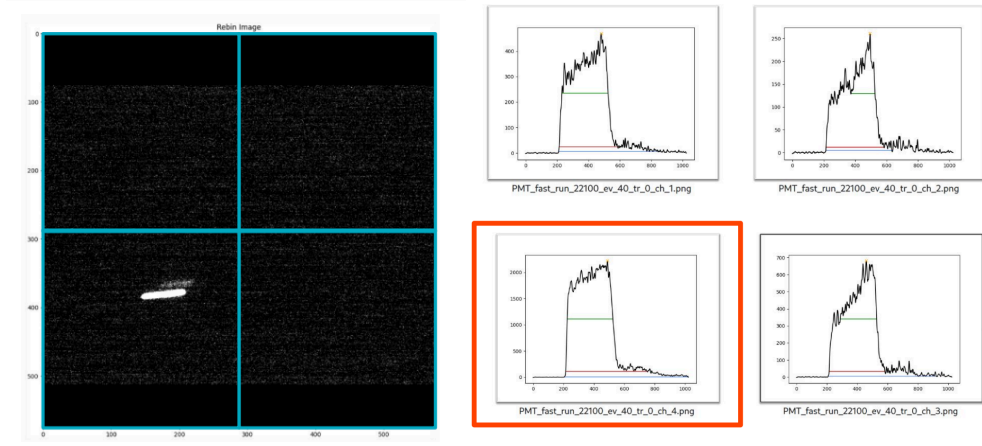


Figure 3.6. Practical explanation of the relative waveform amplitude association method. Since the fourth PMT collects more charge, the track must be closer in space to that PMT. Note that the PMTs’ positioning is reflected in the waveforms’ arrangement in the figure. Picture reproduced from [63].

Clearly, this is useful in presence of very low pile-up and/or when the tracks are separated enough, which indeed is the case most of the time when data is collected underground. If instead two tracks show up in the same “quadrant”, a simplistic approach like the one just described could easily fail.

To overcome this problem, different approaches have been proposed to provide a reliable algorithm which is able to retrieve the track bi-dimensional position in the GEM plane starting from the light seen from the PMTs.

In general, the core idea which guides these approaches has its roots in the theoretical scaling law of light between the GEM plane and the PMTs, which in turn depends on the flux isotropy and PMTs’ geometrical acceptance, which is schematized in figure 3.7.

In principle, denoting with L the true light intensity and with L_i the light seen from the i -th PMT, one should expect $L_i \propto R^{-2}$ (with R being the radial separation between GEMs and PMTs) assuming an isotropical distribution for the light. Moreover, given that the light does not arrive perpendicular to the PMTs, one should also take into account the extra power provided by $\cos \theta = h/R$ where h is the distance between PMTs and the GEM plane, as it can be seen in figure 3.7.

Hence, the scaling of light at the PMTs plane should be with R^{-3} , i.e.:

$$L_i = \frac{L}{R^3} \quad (3.1)$$

Nonetheless, this is not true. In fact, from the results of a study (in which I myself participated) undertaken at LNF [65] and confirmed with underground data [64], the real scaling law is closer to the one reported in equation 3.2 and appreciable in figure 3.8, where the light is converted in charge and r is the distance in the $(x - y)$ plane, while z is the distance from the emitting plane.

$$L_i = \frac{L}{R^4} \quad (3.2)$$

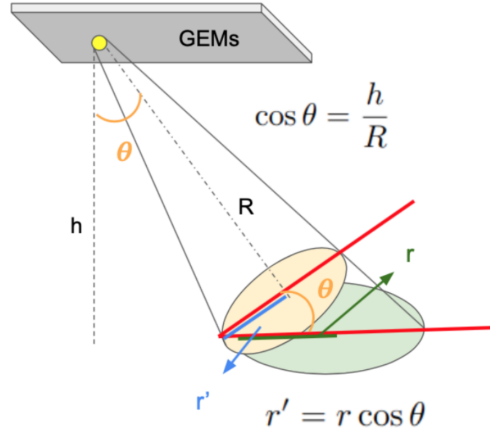


Figure 3.7. Sketch of the PMT geometrical acceptance with respect to the light emitted from the GEM plane. Picture reproduced from [64].

In the same work, the possibility of involving deep learning for position reconstruction from the raw waveforms was tested using two different artificial neural networks architectures achieving $\mathcal{O}(\text{cm})$ precision with spot-like events induced with a LED [65].

At the same time, another promising approach currently under improvement involves a Bayesian fit which tries to retrieve the real light L and the bi-dimensional position from the waveforms' integral. It was tested with spot-like events induced with an ^{55}Fe source and produced promising results [64, 66].

The discussion regarding the PMT data integration with the camera data will be resumed in the next chapter.

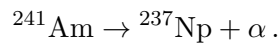
3.3 AmBe data analysis and pre-processing techniques

In this section I will outline my contribution to the data analysis strategy. In particular, my work was focused on the **AmBe sub-campaign** of the RUN 3 data taking campaign (with the description of the different data taking campaigns given in section 2.3.4).

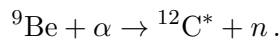
Americium-Beryllium (hereafter $^{241}\text{AmBe}$ or simply AmBe) is a very common calibration source, particularly popular for its neutron production capability.

In fact, the AmBe source is achieved by putting an ^{241}Am radioactive source inside a ^9Be container. The neutron production chain works as follows:

- The ^{241}Am decays in Neptunium, emitting an alpha particle (^4He)



- The alpha particle is caught by the Beryllium atoms in the container, forcing it to emit a neutron



Note that neutrons are emitted with a really broad kinetic energy spectrum, ranging from 0 to 11 MeV, as reported in figure 3.9.

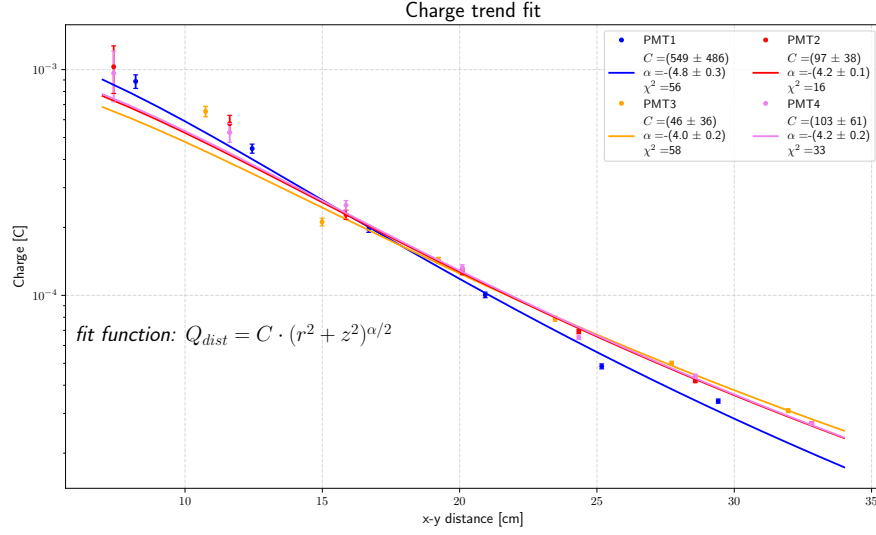


Figure 3.8. Fit of the charge trend vs. $(x - y)$ position of impulsed light. The fit clearly shows a R^{-4} scaling law. Note that in the figure r is the distance in the $(x - y)$ plane, while z is the distance from the emitting plane. Plot from [65].

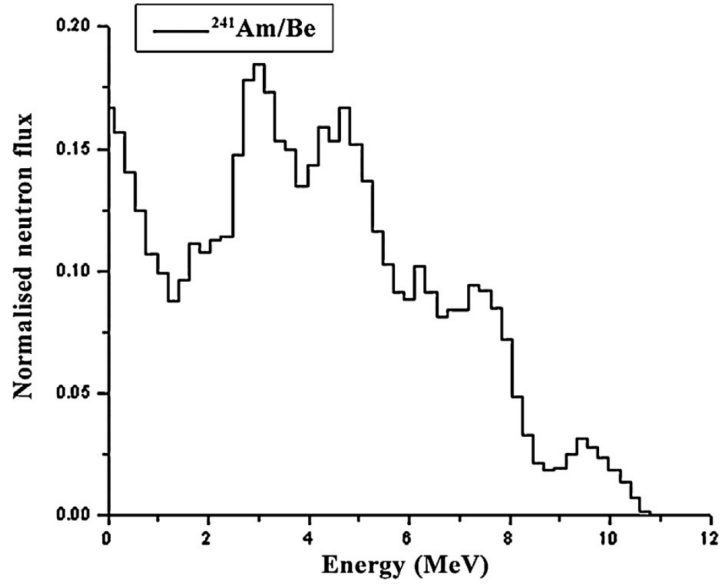


Figure 3.9. $^{241}\text{AmBe}$ neutron spectrum from [67].

Table 3.1. Relevant information about the datasets involved in the analysis. The exposure of the Bkg_stability run is not provided since it is just one calibration run taken with the ^{55}Fe source at the step 3 of the calibration procedure (as described in section 2.3.1). Moreover the year in “date range” is omitted and implied to be 2023.

Alias	Runs range	Date range [mm/dd]	Eff. Exposure [h]
AmBe_p1	23820 ÷ 23984	08/02 ÷ 08/03	16.9
AmBe + Fe	23988 ÷ 24022	08/03 ÷ 08/03	2.1
AmBe_p2	24023 ÷ 24328	08/03 ÷ 08/04	31.0
Stability	25486 ÷ 25772	09/26 ÷ 09/29	10.5
Bkg	27322 ÷ 27844	10/07 ÷ 10/10	56.6
Bkg_stability	27512	10/09 ÷ 10/09	–

The interest in having a neutron source is given by the fact that low-energy neutron can induce low-energy nuclear recoils, which can mimic WIMP-induced nuclear recoils, hence providing a nice test of the detector performance in the region of interest of the parameter space.

In particular, the source provided by LNGS was characterised by a neutron activity $I = 200 \text{ n/s}$.

It is also worth mentioning that the neutron channel, while being the most interesting for our case-study, it is not the only one for the AmBe source. In fact, other relevant emissions are:

- A 4.43 MeV γ ray, emitted from the $^{12}\text{C}^*$ de-excitation, which should produce Compton electrons inside the gas volumes.
- A 59 keV γ ray, emitted from the ^{241}Am alpha decay in ^{237}Np with a probability of 58%, which should be absorbed in large part by the copper shielding.

The first time this dataset was analysed [66], some criticalities emerged, in particular:

- The background data with which the AmBe data was compared was largely contaminated due to a faulty filter line, hence producing strange excesses in various regions of the energy spectrum.
- Due to unstable humidity, the light yield (LY) during the whole data taking was not stable.

While during the first analysis this criticalities have been mitigated with a specialized technique, what I will propose in the following is a more general multi-step procedure with the aim to provide a trustworthy pre-processing technique which can be always applied before the actual data analysis.

Together with the AmBe campaign (which consists in two parts with the AmBe source only, separated by some runs taken with the ^{55}Fe and AmBe source), this analysis will feature also background and stability data taken in a different time period with respect to the AmBe data.

Meaningful information about these sets of runs are reported in table 3.1. Note that in the following “**AmBe**” will be used as alias for the AmBe_p1 + AmBe_p2 dataset.

Note that the background data is not the same which was used in [66]. In fact, the already mentioned contaminated filter line which was used at the beginning of August 2023 was substituted with a new one at the beginning of October 2023. This data taking was much cleaner and it will be easy to see this in the following.

Moreover, all the plots presented from now on will be produced by first removing from the dataset all the clusters which have:

- `sc_tgausssigma-0.152 < 0.5`, to remove the fake clusters and the interactions on the sensor as described in [66].
- $(x - y)$ coordinates sufficiently close to the image borders and, at the same time, `sc_length < 25 mm`, to remove tracks which are produced outside the camera FoV and which are cut from the image frame. If longer than 25 mm, they are probably still interesting tracks to keep and the loss of information generated from than cut can be considered negligible.

The goodness of these cuts will be underlined in the following sections with the ease of appropriate examples.

3.3.1 Data normalisation

The first important point that needs to be clarified is the data normalisation procedure which, in previous works, was pursued by simply dividing for the number of images acquired during the considered run range.

This for sure provides a good estimate of the measured rate, but to do this in a consistent way one needs to normalise the considered measured quantity to the actual time of exposure, which can be computed by extracting the start and stop time of each run from the SQL database which works as a logbook, where all the information about the data taking is stored.

Of course, all the pedestals and daily calibration runs are removed from the calculation, resulting in the “*effective exposure*” reported in table 3.1.

It is evident, from the comparison between raw and normalised data in figure 3.10 (upper panels), that this normalisation procedure nicely aligns the distributions of `sc_integral` of the Bkg and AmBe datasets, enhancing the excess produced by the AmBe source (more on the excess will be treated in the following). Note that the first peak of the histogram seems to be worsened from the normalisation, since it is no longer aligned as it previously was.

In principle this is not a problem, since clusters associated to very low `sc_integral` are known to be related to fake clusters/sensor noise, hence they are not meaningful from the physics point of view. Moreover, we can use the Iron peak extracted from the AmBe+Fe dataset, reported in the lower panels of figure 3.10, as a reference to roughly understand the energy scales: 5.9 keV (the iron peak) is approximately equivalent to 10k Analog-to-Digital Units (**ADU**). A better calibration for the energy scale will come near the end of this chapter and an explanation regarding the evident peak around 20k ADU will be given in the next subsection.

3.3.2 Gain non-uniformity correction

Tracks in different regions of the GEM plane experience different gain. To quantify this difference, a correction map was produced from the study of the response of a non-collimated ^{55}Fe -induced spots, where with “non-collimated” it is intended that the hole under the source (as described in section 2.3.1) is enlarged to make sure that the photons are emitted in a more isotropic way.

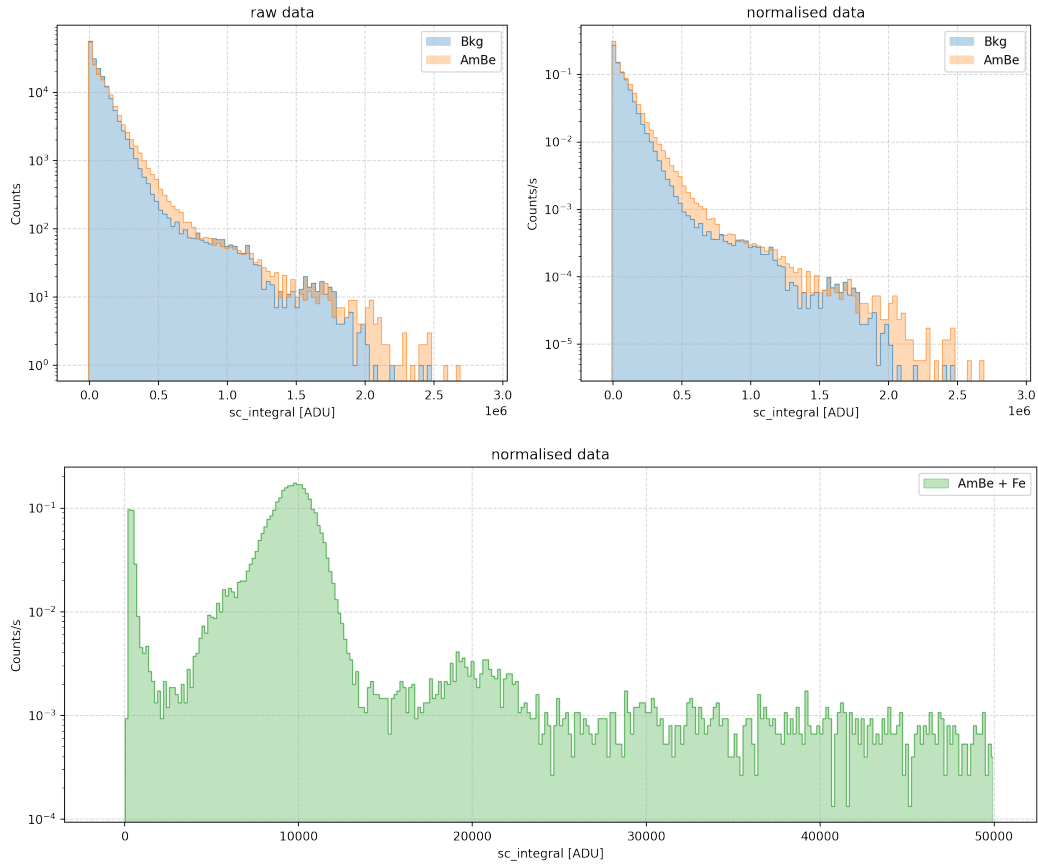


Figure 3.10. Comparison between raw and normalised distribution of `sc_integral` for both Bkg and AmBe datasets (**upper**) and normalised distribution only for the AmBe+Fe dataset (bottom), with the latter reported to have a rough idea of the energy scale (5.9 keV are approximately 10k ADU).

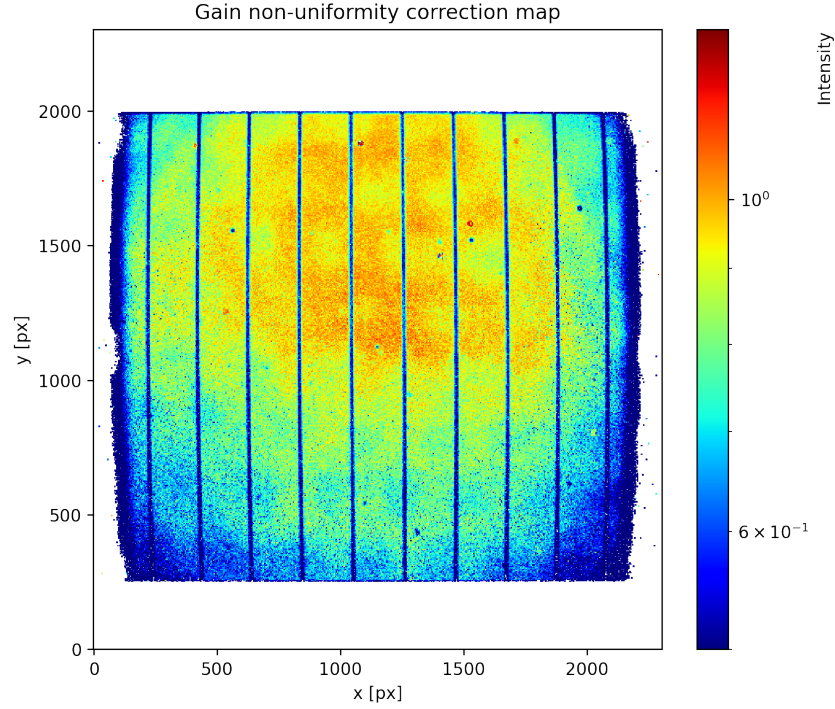


Figure 3.11. Color-map of the gain correction map, which is telling us that, on average, clusters in the upper region of the GEM plane experience an enhanced gain with respect to clusters in the lower region. The vertical lines are the GEM sectors.

From figure 3.11 (where the correction map is reported) it is clear that (on average) tracks produced in the upper region of the GEM plane experience an enhanced gain with respect to tracks produced in the lower region. Moreover, the vertical sectors of the GEMs appear and correspond to blind regions of the detector. A characterisation of the correction map efficiency is reported in appendix A.

To verify the benefits of the non-uniformity correction, figure 3.12 shows the comparison of Bkg, AmBe and AmBe+Fe normalised data before and after the correction, with and without cuts on borders and `sc_tgausssigma`. Actually, the most meaningful comparison is the one regarding the AmBe+Fe data, where the left tail after the correction is definitely more gaussian.

At the same time, looking from top to bottom we can appreciate the benefits provided when removing short tracks cut in half from image borders, the excess around 5k ADU disappears after the cuts, since the majority of this region is populated from parts of Iron spots.

It is also worth to comment the peak around 20k ADU in `sc_integral`: this is related to iron spots which are too close between each other and are reconstructed as a single cluster from the reconstruction algorithm.

At the same time, another excess appear when the cuts described at the end of the introduction to section 3.3 are not applied. This excess is concentrated around 5k ADU and it is generated from iron clusters which are cut in half from the picture borders.

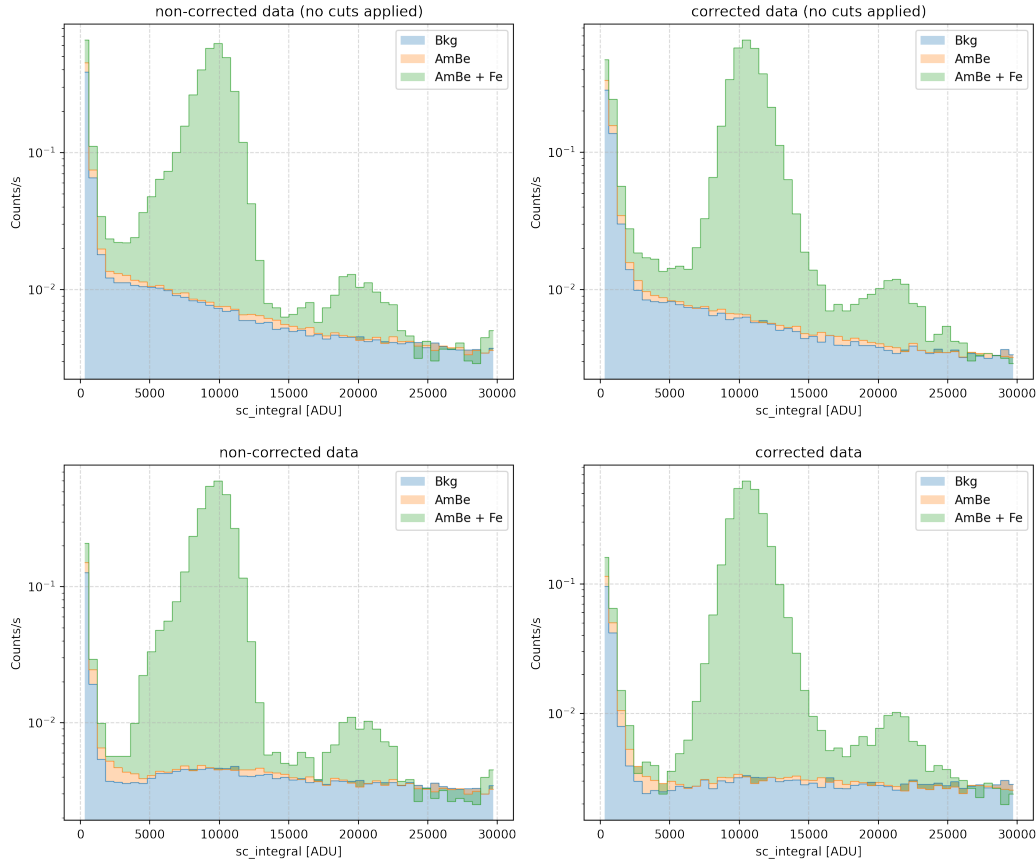


Figure 3.12. Comparison of Bkg, AmBe and AmBe+Fe data before (**left**) and after (**right**) the correction for the gain non-uniformity. A nice check of the correction goodness is given from the iron peak, which after the correction assumes a more gaussian distribution. Moreover the peak around 20k Analog-to-Digital Units (ADU) in `sc_integral` is produced from two iron spots clustered together from the clustering algorithm. At the same time, the upper and lower panels' difference is addressed to the application of the cuts described just before section 3.3.1: the region around 5k ADU is clearly less populated after the cuts, since the Iron spots which are cut from the picture frame are removed.

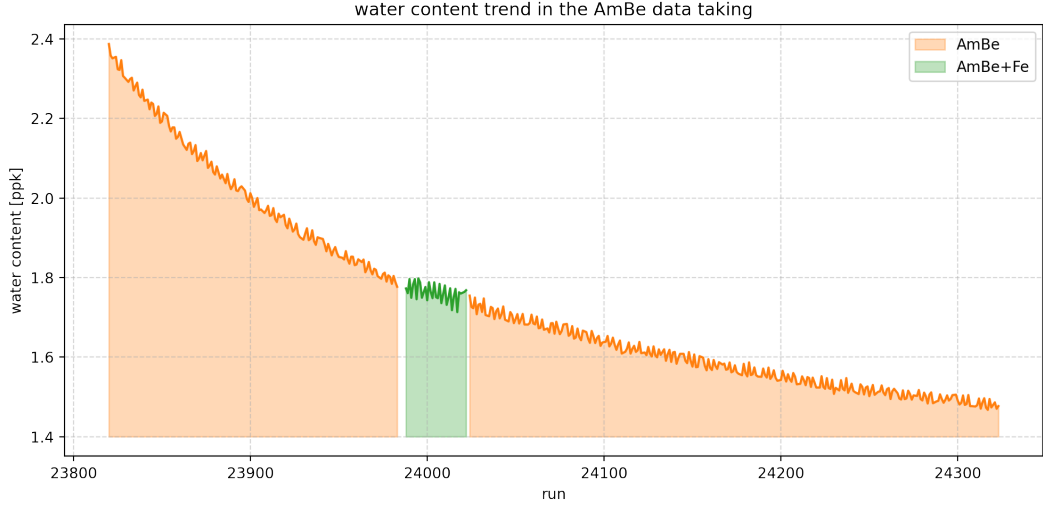


Figure 3.13. Humidity trend during the AmBe data taking. Humidity is reported in parts per thousand (ppk).

3.3.3 Light Yield stability and calibration

Unfortunately, the humidity level inside the detector during the whole AmBe data taking was slowly decreasing, as plotted in figure 3.13. Variations in humidity are reflected in variations of the Light Yield (LY) and the aim of this part of the pre-processing is to quantitatively determine the relation between light yield and humidity in order to correct `sc_integral` for this effect. At the same time this opens the possibility to correctly calibrate ADU to keV.

In fact, the procedure is performed by fitting the Iron peak run by run in the Stability dataset (see table 3.1), from which one can determine the LY vs. humidity relation and, assuming linearity, fit it with a linear function. At this point, from the fit parameters, one can retrieve the calibration factor at each humidity value and the LY can be equalized by dividing for it. Multiplying by 5.9 keV (the real energy of the Iron peak), `sc_integral` is correctly calibrated from ADU to keV. A summary of the just described procedure is reported in figure 3.14 (**top**), together with the same steps applied to the already calibrated energy (**bottom**), which shows how the LY vs. Humidity dependence is flattened after the calibration, as expected. Note that the top panel of each sub-figure reports only a sample of the fitted runs, for the sake of simplicity.

In principle, a single fit of this kind should be enough to obtain a calibration function which can work for every data taking, but in reality (as we will see) the light yield is influenced from various detector parameters in a possibly non-linear way, as e.g. the Oxygen level inside the detector, and it is not always feasible to correct all of them.

We can also reproduce the bottom plot in figure 3.14 with the AmBe+Fe dataset. The results, reported in figure 3.15, show again the flattening of the LY vs. Humidity dependence as in the previous case, but we can also see how the peaks are shifted from the true value. Note also that in this case all the data points are reported.

A further characterization of this fact will be provided in the following sections.

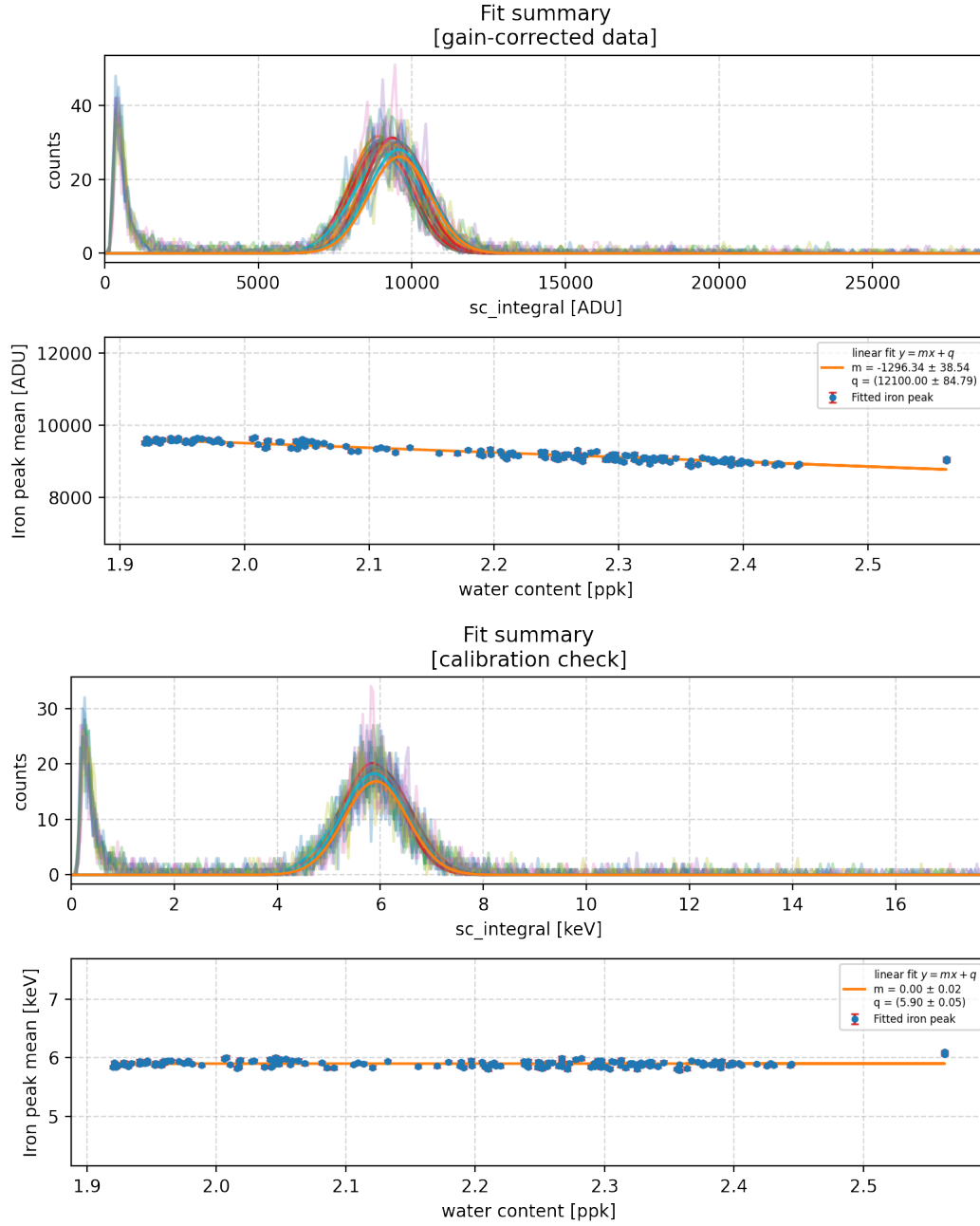


Figure 3.14. Summary of the calibration procedure (**top**) and calibration check (**bottom**) performed on the Stability dataset. The top panel of each sub-figure shows the fitted iron peak for a sample of the actual runs in the dataset (one every ten) to avoid exceeding superposition. The transparent lines are real data, while solid lines are the fitted Gaussian peaks.

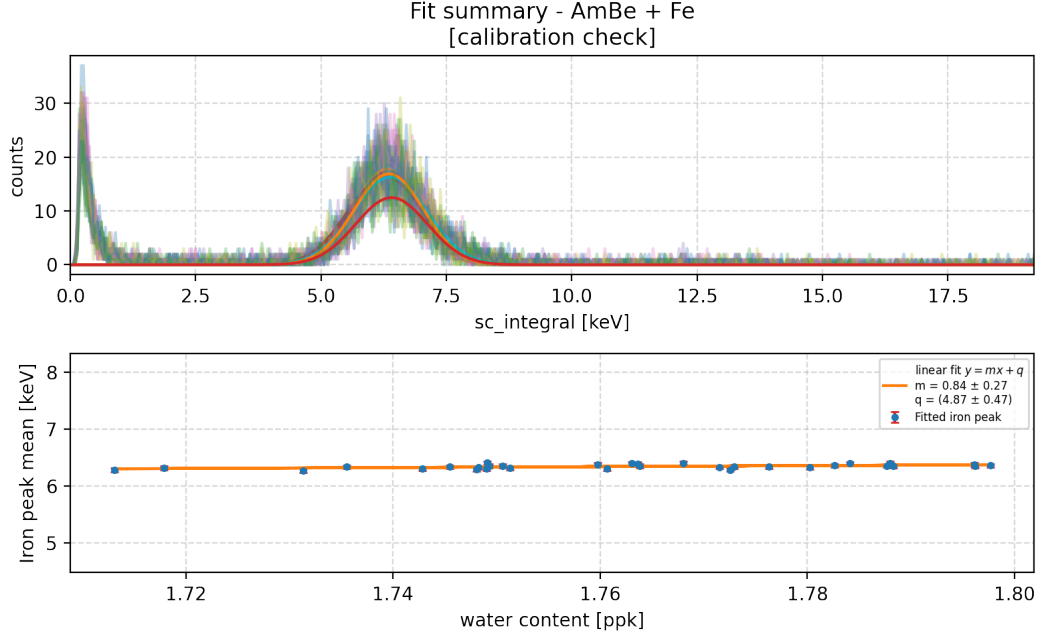


Figure 3.15. Summary of calibration check fit performed on AmBe+Fe data.

AmBe inter-calibration

As already introduced, when applying the calibration fit to the AmBe and AmBe + Fe datasets, it is found that the iron peak from AmBe + Fe is slightly shifted to the right with respect to the expected true value 5.9 keV. As it can be seen in figure 3.17, the distribution is peaked around 6.3 keV, which means a positive difference of 6.7%.

As anticipated, one possible reason of this shift could be related to different conditions in the detector parameters. As reported in [50], a difference in light yield of 0.6% per mbar is expected. Given the fact that the actual pressure difference between AmBe and Stability data taking period was roughly 8 mbar on average (see figure 3.16), this would explain only 4.8% of the difference in the light yield, leaving the rest of the variation unexplained.

With this said, it is clear that the ^{55}Fe peak needs to be centered in 5.9 keV for fundamental reasons, hence both AmBe and AmBe + Fe datasets will be rescaled appropriately, in order to account for the observed discrepancy, with the result of this operation pictured in figure 3.18.

Background inter-calibration

Moving to the Bkg dataset, since in this case there is no Iron source, to check the goodness of the calibration a *daily calibration* run (with Iron in step 3) was used, i.e. the Bkg_stability dataset in table 3.1.

In this case the situation is the opposite with respect to AmBe: the Iron peak is shifted backwards in energy of a factor 8.5%, as reported in figure 3.19, and the observed mean pressure difference of 4 mbar would explain only 2.4% of the variation (see figure 3.16). Note that here the peak is less sharp with respect to the previous due to the lower amount of data inside a single run, but it is enough to give a quantitative idea of the situation, hence this type of check should be done for

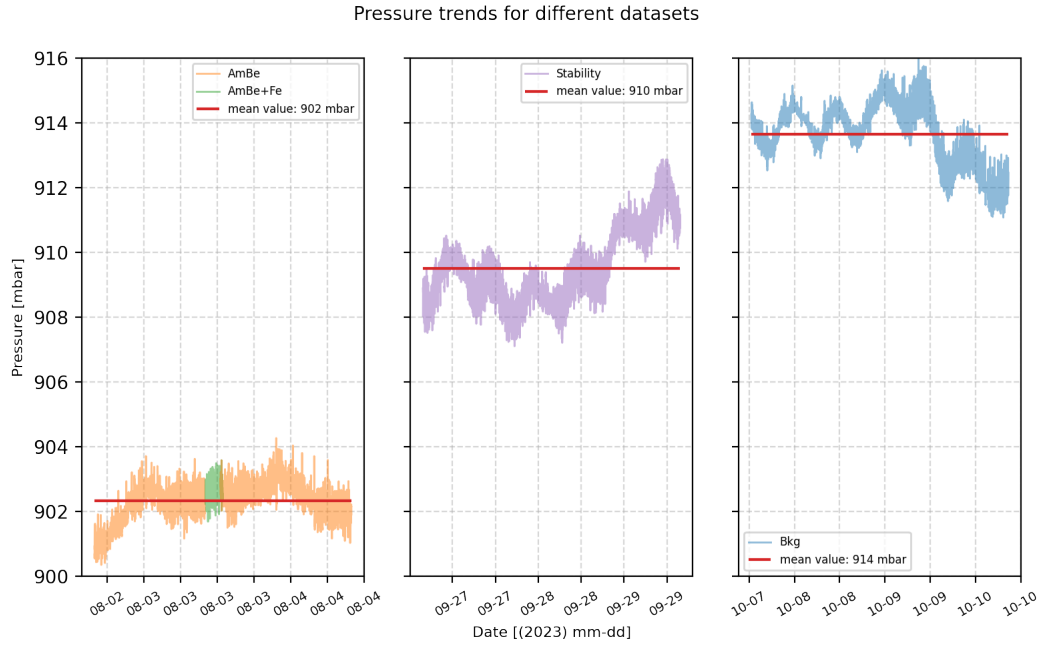


Figure 3.16. Pressure trends during relevant timeframes for considered datasets.

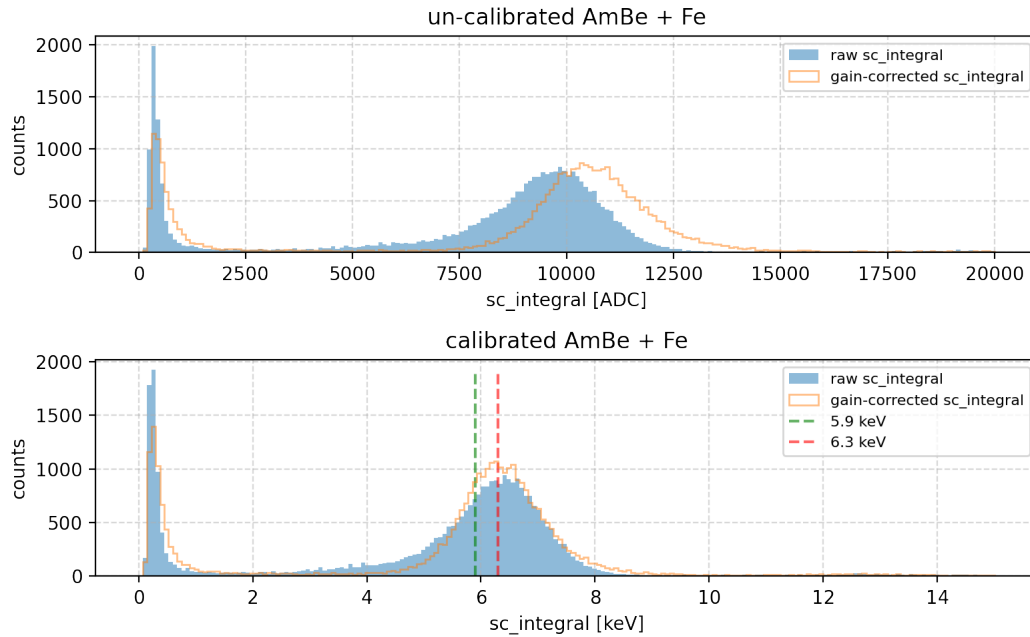


Figure 3.17. AmBe + Fe before (**top panel**) and after (**bottom panel**) calibration using the fit parameters obtained from the Stability dataset. After calibration, the iron peak appears shifted from its true position, probably because of the fact that the data taking was affected by changes in different parameters and not only in the humidity level, as better explained in the text.

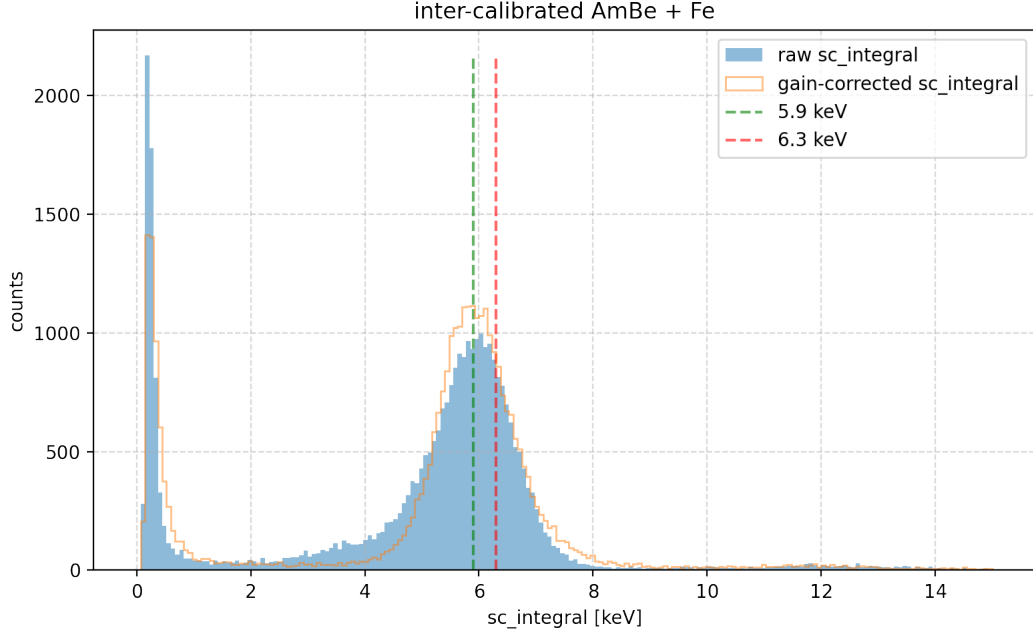


Figure 3.18. AmBe + Fe inter-calibration.

every dataset which is analysed.

The Bkg dataset is then rescaled as described before, using the factor which moves the peak to the true energy position, with the result reported in figure 3.20.

3.4 Pre-processed energy spectra

Before moving to the next chapter, it is worth reporting the comparison between raw data and fully pre-processed energy distributions for AmBe and Bkg datasets (figure 3.21), together with the difference between AmBe and Bkg spectra after the pre-processing procedure in the range 0÷1 MeV (figure 3.22).

In fact, it is interesting to note that already after 200 keV, the two spectra merge in a sort of continuum. At the same time, anyway, a clear excess is present around 80 keV when subtracting the Bkg to AmBe, whose characterisation is still unknown and out of the scope of this work.

Moreover, error bars are reported in figure 3.21 and 3.22, in order to justify the appearance of multiple bins with rate lower than the total time of exposure of the AmBe data taking, which is around 6×10^{-6} Hz (see table 3.1). The error bars show that they are a pure artifact of the subtraction and that they are compatible with the minimum possible measurable value inside 1σ .

Furthermore, the same logic can be applied to the AmBe+Fe dataset to extract the iron peak by subtracting the AmBe spectrum to the AmBe+Fe one. The result of this operation is reported in figure 3.23 and it shows a really sharp Gaussian distribution, a nice test on the goodness of the pre-processing pipeline, also because of the fact that a fit of this peak demonstrates that with this procedure applied we can achieve 10% energy resolution, at least for ^{55}Fe spots.

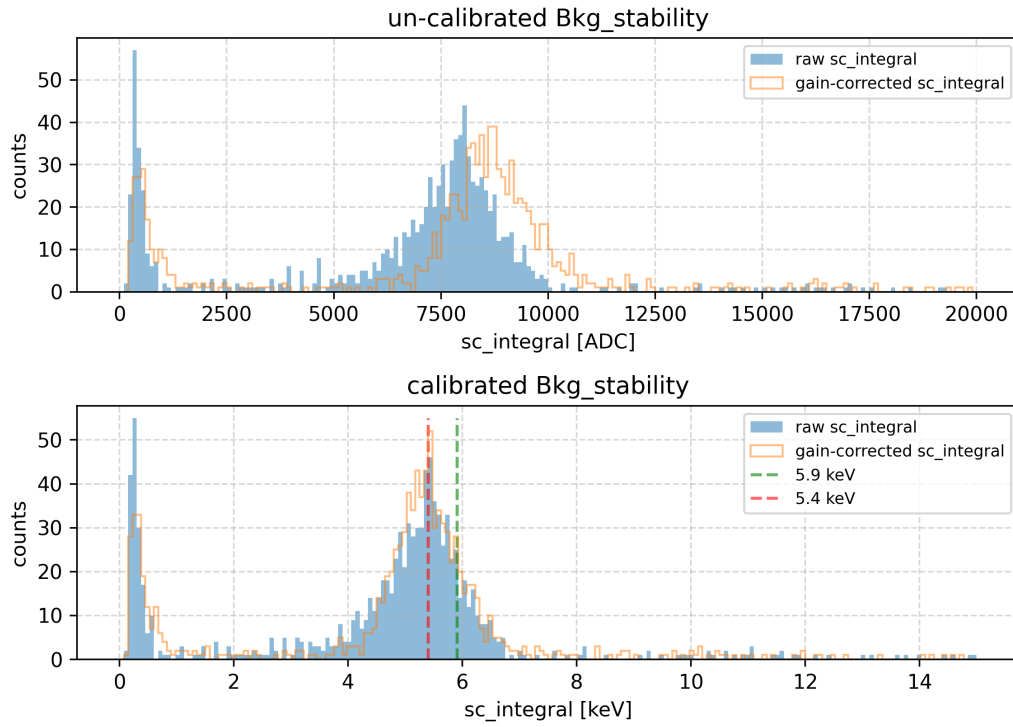


Figure 3.19. Bkg_stability before (**top panel**) and after (**bottom panel**) calibration using the fit parameters obtained from the Stability dataset. The same considerations outlined for figure 3.17 apply.

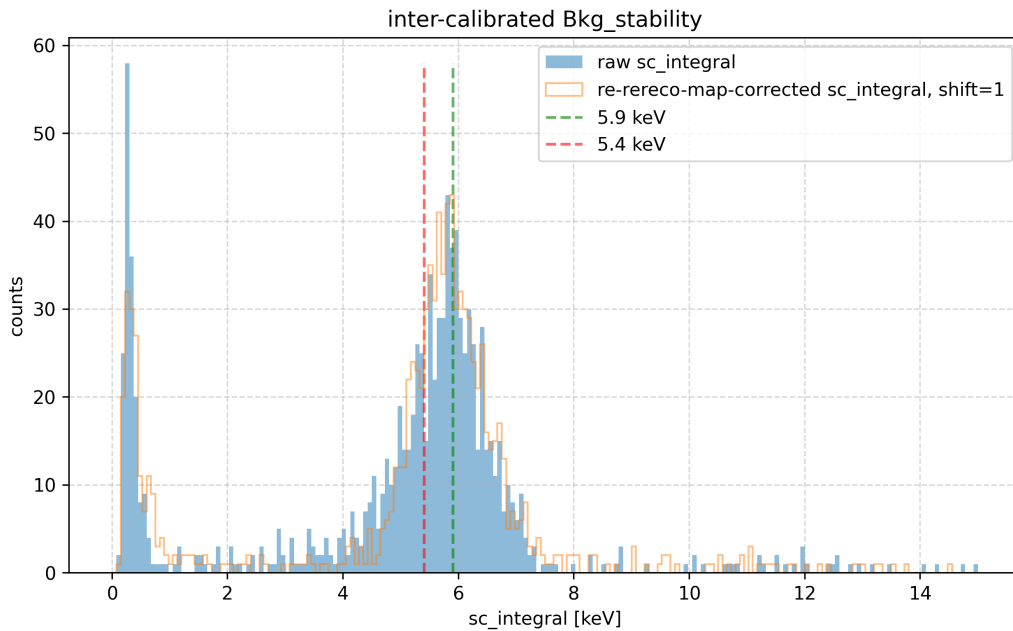


Figure 3.20. Bkg_stability inter-calibration.

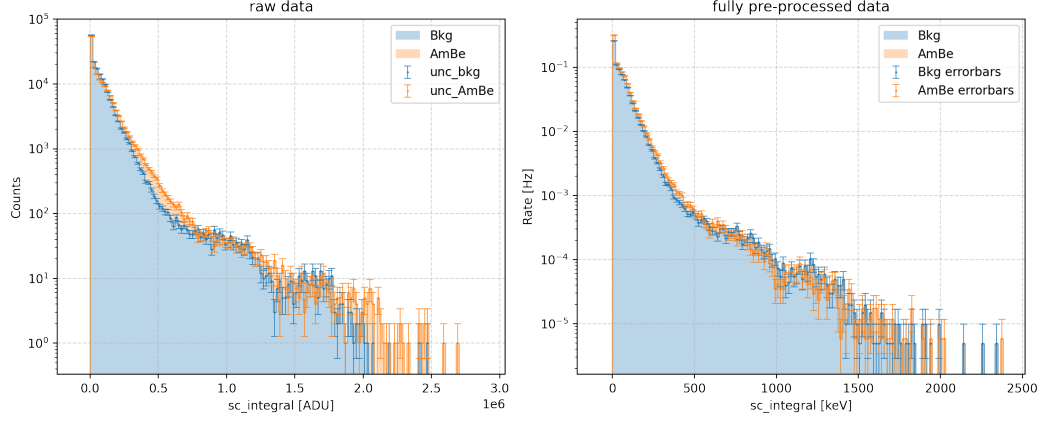


Figure 3.21. Comparison between raw and fully pre-processed energy spectra for Bkg and AmBe datasets. The error bars are computed as the square root of the counts and then normalised to the time of exposure.

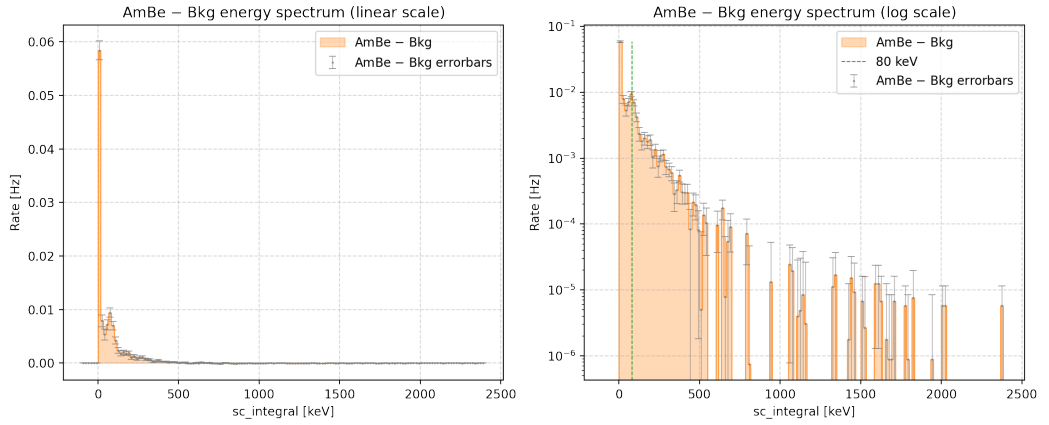


Figure 3.22. AmBe energy spectrum excess obtained from the subtraction of the Bkg spectrum to the AmBe one. The error bars are computed as the sum in quadrature of the AmBe and Bkg errors and then normalised to the time of exposure. They are relevant in order to justify the fact that after the subtraction there are bins with a rate lower than the inverse of the total time of exposure of the AmBe data taking (see table 3.1).

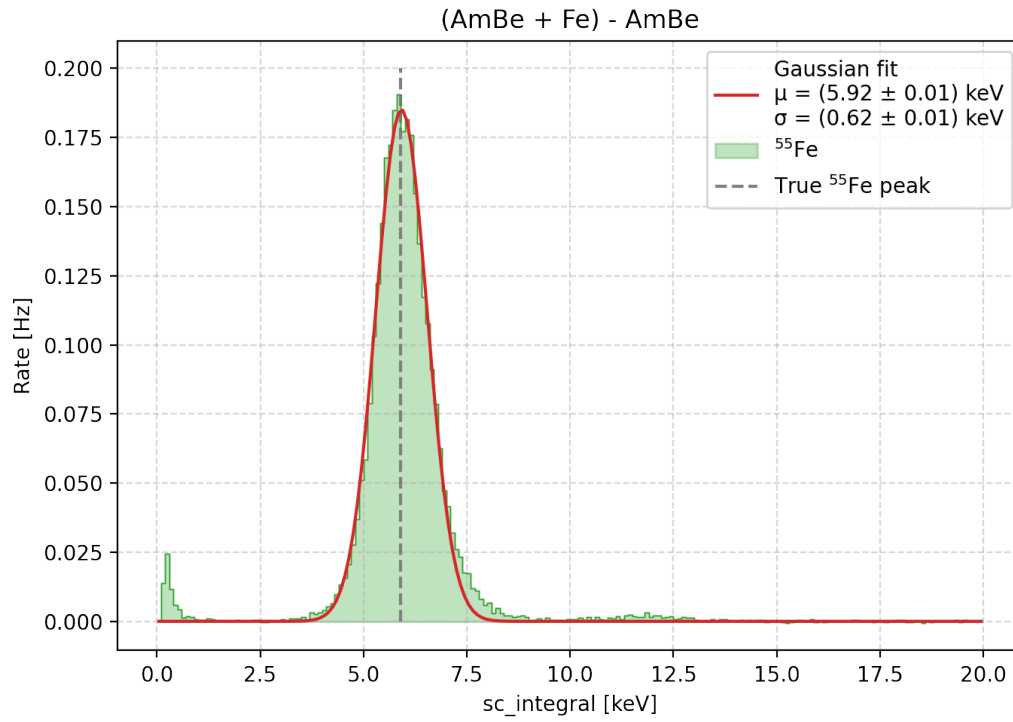


Figure 3.23. ^{55}Fe peak, extracted from the subtraction of the AmBe energy spectrum to the AmBe+Fe one, with a superimposed gaussian fit. It is interesting to note that this plot demonstrates that we have 10% energy resolution, at least on Iron spots.

Chapter 4

Low energy nuclear recoils analysis and directionality

As introduced in the previous chapter, AmBe sources are particularly interesting for their neutron production spectrum, reported in figure 3.9.

In the context of WIMP dark matter direct detection, which aims to detect recoiling nuclei, a neutron source is a perfect calibration tool in the nuclear recoil regime.

This chapter will be focused on expanding the knowledge of CYGNO performance in low-energy nuclear recoil detection by analysing a sample of the AmBe-induced events which is found to stand off from the background data.

In particular, the first part will treat the selection strategy, with a note on techniques to remove degraded data.

The second part will instead focus on the first approach to directionality measurements with the LIME prototype, together with the description of the development of a toy Monte Carlo simulation to validate the results.

Finally, the chapter concludes with a description of the procedure adopted for the first attempt in pursuing 3D length (or range) reconstruction, by means of PMTs data, and with a characterization of the saturation effect on the selected NRs sample energies.

4.1 AmBe data taking reference frame

Before starting with the analysis description, it is important to set the convention for the reference frame of the experimental setup. As it can be seen from figure 4.1, where the real setup photo and its GEANT4 implementation are reproduced from [48], the AmBe source was positioned to the right-hand-side of the APS camera point of view, i.e. the GEM plane.

Since the concept of directionality will be crucial in the following, it is useful to make two key examples to fully understand the reference frame.

Suppose that a track is moving towards the opposite side of the AmBe source (left-hand-side of the GEM plane reference frame), collinear with the horizontal axis: it will have an angle of $\pm 180^\circ$ with respect to the horizontal, depending on whether the angle is calculated from above or below this axis.

On the other hand, tracks moving towards the AmBe source (right-hand-side of the GEM plane reference frame), will be said to have an angle of 0° with respect to the horizontal.

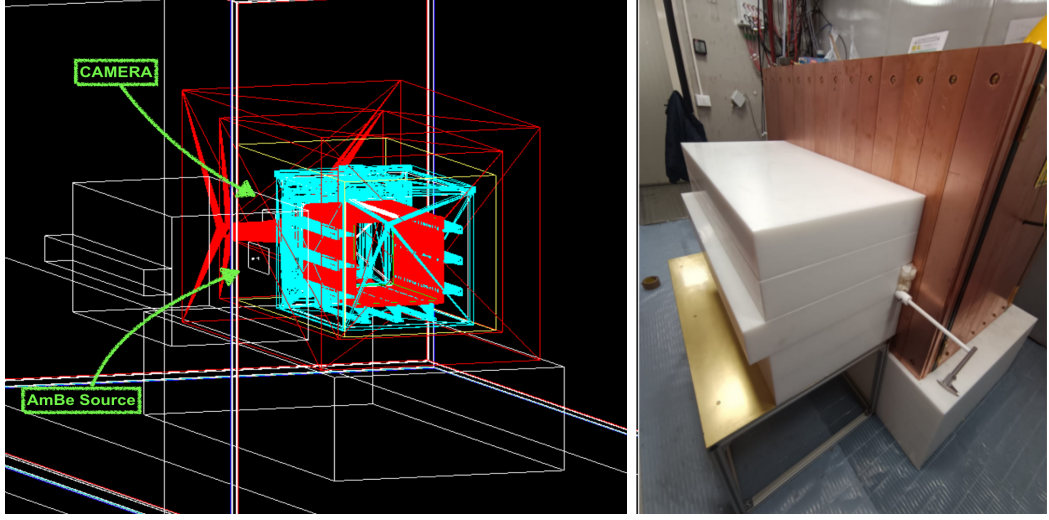


Figure 4.1. LIME setup in GEANT4 (left) and its real counterpart (right) during the data taking with AmBe source. Images from [48]

4.2 Nuclear recoils selection strategy

Nuclear recoils, as already stated during this work, are characterised by their being highly dense in dE/dx with respect to, for example, minimum ionizing particles.

In order to exploit this feature, we can make use of the various number of variables that can be extracted from data. A particularly useful one is $\delta = \text{sc_integral} / \text{sc_nhits}$, which encodes the super-cluster's energy density per hit pixel [68].

If we plot δ vs. calib_sc_integral for the AmBe and Bkg datasets (see table 3.1 for details), we obtain a very peculiar distribution and, more importantly, a very evident excess in the low energy and high density region of AmBe data, as it can be seen in figure 4.2, upper panel. The lower panel shows a zoom-in in energy of the upper plots, with the superimposition of the excess region's boundaries and the events inside this region.

Note that δ is computed from the raw sc_integral , because of the fact that there is no possible calibration for sc_nhits , and the two features are strongly linked together. With calib_sc_integral , instead, I am referring to the result of the pre-processing pipeline. It is in fact reported in keV in all the plots.

The boundaries are parameterized and translated into selection cuts as follows:

- $\text{lower_cut} = \delta > 24$,
- $\text{upper_cut} = \delta < 50$,
- $\text{right_cut} = \delta > 12 + 0.306 \times \text{calib_sc_integral}$,
- $\text{left_cut} = \delta < 24 + 0.578 \times \text{calib_sc_integral}$.

Of course, all the conditions need to be met simultaneously.

In figure 4.3, we can instead see the univariate distributions for both sc_integral and δ that characterise the selected sample of nuclear recoils, from which we can better appreciate the rate difference between the two experimental conditions.

For coherence with the prototypes naming (ORANGE, LEMON, LIME, MANGO), I decided to refer to this sample of events as **oranges' box** (*cassetta delle arance*) or,

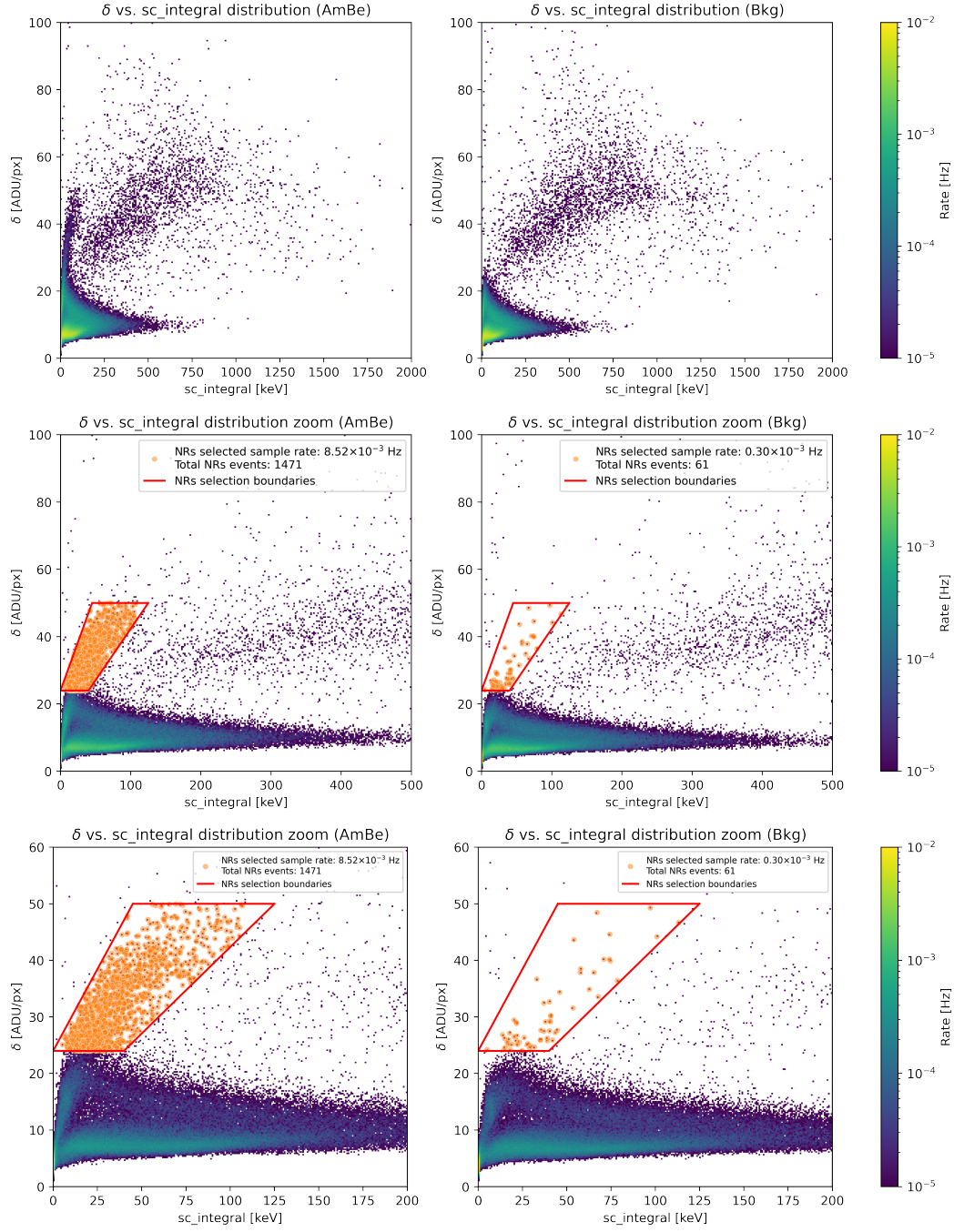


Figure 4.2. δ vs. $calib_sc_integral$ distribution for the AmBe and Bkg datasets in two different ranges of $sc_integral$. The two **lower panels**, in particular, show two different zoom-in in energy and δ of the **upper panel**, with the superimposition of the excess region's boundaries and the events inside this region.

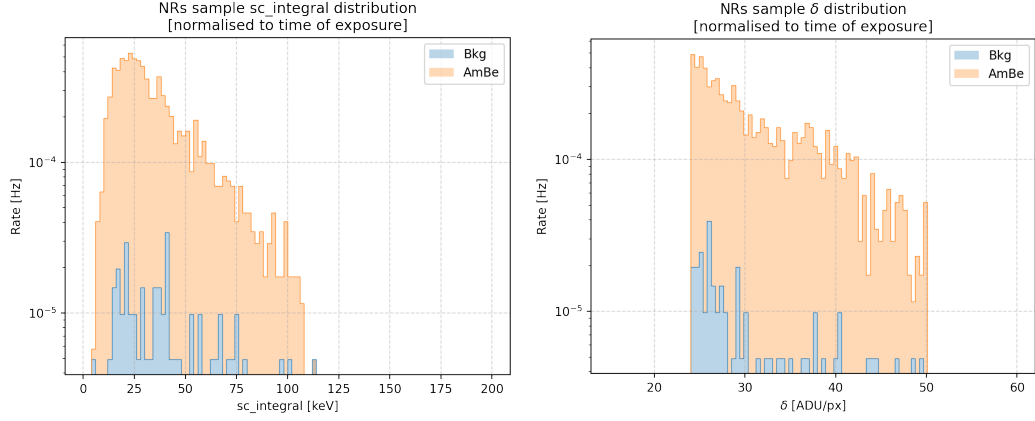


Figure 4.3. Distributions, normalised to the time of exposure, of `sc_integral` (left) and δ (right) for the events in the selection region for AmBe and Bkg datasets.

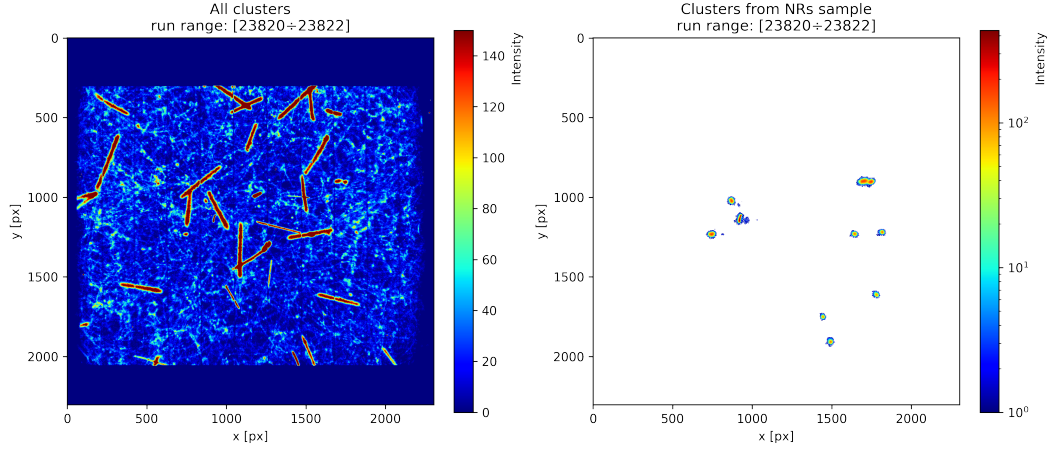


Figure 4.4. Superimposition of all the images acquired during 3 runs (left) and all the clusters in the same runs which are inside the NRs selection region described in the text.

to maintain a more academic tone, **NRs sample**. Their analysis will be the core element of the rest of this chapter.

To have an idea of the shape and number of these nuclear recoils, figure 4.4 shows the comparison between the superimposition of:

- all the images acquired during 3 runs (left plot) and
- the selected events in the same runs (right plot).

They are all short, slightly elliptic clusters and, for these reasons, understanding their directionality will be complicated.

4.2.1 Mowed clusters removal

As a side note, it is worth to talk about the, as I like to call it, *clusters mowing* problem, which is strongly related to the working principles of the sCMOS camera.

As described in section 2.3.4, the camera pixels need to be activated and deactivated at the start/end of every acquisition, and they are all active for a certain amount of time, which we call Global Exposure.

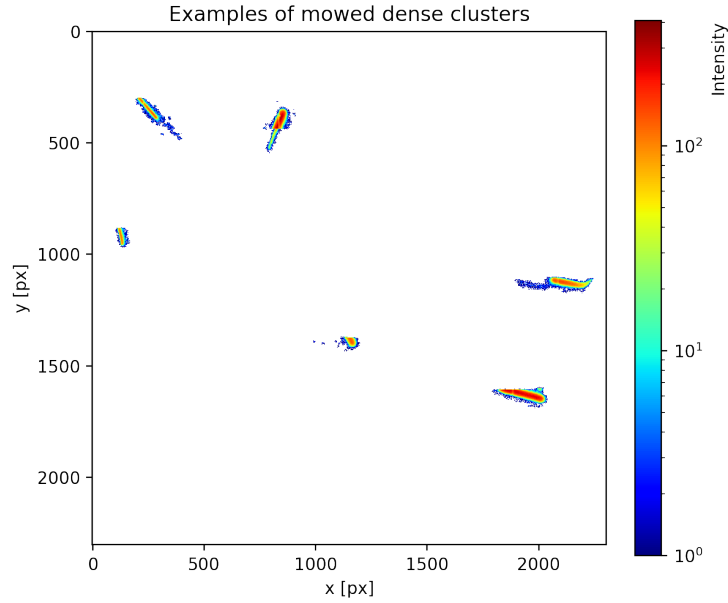


Figure 4.5. Some examples of mowed clusters with high energy density ($\delta > 25$).

If an event is triggered while the pixels are still in the activation or deactivation phase and the track produced is partially inside the deactivated region, only part of the cluster will be acquired. We can say that the cluster was *mowed* from the Global Exposure.

The crucial point is that this fact generates a sort of fake clusters with fake features (e.g. `sc_integral` or `sc_length`) which are not relatable to the physical process that produced them, hence they should be identified and removed from the analysed samples.

Thanks to the PMTs' data integration to the reconstruction algorithm, this objective can be easily achieved, by means of the variable `pmt_wf_TTT`, which refers to the Trigger Time Tag (**TTT**) of the event, as described in section 3.2. We can outline the strategy as follows:

- A given TTT can be converted in the actual number of active pixels (TTT_{eqpx}), counted from the top of the image. We can access to the PMT data associated to the considered image and do this conversion for each recorded event.
- From the camera data, one can retrieve the minimum and maximum extremes of the cluster in the y direction, call them $y_{\min}$ and $y_{\max}$.
- At this point, by checking if $y_{\min} < \text{TTT}_{\text{eqpx}} < y_{\max}$ for at least one TTT_{eqpx} and, in such case, remove the cluster from the sample.

In figure 4.5 a set of energy dense mowed clusters is reported as a reference.

Note that this exclusion method is still very naive: if there is more than one cluster in the image, it is possible that a certain TTT_{eqpx} line intersects a cluster which is not associated to that TTT and, in fact, some clusters appear to be not mowed with respect to others, hence there is room for improvement when a 1-to-1 association between clusters and waveforms is provided.

4.3 Nuclear recoils directionality

This part of the analysis will be conceptually divided into two main blocks:

- The first will describe the strategy adopted to estimate the cluster direction. Since it is the first time that directionality is tested, this will be implemented from scratch.
- The second part will instead describe the validation method, which features a toy Monte Carlo simulation based on simple elastic scattering.

4.3.1 Clusters direction estimation and distribution

To provide an estimation of the clusters' major axis inclination with respect to the horizontal (see one of the previous or following images as a reference), the strategy is to apply by scratch a famous technique named Principal Components Analysis (**PCA**) [69] to extract the axes of the ellipse in which the cluster can be enclosed.

First of all, a dutiful note regards the structure of the clusters. In general, because of diffusion during the drift towards the anode, all the clusters have a brighter and tenuous regions. Some of them (the more energetic) can also induce a sort of “shadow” which seems to be related to light reflection across the PMMA.

Since the sample object of the current analysis is composed from low energy nuclear recoils, hence the shadow problem is not so impactful. Anyway, before applying the PCA to the cluster itself, all the pixels with intensity < 25 counts are removed, to isolate the brightest region and avoid contamination from noisy pixels which could modify the actual shape of the track.

The PCA procedure is then divided in two steps:

- computation of the pixels matrix's covariance matrix;
- diagonalization of the covariance matrix to extract its eigenvalues and their relative eigenvectors.

The biggest eigenvalue correspond to the biggest eigenvector (the direction with higher variance), which in turn is the major axis of the ellipse that encloses the points. The angle with the horizontal is then resolved in the correct quadrant by means of the well known $\arctan2(y, x)$ function, where y and x refer in this case to the respective components of the major eigenvector. Some results of this procedure can be visualized in figure 4.6.

Of course this approach has its limitations, which reside in the convention adopted from the diagonalization algorithm regarding the sense of the eigenvectors.

In particular, since the algorithm I decided to use is developed over the LAPACK routines, the resulting major eigenvectors are always confined under the first and third quadrants' bisector.

At the same time, we can assume the nuclear recoils sample analysed in this section to be the product of a fixed source, whose position is on the right side of the image reference frame. From this assumption we can simply reverse every angle between $\pm 90^\circ$ by adding $\mp 180^\circ$, respectively, in such a way that the head-tail is imposed and the angle is consistent with the reference frame, with 0° on the right and 180° on the left.

The same exact procedure can be applied to the Bkg dataset to evaluate whether there is a match or a discrepancy between the resulting distributions.

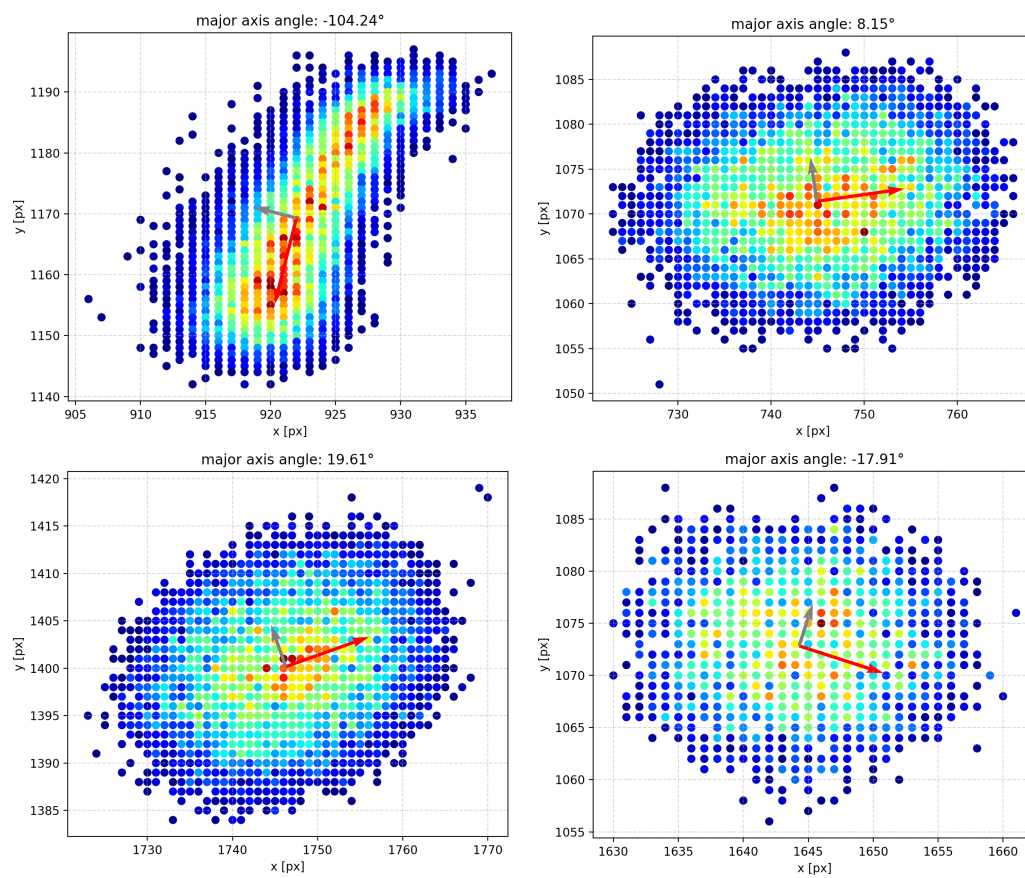


Figure 4.6. Results of the PCA procedure implemented to compute the clusters' angle with respect to the horizontal.

Table 4.1. Relevant information about the Bkg dataset analysed in this section, expanded from 57 to 227 hours of effective exposure, to extract a large number of events in the AmBe excess region. The year in “date range” is omitted and implied to be 2023.

Alias	Runs range	Date range [mm/dd]	Eff. Exposure [h]
Bkg	25735 ÷ 27844	29/09 ÷ 10/10	226.9

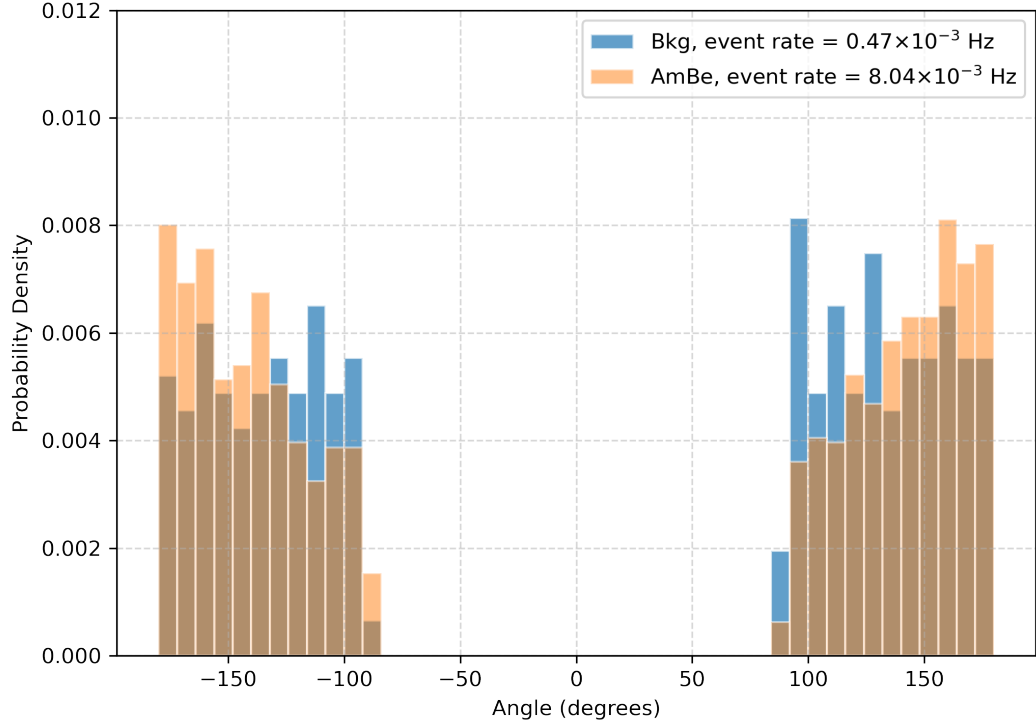


Figure 4.7. Angle distribution comparison between events in the AmBe excess region (see figure 4.2) extracted from the AmBe and expanded Bkg datasets (see tables 3.1 and 4.1).

Since the Bkg dataset described in table 3.1 is too small to have a relevant number of events in the AmBe excess region, for this part of the analysis it is expanded from 57 to 227 hours of effective exposure, as reported in table 4.1.

In the end, as appreciable from figure 4.7, events from the Bkg dataset show a flatter distribution in angle, while for events from the AmBe dataset the distribution is peaked around $\pm 180^\circ$, i.e. there are more events moving straight to the left.

Note that the rate reported in the legend in figure 4.7 is different with respect to the one presented in figure 4.2 and there are multiple reasons to explain this fact:

- for the AmBe dataset the rate is lower because of the deletion of some mowed clusters from the sample, while
- for the Bkg dataset the rate is higher due to the dataset expansion for this part of the analysis.

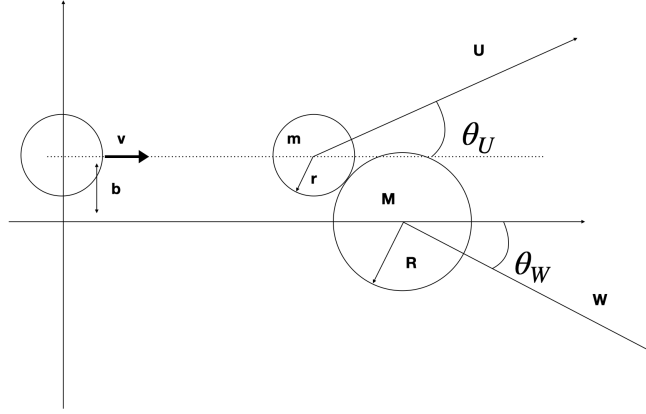


Figure 4.8. Sketch of the elastic scattering in the laboratory reference frame.

4.3.2 Monte Carlo validation

Now, the point is to understand if this peak around $\pm 180^\circ$ is expected or not. To do this, the idea is to develop a simple toy Monte Carlo simulation to find the distribution of simulated source-induced nuclear recoils' angles as seen from the camera point of view.

The idea is based on the simple elastic scattering of two massive spheres (a projectile with mass m and radius r and a fixed target with mass M and radius R), whose kinematics are reported in appendix B. From the same appendix, I recall here the only two important formulas which are needed in the following, i.e. equation B.1 and the second formula in equation B.4:

$$\theta_W = \arctan\left(-\frac{\sin \gamma}{1 + \cos \gamma}\right), \quad \sin \gamma = \frac{b}{r + R} \quad (4.1)$$

where with θ_W it is intended the angle of the target after the scattering in the LAB frame (see figure 4.8), b is impact parameter between the projectile and the target, while γ is the angle after the scattering in the center of mass.

Note that the dependence of γ from the impact parameter is merely fictitious, all is solved by simply sampling $\sin \gamma$ between $[-1, +1]$. Nonetheless, the discussion in terms of impact parameter gives a more formal description of the problem.

Another important remark to make is about the complete blindness of this simulation on the projectile and recoil energies. In order to take into account all these elements, a far more sophisticated simulation should be employed, but this is out of the scope of this thesis.

The simulation of a single nuclear recoil works now as follows:

- The AmBe source and LIME limits are fixed in space in order to reflect the data taking layout. In particular, the source is positioned midway along the $(\hat{z}, \hat{y})$ directions, i.e. (25, 16.5) cm, and 10 cm after the end of the detector in the $\hat{x}$ direction, i.e. 43 cm. Moreover, whenever a triad of coordinates will be displayed in the following plots or text, it will always be reported with the $(\hat{z}, \hat{x}, \hat{y})$ convention.
- An interaction point is sampled inside the detector region and the direction between this point and the source is computed.

- A nuclear recoil is simulated sampling the center of mass angle and boosting it to the LAB frame, using equations 4.1, and then converted to an angle with respect to the direction computed in the previous point. Moreover, the azimuthal angle φ is extracted from a flat distribution in the range $[0, 2\pi]$.
- As a last step, the recoil direction is projected on the spatial planes to compute the angles with respect to (in particular) the GEM plane $(\hat{x}, \hat{y})$ and the “source wall” $(\hat{z}, \hat{x})$.

Examples of events extracted in this way can be seen in figure 4.9. Note that the orange arrow length is not related to the range of the nuclear recoil in the detector.

This simulation is performed in such a way to simulate the same number of events extracted from the AmBe and Bkg expanded datasets. The comparisons are reported in figure 4.10.

There are multiple comments to make from this results:

- First of all, the simulation is confirming the expected peak around $\pm 180^\circ$, i.e. the fact that the majority of the nuclear recoils produced from a source on the right side of the camera should proceed straight to the left.
- Regarding the AmBe-simulation comparison, it is evident that the shape is similar, but the probability density is higher for the simulation in the flatter angles, while it is higher for the AmBe data in the vertical region. More on this will be covered at the end of this section.
- Regarding the Bkg-simulation comparison, the shape mismatch is clear, as it was with the AmBe-Bkg comparison in the previous discussion (figure 4.7).

The differences in the AmBe-simulation comparison in figure 4.10 can be explained from the simple observation that the simulated angles are calculated with full resolution, while we don't know the precision with which the angle is calculated from the cluster's pixels.

We can also try to understand which is the angular resolution where real data are matched by simulated data. This is typically done by adding to the simulated value a number drawn from a Gaussian distribution centered in 0 and with standard deviation equal to 1, multiplied the an error which can be interpreted as an **angular resolution**. In mathematical terms:

$$\text{angle}^{\text{smeared}} = \text{angle}^{\text{sim}} + \text{error} \cdot c, \quad c \in \mathcal{N}(0, 1) \quad (4.2)$$

As it can be seen in figure 4.11 and 4.12, the shapes of the angles distribution for nuclear recoils in the AmBe dataset and the simulated one start to become more and more similar when the resolution approaches $40^\circ/45^\circ$. This is the best indication of the measurement resolution we have on the angle calculated in the previously described way, for nuclear recoils in the 10 to 100 keV range.

From the comparison of Bkg data with the Gaussian-smeared simulation in figure 4.13 and 4.14, instead, it is appreciable that supposing a 40° resolution, the Bkg nuclear recoils angle distribution has always a flatter shape with respect to the simulation, with a clear excess in the vertical region (around $\pm 90^\circ$). It is true, anyway, that the two distributions start to converge when the resolution approaches 60° , equivalent to a resolution >1 rad, which is too big to assure a proper resolving power.

All the measured distributions in figure 4.11 and 4.13 have been statistically compared with their associated simulation by means of the χ^2 -distance test (hereafter

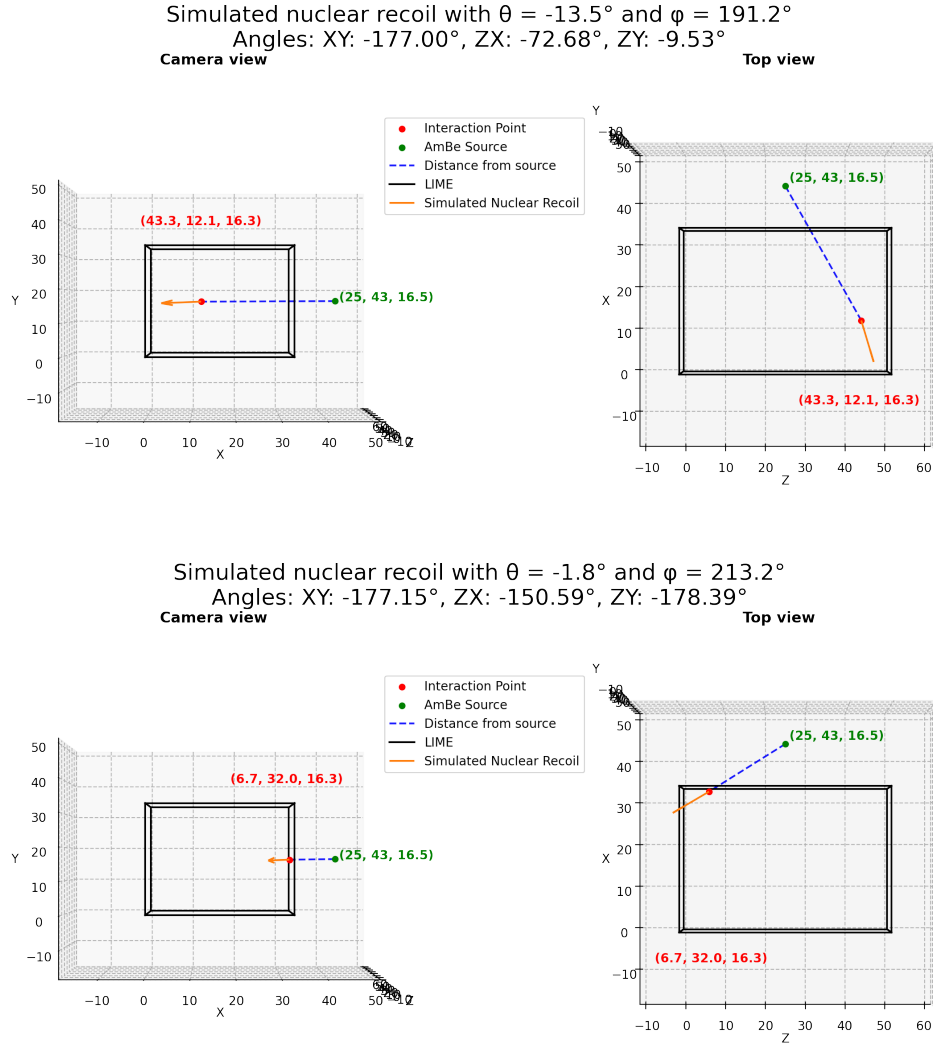


Figure 4.9. Examples of nuclear recoils simulated with the Monte Carlo described in section 4.3.2. Left and right subplots shows different views (camera and top) and the projected angle are reported in the plot titles. Note that the orange arrow length is not related to the range of the nuclear recoil in the detector.

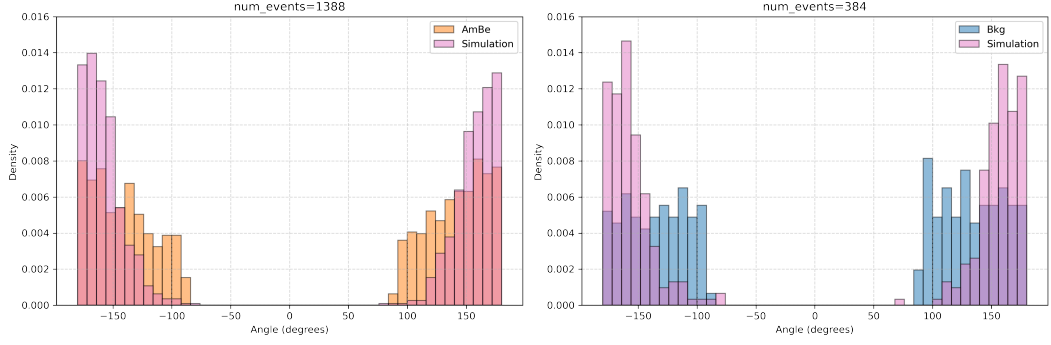


Figure 4.10. Comparison between low energy nuclear recoils angle distributions as seen from the camera point of view, sample from AmBe and Simulation (**left**) and sample from Bkg and Simulation (**right**). Note that the simulations are different, in order to reflect the same number of events, and that the y axis reports the probability density, in such a way that the extremes are the same in the two panels.

only χ^2), which is broadly adopted when dealing with similarity comparisons between histograms and it is directly derived from the classic χ^2 formula applied to this specific case. In particular, it is computed as:

$$\chi^2 = \frac{1}{2} \sum_i \frac{(O_i - E_i)^2}{O_i + E_i} \quad (4.3)$$

where O_i indicates the counts inside each bin of the measured distributions, while E_i represents the same for the simulated distributions.

It is insightful to observe the χ^2 trends when varying the angular resolution for both AmBe and Bkg, reported in figure 4.15. Apart from their relative values, what is interesting is the fact that, while the curve related to the Bkg data is always decreasing with angular resolution, the AmBe curve starts increasing again when the angular resolution surpasses 45° .

Although the achieved angular resolution is not that optimal, this analysis indicates that it is possible to achieve directionality measurements with the LIME setup. The angular resolution can be for sure improved with more statistics and a more sophisticated implementation of the angle measurement algorithm.

4.4 First attempt to nuclear recoils 3D reconstruction

For this part of the analysis, it is necessary to have a 1 to 1 association between PMTs waveforms and the clusters reconstructed from the image, which at the moment is still in development and not integrated in the reconstruction pipeline. In fact, this is the first time that 3D range reconstruction exploiting PMT data is attempted, in the 10-100 keV range, in the CYGNO collaboration.

Hence, in order to overcome this problem, the sample of nuclear recoils that was previously described in this chapter is reduced in number, keeping only the “*lonely clusters*”, i.e. the ones that are alone in the image with precisely four waveforms associated to it. Unfortunately, this is a very stringent constraint which reduces the relevant events from ~ 1400 to ~ 350 .

The idea here is to merge information from the camera with information from the PMTs, in such a way that the full 3D length of the cluster can be reconstructed, by

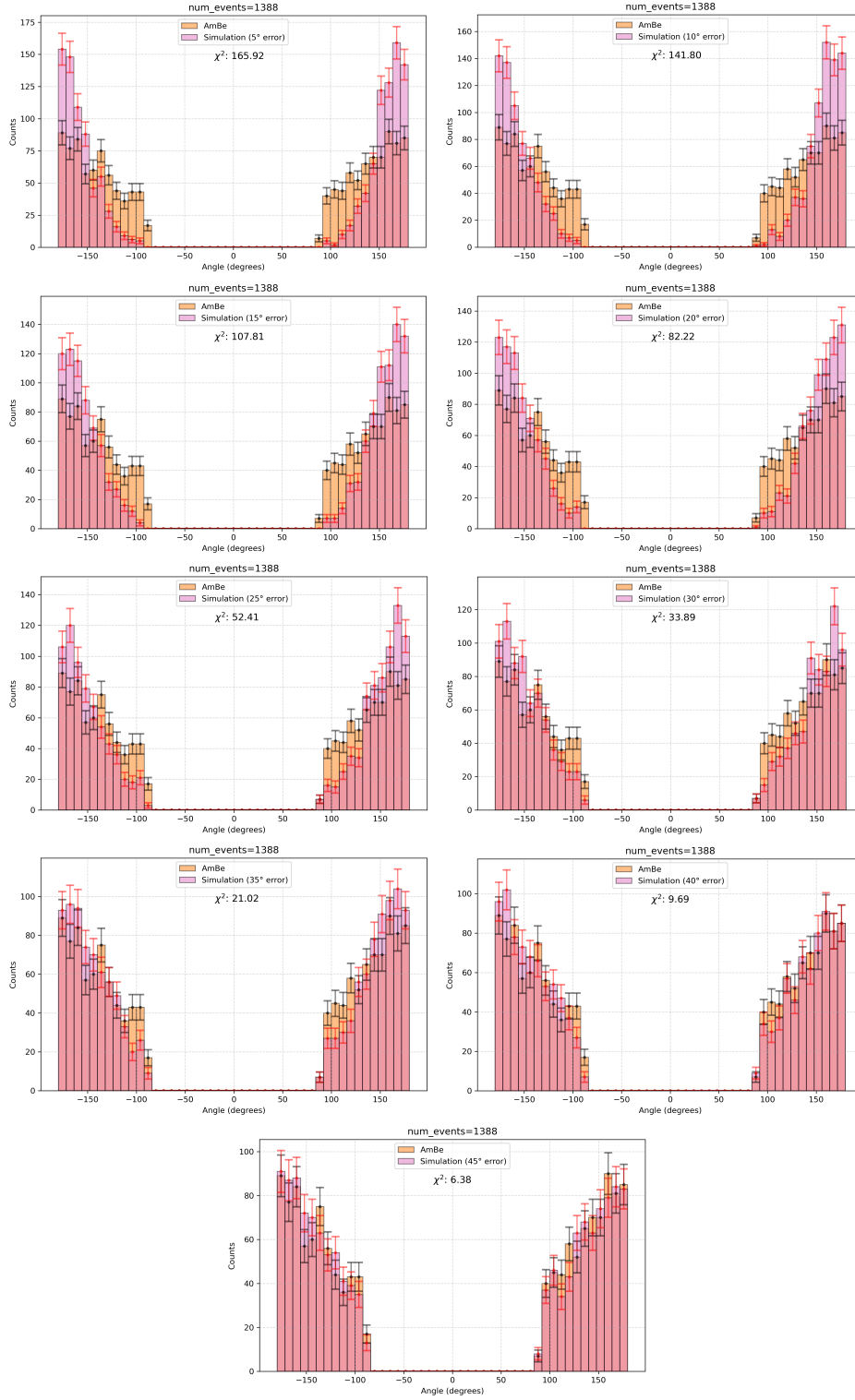


Figure 4.11. Comparison plots between real (from AmBe dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 5° and 45°, whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.

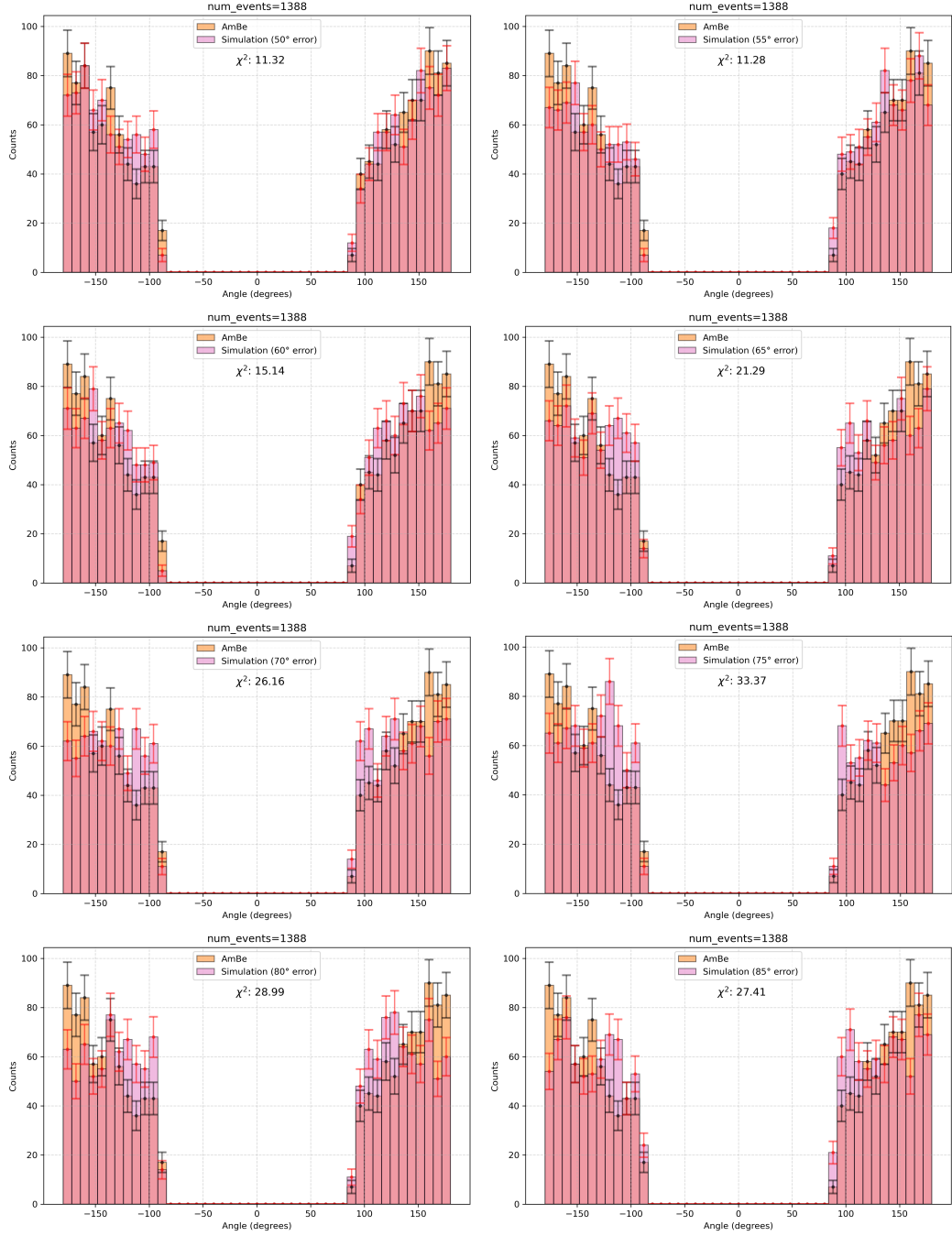


Figure 4.12. Comparison plots between real (from AmBe dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 50° and 85°, whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.

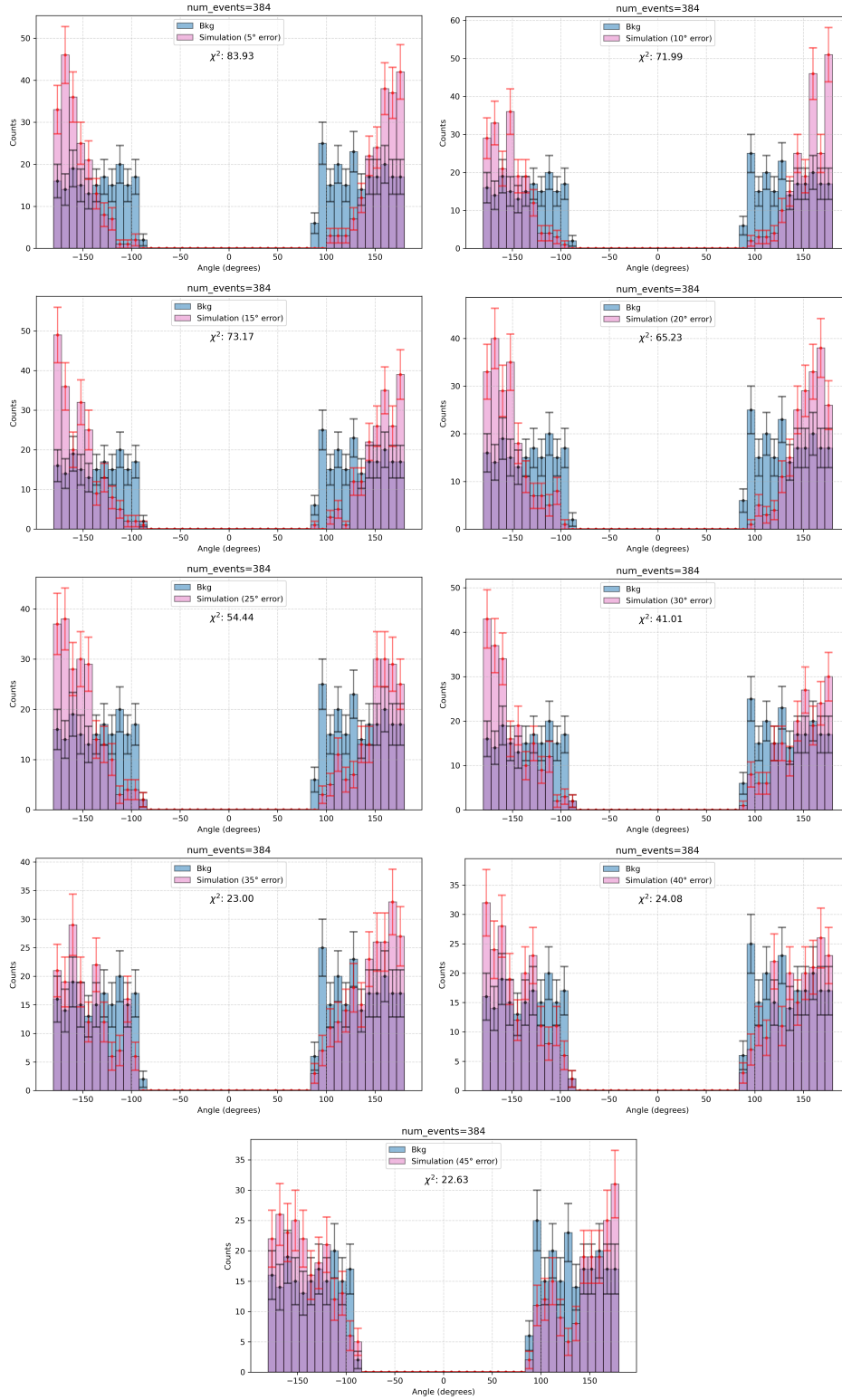


Figure 4.13. Comparison plots between real (from Bkg dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 5° and 45°, whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.

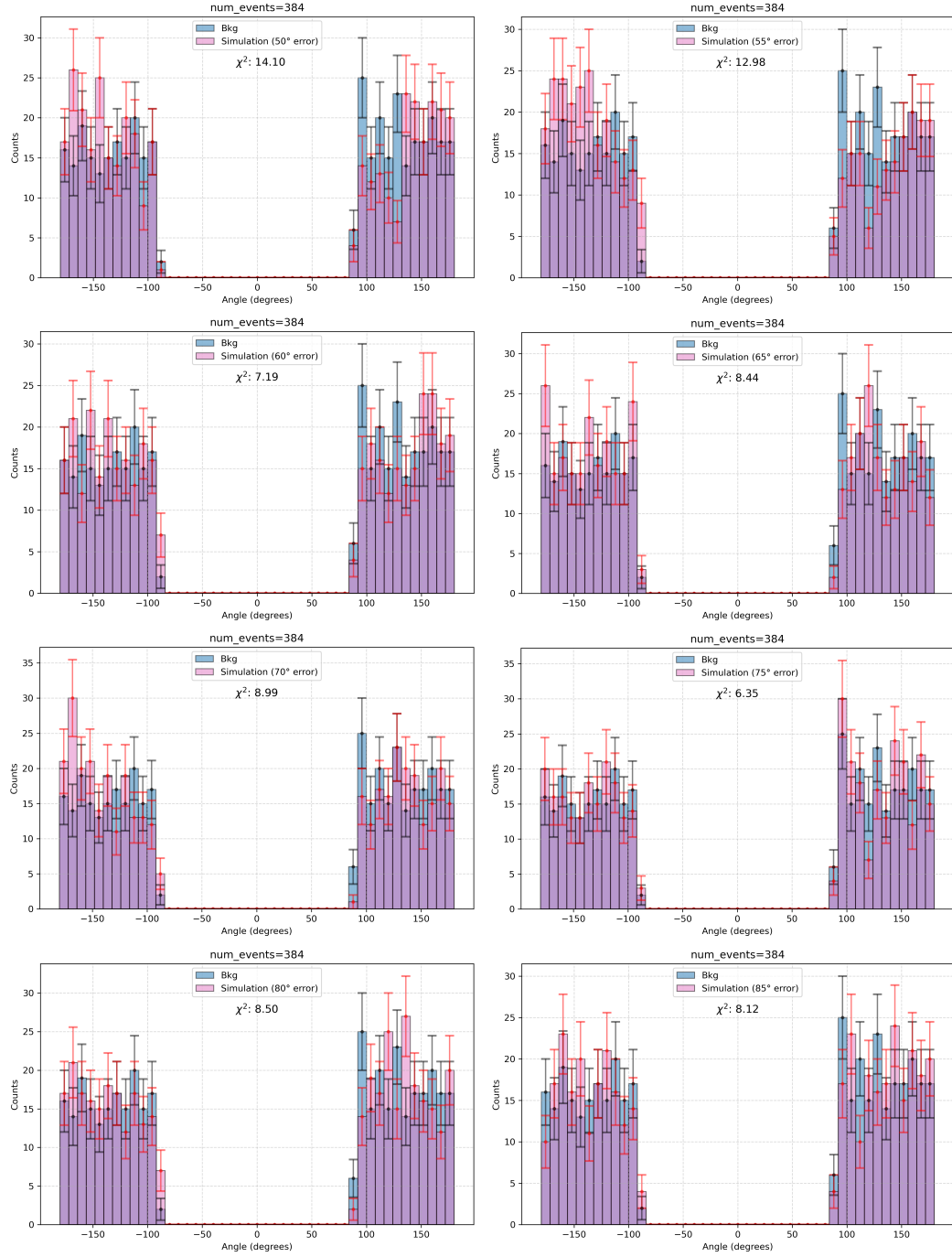


Figure 4.14. Comparison plots between real (from Bkg dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 50° and 85°, whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.

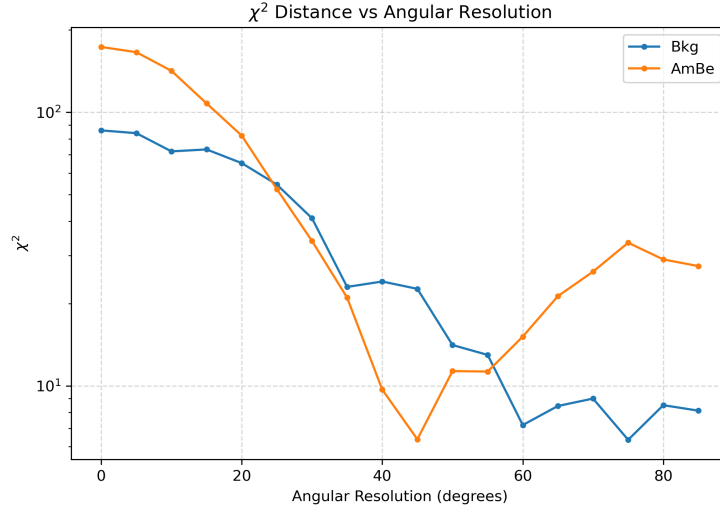


Figure 4.15. Line plot showing the χ^2 distance between AmBe / Bkg NRs' measured angular distributions and their respective simulated angular distributions provided by the toy Monte Carlo simulation, at different angular resolution.

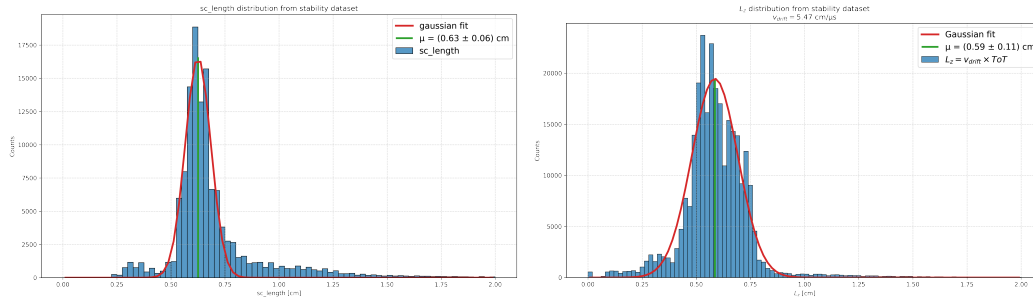


Figure 4.16. Fits of the most prominent peak in `sc_length` and L_z distribution.

means of a simple sum in quadrature between `sc_length` and the distance traveled from the electrons in the drift direction, which will be called L_z .

L_z can be extracted by exploiting the maximum time over threshold (**ToT_{max}**) seen from the PMTs, which is a measure the time difference between the first and the last electron arrived on the GEM plane, multiplied by the drift velocity $\mathbf{v}_{\text{drift}}$, which was estimated to be $v_{\text{drift}} = 5.47 \text{ cm}/\mu\text{s}$ [53].

An important note to make is about the diffusion of the electron cloud inside the gas: looking at the plot in figure 2.10, we can see that an electron produced by the ^{55}Fe source (5.9 keV) should travel approximately 0.5 mm inside the He:CF_4 mixture.

Since the ionization electrons diffuse inside the gas while drifting towards the cathode, we expect this length to be enlarged, and we can verify this fact by recalling data from the Stability dataset (see 3.1).

The mean `sc_length` can be estimated by means of a Gaussian fit of the most prominent peak of the length distribution while, regarding L_z , the same can be done with the distribution of $v_{\text{drift}} \times \text{ToT}_{\text{max}}$.

As it can be seen in figure 4.16, the values obtained are $\text{sc_length}^{\text{mean}} = 6.3 \pm 0.6 \text{ mm}$ and $L_z^{\text{mean}} = 5.9 \pm 1.1 \text{ mm}$, which produce $L_{\text{corr}}^{\text{mean}} = 8.63 \pm 0.9 \text{ mm}$ (with basic error propagation), that is a order of magnitude bigger than the theoretical value.

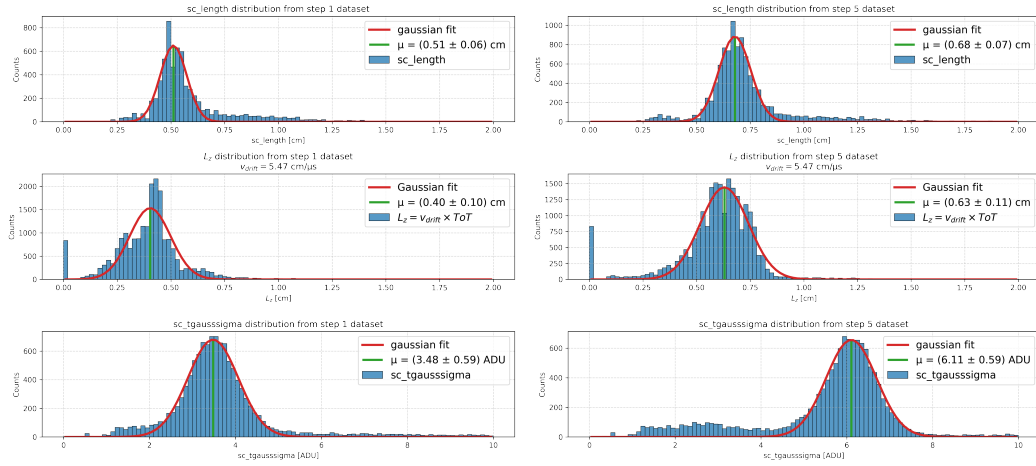


Figure 4.17. Fits of the most prominent peak in sc_length , L_z and $sc_tgaussigma$ distribution for step 1 (left) and step 5 (right) of the daily calibration procedure.

Hence, sc_length^{mean} and L_z^{mean} should be interpreted as offsets for their respective physical quantity. In fact, the ^{55}Fe spots can be interpreted as the zero of our length measurements, since their length in the 3D space (0.5 mm from 2.10) is so small that the length we can measure from the reconstruction is essentially only diffusion.

For a better characterisation of these offsets, a more sophisticated analysis has been performed, by repeating this study at each step of the daily calibration, collecting calibration data from different time periods during RUN 3.

As already introduced in section 3.1.5, a good way to understand the cluster position in the drift direction is given by $sc_tgaussigma$. By fitting the most prominent peak in its distribution at each step, sc_length^{mean} and L_z^{mean} can be associated to a value which varies in the drift direction, to produce a calibration law that permits to evaluate the correct offsets to subtract to each cluster's lengths, depending on its $sc_tgaussigma$.

Since there are 5 steps in the calibration procedure and 3 fits are performed for every step, in figure 4.17 only step 1 and step 5 results are reported, while all the results are summarised in figure 4.18, where the squared data is fitted with a linear law.

Note that both the offsets and $sc_tgaussigma$ are squared because of the intrinsic relation between diffusion and distance by means of a square root law (anticipated at the end of section 2.1.1). In fact, both length and the transverse profile standard deviation, i.e. $sc_tgaussigma$, are affected by diffusion in this way. By squaring them, the dependence becomes linear.

Note also that the agreement between data and the linear fit is a nice confirmation of this known fact and a good signal that the results are meaningful.

Now that we have the laws to reconstruct the diffusion offsets, the 3D range can be computed as already described, from the sum in quadrature of the lengths as seen from camera and PMTs.

In figure 4.19, the range vs energy distribution is plotted, together with the distributions of the single variables on the plot margins.

It is possible to superimpose these data with respect to the simulated range of He recoiling nuclei in figure 2.10, and this is done in figure 4.20, where the simulated energy is extended by an order of magnitude and also the ranges for Carbon and

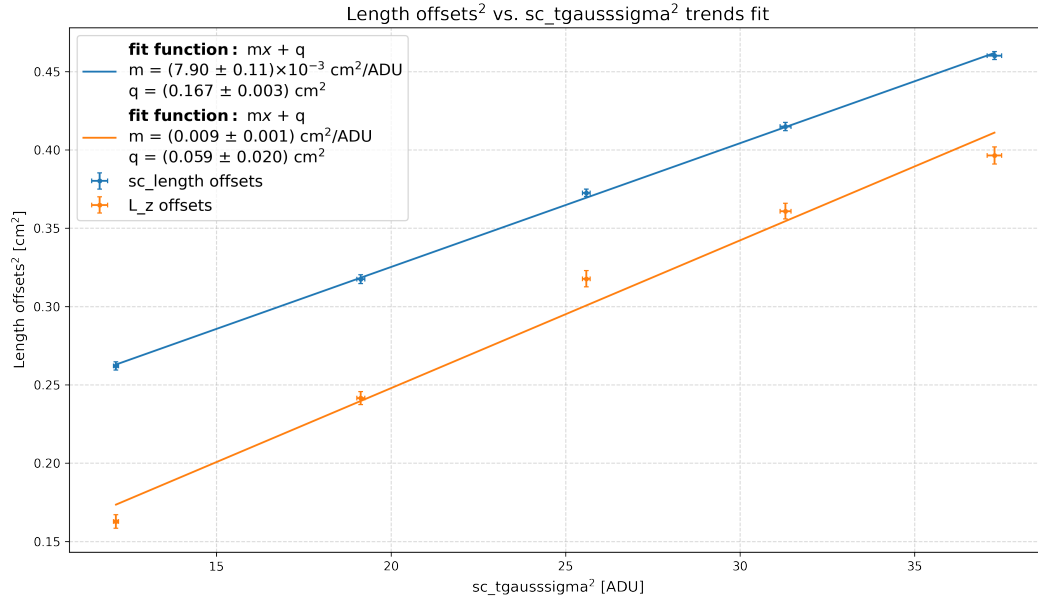


Figure 4.18. Linear fit of the squared length offsets vs. squared $sc_tgausssigma$ trends for both sc_length (blue) and L_z (orange).

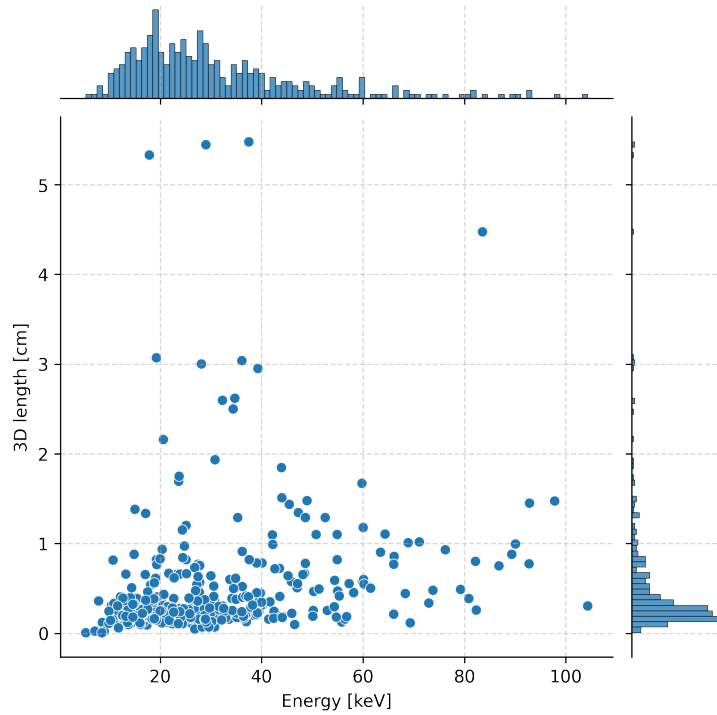


Figure 4.19. 3D range vs energy for the lonely nuclear recoils sample, with the distribution of the single variables on the plot margins.

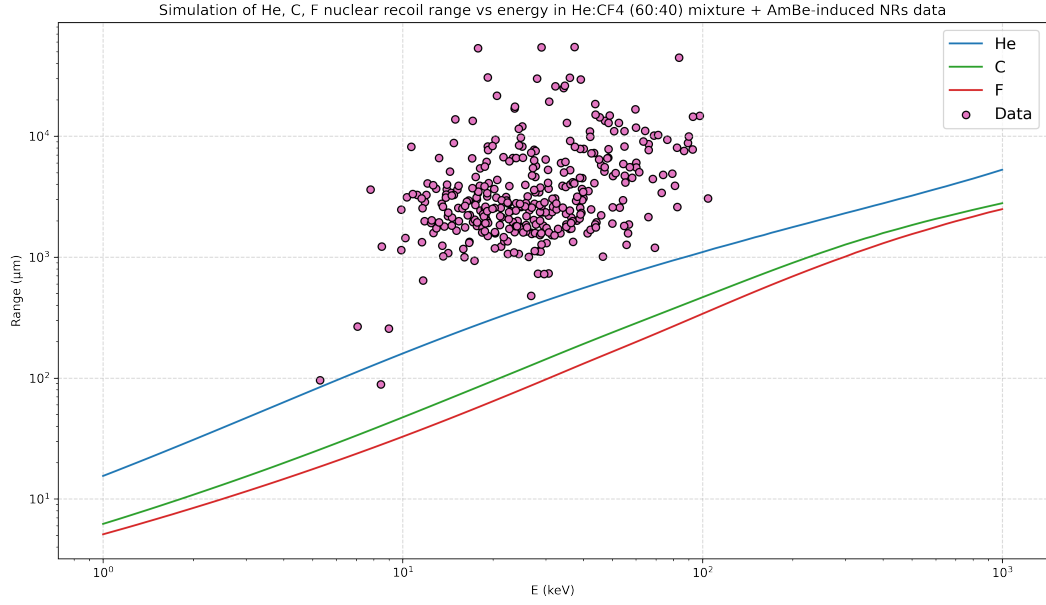


Figure 4.20. 3D range vs energy for He, C and F nuclear recoils in He:CF₄ (60:40) mixture, simulated with SRIM and reproduced from [70] with data in figure 4.19 superimposed.

Fluorine are reproduced using data from [70], which are simulated using SRIM.

A major, already known, problem appear from this analysis: when a high number of electrons reach the GEM plane, distributed on a small area, they generate a large amount of charge in the GEM holes, which effectively reduces the electric field in the amplification regions and this is reflected in a sort of *saturation* of the effective gain. This effect was observed and documented in the LIME prototype with a ^{55}Fe source [50] and it was found that this strongly depends on the charge density, which reduces with the distance from the GEM plane (the closer the cluster, the higher the charge density and saturation, hence the lower the gain). The procedure to correct data for this effect is still in development and it follows an idea developed for the CMS calorimeter [50, 71].

Unfortunately, for very dense tracks (i.e. with large dE/dx) like nuclear recoils, this saturation effect is more evident, and it can be observed by eye from the position of the experimental data with respect to the simulation in figure 4.20: assuming that the measured 3D range is correct, neglecting the data distribution, it is clear that the data points are at least a order of magnitude lower than they should be.

Moreover, we do not have a way to determine the nature of the nuclear recoils, i.e. if they are Helium, Carbon or Fluorine, since no population appear in the data distribution, probably also because of the gain saturation. All we can do is assume that the analysed sample is fully composed by Helium nuclei, since they are the ones with a higher range at equal energy.

Now, since we shifted the focus to the measured 3D range and α nuclear recoils, we can invert the plot axes and fit the Helium simulated range with a polynomial function, in order to extrapolate the hypothetical real energy of the experimental data. This can be visualised in figure 4.21, where the gain saturation effect is even more evident and where the orange line represents the extrapolation of α NRs energy at higher 3D range (hereafter *expected energy*), by means of a 2nd order polynomial fit of the simulated data.

Note that the order of the polynomial fit is chosen in order to avoid the explosion

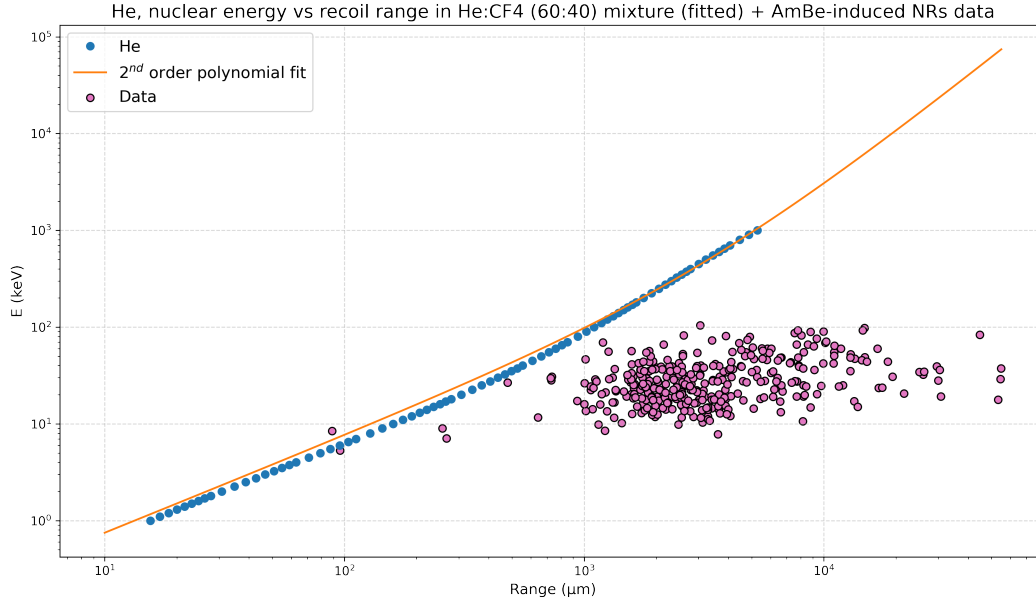


Figure 4.21. SRIM simulated 3D energy vs range for He nuclear recoils in He:CF₄ (60:40) mixture with data in figure 4.19 and a 2nd order polynomial fit of the simulated data superimposed. The fit range of variation is expanded to match the experimental data variation domain.

or fall of the expected energy in the extrapolated range. In fact, it is found that from the third to fourth order, the extrapolated energy falls down to zero very quickly, while from the fifth order the energy explodes to values larger than 10^7 keV.

In figure 4.23 the expected vs. saturated energy distribution is reported, while the *energy saturation factor* (i.e. expected energy/saturated energy) distribution is reported in figure 4.24.

From the expected energy spectrum of the nuclear recoils, reported in figure 4.22, we can see the small amount of outliers above 10^4 keV (i.e. 10 MeV). They are probably fault of the polynomial fit not correctly extrapolating the energy law behaviour after a certain 3D range, or maybe due to a wrong range reconstruction. On the other hand, the majority of the clusters have an expected energy well below 10 MeV.

At the same time, it is clear from the energy saturation factor distribution that most of them are reconstructed with energies that are around a factor 10 lower (as indicated from the peak which appears in figure 4.24) with respect to their expected values. This result appears to be compatible with what is reported in [72], i.e. that nuclear recoils with 1 MeV energy are seen after reconstruction with energies between 30 - 110 keV depending on their distance from the GEMs, which can be translated in an energy saturation factor ranging between 9 and 33.

4.5 Final remarks

From the previously described analysis, for the first time, it is possible to assert two crucial things, i.e. that, with the current LIME hardware and software setup:

- It is possible to reconstruct nuclear recoils angles in 2D with $40^\circ/45^\circ$ angular resolution, in the energy region between 100 and 1000 keV.

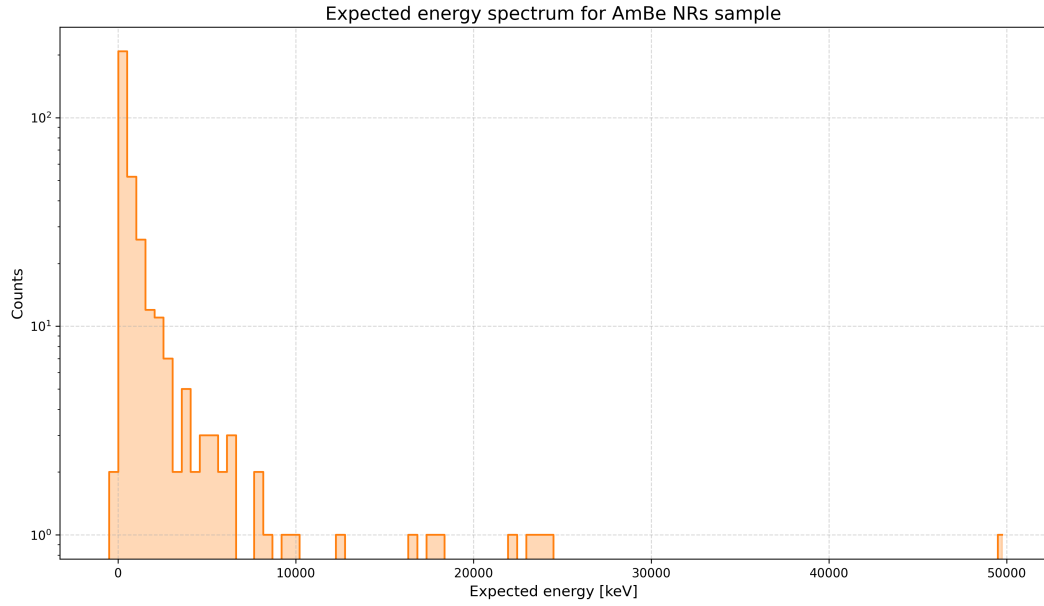


Figure 4.22. Expected energy spectrum computed from the reconstructed 3D range, between 0 and 50 MeV. As it can be seen, there are a small number of outliers above 10 MeV, probably due to a wrong extrapolation of the fitted law reported in figure 4.21 or to an error in the range reconstruction, but the majority of the clusters have expected energies well below 10 MeV.

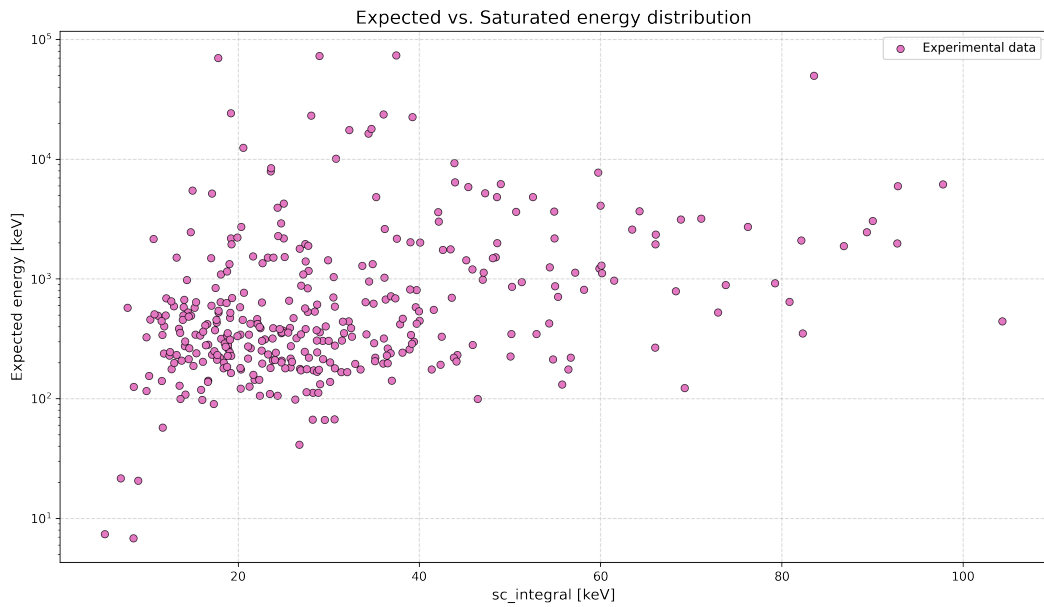


Figure 4.23. Expected vs. saturated energy distribution, which shows what should the real value of the measured energy be, based on a 2nd order polynomial fit of the simulated energy vs. range law for α nuclear recoils in He:CF₄ (reported in figure 4.21).

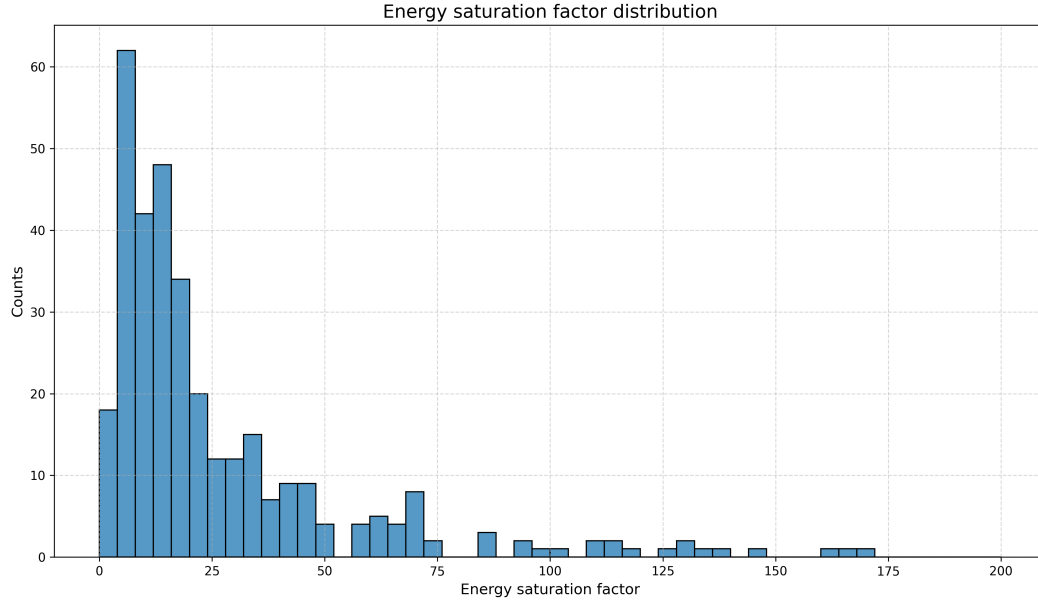


Figure 4.24. Saturation factor distribution computed as the ratio between expected and saturated energy, y and x axis in figure 4.23, respectively.

- It is possible to do 3D range reconstruction exploiting PMTs data, from which it is possible to evaluate the real energy of the nuclear recoils and to make considerations about the physics of their energy spectrum. As a consequence, we can also evaluate the gain saturation which affects the detector readout.

Moreover, as a final remark, it is important to keep in mind that all this analysis (the 3D range reconstruction part in particular) has been performed on a relatively low amount of data, due to the problems that emerged during the data taking, and without a formal simulation, because of the large amount of time required to perform such a simulation.

With this said, the results are promising and their agreement with already available simulations set the stage for a wider application range of the latter.

Of course, further investigation in this area is foreseen in the next months, with the aim to provide an actual simulation and digitization of AmBe-induced nuclear recoils inside the LIME setup, that will be crucial to provide better validation of the results reported here and of information that can be extracted from other analysis channels.

Conclusions

In the WIMP Dark Matter research field, the CYGNO experiment aims to enlarge the knowledge in the low energy region of the phase-space, while also trying to resolve WIMP-induced nuclear recoils (NRs) directionality, in order to eventually disentangle them from the neutrino background and to obtain information regarding possible astrophysical sources of Dark Matter.

This is done by means of an innovative detector design, which features a time projection chamber (TPC) filled with gas and optically read out with an active pixel sensor camera and a system of photomultiplier tubes (PMTs).

The most recent prototype is named Long Imaging Module or LIME, it is a TPC filled with a He:CF₄ (60:40) gas mixture and it has already taken a large amount of data underground at Laboratori Nazionali del Gran Sasso (LNGS), where the background events are minimised.

In particular, a relevant data taking for calibration in the low-energy NRs region of interest is the one that features an Americium-Beryllium (AmBe) neutron source, whose neutrons energy spectrum ranges up to 11 MeV.

This work covered the data analysis of this particular data taking campaign, with the aim of providing a characterisation of the CYGNO experiment performance in low-energy NRs detection.

The first part of the analysis produced a well defined pre-processing pipeline that should be applied to the reconstructed energy, which features data normalization, gain non-uniformity correction and calibration of the light yield as a function of the detector humidity.

Then, a sample of NRs in the 100 - 1000 keV kinetic energy range was identified from the excess of AmBe data with respect to background data. This sample was analysed cluster by cluster to extract a measure of their angles, validated by means of a toy Monte Carlo simulation, from which it was possible to conclude (for the first time) that the angular resolution is around 40°/45°, exploiting a χ^2 distance similarity comparison between measured and simulated angular distributions.

The last part of the analysis was focused on the performance achieved in 3D range reconstruction for a sub-sample of NRs which could be uniquely associated to PMT waveforms. This was the first time that this kind of analysis was applied for the CYGNO experiment and it showed that it is possible to make critical considerations regarding the NRs energy spectrum, while also evaluating the gain saturation which affects the readout. This effect was quantified with an energy saturation factor, mostly distributed around 10, coherent with previous preliminary studies and simulations on this theme.

Although the last parts of this work still have large room for improvement, they provided promising results that could be enhanced by means of a more precise and formal simulation, which will allow for better validation accuracy. Moreover, the agreement between results and already available simulations provide a validation of the latter, which allow the collaboration to trust them for other applications.

Appendix A

Validating the gain non-uniformity correction map

In section 3.11, I described the so-called *gain non-uniformity map* and how its application to `sc_integral` was able to enhance the gaussian shape of the calibration source peak.

In this appendix I will present some quantitative results which justify its application to data and why we should expect it to work also on long tracks like minimum ionizing particles (MIPs), even if the map is produced from iron spots.

A.1 Strategy

Three different types of analysis have been conducted:

- The first is more quantitative and aims to test agnostic power of the map both in the z .
- The second is focused on the possible explanation of the observed gain non-uniformity between the upper and lower part of the detector.
- The third is instead focused on the results of the map application to the MIP band.

A.2 Daily calibration tests

To pursue this analysis, random daily scans were extracted from different months of data taking and the iron peak was fitted for every calibration step. Then, from the fitted parameters, the resolution σ/μ (where σ is the standard deviation and μ is the mean value of the gaussian) is computed, effectively producing a resolution characterization for spots produced at different positions in the z direction.

In figure A.1 is reported an example of the most detailed output of the analysis, showing `sc_integral` distributions before and after the correction map application, at every calibration step (number increases with increasing distance from the anode), together with their gaussian fit.

To avoid redundancy, the relevant information are encoded in the resolution variable and are plotted in a bar plot which makes it easy to compare the goodness of the results step by step. A summary of bar plots of such type for different months

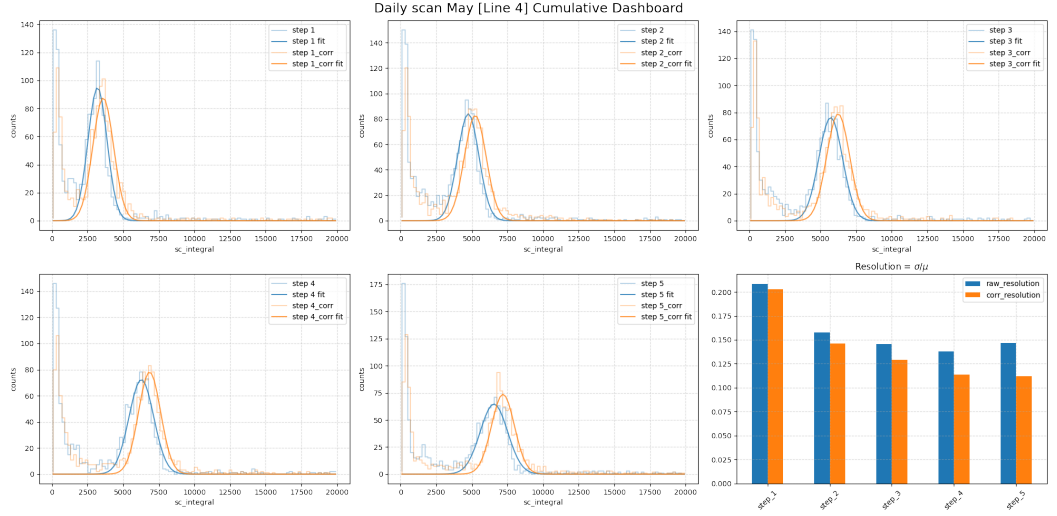


Figure A.1. Example of one of the cumulative dashboards produced for a random daily scan extracted from the data taking in May, showing `sc_integral` distributions before and after the correction map application, at every calibration step (number increases with increasing distance from the anode), together with their gaussian fit. In the end, a bar plot is produced in order to easily compare the resolution σ/μ at different steps.

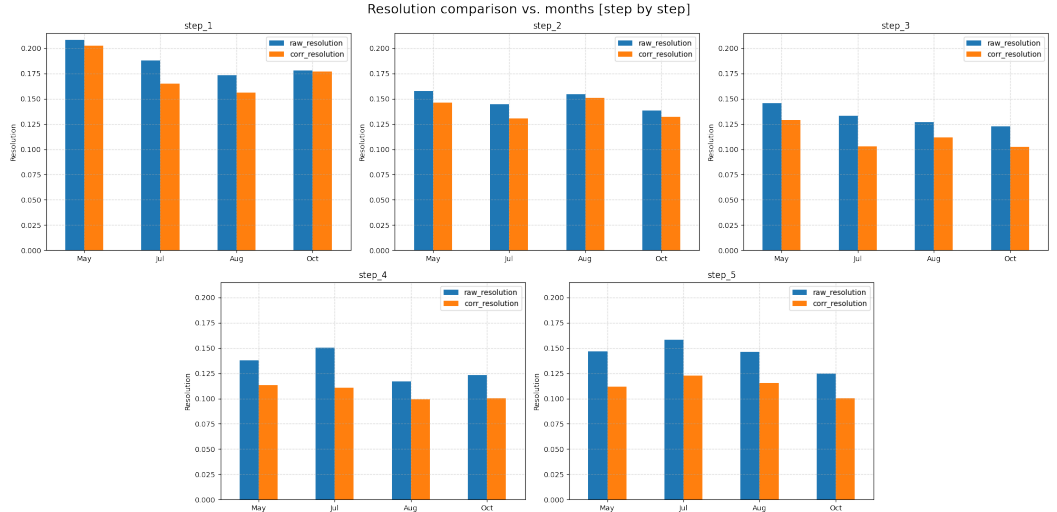


Figure A.2. Aggregated version of multiple dashboards (one per month) of the kind reproduced in figure A.1 which shows only the resolution bar plots for every step and every month.

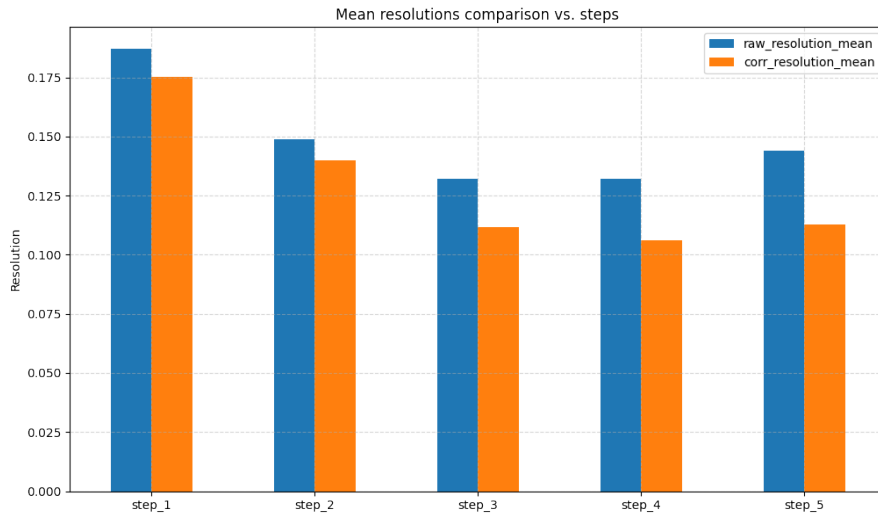


Figure A.3. Aggregated version of the plot reported in figure A.2, obtained by taking the mean value of the resolution for each step in the various months.

is reported in figure A.2, conveniently re-shuffling data in order to have different panels for each step including every month inside each panel.

The results are crystal clear and there is no cherry-picking in the choice of the daily calibrations: *the resolution after correction is always lower than before, at every distance from the anode and this is consistent between all the months of RUN 3 data taking.*

A more aggregated version of figure A.2 is obtained by taking the mean resolution value at each step between months, which produces, of course, the same considerations. This plot is reported in figure A.3

Appendix B

Elastic scattering kinematics from the angular point-of-view

The aim of this appendix is to provide a simple derivation of the formulas which have been applied to model the angular behaviour of a sphere-like particle, assuming elastic scattering, inside the Monte Carlo simulation described in chapter 4.

First of all, let's setup the stage in the center of mass (**COM**) frame, with graphical reference in figure B.1:

- Let our particles be spheres of radii r and R .
- Let the two sphere's centers be separated from an "impact parameter" b .
- The two particles have a mass m and M , respectively .
- The two particles are moving towards each other with a velocity whose magnitude is v

From simple geometric arguments, we can deduce that the angle γ between the recoil direction and the horizontal axis is related to b , r and R from the equation B.1:

$$\sin \gamma = \frac{b}{r + R} \quad (\text{B.1})$$

And, from purely classical mechanics reasoning, we can deduce that the COM velocity is $v_c = mv/(m + M)$, while the two velocities after the scattering are, in magnitude:

$$v_m^* = \frac{M}{m + M}v \quad , \quad v_M^* = \frac{m}{m + M}v \quad (\text{B.2})$$

with the superscript $\star$ to represent the COM frame.

If we now move the discussion to the laboratory (**LAB**) frame, we are in front of the situation depicted in figure B.2.

In particular:

- The smaller particle is the projectile (a neutron, for example), impinging with a velocity v on the bigger particle which is at rest (a nucleus).
- After the scattering, the two particles recoil away with velocity U and W , and angle θ_U and θ_W with respect to the horizontal, respectively.

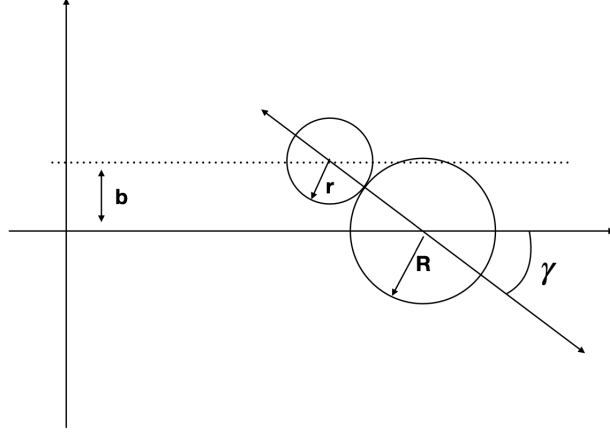


Figure B.1. Sketch of the elastic scattering in the center of mass reference frame.

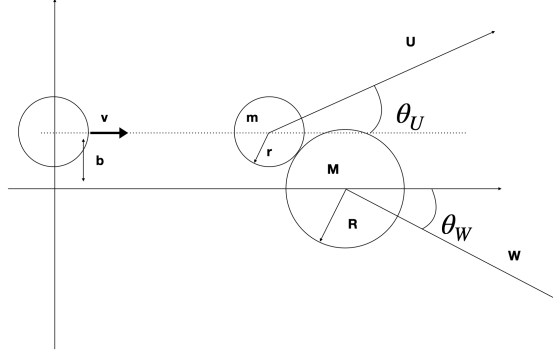


Figure B.2. Sketch of the elastic scattering in the laboratory reference frame.

Transforming equations B.2 with the Galileo transformations, we can obtain the following equations for the velocities in the LAB frame.

In particular we need to sum v_c to the x components of v_m^* , v_M^* , while the y components are unchanged. Note also that the $v_{m,x}^*$ is taken with a minus sign, since the positive direction for the horizontal axis is facing towards the right side of the plot in figures B.1 and B.2.

In the end, we obtain:

$$\begin{cases} U_x = \frac{v}{m+M}(m - M \cos \gamma) \\ U_y = \frac{M}{m+M}v \sin \gamma \end{cases}, \begin{cases} W_x = \frac{mv}{m+M}(1 + \cos \gamma) \\ W_y = -\frac{mv}{m+M} \sin \gamma \end{cases} \quad (\text{B.3})$$

and we can now extract the formulas for the recoiling angles by simply exploiting the fact that $y/x = \tan \theta$ for two vectors x and y separated by an angle θ .

What we obtain is:

$$\theta_U = \arctan \left(\frac{M \sin \gamma}{m - M \cos \gamma} \right), \quad \theta_W = \arctan \left(-\frac{\sin \gamma}{1 + \cos \gamma} \right) \quad (\text{B.4})$$

from which we can see that the recoiling nucleus angle is completely independent from the particles masses.

List of Figures

1.1	The Coma Cluster in the visible range [SkyCenter/University of Arizona].	2
1.2	The Coma Cluster in the x-ray range [ESA].	2
1.3	A composite image of the Bullet Cluster, in pink the X-Ray emission from the inter-cluster hot gas (slowed down from the collision), in blue the computed matter distribution required to produce the observed lensing effect. [ESA].	5
1.4	Vera Rubin operating the Kitt Peak National Observatory's telescope [NOAO/AURA/NSF].	6
1.5	M33 Doppler shift from the 21 cm Hydrogen line, which is telling us the rotation direction of the galaxy.	6
1.6	Superimposition of rotational curves for the NGC 3198 galaxy from [10]. The points are observed data, “disk” is the rotational curve expected from the visible mass alone, “halo” is the rotational curve expected from the dark matter halo alone, while the solid curve consistent with the data is the combination disk + halo.	7
1.7	upper panel: CMB power spectrum produced from Planck2018 with data superimposed to the best prediction from the Λ CDM model. lower panel: residuals. [12]	9
1.8	Large Scale Structures as simulated from the Millennium Project [14], also known as “ <i>cosmic web</i> ”.	11
1.9	A not-in-scale overview of the most important proposed dark matter candidates [Source Farinaldo Queiroz IIP-Natal].	12
1.10	A quantitative overview of ranges for dark matter candidates masses, anomalies, and search techniques [16].	13
1.11	Dark matter yield or density parameter vs. Universe temperature or time for a 100 GeV WIMP. Shaded regions account for different cross-sections, the solid line is produced from the cross-section which yields the correct density parameter, while the dashed line is produced considering a particle which stays in thermal equilibrium. Figure from [17].	15
1.12	Macroscopic ($\sim$ km scale) and microscopic ($\sim$ fm scale) description of the main concepts of dark matter direct detection, from [18]. . . .	17
1.13	Annual average rate spectra for different configurations of WIMP - target masses in the spin-independent case, assuming isosinglet approximation and fixing $\sigma_p = 10^{-46}$ cm ² . Note the dips visible in the W and Xe case, related to the Helm form factor vanish at specific values of momentum transfer. Dashed lines are computed neglecting the form factor contribution, while solid lines takes it into account. Plot form [18].	20

1.14	A schematic representation of the solar system orbiting around the galactic center always behind the Cygnus constellation (left) [1] and of the Earth inclination with respect to the galactic plane and Cygnus constellation (right) [24].	21
1.15	Expected differential recoil rate galactic map for a 100 GeV WIMPs population and for $E_R > 20$ keV [23].	22
1.16	Very realistic representation of the apparent wind of WIMPs generated by the Solar System motion following the Cygnus constellation. . . .	22
1.17	Current exclusion limits for spin-independent interactions from state-of-the-art experiments in the WIMP dark matter direct detection field. The excluded region (at 90% C.L.) is shaded in green, the individual contribution from the different experiments is reported in solid lines, the neutrino-floor region is delimited by the dashed orange line. Plot from [25].	23
2.1	ALICE results in particle identification based on their energy deposition density dE/dx vs momentum with superimposed the corresponding Bethe-Bloch theoretical curve. Plot from [38]	27
2.2	Time projection chamber geometry from [39].	27
2.3	Directional time projection chamber geometry from [1], where with χ we indicate a WIMP DM particle.	28
2.4	Schematic representation of a photomultiplier tube.	29
2.5	PMT waveforms for a track parallel (left) or tilted (right) with respect to the GEM plane.	30
2.6	50 μm thick GEM section pictured with electron microscope showing holes with 70 μm diameter and 140 μm pitch (left) and sketch of the electric field configuration inside GEM holes after the application of a large difference of potential (right).	31
2.7	Light intensity vs. wavelength for a He:CF ₄ (60:40) mixture. Plot from [41]	32
2.8	Diffusion parameters (upper panel) and discharge probability (upper panel) for a 60:40 and 70:30 He:CF ₄ gas mixture. Plots from [47]. .	33
2.9	Garfield simulation for transversal and longitudinal diffusion coefficients per unit length (left) and drift velocity (right) vs. drift field for He:CF ₄ (60:40). Plots from [48].	33
2.10	Average 3D range simulation for both He and electron recoils in a He:CF ₄ (60:40) gas mixture.	34
2.11	CYGNO timeline with all the steps completed and foreseen. As of writing this thesis, we are at the end of PHASE 0, with the CYGNO_04 demonstrator already commissioned [49] and its final installation expected in 2025.	35
2.12	LIME schematics as seen from above, with notes indicating each part of the detector. Picture from [50].	35
2.13	Picture of the LIME prototype from a diagonal perspective. From this view, one can see the cathode plane and the copper rings which ensure the uniform drift field, all enclosed in the transparent acrylic vessel. The GEM plane and the optical readout are covered by the rest of the detector, on the right [50].	36

2.14	Detailed view of the automated mechanism which is used to move the trolley with the Iron source to calibrate the detector. The wheel on the left can be rotated with a small electric engine managed from an Arduino board in order to move the source from the start to the end of the upper window by means of a series of pulleys. When not required, the source can be put in a “parking” position, over a 5 cent coin.	37
2.15	Quantum efficiency for both Orca Fusion and PMT R7378 (taken from Hamamatsu), with the He:CF ₄ (60:40) scintillation spectrum superimposed [41]. The camera QE maximum precisely matches the He:CF ₄ peak in the visible range.	38
2.16	Visual representation of the camera data acquisition which is described in the text. The red region is called Global Exposure and it corresponds to the time window in which all the rows of the sensor are simultaneously exposed. In the lower part, the blue trigger line simulates two hypothetical events triggered and acquired from the PMTs, with the first one possibly not linked to any track in the corresponding image.	39
2.17	⁵⁵ Fe differential rate with respect to the triggering threshold. Plot from [55].	40
2.18	Example of a full event with only one track and four waveforms. The PMTs positions are labeled in the image reference frame, keeping coherence with the colors in the waveforms plot. From the waveforms relative intensity, it is easy to understand the position of the associated track by eye.	41
3.1	Examples of tracks pictured from the LIME prototype.	44
3.2	sCMOS image with the excluded regions delimited from the red lines (left) and noise rms distribution extracted from a set of pedestal runs (right) [50].	45
3.3	Measure of the vignetting effect, obtained by imaging a uniformly illuminated white sheet of paper (left) [50], together with the vignetting map (right) [48].	46
3.4	Visual description of the directional super-clustering procedure, taken from [56].	47
3.5	Examples of results produced by the iDBSCAN (left) and iDDBSCAN (right). It is clear how the additional directional super-clustering step helps in connecting correlated clusters while also disentangling eventual superpositions. Pictures from [56].	47
3.6	Practical explanation of the relative waveform amplitude association method. Since the fourth PMT collects more charge, the track must be closer in space to that PMT. Note that the PMTs’ positioning is reflected in the waveforms’ arrangement in the figure. Picture reproduced from [63].	50
3.7	Sketch of the PMT geometrical acceptance with respect to the light emitted from the GEM plane. Picture reproduced from [64].	51
3.8	Fit of the charge trend vs. $(x - y)$ position of impulsed light. The fit clearly shows a R^{-4} scaling law. Note that in the figure r is the distance in the $(x - y)$ plane, while z is the distance from the emitting plane. Plot from [65].	52
3.9	²⁴¹ AmBe neutron spectrum from [67].	52

3.10	Comparison between raw and normalised distribution of <code>sc_integral</code> for both Bkg and AmBe datasets (upper) and normalised distribution only for the AmBe+Fe dataset (bottom), with the latter reported to have a rough idea of the energy scale (5.9 keV are approximately 10k ADU).	55
3.11	Color-map of the gain correction map, which is telling us that, on average, clusters in the upper region of the GEM plane experience an enhanced gain with respect to clusters in the lower region. The vertical lines are the GEM sectors.	56
3.12	Comparison of Bkg, AmBe and AmBe+Fe data before (left) and after (right) the correction for the gain non-uniformity. A nice check of the correction goodness is given from the iron peak, which after the correction assumes a more gaussian distribution. Moreover the peak around 20k Analog-to-Digital Units (ADU) in <code>sc_integral</code> is produced from two iron spots clustered together from the clustering algorithm. At the same time, the upper and lower panels' difference is addressed to the application of the cuts described just before section 3.3.1: the region around 5k ADU is clearly less populated after the cuts, since the Iron spots which are cut from the picture frame are removed.	57
3.13	Humidity trend during the AmBe data taking. Humidity is reported in parts per thousand (ppk).	58
3.14	Summary of the calibration procedure (top) and calibration check (bottom) performed on the Stability dataset. The top panel of each sub-figure shows the fitted iron peak for a sample of the actual runs in the dataset (one every ten) to avoid exceeding superposition. The transparent lines are real data, while solid lines are the fitted Gaussian peaks.	59
3.15	Summary of calibration check fit performed on AmBe+Fe data. . . .	60
3.16	Pressure trends during relevant timeframes for considered datasets. .	61
3.17	AmBe + Fe before (top panel) and after (bottom panel) calibration using the fit parameters obtained from the Stability dataset. After calibration, the iron peak appears shifted from its true position, probably because of the fact that the data taking was affected by changes in different parameters and not only in the humidity level, as better explained in the text.	61
3.18	AmBe + Fe inter-calibration.	62
3.19	Bkg_stability before (top panel) and after (bottom panel) calibration using the fit parameters obtained from the Stability dataset. The same considerations outlined for figure 3.17 apply.	63
3.20	Bkg_stability inter-calibration.	63
3.21	Comparison between raw and fully pre-processed energy spectra for Bkg and AmBe datasets. The error bars are computed as the square root of the counts and then normalised to the time of exposure. . . .	64
3.22	AmBe energy spectrum excess obtained from the subtraction of the Bkg spectrum to the AmBe one. The error bars are computed as the sum in quadrature of the AmBe and Bkg errors and then normalised to the time of exposure. They are relevant in order to justify the fact that after the subtraction there are bins with a rate lower than the inverse of the total time of exposure of the AmBe data taking (see table 3.1).	64

3.23	^{55}Fe peak, extracted from the subtraction of the AmBe energy spectrum to the AmBe+Fe one, with a superimposed gaussian fit. It is interesting to note that this plot demonstrates that we have 10% energy resolution, at least on Iron spots.	65
4.1	LIME setup in GEANT4 (left) and its real counterpart (right) during the data taking with AmBe source. Images from [48]	68
4.2	δ vs. <code>calib_sc_integral</code> distribution for the AmBe and Bkg datasets in two different ranges of <code>sc_integral</code> . The two lower panels , in particular, show two different zoom-in in energy and δ of the upper panel , with the superimposition of the excess region's boundaries and the events inside this region.	69
4.3	Distributions, normalised to the time of exposure, of <code>sc_integral</code> (left) and δ (right) for the events in the selection region for AmBe and Bkg datasets.	70
4.4	Superimposition of all the images acquired during 3 runs (left) and all the clusters in the same runs which are inside the NRs selection region described in the text.	70
4.5	Some examples of mowed clusters with high energy density ($\delta > 25$).	71
4.6	Results of the PCA procedure implemented to compute the clusters' angle with respect to the horizontal.	73
4.7	Angle distribution comparison between events in the AmBe excess region (see figure 4.2) extracted from the AmBe and expanded Bkg datasets (see tables 3.1 and 4.1).	74
4.8	Sketch of the elastic scattering in the laboratory reference frame.	75
4.9	Examples of nuclear recoils simulated with the Monte Carlo described in section 4.3.2. Left and right subplots shows different views (camera and top) and the projected angle are reported in the plot titles. Note that the orange arrow length is not related to the range of the nuclear recoil in the detector.	77
4.10	Comparison between low energy nuclear recoils angle distributions as seen from the camera point of view, sample from AmBe and Simulation (left) and sample from Bkg and Simulation (right). Note that the simulations are different, in order to reflect the same number of events, and that the y axis reports the probability density, in such a way that the extremes are the same in the two panels.	78
4.11	Comparison plots between real (from AmBe dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 5° and 45° , whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.	79
4.12	Comparison plots between real (from AmBe dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 50° and 85° , whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.	80
4.13	Comparison plots between real (from Bkg dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 5° and 45° , whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.	81

4.14	Comparison plots between real (from Bkg dataset) and simulated data after a Gaussian smearing on the latter with different angular resolution varying on each panel between 50° and 85° , whose value can be seen from the legend. The uncertainties are computed as the square root of the counts.	82
4.15	Line plot showing the χ^2 distance between AmBe / Bkg NRs' measured angular distributions and their respective simulated angular distributions provided by the toy Monte Carlo simulation, at different angular resolution.	83
4.16	Fits of the most prominent peak in sc_length and L_z distribution. .	83
4.17	Fits of the most prominent peak in sc_length , L_z and sc_tgausssigma distribution for step 1 (left) and step 5 (right) of the daily calibration procedure.	84
4.18	Linear fit of the squared length offsets vs. squared sc_tgausssigma trends for both sc_length (blue) and L_z (orange).	85
4.19	3D range vs energy for the lonely nuclear recoils sample, with the distribution of the single variables on the plot margins.	85
4.20	3D range vs energy for He, C and F nuclear recoils in He:CF ₄ (60:40) mixture, simulated with SRIM and reproduced from [70] with data in figure 4.19 superimposed.	86
4.21	SRIM simulated 3D energy vs range for He nuclear recoils in He:CF ₄ (60:40) mixture with data in figure 4.19 and a 2^{nd} order polynomial fit of the simulated data superimposed. The fit range of variation is expanded to match the experimental data variation domain.	87
4.22	Expected energy spectrum computed from the reconstructed 3D range, between 0 and 50 MeV. As it can be seen, there are a small number of outliers above 10 MeV, probably due to a wrong extrapolation of the fitted law reported in figure 4.21 or to an error in the range reconstruction, but the majority of the clusters have expected energies well below 10 MeV.	88
4.23	Expected vs. saturated energy distribution, which shows what should the real value of the measured energy be, based on a 2^{nd} order polynomial fit of the simulated energy vs. range law for α nuclear recoils in He:CF ₄ (reported in figure 4.21).	88
4.24	Saturation factor distribution computed as the ratio between expected and saturated energy, y and x axis in figure 4.23, respectively.	89
A.1	Example of one of the cumulative dashboards produced for a random daily scan extracted from the data taking in May, showing sc_integral distributions before and after the correction map application, at every calibration step (number increases with increasing distance from the anode), together with their gaussian fit. In the end, a bar plot is produced in order to easily compare the resolution σ/μ at different steps.	94
A.2	Aggregated version of multiple dashboards (one per month) of the kind reproduced in figure A.1 which shows only the resolution bar plots for every step and every month.	94
A.3	Aggregated version of the plot reported in figure A.2, obtained by taking the mean value of the resolution for each step in the various months.	95
B.1	Sketch of the elastic scattering in the center of mass reference frame.	98

B.2 Sketch of the elastic scattering in the laboratory reference frame. . .	98
---	----

List of Tables

3.1	Relevant information about the datasets involved in the analysis. The exposure of the Bkg_stability run is not provided since it is just one calibration run taken with the ^{55}Fe source at the step 3 of the calibration procedure (as described in section 2.3.1). Moreover the year in “date range” is omitted and implied to be 2023.	53
4.1	Relevant information about the Bkg dataset analysed in this section, expanded from 57 to 227 hours of effective exposure, to extract a large number of events in the AmBe excess region. The year in “date range” is omitted and implied to be 2023.	74

Bibliography

- ¹M. Schumann, “Direct detection of wimp dark matter: concepts and status”, [Journal of Physics G: Nuclear and Particle Physics](#) **46**, 103003 (2019) (cit. on pp. 1, 19, 21, 28).
- ²B. Ryden, *Introduction to cosmology* (Cambridge University Press, 1970) (cit. on pp. 1, 4, 8, 10, 12).
- ³S. Burles, K. M. Nollett, and M. S. Turner, “Big bang nucleosynthesis predictions for precision cosmology”, [The Astrophysical Journal](#) **552**, L1–L5 (2001) (cit. on p. 2).
- ⁴G. Gavazzi, “21 Centimeter Study of Spiral Galaxies in the Coma Supercluster”, [The Astrophysical Journal](#) **320**, 96 (1987) (cit. on p. 2).
- ⁵F. Zwicky, ““the redshift of extragalactic nebulae”, english translation (2009)”, [originally published in Helvetica Physica Acta](#) (1933) (cit. on p. 3).
- ⁶M. Fitchett and R. Webster, “Substructure in the Coma Cluster”, [The Astrophysical Journal](#) **317**, 653 (1987) (cit. on p. 4).
- ⁷J. M. Kubo et al., “The mass of the coma cluster from weak lensing in the sloan digital sky survey”, [The Astrophysical Journal](#) **671**, 1466–1470 (2007) (cit. on p. 4).
- ⁸D. Clowe, A. Gonzalez, and M. Markevitch, “Weak-lensing mass reconstruction of the interacting cluster 1e 0657-558 : direct evidence for the existence of dark matter”, [The Astrophysical Journal](#) **604**, 596–603 (2004) (cit. on p. 5).
- ⁹V. C. Rubin and J. Ford W. Kent, “Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions”, [The Astrophysical Journal](#) **159**, 379 (1970) (cit. on p. 6).
- ¹⁰T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi, “Distribution of dark matter in the spiral galaxy NGC 3198.”, [The Astrophysical Journal](#) **295**, 305–313 (1985) (cit. on pp. 6, 7).
- ¹¹D. J. Fixsen, “The temperature of the cosmic microwave background”, [The Astrophysical Journal](#) **707**, 916–920 (2009) (cit. on p. 8).
- ¹²N. Aghanim et al., “Planck2018 results: vi. cosmological parameters”, [Astronomy & Astrophysics](#) **641**, A6 (2020) (cit. on pp. 8, 9, 16).
- ¹³R. L. Workman et al. (Particle Data Group), “Review of Particle Physics”, [PTEP](#) **2022**, 083C01 (2022) (cit. on pp. 9, 16).
- ¹⁴V. Springel et al., “Simulations of the formation, evolution and clustering of galaxies and quasars”, [Nature](#) **435**, 629–636 (2005) (cit. on p. 11).
- ¹⁵C. Frenk and S. White, “Dark matter and cosmic structure”, [Annalen der Physik](#) **524**, 507–534 (2012) (cit. on p. 12).

- ¹⁶M. Battaglieri et al., *Us cosmic visions: new ideas in dark matter 2017: community report*, 2017 (cit. on p. 13).
- ¹⁷J. L. Feng, “Dark matter candidates from particle physics and methods of detection”, [Annual Review of Astronomy and Astrophysics](#) **48**, 495–545 (2010) (cit. on pp. 14, 15).
- ¹⁸E. Del Nobile, *The theory of direct dark matter detection: a guide to computations* (Springer International Publishing, 2022) (cit. on pp. 17–20).
- ¹⁹P. de Salas et al., “On the estimation of the local dark matter density using the rotation curve of the milky way”, [Journal of Cosmology and Astroparticle Physics](#) **2019**, 037–037 (2019) (cit. on p. 17).
- ²⁰Y. Zhou, X. Li, Y. Huang, and H. Zhang, “The circular velocity curve of the milky way from 5–25 kpc using luminous red giant branch stars”, [The Astrophysical Journal](#) **946**, 73 (2023) (cit. on p. 17).
- ²¹M. C. Smith et al., “The rave survey: constraining the local galactic escape speed”, [Monthly Notices of the Royal Astronomical Society](#) **379**, 755–772 (2007) (cit. on p. 17).
- ²²J. Lewin and P. Smith, “Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil”, [Astroparticle Physics](#) **6**, 87–112 (1996) (cit. on pp. 18, 19).
- ²³D. G. Cerdeño and A. M. Green, “Direct detection of wimps”, in [Particle dark matter](#) (Cambridge University Press, Jan. 2010), pp. 347–369 (cit. on pp. 18, 22).
- ²⁴R. W. SCHNEE, “Introduction to dark matter experiments”, in [Physics of the large and the small](#) (Mar. 2011) (cit. on p. 21).
- ²⁵J. Billard et al., “Direct detection of dark matter—appec committee report*”, [Reports on Progress in Physics](#) **85**, 056201 (2022) (cit. on p. 23).
- ²⁶C. Savage, G. Gelmini, P. Gondolo, and K. Freese, “Compatibility of dama/libra dark matter detection with other searches”, [Journal of Cosmology and Astroparticle Physics](#) **2009**, 010–010 (2009) (cit. on p. 23).
- ²⁷R. BERNABEI et al., “Dark matter particles in the galactic halo: results and implications from dama/nai”, [International Journal of Modern Physics D](#) **13**, 2127–2159 (2004) (cit. on p. 23).
- ²⁸R. Bernabei et al., “Final model independent result of dama/libra–phase1”, [The European Physical Journal C](#) **73**, 10.1140/epjc/s10052-013-2648-7 (2013) (cit. on p. 23).
- ²⁹R. Bernabei et al., “First model independent results from dama/libra-phase2”, [Nuclear Physics and Atomic Energy](#) **19**, 307–325 (2018) (cit. on p. 23).
- ³⁰E. Aprile et al., “Search for electronic recoil event rate modulation with 4 years of xenon100 data”, [Physical Review Letters](#) **118**, 10.1103/physrevlett.118.101101 (2017) (cit. on p. 23).
- ³¹D. S. Akerib et al., “Search for annual and diurnal rate modulations in the lux experiment”, [Physical Review D](#) **98**, 10.1103/physrevd.98.062005 (2018) (cit. on p. 23).
- ³²E. Aprile et al., “Dark matter search results from a one ton-year exposure of xenon1t”, [Physical Review Letters](#) **121**, 10.1103/physrevlett.121.111302 (2018) (cit. on p. 24).

- ³³P. Agnes et al., “Low-mass dark matter search with the darkside-50 experiment”, [Physical Review Letters](#) **121**, [10.1103/physrevlett.121.081307](#) (2018) (cit. on p. 24).
- ³⁴A. H. Abdelhameed et al., “First results from the cresst-iii low-mass dark matter program”, [Physical Review D](#) **100**, [10.1103/physrevd.100.102002](#) (2019) (cit. on p. 24).
- ³⁵D. R. Nygren, “The Time Projection Chamber: A New 4 pi Detector for Charged Particles”, eConf **C740805**, edited by J. Kadyk, D. Nygren, W. Wenzel, and F. Winkelmann, 58 (1974) (cit. on p. 25).
- ³⁶J. Alme et al., “The alice tpc, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events”, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **622**, 316–367 (2010) (cit. on p. 25).
- ³⁷L. Baudis et al., “A dual-phase xenon tpc for scintillation and ionisation yield measurements in liquid xenon”, [The European Physical Journal C](#) **78**, [10.1140/epjc/s10052-018-5801-5](#) (2018) (cit. on p. 25).
- ³⁸ALICEcollaboration, “First results of the ALICE detector performance at 13 TeV”, (2015) (cit. on pp. 26, 27).
- ³⁹I. Fleck, M. Titov, C. Grupen, and I. Buvat, *Handbook of Particle Detection and Imaging* (Springer, 2021) (cit. on p. 27).
- ⁴⁰F. Sauli, *Principles of operation of multiwire proportional and drift chambers*, tech. rep., CERN, Geneva, 1975 - 1976 (CERN, Geneva, 1977) (cit. on p. 26).
- ⁴¹M. Fraga et al., “The gem scintillation in he–cf₄, ar–cf₄, ar–tea and xe–tea mixtures”, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **504**, [Proceedings of the 3rd International Conference on New Developments in Photodetection](#), 88–92 (2003) (cit. on pp. 28, 32, 38).
- ⁴²R. Campagnola, “Study and optimization of the light-yield of a triple-GEM detector”, [Rome U., Presented 20 Feb 2018](#) (2018) (cit. on pp. 29, 32).
- ⁴³V. Antochi et al., “Combined readout of a triple-gem detector”, [Journal of Instrumentation](#) **13**, P05001 (2018) (cit. on p. 29).
- ⁴⁴F. Sauli, “The gas electron multiplier (gem): operating principles and applications”, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **805**, [Special Issue in memory of Glenn F. Knoll](#), 2–24 (2016) (cit. on pp. 30, 31).
- ⁴⁵D. Pinci, “A triple GEM detector for the muon system of the LHCb experiment”, [Cagliari U., Presented on 14 Dec 2006](#) (2006) (cit. on p. 31).
- ⁴⁶F. Amaro et al., “Secondary scintillation yield from gem electron avalanches in image 1-image 2 and image 1-image 2-isobutane for cygno — directional dark matter search with an optical tpc”, [Physics Letters B](#) **855**, 138759 (2024) (cit. on p. 32).
- ⁴⁷E. Baracchini et al., “Stability and detection performance of a gem-based optical readout tpc with he/cf₄ gas mixtures”, [Journal of Instrumentation](#) **15**, P10001–P10001 (2020) (cit. on pp. 32, 33).
- ⁴⁸F. D. Giambattista, “Background study of the lime prototype at underground lngs for the cygno experiment”, PhD thesis (Gran Sasso Science Institute, 2024) (cit. on pp. 33, 46, 67, 68).

- ⁴⁹G. Mazzitelli et al., *Technical design report - tdr cygno-04/initium* (Feb. 2023) (cit. on p. 35).
- ⁵⁰F. D. Amaro et al. (CYGNO), *A 50 liter cygno prototype overground characterization*, 2023 (cit. on pp. 34–36, 45, 46, 60, 86).
- ⁵¹M. Marafini et al., “Orange: a high sensitivity particle tracker based on optically read out gem”, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **845**, [Proceedings of the Vienna Conference on Instrumentation 2016](#), 285–288 (2017) (cit. on p. 34).
- ⁵²V. Antochi et al., “Performance of an optically read out time projection chamber with ultra-relativistic electrons”, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **999**, 165209 (2021) (cit. on p. 34).
- ⁵³G. Dho, “Optimisation of amplification and gas mixture for directional Dark Matter searches with the CYGNO/INITIUM project”, [Gran Sasso Science Institute](#) (2023) (cit. on pp. 38, 83).
- ⁵⁴A. Henelius and J. Tornaiainen, “Midas: open-source framework for distributed online analysis of data streams”, [SoftwareX](#) **7**, 156–161 (2018) (cit. on p. 39).
- ⁵⁵D. Pinci, “Cygno meeting news”, [INFN Indico Agenda](#) (2024) (cit. on p. 40).
- ⁵⁶F. D. Amaro et al., “Directional idbscan to detect cosmic-ray tracks for the cygno experiment”, [Measurement Science and Technology](#) **34**, 125024 (2023) (cit. on pp. 43, 47).
- ⁵⁷Y. Dong and S. Xu, “A new directional weighted median filter for removal of random-valued impulse noise”, [IEEE Signal Processing Letters](#) **14**, 193–196 (2007) (cit. on p. 45).
- ⁵⁸A. Rowlands, *Physics of digital photography*, 2053-2563 (IOP Publishing, 2017) (cit. on p. 45).
- ⁵⁹M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise”, in *Proceedings of the second international conference on knowledge discovery and data mining, KDD’96* (1996), pp. 226–231 (cit. on p. 45).
- ⁶⁰E. Baracchini et al., “A density-based clustering algorithm for the cygno data analysis”, [Journal of Instrumentation](#) **15**, T12003–T12003 (2020) (cit. on p. 46).
- ⁶¹M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”, [Commun. ACM](#) **24**, 381–395 (1981) (cit. on p. 46).
- ⁶²D. Marques, “Pmt reco integration”, [INFN Indico Agenda](#) (2024) (cit. on p. 49).
- ⁶³D. Marques, “Pmt reco integration”, [INFN Indico Agenda](#) (2024) (cit. on p. 50).
- ⁶⁴F. Borra, “Study of the PMTs signals during the first underground run of the LIME prototype for the CYGNO experiment”, Presented 16 Jan 2023 (U. Rome La Sapienza, 2023) (cit. on pp. 50, 51).
- ⁶⁵P. Martini, D. Migliorati, G. Muco, and L. Zappaterra, “Physics laboratory II report” (U. Rome La Sapienza) (cit. on pp. 50–52).
- ⁶⁶M. Folcarelli, “Characterization of the CYGNO experiment prototype during the underground campaign at LNGS”, Presented 23 Oct 2023 (U. Rome La Sapienza, 2023) (cit. on pp. 51, 53, 54).

- ⁶⁷J. Fantidis, “Comparison of different geometric configurations and materials for neutron radiography purposes based on a $^{241}\text{Am}/\text{Be}$ neutron source”, [Journal of Taibah University for Science](#) **11**, 1214–1220 (2017) (cit. on p. 52).
- ⁶⁸E. Baracchini et al., “Identification of low energy nuclear recoils in a gas time projection chamber with optical readout”, [Measurement Science and Technology](#) **32**, 025902 (2020) (cit. on p. 68).
- ⁶⁹I. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments”, [Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences](#) **374**, 20150202 (2016) (cit. on p. 72).
- ⁷⁰E. Marconato, “Studio del segnale di una TPC a lettura ottica per la ricerca di Dark Matter” (U. Rome La Sapienza, 2019) (cit. on p. 86).
- ⁷¹CMS Collaboration, “Energy calibration and resolution of the cms electromagnetic calorimeter in pp collisions at 7 tev of invariant mass”, [Journal of Instrumentation](#) **8**, P09009 (2013) (cit. on p. 86).
- ⁷²A. Prajapati, “Multivariate analysis for background rejection in cygno/initium experiment”, PhD thesis (Gran Sasso Science Institute, 2024) (cit. on p. 87).