Identification of nuclear recoils in a gas TPC with optical readout

E Baracchini[1,2], L Benussi[3], S Bianco[3], C Capoccia[3], M Caponero[3,4], G Cavoto[5,6], A Cortez[1,2], I A. Costa[7], E Di Marco[5], G D'Imperio[5], G Dho[1,2], F Iacoangeli[5], G Maccarrone[3], M Marafini[5,8], G Mazzitelli[3], A Messina[5,6], R A. Nobrega[7], A Orlandi[3], E Paoletti[3], L Passamonti[3], F Petrucci[9,10], D Piccolo[3], D Pierluigi[3], D Pinci[5], F Renga[5], F Rosatelli[3], A Russo[3], G Saviano[3,11] and S Tomassini[3]

[1]Gran Sasso Science Institute, L'Aquila, I-67100, Italy
[2]Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali del Gran Sasso, Assergi, Italy
[3]Istituto Nazionale di Fisica Nucleare , Laboratori Nazionali di Frascati, I-00044, Italy
[4]ENEA Centro Ricerche Frascati, Frascati, Italy
[5]Istituto Nazionale di Fisica Nucleare, Sezione di Roma, I-00185, Italy
[6]Dipartimento di Fisica Sapienza Università di Roma, I-00185, Italy
[7]Universidade Federal de Juiz de Fora, Juiz de Fora, Brasil
[8]Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi", Piazza del Viminale 1, Roma, I-00184, Italy
[9]Dipartimento di Matematica e Fisica, Università Roma TRE, Roma, Italy
[10]Istituto Nazionale di Fisica Nucleare, Sezione di Roma TRE, Roma, Italy
[11]Dipartimento di Ingegneria Chimica, Materiali e Ambiente, Sapienza Università di Roma, Roma, Italy

E-mail: emanuele.di.marco@roma1.infn.it

May 2020

Abstract. The search for a novel technology able to detect and reconstruct nuclear recoil events in the keV energy range has become more and more important as long as vast regions of high mass WIMP-like Dark Matter candidate have been excluded. Gaseous Time Projection Chambers (TPC) with optical readout are very promising candidate combining the complete event information provided by the TPC technique to the high sensitivity and granularity of last generation scientific light sensors. A TPC with an amplification at the anode obtained with Gas Electron Multipliers (GEM) was tested at the Laboratori Nazionali di Frascati. Photons and neutrons from radioactive sources were employed to induce recoiling nuclei and electrons with kinetic energy in the range from 1 to few tens of keV. A He-CF_4 (60/40) gas mixture was used and the light produced during the multiplication in the GEM channels was acquired by a high position resolution and low noise scientific CMOS camera and a photomultiplier. A multi-stage pattern recognition algorithm based on an advanced clustering technique is presented here. A number of cluster shape observables are used to identify nuclear

recoils induced by neutrons originated from a AmBe source against X-ray ^{55}Fe photo-electrons. An efficiency of 18% to detect nuclear recoils with a 96% ^{55}Fe photo-electrons suppression with an energy of 5.6 keV is obtained. This makes this optically readout gas TPC a very promising candidate for future investigations of ultra-rare events as directional direct Dark Matter searches.

1. Introduction

The advent of a market of high position resolution and single photon light sensors can open new opportunity to investigate ultra-low rate phenomena as Dark Matter (DM) particle scattering on nuclei in a gaseous target.

The nature of DM is still one of the key issues to understand our Universe [1, 2]. Different models predicts the existence of neutral particles with a mass of GeV or higher that would fill our Galaxy. They could interact with the nuclei present in ordinary matter producing highly ionizing nuclear recoils but with a kinetic energy as small as few keV. Moreover, given the motion of the Sun in the Milky Way towards the Cygnus constellation such nuclear recoils would exhibit a dipole angular distribution in a terrestrial detector. In this paper the use of a scientific CMOS camera to capture the light emitted by Gas Electron Multipliers (GEMs) in a Time Projection Chamber (TPC) device is described. The GEMs are located in the TPC gas volume at the anode position and are used to convert the ionization produced in the gas by the nuclear recoils into flashes of visible light. The emitted light and its spatial distribution is located in the detector geometry by using a clustering algorithm. Neutron and γ radiation emitted by radioactive sources are used to set in motion atomic electrons and nuclei, respectively, in the gas volume. Moreover, natural radiation as cosmic rays is leaving a trail of ionization in the gas.

All types of interactions produce distinctive patterns of light emission from the GEMs that can be reconstructed and analyzed. Therefore, nuclear recoils can be efficiently identified and separated from different kinds of background down to a few keV kinetic energy. The study of the optical readout of a TPC has been recently conducted with several small size prototypes (NITEC [3], ORANGE [4, 5], LEMON [6–8]) with various particle sources, in the context of the CYGNO project. In the following, the study of nuclear recoils excited by neutrons from an AmBe source and electron recoils from a ^{55}Fe source in the gas volume of the LEMON prototype is presented.

2. Experimental layout

A 7 liter active sensitive volume TPC (named LEMON) was employed to detect the particle recoils. A sketch (not to scale) of the detector setup is shown in Fig. 1 (left), while an image of the detector in the experimental area is shown in Fig. 1 (right). The sensitive volume where the ionization electrons are drifting features a $200\times240\,\mathrm{mm}^2$ elliptical field cage with a 200 mm distance between the anode and the cathode. The

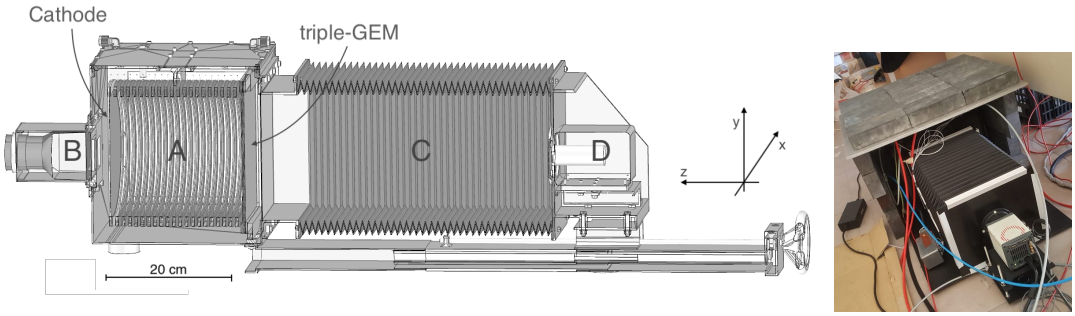

Figure 1. Left: the LEMON prototype with its 7 liter sensitive volume (A), the PMT (B), the adjustable bellow (C) and the sCMOS camera with its lens (D). Right: LEMON with the lead shield around the drift volume cage. The sCMOS camera (on the front) is looking at the GEMs through a blackened bellow.

anode side is instrumented with a $200\times240\,\mathrm{mm}^2$ rectangular triple GEM structure. Standard LHCb-like [9] GEMs ($70\,\mu$m diameter holes and $140\,\mu$m pitch) were used with two 2 mm wide transfer gaps between them. The light emitted from the GEMs is detected with an ORCA-Flash 4.0 camera [10] through a $203\times254\times1\,\mathrm{mm}^3$ transparent window and a bellow of adjustable length. This camera is positioned at a 52 cm distance from the outermost GEM layer and is based on a sCMOS sensor with high granularity (2048×2048 pixels), very low noise (around two photons per pixel), high sensitivity (70% quantum efficiency at 600 nm) and good linearity. This camera is instrumented with a Schneider lens (with an aperture f/0.95 and a focal length of 25 mm). The lens is placed at a distance $d = 50.6\,\mathrm{cm}$ from the last GEM in order to obtain a de-magnification $\Delta = (d/f)-1 = 19.25$ to image a surface $25.6\times25.6\,\mathrm{cm}^2$ onto the $1.33\times1.33\,\mathrm{cm}^2$ sensor. In this configuration, each pixel is therefore imaging an effective area of $125\times125\,\mu\mathrm{m}^2$ of the GEM layer. The fraction of the light collected by the lens is evaluated [11] to be 1.7×10^{-4}. A semi-transparent mesh was used as a cathode in order to collect light on that side also with a $50\times50\,\mathrm{mm}^2$ *HZC Photonics XP3392* photomultiplier [12] (PMT) detecting light through a transparent $50\times50\times4\,\mathrm{mm}^3$ fused silica window. More details can be found in Ref. [13].

A 5 cm thick lead shielding was mounted around the LEMON field cage to reduce the environmental natural radioactivity background. From the measurements of the GEM current with and without the lead shielding, a factor two reduction in the total ionization within the sensitive volume, very likely due to environmental radioactivity, was estimated.

3. Particle images in LEMON gas volume

The LEMON detector was operated in an overground location at Laboratori Nazionali di Frascati (LNF) with a He-CF_4 (60/40) gas mixture, the triple GEM system set at a voltage across the GEM sides of 460 V and an electric field between the GEM layers of 2.5 kV/cm. A six independent HV channels *CAEN A1257* module ensured stability

and monitored the bias currents with a precision of 20 nA. The gas mixture was kept at atmospheric pressure under continuous flow of about 200 cc/min and with the GEMs operated at 2.0×10^6 gain. The typical photon yield for this type of gas mixtures has been measured to be around 0.07 photons per avalanche electron [11, 14, 15]. The field cage was powered by a *CAEN N1570* [16], generating an electric field of 0.5 kV/cm.

The motion of particles within the gas mixtures was studied by means of different simulation tools. In particular, GARFIELD [17, 18] program was used to evaluate the transport properties for ionization electrons in the sensitive volume for an electric field of 500 V/cm.

Given the diffusion in the gas, ionization electrons produced at a distance z from the GEM will distribute over a region on the GEM surface, having a Gaussian transverse profile with a σ given by:

$$\sigma = \sqrt{\sigma_0^2 \oplus D^2 \cdot z}, \tag{1}$$

where D is the transverse diffusion coefficient, calculated to be 140 $\mu\text{m}/\sqrt{\text{cm}}$. The value of σ_0^2 was measured to be about 300 μm [19, 20]. Therefore, in average, a point-like ionization will result in a spot of 3–4 mm^2.

The expected effective ranges of electron and He-nucleus recoils were evaluated respectively with GEANT4 [21] and with SRIM [22] simulation programs. The recoil range estimated from simulation, as a function of the impinging particle kinetic energy, is shown in Fig. 2. These results show that:

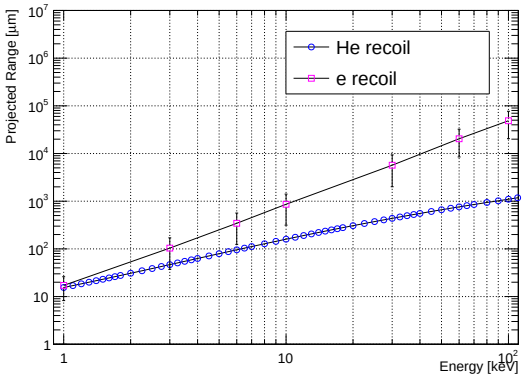

Figure 2. Average ranges for electron and He-nucleus recoils as a function of their kinetic energy. Range proiettato sulla direzione iniziale della traccia (parallela al piano delle GEM). Poco significativo. Spero che Giulia faccia un plot sul range 3D

- He-nuclei recoils have a sub-millimetre range up to energies of 100 keV and are thus expected to produce bright spots with sizes mainly dominated by diffusion;
- low energy (less than 10 keV) electron recoils are in general longer than He-nucleus recoils with same energy and are expected to produce less intense spot-like signals. For a kinetic energy of 10 keV, the electron range becomes longer than 1 mm and for few tens of keV, tracks of few cm are expected.

The sCMOS sensor was operated in continuous mode with a global exposure time of 30 ms. Typical images are shown in Fig. 3.

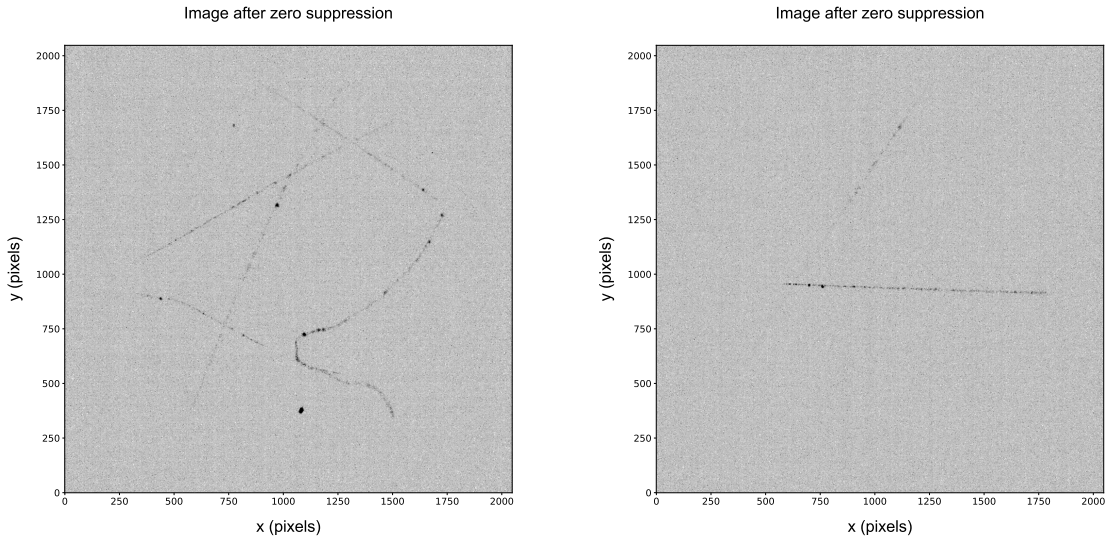

Figure 3. Two typical pictures taken with the sCMOS camera with a 30 ms exposure time. Left: cosmic tracks and natural radioactivity signals are present. Right: two long cosmic rays tracks are present, observed in a data taking run without any artificial source.

The PMT waveform was sent to a digitizer board with a sampling frequency of 4 GS/s. The experiment trigger scheme is based on the PMT signal: if, during the exposure time window, this exhibits a peak exceeding a threshold of 80 mV, it is acquired in a time window of 25 μs and the corresponding sCMOS image is stored. The digitizer is operated in single-event mode. No more than one signal is recorded in each sCMOS exposure time. Therefore, the PMT information was mainly exploited to select events with a cosmic ray track.

Several light spots are visible with different ionization patterns due to different types of particles interacting in the gas. Figure 3 (left) shows an image with typical long tracks from cosmic rays traveling through the full gas volume, where clusters of light with larger energy deposition are clearly visible, superimposed to radioactivity events likely due to natural origin. Figure 3 (right) shows an example of a cleaner event with one straight cosmic ray track, that can be used for energy calibration purposes.

Two different artificial radioactive sources were employed for testing and studying the detector responses.

A **neutron** source, based on a 3.5×10^3 MBq activity ^{241}Am source contained in a Beryllium capsule (AmBe) was placed at a distance of 50 cm from the sensitive volume side. Because of the interactions between α particles produced by the ^{241}Am and the Beryllium nuclei, the AmBe source isotropically emits:

- photons with an energy of 59 keV produced by ^{241}Am;
- neutrons with a kinetic energy mainly in a range 1–10 MeV
- photons with an energy of 4.4 MeV produced along with neutrons in the interaction between α and Be nucleus.

The presence of a lead shield around the sensitive volume absorbed almost completely the 59 keV photon component. A small faction of them reached the gas through small gaps accidentally present between the lead bricks.

A ^{55}Fe source emitting **X-rays** with a main energy peak at 5.9 keV. This is the standard candle for calibration and performance evaluation of LEMON, and its extensive use is documented in Ref. [23].

The events in Fig. 4 show images recorded with the same 30 ms exposure time, in presence of the two sources. The left panel shows an example of several light spots, characteristic of energy deposits due to ^{55}Fe low energy photons. The right panel shows a frame recorded in presence of the AmBe radioactive source: the short and bright track well visible in the center is very likely due to a nuclear recoil induced by a neutron scattering.

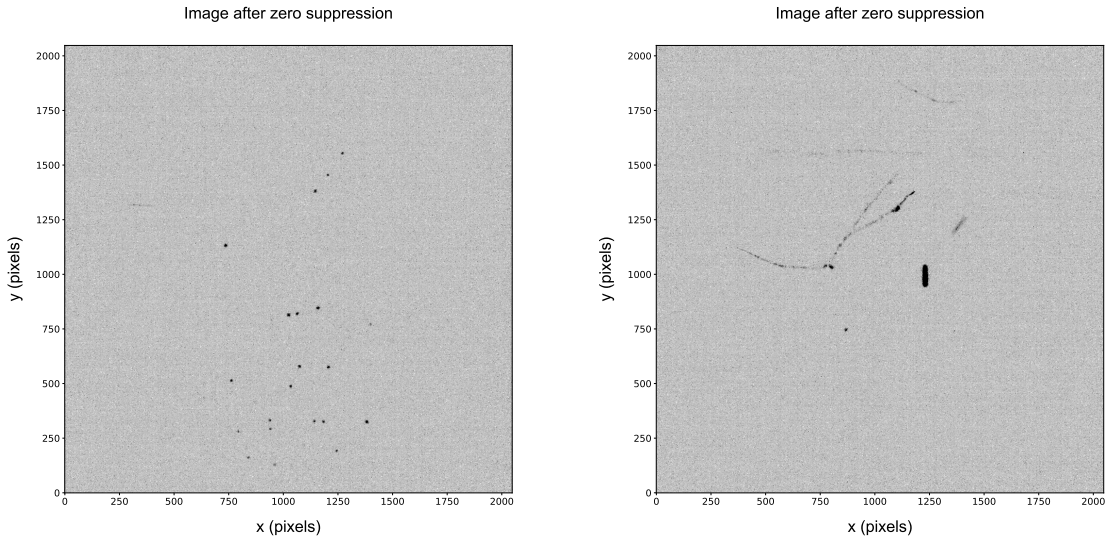

Figure 4. Two pictures taken with the sCMOS camera with a 30 ms exposure time. Left: picture taken in presence of ^{55}Fe radioactive source. Right: a nuclear recoil candidate is present, in an image with AmBe radioactive source, together with signals from natural radioactivity.

4. Cluster pattern recognition

The light produced in the multiplication process in the GEM and detected by the sCMOS sensor is associated in clusters of neighboring pixels. This is done by following the trail of energy deposition of the particle traveling through the gas of the sensitive volume. The deposited energy (that for stopped particles is equivalent to the total energy of the particle) is estimated by the amount of the light collected by the sensor. Therefore, it is of primary importance to have a reconstruction algorithm that includes all the camera pixels hit by the real photons originating from the energy deposits, while rejecting most of the electronic noise. This can either create fake clusters or, more likely, add pixels in the periphery of clusters of real photons, biasing the energy estimate.

The energy reconstruction follows a three-steps procedure: the single-pixel noise suppression is briefly described in Section 4.1. This is followed by the proper clustering: first the algorithm to form basic clusters from single small deposits is described in Section 4.2, then the supercluster method, aiming to follow the full track pattern, and seeded by the basic clusters, is described in Section 4.3.

The results of this paper are based on the properties of the reconstructed superclusters and are described in Section 5.

4.1. Noise suppression

The electronic noise of the sensor was estimated in data-taking runs acquired with the sensor in complete dark (*pedestal* runs). For each pixel, the pedestal was computed as the average of the counts over many frames, while the electronic noise was estimated as their standard deviation (SD). The distribution of the pixels SD is shown in Fig. 5. The mode of this distribution is about 1.8 photons per pixels, but a tail is present, with pixels having a noise of more than 5 photons per pixels. For such pixels, a very non-Gaussian distribution was observed, while for the pixels in the bulk of the distribution, the pedestal distribution followed a Gaussian shape. To form the pedestal-subtracted image, the pedestal mean μ_i was subtracted to the image for each i^{th} pixel. An initial noise suppression was applied by neglecting the pixels with counts less than $1.3 \times \mathrm{SD}_i$. On such pedestal-subtracted zero-suppressed images an upper threshold was applied to reject hot pixels, which are more likely due to sensor instabilities than to energy deposition. These were found to be not malfunctioning pixels since they disappeared after a power cycle of the camera: therefore a dynamic (run-by-run) suppression was needed. They were efficiently identified as high-intensity, isolated pixels, and distinguished by a true energy deposit, for which each pixel is surrounded by some other active pixels. A threshold was applied on the ratio R_9 between the pixel and the average of the counts in a 3×3 pixels matrix surrounding it, and a minimum number of two pixels above noise in that matrix was required to discriminate good from hot pixels.

The resolution of the resulting image was then reduced by forming *macro-pixels*, by averaging the counts in 4×4 pixel matrices. This was needed to reduce the combinatorics of the subsequent clustering algorithm to be executed in a reasonable time for each

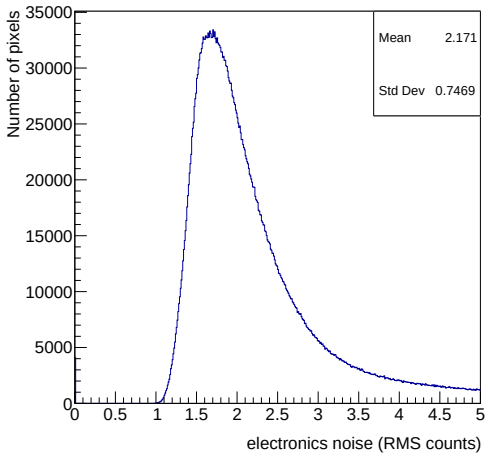

Figure 5. Distribution of the electronic noise of the sensor, estimated in images taken with sensor in complete dark, and evaluated as the SD of the distribution of the counts for each pixel.

image. On such 512×512 pixel map, a median filtering [24] was applied, as described in more details in Ref. [25]. The output image is passed to the basic clustering algorithm, described in the following.

4.2. Basic clusters reconstruction

The basic clustering algorithm, called IDBSCAN and described in details in Ref. [26], represents an evolution of the neighboring pixels clusters, called NNC, previously used to study the performances of the LEMON detector with ^{55}Fe radioactive source [23]. It is briefly described also here, since it represents the seeding for the final clustering algorithm.

The energy deposition in the sensitive volume of the TPC was estimated from the two-dimensional (2D) projection on the x–y axes of the light emitted in the multiplication process within the GEMs planes. The pattern showed a large variation, depending on the interacting particle. For events of ^{55}Fe calibration source, the signature of the typical 5.9 keV photons was a spot of few mm^2 with the exact size depending on the diffusion in the gas, i.e., on the distance from the anode along z of the energy deposition (see Fig. 4 left). Cosmic rays travel across the volume and leave a typical signature of a straight track, shown in Fig. 3 (right), but with several agglomerations with larger density along the path. Finally, natural radioactivity and the signal from nuclear recoils due to neutrons originated by the AmBe source showed an irregular pattern, sometimes curly, with several kinks along the path. Their track length and their size was found to depend a lot on the initial energy of the impinging neutron, and also on the mass of the recoiling nucleus.

Thus, the clustering algorithm needs to be flexible enough to efficiently reconstruct

a diverse set of patterns, from small round spots to long and kinky tracks. A first step of the clustering, called *seeding*, was used: it focused in the clustering of spot-like neighboring pixels. The method applied for the LEMON detector is an evolution of the classic DBSCAN algorithm [27]. This is a non-parametric, density-based clustering, which groups together pixels above threshold with many neighbors. Its distinctive characteristics making this method very suitable to the LEMON case is its ability to label as outliers, and so not to include in the clusters, pixels that lie isolated in low-density regions, i.e., pixels from electronic noise of the sensor surviving the zero suppression. The extension of DBSCAN used for LEMON data analysis consists in a larger phase space for the points that includes not only the x–y plane, but also the number of photons in each pixel N_{ph} (i.e., the light intensity measured in each pixel).

To be as inclusive as possible, and since different interactions may have vastly different intensities, even varying along the track, the clustering procedure was iterated three times. First, the DBSCAN parameters were tuned to form clusters of dense (in x–y dimension) and intense (in the N_{ph} dimension) pixels. The density in 3D was called *sparsity*. This step typically identifies either rare hot spots of the GEMs, or, efficiently, short nuclear recoils. The pixels belonging to the reconstructed clusters were then removed from the image, and the DBSCAN procedure was repeated, with looser sparsity parameters. The second iteration was tuned to efficiently reconstruct ^{55}Fe round spots and slices of tracks from nuclear recoils with lower intensity. It also collected the agglomerations with larger density along cosmic tracks, clearly visible in the example in Fig. 3 (right). A third iteration of DBSCAN with even looser parameters was finally executed, targeting faint portions of a cluster. These were especially used as a proxy for the characterization of clustered noisy pixels.

To be computationally viable, the IDBSCAN basic clustering was performed on the image with reduced resolution, 512×512. In typical images this allows the basic clusters reconstruction to be run in approximately 1 s on an *Intel Xeon E5-2620 2.00 GHz* and 64 GB RAM. The reconstruction algorithm is implemented in PYTHON3 [28], and interfaced with the CERN ROOT6 v.6 [29].

Examples of clustered pixels in two cases are shown in Fig. 6. The left panel shows an example of clusters reconstructed on the low-resolution image of one event with ^{55}Fe source. Three spots are clearly visible: one, as typical for events with this calibration source with a moderate activity, is reconstructed by a single cluster of the second iteration. The other two are close enough that are merged in a single cluster of the same iteration. The right panel shows the outcome of the IDBSCAN algorithm on a longer track presumably from natural radioactivity and one possible short nuclear recoil. The nuclear recoil candidate is very dense, high-energetic, and isolated, and it is reconstructed as a single cluster in the first iteration. The long track shows several clusters with higher intensity. One of them has a large energy, and it is reconstructed as an isolated single iteration-1 cluster. The rest of the track is reconstructed by multiple iteration-2 clusters, which are split where the energy deposition has a minimum for too many pixels to be joined together in the same clusters. Events like these, which

are frequent for cosmic rays, natural radioactivity, but also signals from nuclear recoils with higher energy, justify the need of the subsequent step of the *superclustering*, which follows the track pattern without splitting it in parts. This is described in the following section.

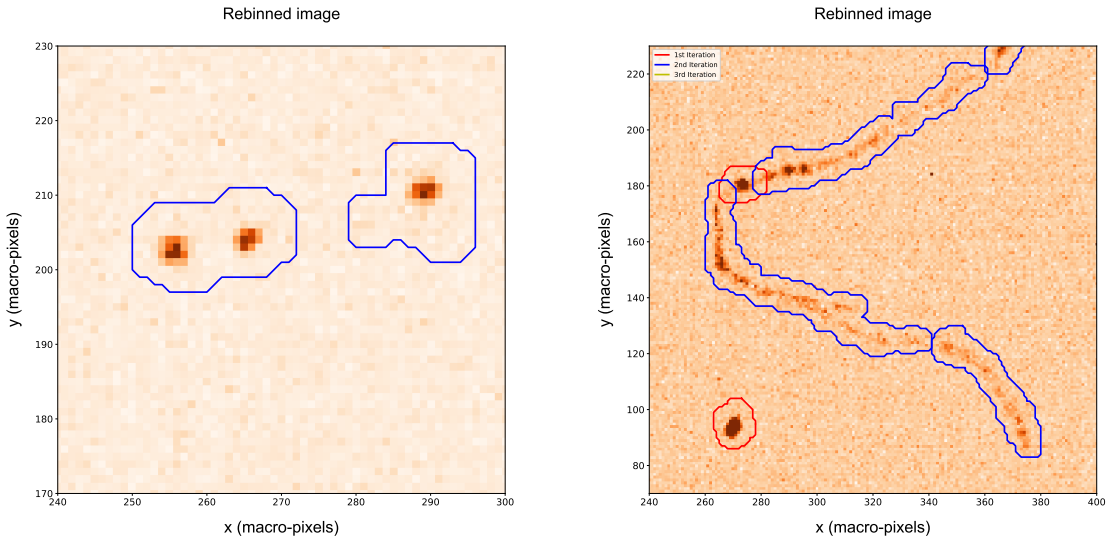

Figure 6. Basic clusters reconstructed with the IDBSCAN algorithm in the low resolution (512×512) image for two example events with very different patterns. Left: clusters on spots from ^{55}Fe source, two of which are merged together. Right: Track from natural radioactivity and a nuclear recoil candidate in an event with AmBe source. The long track is split in several basic clusters of different IDBSCAN iteration.

4.3. Superclusters reconstruction

The aim of the superclustering procedure is to collect the majority of the pixels belonging to a track which is long and, eventually, with an irregular pattern. The main limitation of IDBSCAN to follow a long track is mainly originated by the non uniform energy deposition along the path length. As can be clearly seen in Fig. 6 (right), or even in the example of a raw image of an event with two long cosmic rays in Fig. 3 (right), clusters with larger energy release are followed by regions along the path with a lower or even a zero release. These local minima are sometimes as large, in the 2D space, as the typical size of the ϵ parameter of DBSCAN [27]. Despite the low electronic noise of the ORCA-Flash 4.0 camera sensor, the energy releases in these local minima are similar in magnitude to the average single-pixel noise. The IDBSCAN is limited in connecting the full length of an extended path, because of two reasons. First, inflating ϵ parameter as much as needed to cover the areas of local minima conflicts with the need to reject noise around the cluster. The IDBSCAN parameters were optimized for the LEMON running conditions to collect most of the signals of $E \approx 5\,\mathrm{keV}$ and to reject the typical noise of ≈ 1 photon per pixel. This avoids collecting extra noise in the cluster, biasing

the energy scale and worsening its resolution, and keeps the rate of fake clusters at a negligible level. This is studied in great detail in Ref. [26]. Second, the iterative nature of the algorithm with very different parameters for each iteration, each tuned for very different intensity, makes it convenient and efficient for a deposition of a fixed energy density (like the spots originating from the ^{55}Fe source), but not for the cases as in Fig. 6 (right), where the same track is split in several parts, with some of them in different iterations. This requires a method that can continuously follow the pattern of the track, profiting of the full resolution image, where the *gradients* of the energy deposition along the track trajectory are smaller than the ones in the transverse direction. Moreover, executing any of the most common clustering methods on the full 1024×1024 image is not manageable CPU-wise, due to the huge pixel combinatorics.

The procedure adopted for the final supercluster reconstruction in the LEMON detector started from defining the *interesting regions* in the image that may contain pixels from an energy deposit. These are identified by the basic cluster algorithm IDBSCAN previously described, which is applied on the 512×512 reduced-resolution image. In order to gather the peripheral pixels, especially along the track trajectory where breaks into small basic clusters may have happened, a window of 5×5 pixels is considered, around each pixel belonging to a macro-pixel clustered in a basic cluster. A full resolution image formed only by the interesting pixels passing the simple initial filtering described in Sec. 4.1 was created. The gradients of the intensity N_{ph} in such image were computed pixel-by-pixel to look for the edge region where the image turns from signal to noise-only:

$$||\nabla(N_{ph})|| = \sqrt{\left(\frac{\partial N_{ph}}{\partial x}\right)^2 + \left(\frac{\partial N_{ph}}{\partial y}\right)^2}, \tag{2}$$

while the gradient direction is given by:

$$\theta = \tan^{-1}\left(\frac{\partial N_{ph}}{\partial y} \Big/ \frac{\partial N_{ph}}{\partial x}\right). \tag{3}$$

In order to reduce the effect of the noise which makes the first derivatives in Eq. 2 to fluctuate, a Gaussian filter is applied, with a 5σ threshold, where σ is the SD of the intensities of the pixels considered.

The superclustering algorithm, applied on the filtered image, is an application of the *morphological geodesic active contours* [30,31], called GAC in the following. This method uses an active contour finding, widely used in computer vision, where the boundary curve $\mathcal{C}$ of an object is detected by minimizing the *energy* E associated to $\mathcal{C}$:

$$E(\mathcal{C}) = \int_0^1 g(N_{ph})(\mathcal{C}(p)) \cdot |\mathcal{C}_p| dp, \tag{4}$$

where $ds = |\mathcal{C}_p|dp$ is the arc-length parameterization of the curve in the 2D space, and g is the stopping edge function, which allows to select the boundary of the cluster. In the GAC method used for the LEMON images, the g function is purely geometrical, and

uses the geodesics of the image, i.e., the local minimal distance path between points with the same gradient, defined before. The function $g(N_{ph})$ is given by:

$$g(N_{ph}) = \frac{1}{\sqrt{1 + \alpha|\nabla G_\sigma * N_{ph}|}}, \tag{5}$$

which is minimal in the edges of the image. The $G_\sigma * N_{ph}$ is the aforementioned 5σ Gaussian filter, and the parameter α, which regulates the strength of the filtering was tuned on typical LEMON images to be $\alpha = 100$.

This method was chosen because it allows to follow track patterns that may vary from convex to concave shape, eventually with kinks, e.g. in cases of δ-ray emissions. To improve the shrinking of the cluster boundary in the cases of tracks turning from concave to convex along their trail, the *balloon* force [31] is set to -1, in order to push the contour towards a border in the areas where the gradient is too small. A number of 300 iterations is used to evolve the supercluster contour.

The example track shown in Fig. 6 (right) after the basic clustering step, is shown again in full resolution, zoomed around the cluster, in Fig. 7 (left). The output of the superclustering with the GAC algorithm is shown on the right panel of the same figure. The splitting of the clustering, present after the basic cluster step, was recovered. The portions with high density and low density along the path of the energy deposition were joined together. Other three examples of superclustered images are shown in Fig. 8, in runs without any artificial radioactive source. The top left panel shows an example of a cosmic ray track fully reconstructed by the GAC superclustering, which also includes a δ-ray in the middle of the track length. The top right panel shows an example of curly track from a candidate of natural radioactivity interaction; bottom panel shows an example where both a cosmic ray and a curly track are present. In this case, the extremes of the long and straight track are still split, but this is much rarer than after the basic clustering, and it happens when the local minimal along the trajectories are compatible with noise-only for more than $\approx$1 cm.

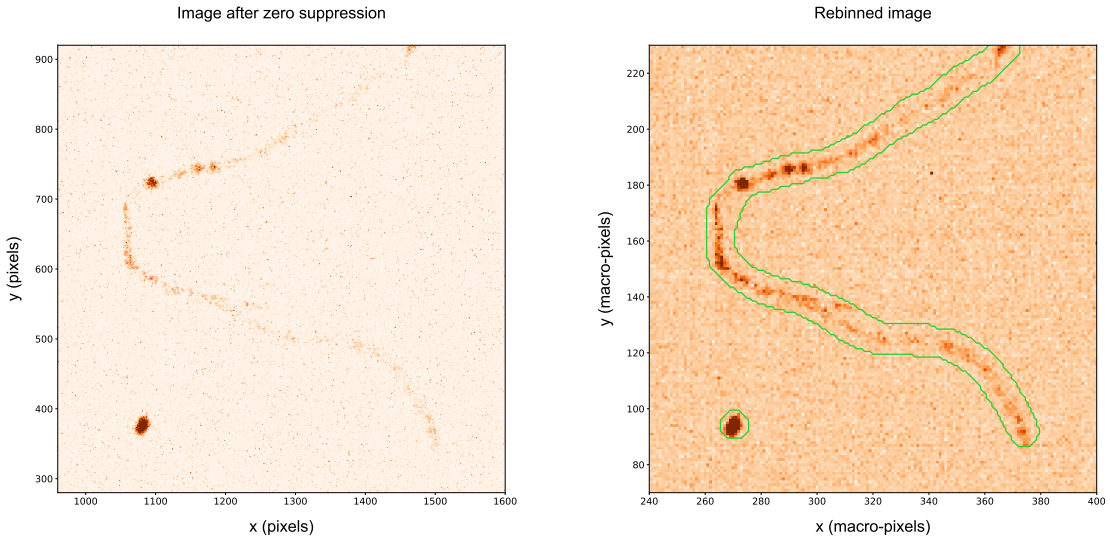

Figure 7. Left: zoom on the full-resolution image of a track candidate in a run with the AmBe radioactive source. Right: output of the superclustering on the rebinned image.

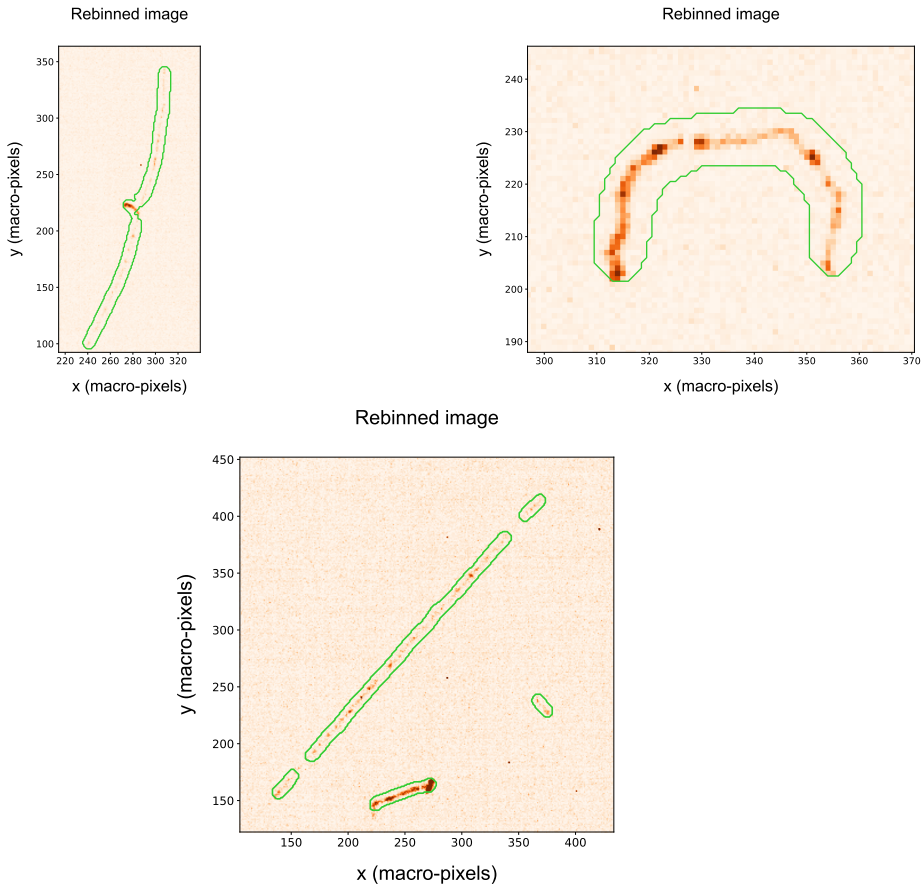

Figure 8. Superclusters reconstructed in a run without artificial radioactive sources. Top left: cosmic ray track fully reconstructed by the GAC superclustering. A δ-ray is included in the supercluster. Top right: curly track from a candidate of natural radioactivity interaction. Bottom: a cosmic ray with the extremes not joined to the main track, plus a curly track from natural radioactivity.

4.4. Superclusters calibration

The containment of the energy in the supercluster was verified with simulations performed with SRIM [22] of nuclear and electron recoils within the gas mixture of the LEMON detector. For both types of recoils, for the energy range of interest for DM search, e.g. $E < 10\,\text{keV}$, when considering deposits without electronics noise and no diffusion in the gas, the peak of the $|E - E_{true}|/E_{true}$ is within 5%. Adding a Gaussian noise distribution with a mean equal to the one observed in the pedestal runs, and a diffusion following the parameterization in Eq. 1, the fraction of the true energy contained in the supercluster decreases to about 80%, as shown in Fig. 9. The distributions were obtained for nuclear recoils with $E = 6\,\text{keV}$ generated at the exact center of the LEMON detector. The mean and the Gaussian width of the peak are estimated by fitting the distributions with a Crystal Ball shape [32, 33], which includes a tail to consider a non-Gaussian asymmetric tail due to partial containment in the supercluster.

The decrease in the energy containment in the supercluster is due to the smearing of the 2D track pattern around the periphery of the cluster. This decreases the gradients in Eq. 2 around the edges, and so the supercluster can shrink more around the crest, loosing part of the tails that can be confused more easily with the noise. A more realistic noise description, and an improved diffusion model, based on the one measured in data is necessary to tune the supercluster parameters in simulation to recover part of the containment. Despite this, the resolution in presence of noise and diffusion is estimated in this simulated nuclear recoils with $E = 6\,\text{keV}$ to be around 4%. This energy resolution is expected to be very optimistic because the absence, in the simulation, of the dominant fluctuations of the GEM gain.

The absolute energy scale was then calibrated with the energy distribution measured in data with the ^{55}Fe source, which provides monochromatic photons of 5.9 keV, with the procedure described in Ref. [23]. The supercluster integral is defined as:

$$I_{SC} = \sum_{i}^{cluster} N_i, \tag{6}$$

where N_i is the number of counts (photons) in the i^{th} pixel, and the sum runs over all the pixels of the supercluster. While to perform the basic- and super-clustering only pixels passing the zero suppression are considered, for the energy estimate in Eq. 6 all the pixels within the cluster contours are counted, eventually having negative N_i, after the pedestal subtraction. This is done to avoid a bias on the energy estimate. The distribution of I_{SC}, for a run taken in presence of ^{55}Fe source, is shown in Fig. 10. The position of the maximum in the single-spot distribution in runs with ^{55}Fe source allowed to calibrate the absolute energy scale of the LEMON detector. The energy resolution for the reconstructed GAC superclusters is about 18%, similar to the one that can be obtained with only the basic clustering step with IDBSCAN [26], and improving the one with the simple NNC algorithm previously used [23]. This value is still much larger than the one obtained with the simulation of nuclear recoils at the same energy, but an

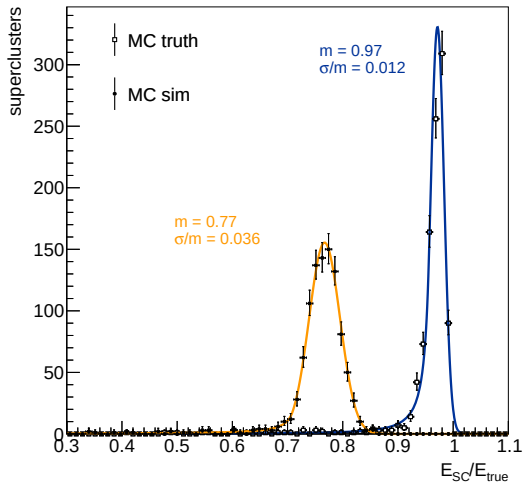

Figure 9. Distribution of the ratio of reconstructed supercluster energy, E, and the true energy of nuclear recoils $E_{true} = 6\,\text{keV}$, generated in the center of LEMON and simulated with GEANT4 Monte Carlo (MC). The hollow points show the MC events generated without electronics noise and diffusion in the gas, while the filled circles represent MC events with both effects included. The curves represent parametric fits with a Crystal Ball function.

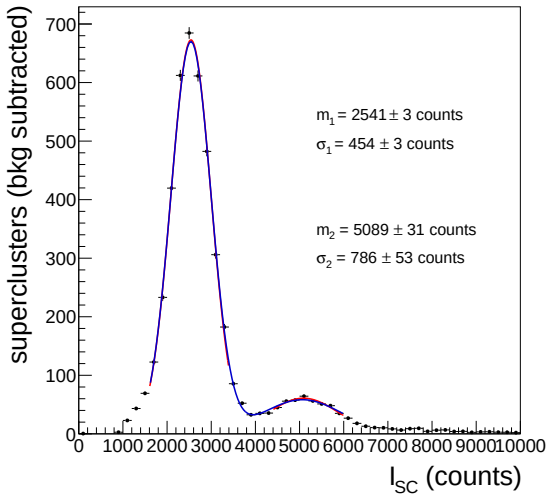

Figure 10. Distribution of the supercluster integral, before the absolute energy scale calibration is applied, in events with the ^{55}Fe source. Clearly visible is the large peak of a single spot, and, at around twice the energy, a broader peak for the case of two neighboring spots merged in a single supercluster.

396 improved noise and gas diffusion model, and a simulation of the gas gain fluctuations
397 are needed to improve the data-MC agreement.

Using runs with this monochromatic, high rate source, positioned at different distances from the GEM planes, a decrease of the light response for lower distances from the GEM was observed. This effect is opposite to the expected behavior of a decreasing light yield at larger distances. Indeed, it is expected that during the drift along the z-direction the ionization charge undergoes a diffusion in the TPC gas and some electrons are removed by attachment to the gas molecule. Consequently some loss in the light collection may be expected. The opposite behavior, instead, is clearly observed. While this effect is currently under study in more detail, it was attributed to a possible saturation effect of the GEMs, especially in the third stage of multiplication, for which the charge density in one GEM hole is maximal. Under this hypothesis, an effective, empirical correction was developed, which relies on the charge density of a cluster from a ^{55}Fe deposit. The light density, δ, is defined as:

$$\delta = I_{SC}/n_p, \tag{7}$$

where n_p is the number of pixels passing the zero-suppression threshold (differently from the numerator, where all the pixels in the supercluster are considered). This effective calibration provides the absolute energy of a spot-like region similar to the ^{55}Fe ones, as a function of the supercluster density, δ: $E = c(\delta) \cdot I_{SC}$. In the hypothesis of saturation, the *local* density along the track is the parameter which regulates the magnitude of the effect, thus the correction has to be applied dynamically for slices of the supercluster having a size similar to the ^{55}Fe spots. This is achieved with the procedure described in the following.

First, the supercluster *skeleton*, i.e., the 1 pixel wide representation along the energy deposition path, is reconstructed. This is performed with a morphological thinning of the superclusters with the iterative algorithm from Ref. [34,35]. Second, a pruning of the obtained skeleton is done, to remove residual branches along the main pattern, using a hit-or-miss transform. The output of the *skeletonization* for the track of Fig. 7 is shown in Fig. 11. For the calibration procedure, the skeleton was followed, starting from one of the two end points, and circles having their center on a pixel of the skeleton and their radius equal to the average spot size of the ^{55}Fe clusters were defined. The local density δ_s of the slice s is computed, and its integral I_s is calibrated to an absolute energy through the effective correction $E_s = c(\delta_s) \cdot I_s$. The pixels of the supercluster used for the slice calibration are removed (including the skeleton ones), and the procedure is iterated, until having included all the pixels. The sum of the energies of all the slices is the estimate of the calibrated energy of the supercluster:

$$E_{SC} = \sum_{s}^{slices} E_s \tag{8}$$

As a closure test of this procedure, the calibrated energy of the superclusters reconstructed in the runs with the ^{55}Fe source is obtained. The value of the energy peak was obtained by fitting the distribution with the same function used in Fig. 10, and equals to $m_1 = 5.93 \pm 0.01$ keV, compatible with the expected value. The calibration procedure is an overkill for the case of the small ^{55}Fe spots, but it is necessary for very

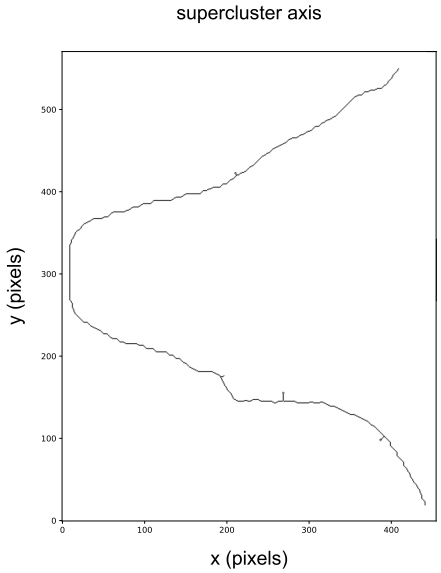

Figure 11. Output of the skeletonization and pruning of the branches for one example supercluster extended in space.

long cosmic ray tracks or even for medium-length superclusters from nuclear and electron recoils. The energy resolution worsen after the calibration ($\sigma_1 = 1.48 \pm 0.01$, i.e., 25% energy resolution), as a sign that the empiric correction is still suboptimal.

The skeletonization procedure provides a general method to estimate the track length (l_p), which is accurate both in the case of straight and curving track. In the case of straight tracks, the length extracted in this way coincides with the major axis estimated with a principal component analysis (PCA), described in the following section. For exactly round spots, the skeleton would collapse in the center of the cluster and the resulting length would be 1 pixel, but this completely symmetric case never happens in the considered samples.

5. Cluster shape observables

The interaction of different particles with the nuclei or the electrons in the gas of the TPC produce different patterns of the 2D projection of the initial 3D particle trajectory. These characteristics, to which we refer generically as "cluster shapes observables", are useful to discriminate different ionizing particles. In particular, they were used to select a pure sample of nuclear recoil candidates produced by the interaction of the neutrons originating from the AmBe source and to identify various sources of backgrounds. The main cluster shape observables are described in the following:

- *projected length and width:* a singular value decomposition (SVD) on the $x \times y$ matrix of the pixels belonging to the supercluster is performed. The eigenvectors found can be interpreted as the directions of the two axes of an ellipse in 2D. The

eigenvalues represent the magnitudes of its semiaxes: the major one is defined as *length*, l the minor one as *width*, w. These are well defined for elliptic clusters, or for long and straight tracks. The directions along the major and the minor axis are defined as *longitudinal* and *transverse* in the following. The longitudinal and transverse supercluster profiles, for the cosmic ray track candidates shown as an example in Fig. 8 (bottom) are shown in Fig. 12. The longitudinal profile shows the typical pattern of energy depositions in clusters, while the transverse profile, dominated by the diffusion in the gas, shows a Gaussian shape. It has to be noted that the cluster sizes represent only the projection of the 3D track in the TPC on the 2D x–y plane;

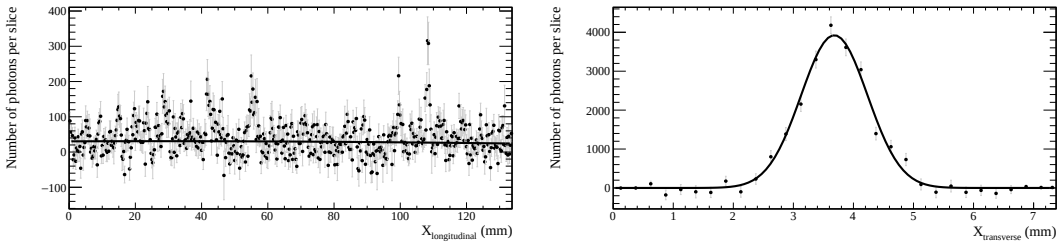

Figure 12. Supercluster profile in the longitudinal (top) or transverse (bottom) direction, for a long and straight cosmic ray track candidate shown in Fig. 8 (bottom). The longitudinal profile shows an energy deposition in sub-clusters, while the transverse direction shows the typical width of the diffusion in the gas. For the longitudinal profile, the line represent the average number of photons per slice. For the transverse profile, it represents a fit with a Gaussian PDF.

- *projected path length:* for curly and kinky tracks the values returned by the SVD of the supercluster are not an accurate estimates of their size. While the width is dominated by the diffusion, the length for patterns like the one shown in the example of Fig. 7 is ill-defined. In these cases, the path length, l_p, computed with the skeletonization procedure in Fig. 11 is used;
- *Gaussian width:* the original width of the track in the transverse direction is expected to be much lower than the observed width induced by the diffusion in the gas. Thus, as shown in Fig. 12 (right), the standard deviation, σ^T_{Gauss}, can estimated by a fit with a Gaussian probability density function (PDF);
- *slimness:* the ratio of the width over length, $\xi = w/l$, represents the aspect ratio of the cluster. It is very useful to discriminate between cosmic rays-induced background (long and thin) from low energy nuclear or electron recoils (more elliptical or round, as the ^{55}Fe spots);
- *integral:* the total number of photons detected by all the pixels gathered in the supercluster, I_{SC}, as defined in Eq. 6;
- *pixels over threshold:* the number of pixels in the supercluster passing the zero-suppression threshold, n_p;

- *density:* the ratio δ of I_{SC}, divided by n_p, as defined in Eq. 7;
- *energy:* the calibrated energy, expressed in keV. The calibration method simultaneously performs both the per-slice correction as a function of the local δ, and the absolute energy scale calibration, which corrects the non perfect containment of the cluster, using with ^{55}Fe source.

The projected supercluster path length, l_p, and Gaussian transverse size, σ^T_{Gauss}, are shown in Fig. 13, for data taken in different types of runs. During the data-taking approximately 3000 frames were recorded in absence of any external artificial source (*no-source* sample). In these frames the interaction of ultra-relativistic cosmic ray particles (mostly muons) are clearly visible as very long clusters. Internal radioactivity of the LEMON materials also contribute with several smaller size clusters. About 1500 frames were acquired with the AmBe source, and approximately 10^4 calibration images with ^{55}Fe source. In Fig. 13, as well as in the following ones showing other cluster properties, the distributions obtained in runs without artificial radioactive sources are normalized to the AmBe data total CMOS exposure time. For the data with ^{55}Fe source, since the activity of the source is such to produce about 15 clusters/event, the data are scaled by a factor one-tenth with respect to the AmBe exposure time for clearness. Considerations about the trigger efficiency scale factor between data with and without an artificial source are detailed later. The distributions in this section aim to show the different cluster shape observables among the different kinds of events.

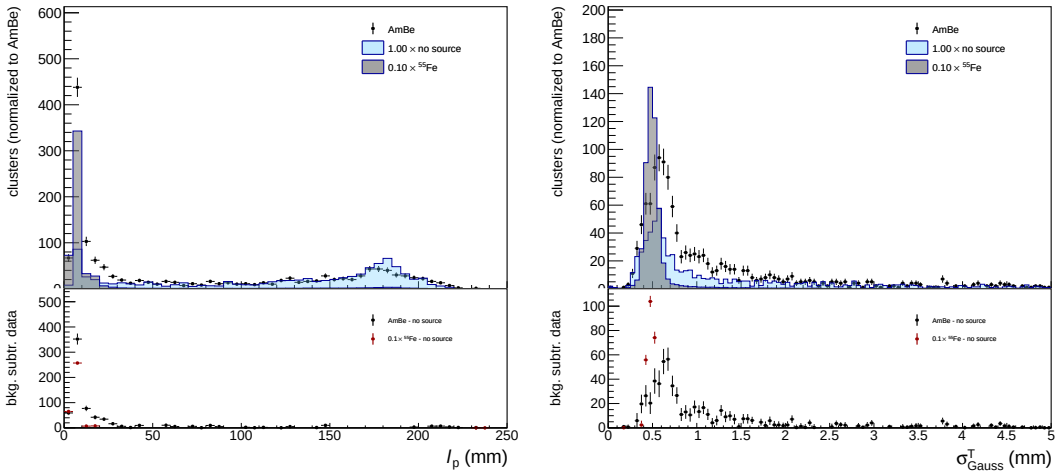

Figure 13. Supercluster sizes projected onto the x–y plane. Left: longitudinal path length, l_p. Right: transverse Gaussian spread, σ^T_{Gauss}. Filled points represent data with AmBe source, dark gray (light blue) distribution represents data with ^{55}Fe source (no source). The normalization of data without any artificial source is scaled to the same exposure time of the AmBe one. For the data with ^{55}Fe source , a scaling factor of one tenth is applied for clearness, given the larger activity of this source.

As described in Sec. 2, each pixel images an area of $125 \times 125\,\mu$m^2. Thus the cluster

sizes distributions show an average Gaussian width for the ^{55}Fe spots $\sigma^{T}_{Gauss} \approx 500\,\mu$m (dominated by the diffusion in the gas), while it is larger, approximately 625 μm, for data with AmBe source. The contribution of cosmic rays, present in all the data, is clearly visible in the data without any artificial radioactive source, corresponding to clusters with a length similar to the detector transverse size (22 cm).

Other observables are the slimness, ξ, and the light density, δ, shown in Fig. 14. The former is a useful handle to reject tracks from cosmic rays, which typically have a slim aspect ratio, i.e., low values of ξ, while the clusters from ^{55}Fe are almost round, with values of $\xi \approx 1$. Data with AmBe source, which contains a component of nuclear recoils, show a component of round spots, similar in size to the ones of ^{55}Fe, and a more elliptical component, with $0.4 < \xi < 0.8$ values. Finally, the light density, δ, is the variable expected to better discriminate among different candidates: cosmic rays induced background, electron recoils and nuclear recoil candidates. This is the variable used for the final particle identification.

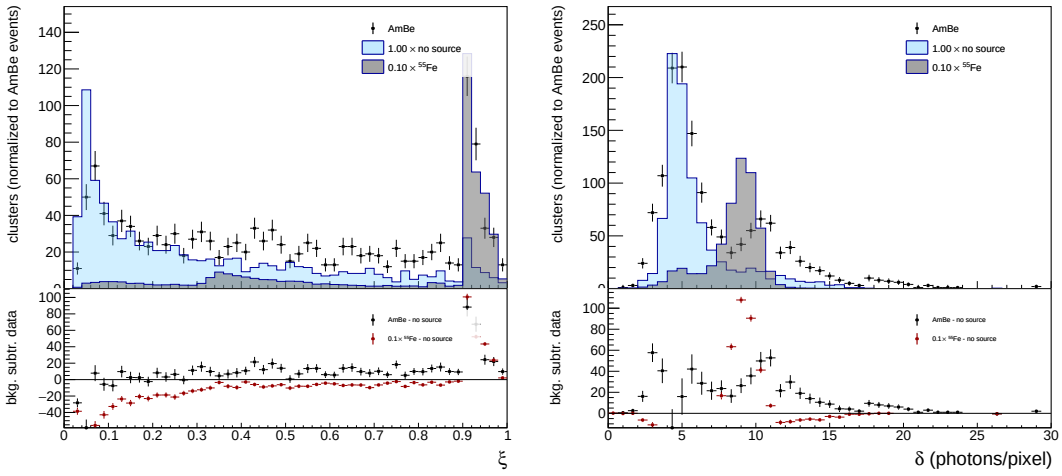

Figure 14. Supercluster variables. Left: slimness ξ; right: light density δ. Filled points represent data with AmBe source, dark gray (light blue) distribution represents data with ^{55}Fe source (no source). The normalization of data without source is to the same exposure time of the AmBe one. For the data with ^{55}Fe, a scaling factor of one tenth is applied for clearness, given the larger activity of this source.

Finally, Fig. 15 shows the calibrated energy (E) spectrum for the reconstructed superclusters and the average projected $\frac{dE}{dl_p}$. The energy spectrum shows the $E = 5.9$ keV for data with ^{55}Fe source, and a characteristic broad peak for cosmic rays tracks at around 60 keV. The distribution of the observed $\frac{dE}{dl_p}$ for the no-source sample and for the AmBe samples. The broadening of the distribution is mainly due to the specific energy loss fluctuation in the gas mixture of the cosmic ray particles. Its modal value, corrected for the effect of the angular distribution (an average inclination of 56° was measured from track reconstruction) is 2.5 keV/cm, in good agreement with the GARFIELD prediction of 2.3 keV/cm.

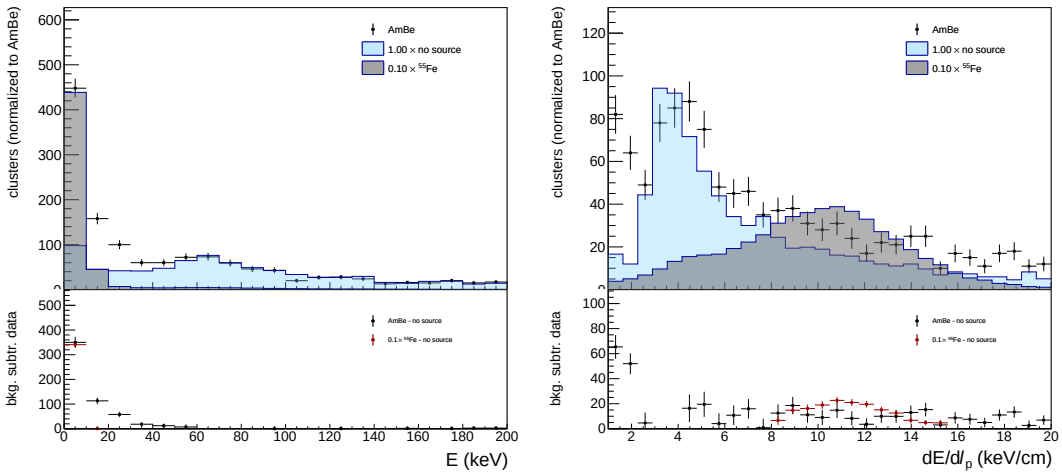

Figure 15. Supercluster calibrated energy spectrum (left) and their average $\frac{dE}{dl_p}$. Filled points represent data with AmBe source, dark gray (light blue) distribution represents data with ^{55}Fe source (no source). The normalization of data without source is to the same exposure time of the AmBe one. For the data with ^{55}Fe, a scaling factor of one tenth is applied for clearness, given the larger activity of this source.

5.1. Background normalization

The data with AmBe source, taken on the Earth surface, suffers from a large contribution of interactions of cosmic rays. The cluster shape observables provide a powerful handle to discriminate them from nuclear recoils candidates, but the small residual background needs to be statistically subtracted. The distributions shown earlier, where the different types of data are normalized to the same exposure time, demonstrate that the live-time normalization provides already a good estimate of the amount of cosmic rays in data with artificial radioactive sources. This approach does not account for a possible bias from the trigger, which is generated by the PMT signals, as described in Sec. 2. Indeed, in runs with the AmBe source, the PMT can trigger both on signals from neutron recoils or photons produced by the ^{241}Am, and on ubiquitous signals from cosmic rays, while in the sample without source only the latter is possible. This implies that, during the same exposure time, the probability to trigger on cosmic rays is lower in events with AmBe than in no-source events. The trigger efficiency scale factor, ε_{SF}, can be obtained as the ratio of the number of clusters selected in pure control samples of cosmic rays (CR) obtained on both types of runs:

$$\varepsilon_{SF} = \frac{N_{CR}^{AmBe}}{N_{CR}^{no-source}}. \tag{9}$$

The CR control region is defined by selecting clusters with $l > 13\,\text{cm}$, $\xi < 0.1$, $\sigma_{Gauss}^{T} < 6\,\text{mm}$, having an energy within a range dominated by the cosmic rays contribution, $50 < E < 80\,\text{keV}$. The selected clusters show small values of $\delta \approx 5$, well compatible with the small specific ionization of ultra-relativistic particles. This

sample is limited in statistics, but it is expected to be almost 100% pure. The scale factor obtained is $\varepsilon_{SF} = 0.75 \pm 0.02$.

In Fig. 16 the typical light density and polar angle (with respect the horizontal axis) distributions for long clusters of any energy, still dominated by cosmic rays, are shown for the AmBe and for the no-source sample, after having applied the ε_{SF} scale factor to the latter. Clusters with $\delta < 6$ are thus expected to be mostly coming from cosmic tracks, and they show indeed a polar angle which is shifted at values towards 90°.

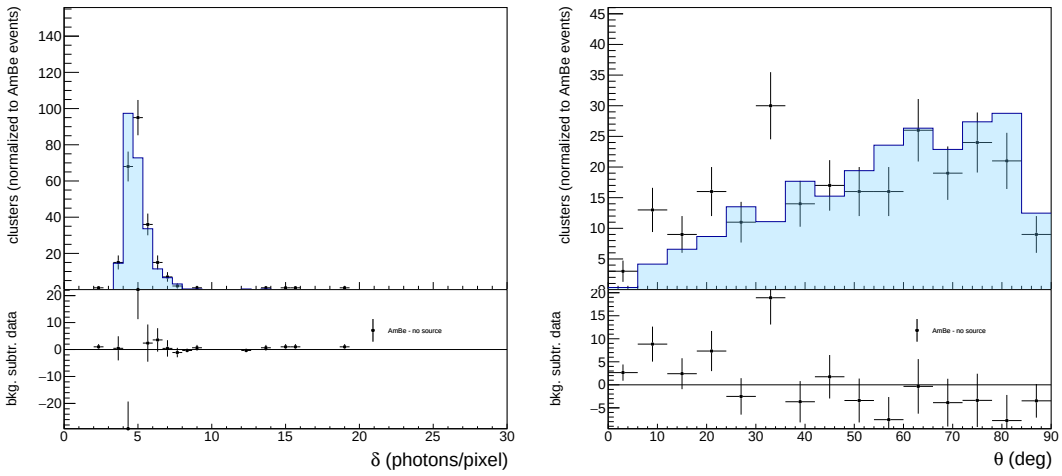

Figure 16. Supercluster light density δ (left) and polar angle (right) - with respect the horizontal axis - distributions for long clusters, dominated by cosmic rays tracks. Filled points represent data with AmBe source, light blue distribution represents data without any artificial source. The normalization of data without source is to the same exposure time of the AmBe one, accounting for the trigger scale factor ε_{SF}, as defined in the text.

6. Nuclear recoil identification results

The main observable to distinguish the signal of nuclear recoils from the various types of background, is the energy density δ of the cluster.

6.1. Signal Preselection

To enhance the purity of the signal sample, a pre-selection was applied, prior to a tighter selection on δ: clusters with $l_p > 6.3\,\text{cm}$ or $\xi < 0.3$ were rejected to primarily suppress the contribution from cosmic rays. A further loose requirement $\delta > 5$ photons/pixel was also applied to remove the residual cosmic rays background based on their low specific ionization. Considering that the applied thresholds are very loose for nuclear recoils with $E < 1\,\text{MeV}$ energies, as shown in Fig. 2, the preselection efficiency is assumed to

be 100%. For electron recoils it is estimated on the ^{55}Fe data sample, and is measured to be $\varepsilon_B^{presel} = 70\%$.

With this preselection, the distribution in the 2D plane δ–l_p is shown in Fig. 17 for AmBe source and no-source data and for the resulting background-subtracted AmBe data. The latter distribution shows a clear component of clusters with short length ($l_p \lesssim 1\,$cm) and high density ($\delta \gtrsim 10$), expected from nuclear recoils deposits.

In addition, it shows a smaller component, also present only in the data with AmBe source, of clusters with a moderate track length, $1.5 \lesssim l_p \lesssim 3.0\,$cm, and a lower energy density than the one characteristic of the nuclear recoils ($9 \lesssim \delta \lesssim 12$). Since the density is inversely proportional to the number of active pixels n_p, which is correlated to the track length, the almost linear decrease of δ as a function of l_p points to a component with fixed energy. The ^{241}Am is expected to produce photons with $E = 59\,$keV. This hypothesis is verified by introducing an oblique selection in the $\delta - l_p$ plane: $|\delta - y| < 2$, where $y = 14 - p_l/50$, for the clusters with $120 < l_p < 250\,$pixels, defining the control region PR. The obtained energy spectrum for these clusters is shown in Fig. 18, which indeed shows a maximum at $E = 60.9 \pm 3.6\,$keV, within the expected resolution. These events are thus rejected from the nuclear recoils candidates by vetoing the PR phase space.

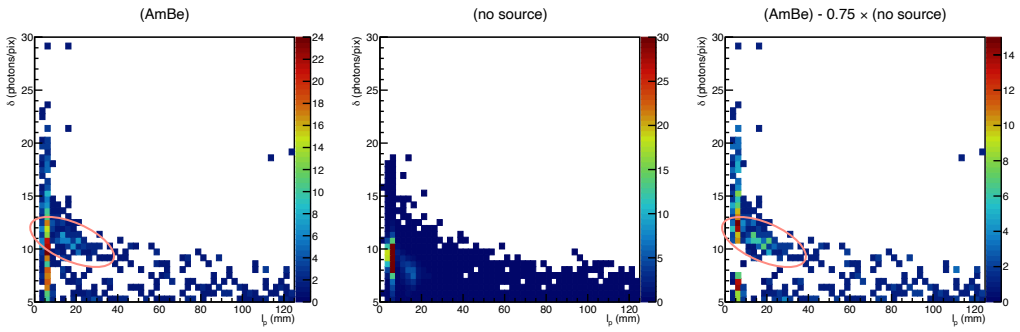

Figure 17. Supercluster light density δ versus length l_p, for data with AmBe source (left), data without any artificial source (middle), and the resulting background-subtracted AmBe data. The normalization of data without source is to the same exposure time of the AmBe one, accounting for the trigger scale factor ε_{SF}, as defined in the text.

6.2. PMT-based cosmic ray suppression

An independent information to the light detected by the sCMOS sensor of the camera is obtained from the PMT pulse, used to trigger the image shooting. For each image acquired, the corresponding PMT pulse waveform is recorded. Tracks from cosmic rays, which typically have a large angle with respect the cathode plane, as shown in Fig. 16 (right), show a broad waveform, characterized by the different arrival times of the several ionization clusters produced along the track at different z. Conversely, spot-like signals

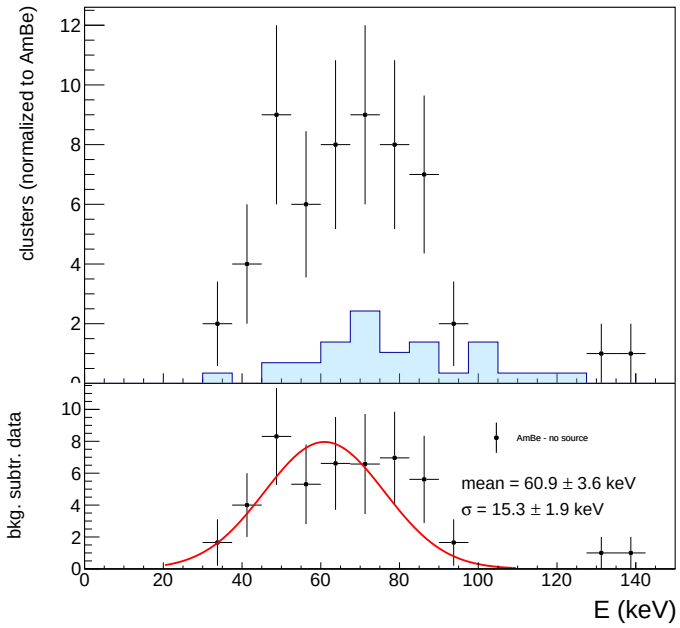

Figure 18. Calibrated energy spectrum for candidates in the control region PR, defined in the text. The background-subtracted distribution is fitted with a Gaussian PDF, which shows a mean value compatible with $E = 59\,\text{keV}$ originated from the ^{241}Am γs interaction within the gas.

like ^{55}Fe deposits or nuclear recoils are characterized by a short pulse, as shown in Fig. 19.

The Time Over Threshold (TOT) of the PMT pulse was measured, and is shown in Fig. 20. It can be seen from the region around 270 ns, dominated by the cosmic rays also in the data with the AmBe source, that the trigger scale factor ε_{SF} also holds for the PMT event rate. As expected, spot-like clusters (in 3D) correspond to a short pulse in the PMT, while cosmic ray tracks have a much larger pulse. The contribution of cosmic ray tracks is clearly visible in the data with radioactive sources. A selection on this variable is helpful to further reject residual cosmic rays background present in the AmBe or ^{55}Fe data, in particular tracks which may have been split in multiple superclusters, like the case shown in Fig. 8 (bottom), and thus passing the above preselection on the cluster shapes. A selection $TOT < 250\,\text{ns}$ is then imposed. It has an efficiency of 98% on cluster candidates in AmBe data (after background subtraction), while it is only 80% efficient on data with ^{55}Fe source. The light density, and the energy spectrum, after the full preselection, is shown in Fig. 21.

6.3. Light density and ^{55}Fe events rejection

The light density distribution, after the above preselection and cosmic ray suppression, appear to be different among the data with AmBe source, data with ^{55}Fe source, and

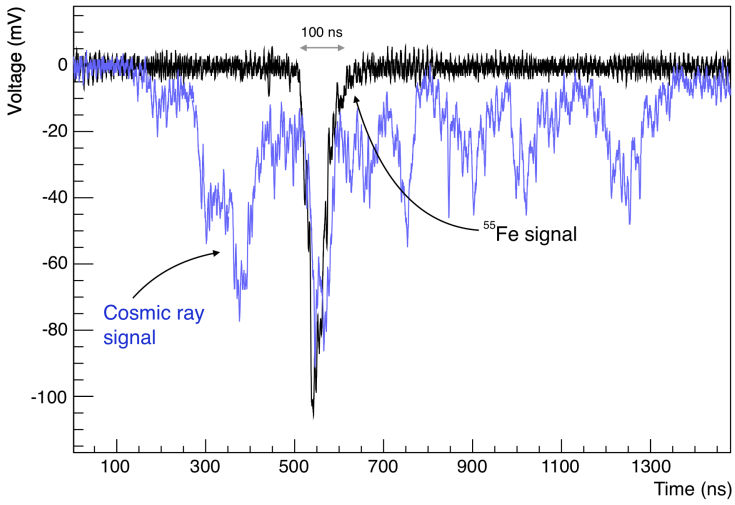

Figure 19. Example of two acquired waveforms: one short pulse recorded in presence of ^{55}Fe radioactive source, together with a long signal very likely due to a cosmic ray track.

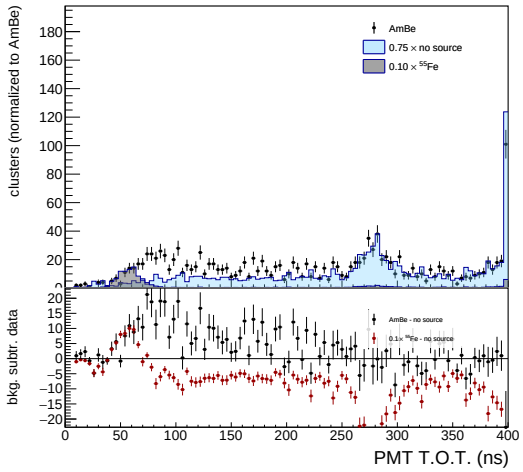

Figure 20. PMT waveform time over threshold (TOT). The last bin integrates all the events with $TOT > 400$ ns. Filled points represent data with AmBe source, dark gray (light blue) distribution represents data with ^{55}Fe source (no source). The normalization of data without source is to the same exposure time of the AmBe one, with trigger scale factor ε_{SF} applied. For the data with ^{55}Fe, a scaling factor of one tenth is applied for clearness, given the larger activity of this source.

data without any artificial source. The cosmic-background-subtracted distributions of δ in AmBe data and ^{55}Fe data, shown in the bottom panel of Fig. 21 (left), are used to evaluate a curve of electron recoils rejection ($1-\varepsilon_B^\delta$) as a function of signal efficiency (ε_S^δ), obtained varying the selection on δ. This is shown in Fig. 22. The same procedure could

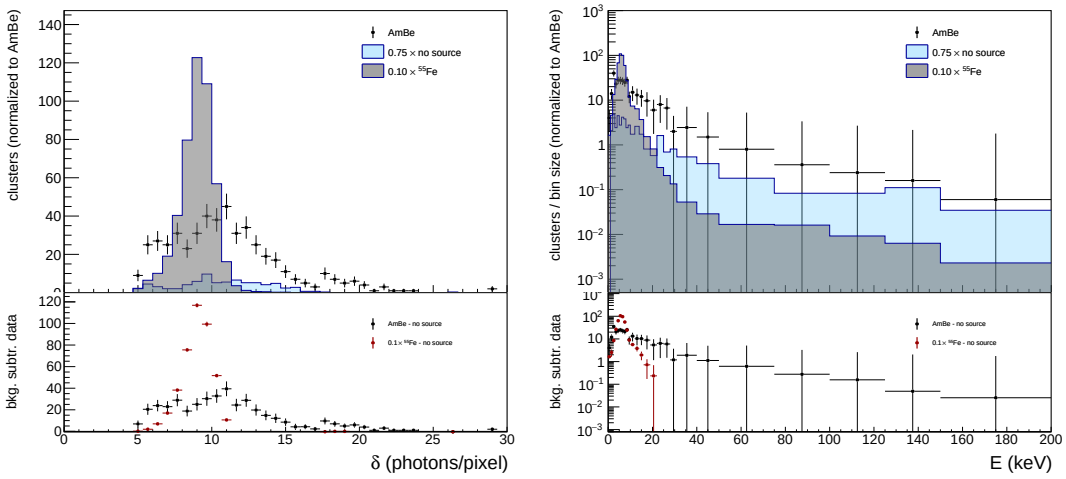

Figure 21. Supercluster light density δ (left) and calibrated energy E (right), after the preselection and cosmic ray suppression described in the text to select nuclear recoil candidates. Filled points represent data with AmBe source, dark gray (light blue) distribution represents data with ^{55}Fe source (no-source). The normalization of no-source data is to the same exposure time of the AmBe data, with the trigger scale factor ε_{SF} applied. For the data with ^{55}Fe, a scaling factor of one tenth is applied for clearness, given the larger activity of this source.

Table 1. Signal (nuclear recoils) and background (electron recoils) efficiency for two different selections on δ.

working point	Signal efficiency			Background efficiency		
	ε_S^{presel}	ε_S^{δ}	ε_S^{total}	ε_B^{presel}	ε_B^{δ}	ε_B^{total}
WP_{50}	0.98	0.51	0.50	0.70	0.050	0.035
WP_{40}	0.98	0.41	0.40	0.70	0.012	0.008

be applied to estimate the rejection factor against the cosmic ray induced background, but this is not shown because of the limited size of the no-source data. This kind of background will however be negligible when operating the detector underground. This is shown in Fig. 22. While this cut-based approach is minimalist, and could be improved by profiting of the correlations among δ and the observables used in the preselection in a more sophisticated multivariate analysis, it shows that a good rejection factor of electron recoils at $E = 5.9\,\mathrm{keV}$ can be obtained.

Table 1 shows then the full signal efficiency and electrons rejection factor for two example working points, WP_{40} and WP_{50}, having 40% and 50% signal efficiency for the selection on δ. They correspond to a selection $\delta > 11$ and $\delta > 10$, respectively.

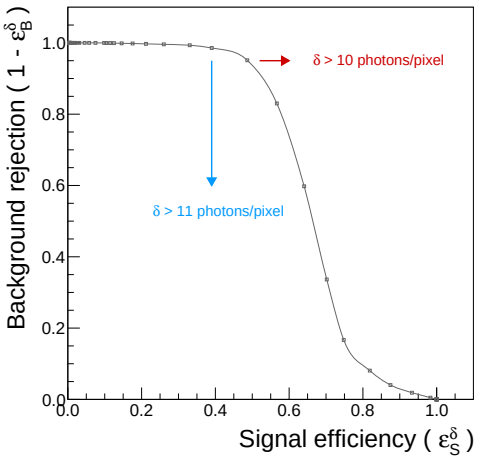

Figure 22. Background rejection as a function of the signal efficiency, varying the selection on the δ variable in data with either ^{55}Fe (background sample) or AmBe (signal sample) sources.

6.4. Signal Energy Spectrum and efficiency

The energy spectrum for the candidates with $\varepsilon_S^{total} = 50\%$ in the AmBe sample is shown in Fig. 23 (left). With the full WP_{50} selection, the signal efficiency is computed for both the example working points in bins of energy. The electron recoil efficiency, ε_B^{total}, represents a γ background efficiency at a fixed energy $E = 6\,\text{keV}$, that is close to the ^{55}Fe emitted photon energy. For the WP_{50}, the efficiency for very low-energy recoils, $E = 6\,\text{keV}$, is still 18%, dropping to almost zero at $E \lesssim 4\,\text{keV}$.

Two candidate nuclear recoils images, fulfilling the WP_{50} selection (with a light density $\delta \gtrsim 10$ photons/pixels and with energies of 5.2 and 6.0 keV) are shown in Fig. 24. The displayed images are a portion of the full-resolution frame, after the pedestal subtraction.

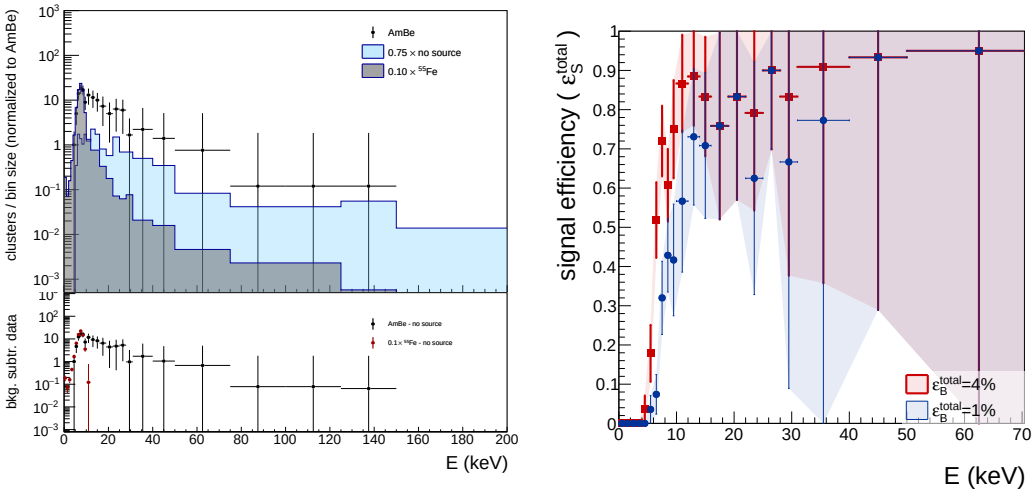

Figure 23. Left: supercluster calibrated energy E (left), after the full selection, which includes $\delta > 10$, 50% efficient on signal, to select nuclear recoil candidates. Filled points represent data with AmBe source, dark gray (light blue) distribution represents ^{55}Fe source (no-source) data. The normalization of no-source data is to the same exposure time of the AmBe data, with the trigger scale factor ε_{SF} applied. For the ^{55}Fe data, a scaling factor of one tenth is applied for clearness, given the larger activity of this source. Right: efficiency for nuclear recoil candidates as a function of energy, estimated on AmBe data, for two example selections, described in the text, having either 4% or 1% efficiency on electron recoils at $E = 6\,\text{keV}$.

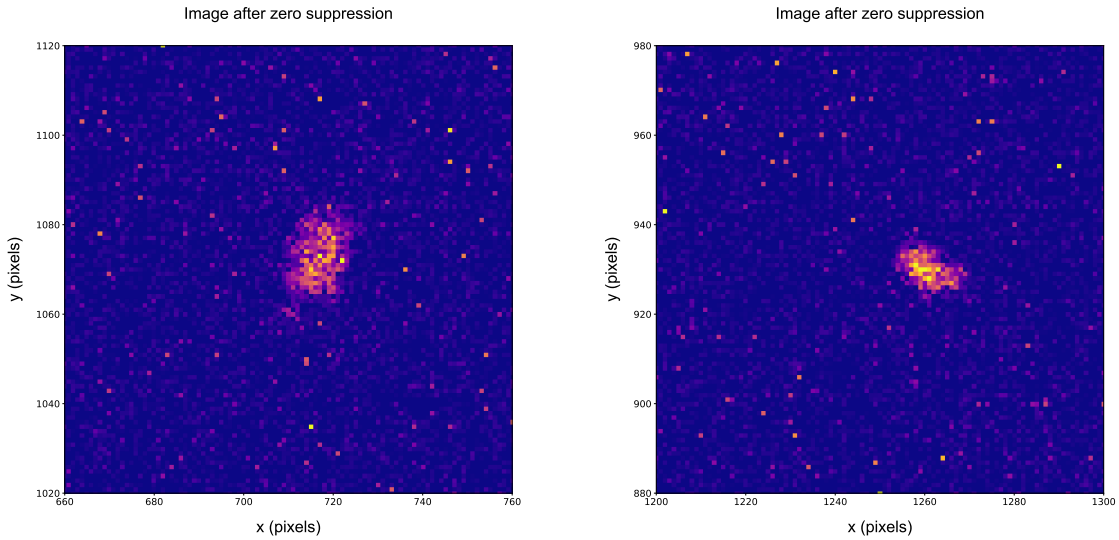

Figure 24. Examples of two nuclear recoil candidates, selected with the full selection, shown in a portion of 100 × 100 pixel matrix, after the zero suppression of the image. Left: a candidate with $E = 5.2\,\text{keV}$ and $\delta = 10.5$, right: a candidate with $E = 6.0\,\text{keV}$ and $\delta = 10$.

7. Conclusion and outlook

A method to efficiently identify recoiling nuclei after an elastic scattering with fast neutrons with an optically readout TPC was presented in this paper. A 7 liter prototype was employed by exposing its sensitive volume to two kinds of neutral particles in an overground location:

- photons with energy of 5.9 keV and 59 keV respectively provided by a radioactive source of ^{55}Fe and by one of ^{241}Am able to produce electron recoils with equal energy by means of photoelectric effect;
- neutrons with kinetic energy of few MeV produced by an AmBe source that can create nuclear recoils with kinetic energy lower than the neutron ones.

The high sensitivity of the adopted sCMOS optical sensor allowed a very good efficiency in detecting events with an energy released in gas even below 10 keV.

Moreover, the possibility of exploiting the topological information (shape, size and more) of clusters of emitted light permitted to develop algorithms able to reconstruct not only the total released energy, but also to identify the kind of the recoiling ionizing particles in the gas (either an electron or a nucleus). Cosmic ray long tracks are also clearly separated.

Because of their larger mass and electric charge, nuclear recoils are expected to release their energy by ionizing the gas molecules in few hundreds μm while the electrons are able to travel longer paths. For this reason, by exploiting the spatial distribution of the collected light, it was possible to identify 5.9 keV electron recoils with an efficiency of 96.5% (99.2%) against nuclear recoils by retaining a capability of detecting them with an efficiency of 50% (40%), averaged across the measured AmBe spectrum.

In average, nuclear recoils with a kinetic energy lower than 10 keV can be detected with an efficiency of about 14%.

The results obtained in the studies presented in this paper can be improved by means of more sophisticated analyses exploiting a multivariate approach, which combines a more complete topological information about the light distribution along the tracks. Additional enhancement of sensitivity can be achieved with a DAQ system collecting single PMT waveforms to be correlated with the track reconstructed in the sCMOS images.

8. Acknoledgements

We are grateful to Servizio Sorgente LNF... This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 818744)".

[1] B. W. Lee and S. Weinberg, "Cosmological lower bound on heavy-neutrino masses," *Phys. Rev. Lett.*, vol. 39, pp. 165–168, Jul 1977.

[2] T. M. Undagoitia and L. Rauch, "Dark matter direct-detection experiments," *Journal of Physics G: Nuclear and Particle Physics*, vol. 43, p. 013001, dec 2015.

[3] E. Baracchini, G. Cavoto, G. Mazzitelli, F. Murtas, F. Renga, and S. Tomassini, "Negative ion time projection chamber operation with SF_6 at nearly atmospheric pressure," *Journal of Instrumentation*, vol. 13, pp. P04022–P04022, Apr. 2018.

[4] M. Marafini, V. Patera, D. Pinci, A. Sarti, A. Sciubba, and E. Spiriti, "ORANGE: A high sensitivity particle tracker based on optically read out GEM," *Nucl. Instrum. Meth.*, vol. A845, pp. 285–288, 2017.

[5] V. C. Antochi, E. Baracchini, G. Cavoto, E. D. Marco, M. Marafini, G. Mazzitelli, D. Pinci, F. Renga, S. Tomassini, and C. Voena, "Combined readout of a triple-GEM detector," *JINST*, vol. 13, no. 05, p. P05001, 2018.

[6] D. Pinci, E. Di Marco, F. Renga, C. Voena, E. Baracchini, G. Mazzitelli, A. Tomassini, G. Cavoto, V. C. Antochi, and M. Marafini, "Cygnus: development of a high resolution TPC for rare events," *PoS*, vol. EPS-HEP2017, p. 077, 2017.

[7] G. Mazzitelli, V. A. Antochi, E. Baracchini, G. Cavoto, A. De Stena, E. Di Marco, M. Marafini, D. Pinci, F. Renga, S. Tomassini, and C. Voena, "A high resolution TPC based on GEM optical readout," in *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–4, Oct 2017.

[8] D. Pinci, E. Baracchini, G. Cavoto, E. Di Marco, M. Marafini, G. Mazzitelli, F. Renga, S. Tomassini, and C. Voena, "High resolution TPC based on optically readout GEM," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018.

[9] D. Pinci, *A triple-GEM detector for the muon system of the LHCb experiment.* PhD thesis, Cagliari University, CERN-THESIS-2006-070, 2006.

[10] Hamamatsu, *ORCA-Flash4.0 V3 Digital CMOS camera*, 2018.

[11] M. Marafini, V. Patera, D. Pinci, A. Sarti, A. Sciubba, and E. Spiriti, "High granularity tracker based on a Triple-GEM optically read by a CMOS-based camera," *JINST*, vol. 10, no. 12, p. P12010, 2015.

[12] HZC Photonics, *XP3392 Photomultiplier.*

[13] V. C. Antochi, G. Cavoto, I. Abritta Corsta, E. Di Marco, G. D'Imperio, F. Iacoangeli, M. Marafini, A. Messina, D. Pinci, F. Renga, C. Voena, E. Baracchini, A. Cortez, G. Dho, L. Benussi, S. Bianco, C. Capoccia, M. Caponero, G. Maccarrone, G. Mazzitelli, A. Orlandi, E. Paoletti, L. Passamonti, D. Piccolo, D. Pierluigi, F. Rosatelli, A. Russo, G. Saviano, S. Tomassini, R. A. Nobrega, and F. Petrucci, "A GEM-based optically readout time projection chamber for charged particle tracking," 2020.

[14] R. Campagnola, "Study and optimization of the light-yield of a triple-GEM detector ," Master's thesis, Sapienza University of Rome, 2018.

[15] N. Torchia, "Development of a tracker based on GEM optically readout ," Master's thesis, Sapienza University of Rome, 2016.

[16] CAEN, *2 Channel 15 kV/1 mA (10 W) NIM HV Power Supply Module*, 2017.

[17] R. Veenhof, "GARFIELD, recent developments," *Nucl. Instrum. Meth. A*, vol. 419, pp. 726–730, 1998.

[18] R. Veenhof, "GARFIELD, a drift chamber simulation program," *Conf. Proc. C*, vol. 9306149, pp. 66–71, 1993.

[19] G. Mazzitelli, A. Ghigo, F. Sannibale, P. Valente, and G. Vignola, "Commissioning of the DAϕNE beam test facility," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 515, no. 3, pp. 524 – 542, 2003.

[20] I. Abritta Costa, E. Baracchini, F. Bellini, L. Benussi, S. Bianco, M. A. Caponero, G. Cavoto, G. D'Imperio, E. Di Marco, G. Maccarrone, M. Marafini, G. Mazzitelli, A. Messina, F. Petrucci, D. Piccolo, D. Pinci, F. Renga, F. Rosatelli, G. Saviano, and S. Tomassini, "Stability and detection performance of a GEM-based optical readout TPC with He/CF_4 gas mixtures," *Journal of Instrumentation*, vol. xx, p. xxx, jul 2020.

[21] S. Agostinelli *et al.*, "GEANT4—a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.

[22] J. Ziegler, "Srim – 2003," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 219-220, no. 3, pp. 1027–1036, 2004.

[23] I. Abritta Costa, E. Baracchini, F. Bellini, L. Benussi, S. Bianco, M. A. Caponero, G. Cavoto, G. D'Imperio, E. Di Marco, G. Maccarrone, M. Marafini, G. Mazzitelli, A. Messina, F. Petrucci, D. Piccolo, D. Pinci, F. Renga, F. Rosatelli, G. Saviano, and S. Tomassini, "Performance of optically readout GEM-based TPC with a ^{55}Fe source," *Journal of Instrumentation*, vol. 14, pp. P07011–P07011, jul 2019.

[24] Y. Dong and S. Xu, "A new directional weighted median filter for removal of random-valued impulse noise," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 193–196, 2007.

[25] G. S. P. Lopes, E. Baracchini, F. Bellini, L. Benussi, S. Bianco, G. Cavoto, I. A. Costa, E. Di Marco, G. Maccarrone, M. Marafini, G. Mazzitelli, A. Messina, R. A. Nobrega, D. Piccolo, D. Pinci, F. Renga, F. Rosatelli, D. M. Souza, and S. Tomassini, "Study of the impact of pre-processing applied to images acquired by the cygno experiment," in *Pattern Recognition and Image Analysis* (A. Morales, J. Fierrez, J. S. Sánchez, and B. Ribeiro, eds.), (Cham), pp. 520–530, Springer International Publishing, 2019.

[26] I. Abritta *et al.*, "A density-based clustering algorithm for the cygno data analysis," *In preparation*, vol. 00, no. 0, pp. 00–00, 2020.

[27] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," pp. 226–231, AAAI Press, 1996.

[28] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[29] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth. A*, vol. 389, pp. 81–86, 1997.

[30] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic Active Contours," *International Journal of Computer Vision*, vol. 22, pp. 61–79, 1997.

[31] P. Márquez-Neila, L. Baumela, and L. Alvarez, "A morphological approach to curvature-based evolution of curves and surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 2–17, 2014.

[32] M. Oreglia, *A Study of the Reactions $\psi' \to \gamma\gamma\psi$*. PhD thesis, SLAC, 1980.

[33] J. E. Gaiser, *Charmonium spectroscopy from radiative decays of the J/ψ and ψ'*. PhD thesis, SLAC, 1982.

[34] Z. Guo and R. W. Hall, "Parallel thinning with two-subiteration algorithms," *Commun. ACM*, vol. 32, p. 359–373, Mar 1989.

[35] L. Lam, S. Lee, and C. Y. Suen, "Thinning methodologies - a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869–885, 1992.