PAPER

Topological background discrimination in the PandaX-III neutrinoless double beta decay experiment

To cite this article: J Galan *et al* 2020 *J. Phys. G: Nucl. Part. Phys.* **47** 045108

View the article online for updates and enhancements.

Topological background discrimination in the PandaX-III neutrinoless double beta decay experiment

J Galan[1,2,14] ⓘ, X Chen[1], H Du[1], C Fu[1] ⓘ, K Giboni[1],
F Giuliani[1], K Han[1], B Jiang[1], X Ji[1,3], H Lin[1], Y Lin[1], J Liu[1],
K Ni[1], X Ren[1], S Wang[1], S Wu[1], C Xie[1], Y Yang[1], T Zhang[1],
L Zhao[1], S Aune[2], Y Bedfer[2], E Berthoumieux[2], D Calvet[2],
N d'Hose[2], E Ferrer-Ribas[2], F Kunne[2], B Manier[2], D Neyret[2],
T Papaevangelou[2], L Chen[4], S Hu[4], P Li[4], X Li[4], H Zhang[4],
M Zhao[4], J Zhou[4], Y Mao[5], H Qiao[5], S Wang[5], Y Yuan[5],
M Wang[6], Y Chen[7], A N Khan[7], J Tang[7], W Wang[7],
H Chen[8,9], C Feng[8,9], J Liu[8,9], S Liu[8,9], X Wang[8,9], D Zhu[8,9],
J F Castel[10], S Cebrián[10], T Dafni[10], I G Irastorza[10],
G Luzón[10], H Mirallas[1,10], X Sun[11], A Tan[3], W Haxton[12],
Y Mei[12], C Kobdaj[13] and Y Yan[13]

[1] INPAC and School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai Laboratory for Particle Physics and Cosmology, Shanghai 200240, People's Republic of China
[2] IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
[3] Department of Physics, University of Maryland, College Park, MD 20742, United States of America
[4] Science and Technology on Nuclear Data Laboratory, China Institute of Atomic Energy, Beijing 102413, People's Republic of China
[5] School of Physics, Peking University, Beijing 100871, People's Republic of China
[6] School of Physics and Key Laboratory of Particle Physics and Particle Irradiation (MOE), Shandong University, Jinan 250100, People's Republic of China
[7] School of Physics and Engineering, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China
[8] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, People's Republic of China
[9] State Key Laboratory of Particle Detection and Electronics, University of Science and Technology of China, Hefei 230026, People's Republic of China
[10] University of Zaragoza, C/P Cerbuna 12, E-50009, Zaragoza, Spain
[11] College of Physical Science and Technology, Central China Normal University, Wuhan 430079, People's Republic of China
[12] Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America
[13] Center for Excellence in High Energy Physics and Astrophysics, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

E-mail: javier.galan.lacarra@cern.ch

CrossMark

Abstract

The PandaX-III experiment plans to search for neutrinoless double beta decay
($0\nu\beta\beta$) of ^{136}Xe in the China JinPing underground Laboratory (CJPL). The
experiment will use a high pressure gaseous time projection chamber (TPC) to
register both the energy and the electron track topology of an event. This article is
devoted to demonstrate our particular detector setup capabilities for the identi-
fication of $0\nu\beta\beta$ and the consequent background reduction. As software tool we
use REST, a framework developed for the reconstruction and simulation of TPC-
based detector systems. We study the potential for background reduction by
introducing appropriate parameters based on the properties of $0\nu\beta\beta$ events. We
exploit for the first time not only the energy density of the electron track-ends,
but also the electron scattering angles produced by an electron near the end of its
trajectory. To implement this, we have added new algorithms for detector signal
and track processing inside REST. Their assessment shows that background can
be reduced by about 7 orders of magnitude while keeping $0\nu\beta\beta$ efficiency above
20% for the PandaX-III baseline readout scheme, a two-dimensional 3 mm pitch
stripped readout. More generally, we use the potential of REST to handle 2D/3D
data to assess the impact on signal-to-background significance at different
detector granularities, and to validate the PandaX-III baseline choice. Finally, we
demonstrate the additional potential to discriminate surface background events
generated at the readout plane in the absence of t_o, by making use of event
parameters related with the diffusion of electrons.

Keywords: neutrino, PandaX-III, topology, pattern recognition, Majorana,
mass, double beta decay

(Some figures may appear in colour only in the online journal)

1. Introduction

The neutrino sector has been one of the most promising research areas for searching for new
physics beyond the standard model (SM). Neutrino oscillations and the inferred non-zero neutrino
mass offered a concrete evidence of new physics, as demonstrated by experiments including
Super-Kamiokande [1], SNO [2], and etc. Early in 1937, the possibility for an electrically neutral
fermion to be the anti-particle of itself was raised by Majorana [3], and neutrinos are the most
promising candidates for the so-called Majorana fermions. Soon after that, physicists started to use
the neutrinoless double beta decay ($0\nu\beta\beta$) to search for the existence of Majorana neutrinos [4]. In
a SM-allowed two-neutrino double beta decay process, two neutrons in a candidate nucleus decay
into two protons simultaneously and emit two electrons together with two electron antineutrinos.
Such a process has been observed in a dozen or so isotopes, including ^{76}Ge, ^{130}Te, ^{136}Xe, and etc

---

[14] Author to whom any correspondence should be addressed.

(see, for example [5]). If neutrinos are Majorana fermions, no electron antineutrinos are released in the double beta decay process. It has also been argued with the Black Box theorem that an observation of $0\nu\beta\beta$ directly indicates the Majorana nature of neutrinos [6]. The experimental observation of $0\nu\beta\beta$ will also distinctly violate the conservation of the lepton number, and have far-reaching impact beyond the neutrino sector [7]. After about 80 years of experimental effort, no firm evidence for $0\nu\beta\beta$ has been obtained. Several major experiments are taking data or are under construction and many more smaller-scale efforts are in R&D phase. Current status and expected sensitivity can be seen in [8] and references within.

The PandaX-III experiment aims to search for $0\nu\beta\beta$ of ^{136}Xe using a high pressure gaseous time projection chamber (TPC) [9]. The experiment will be located in the newly excavated China JinPing underground Laboratory Phase II (CJPL-II) in Sichuan Province, China [10]. As a calorimeter, PandaX-III measures the energy of two electrons from $0\nu\beta\beta$ and constructs an energy spectrum at the region of interest (ROI) around the $Q_{\beta\beta}$ value of ^{136}Xe, 2,457.83 keV [11]. The expected energy resolution is better than 3% full-width-half-maximum (FWHM) at the $Q_{\beta\beta}$ value [12, 13]. PandaX-III follows the standard recipe of a low background experiment, including deep underground lab environment, passive shielding, and careful screening of detector as well as electronics material to minimize the number of dubious background events in the ROI. The descendants of ^{238}U and ^{232}Th, mainly ^{214}Bi and ^{208}Tl, contribute most significantly to the background in the ROI. One of the gamma lines from ^{214}Bi at 2448 keV, only 10 keV below the $0\nu\beta\beta$ Q-value, poses a major challenge to event identification with energy alone. The expected background level at the ROI including a realistic detector response is on the order of 10^{-3} keV^{-1} kg^{-1} yr^{-1} when considering only the event energy.

Additional background suppression can be realized with event tracking information, which is the main topic of this paper. Different from other detector technologies, a gaseous TPC will record detailed trajectories of the two electrons from a potential $0\nu\beta\beta$ event. PandaX-III will exploit the electron tracking potential using fine-pitched Microbulk Micromegas [14] readout planes to reach millimeter level spatial resolution, in comparison with the typical $0\nu\beta\beta$ track length of 10cm scale at 10 bar. The signal identification will exploit the features of the electron tracks topology produced by $0\nu\beta\beta$ decays. The two electrons generated by a $0\nu\beta\beta$ decay will produce two bright Bragg peaks, occurring on at least 95% of the decays [15]. On the other hand, gamma background events in the ROI normally have only one Bragg peak. We will review the details on the main characteristics of signal and background events, previously investigated in [15], and we will exploit them to perform a discrimination analysis to demonstrate the ultimate background level achievable by the PandaX-III setup.

The work presented in this paper includes an accurate Monte Carlo simulation of the operation of the PandaX-III detector setup from the point of view of $0\nu\beta\beta$ event identification. Previous work, published in [9], provided a rough estimation of the final background achievable. This paper provides a complete and detailed analysis for PandaX-III including a realistic two-dimensional detector readout.

PandaX-III adopts the event-oriented software framework REST[15] for simulation and data analysis, especially to facilitate the topological analysis. With REST and the underlying Geant4 framework, we simulate $0\nu\beta\beta$ events and background events with energy deposition and tracking information in the gas medium. Realistic detector response, including the strip readout scheme, is then added to generate mock data, which are then fed into our data analysis chain for energy and track reconstruction. We then optimize the efficiency of the

---

[15] We include in appendix A necessary technical details on the software used, and the description of the *event data* and *event process* routines used in our analysis. Any keyword starting by *TRest* will do reference to REST and it will be detailed in one of the appendices of this document.

discrimination between signal and background by analyzing the topological information of the events.

This paper is organized as follows. In section 2 we present the results obtained in terms of signal-to-background enhancement for the PandaX-III baseline readout scheme. In section 3, we explore the impact of different readout granularities and assess the goodness of choice for the PandaX-III baseline readout. Finally, in section 4, we go a step forward in our analysis to demonstrate the capability of our detector to fiducialize in the drift direction of the TPC and reject background from our charge readout plane effectively.

2. Background rejection on the PandaX-III baseline readout scheme

We present here the results obtained using a realistic detector readout description based on the existing PandaX-III baseline Microbulk Micromegas design [16]. For the sake of completeness, we provide in appendix B details on how we generate our Monte Carlo *event data* including any necessary details such as physics considerations and detailed detector readout description for PandaX-III, based on a two-dimensional charge readout, which represents a major milestone on the event reconstruction and data analysis of the experiment. In section 2.1, we define the parameters, extracted during the *event data* processing, that will be used for event pattern recognition, leading to our final background discrimination results reported in section 2.2.

2.1. Topological parameters used for pattern recognition

This section does not focus on the nature, or the source, of background events, but mostly on the type of background events that might mimic a $0\nu\beta\beta$ in the ROI at the $Q_{\beta\beta}$ energy of ^{136}Xe. Such background events, producing a topological structure similar to a $0\nu\beta\beta$ decay, can only be produced in the active detector volume by gammas with an energy equal to or above the $Q_{\beta\beta}$ value. Some of those gammas end up producing a long electron track of characteristics similar to the $0\nu\beta\beta$ signal. Still, a good choice of parameters extracted from the *physical track* provides an excellent instrument to differentiate between this kind of events and electron tracks produced by $0\nu\beta\beta$ decays.

Previous studies on $0\nu\beta\beta$ pattern recognition [17–19] have already shown the potential of certain parameters, such as the *number of secondary tracks* or *end-track energies* to differentiate background and signal events. We reach the same conclusions, and we obtain comparable results on their discrimination power. Furthermore, we will investigate the potential of novel parameters never used in previous studies, *viz.* as the end-track *twist parameter* and the *track length*. Below we discuss the respective merits of both, previously mentioned and novel parameters.

- *Number of secondary tracks, or track energy ratio.* One of the most powerful discriminators is related to the number of independent tracks identified inside an event taking place in the active detector volume[16]. Background events are originated by high energy gammas, and many of those gammas interacting in the detector will produce multi-track events due to Compton scattering combined with a photoelectric interaction absorbing the remaining gamma energy, therefore producing a main long energetic track. Additionally, the higher initial energy of the single electron background compared to the

---

[16] The number of tracks is found after using *TRestHitsToTrackProcess*, described in appendix A.2, while the conditions in which those are obtained is described in appendix B.1.

initial energy of each of the two electrons generated in the $0\nu\beta\beta$ decay makes so that the probability of producing secondary gamma radiation is higher for single electron background events.

In practice, it is convenient to define the *track energy ratio*, θ, as the ratio between the total energy of secondary tracks and the total energy of the event, E_{tot}, expressed as

$$\theta = \frac{1}{E_{\text{tot}}} \sum_{n_i=1}^{N_i} E_{i,n}, \tag{1}$$

where $E_{i,n}$ is the energy of track number n from projection i, and $n_i = 0$ corresponds to the most energetic track of each projection. If the event does not contain secondary tracks, as it is frequently the case in $0\nu\beta\beta$ events, the value of θ will be zero. Such a definition accommodates the 2D and 3D detector readouts in a single observable, and provides a probability density function of the θ parameter that can be exploited for signal and background separation. It is important to remark that the definition of θ is done for convenience on the flexibility to tune the signal-to-background ratio by selecting an appropriate θ-value. We have compared ourselves this observable with other approaches used in other studies [17], obtaining very similar results. It must be remarked, that using the θ parameter does not constrain the number of tracks. However, typically low values of θ will accept mostly events containing only one or two tracks, and rarely three.

- *End-track energies, or blobs charge.* An electron traveling in a gas medium with energies above $\sim$MeV experiences a constant energy loss, dE/dx, along its trajectory in the medium. Once the electron loses most of its energy, at levels well below 1 MeV, its dE/dx increases suddenly until it loses all of its kinematic energy. This phenomenon produces a high density charge region at the electron track end, or Bragg peak, commonly called *blob*. Obviously, a $0\nu\beta\beta$ track emitting two electrons with a common origin will be similar to a single track with two *blobs* at the track ends, while single electron background tracks will only produce one Bragg peak. The end of single electron tracks where no *blob* is found will be called *tail*.

  The energy, or charge, Q, in a certain *blob* is determined by summing up the energy deposited around a sphere of radius R_o centered at each of the high density track end coordinates[17]. The following mathematical expression summarizes our *blob* charge definition

$$Q_j = \sum_{d=r_{j,n}<R_o} q_n \quad \text{and} \quad Q_l = \min(Q_1, Q_2), \tag{2}$$

where the index $j = 1, 2$ represents each of the track ends, and the index n corresponds to the hit number inside the *TRestTrackEvent* definition (given in appendix A). The charge integration at each end-track, $Q_{1,2}$, is performed for those hits, q_n, that satisfy that the hit distance d to the blob coordinates, at each track end, j, is below R_o. Different combinations of Q_1 and Q_2 may be chosen to construct an observable to be used for pattern recognition. In our posterior analysis we will use the lowest charge blob definition, Q_l, a value which, in principle, is necessarily low for background events in which the electron *tail* is properly identified, while for $0\nu\beta\beta$ events cannot be low. In the case of two-dimensional readouts this observable is independently calculated for each projection, obtaining two values, Q_l^X and Q_l^Y, to be combined later.

- *End-track twist.* Another feature characterizing the behavior of the electron trajectory near the *blob* is the erratic nature of its trajectory. This erratic behavior, due to higher electron

---

[17] The blob coordinates are obtained by *TRestFindTrackBlobsProcess*, detailed in appendix B.3.

scattering angles, does not manifest at the initial interaction vertex, or *tail*, at single electron background events. On the contrary, energetic electrons, with initial energies of the order of the $Q_{\beta\beta}$ value, will usually produce a clear straight *tail*.

In order to quantify this effect we introduce the *twist parameter* that measures the angle between consecutive hits belonging to a *top-level track*[18] near the track ends. The *twist parameter*, ξ, is expressed as follows

$$\xi_j = \frac{1}{N_j} \sum_{i=0}^{N_j} (1 - \hat{n}_i \cdot \hat{n}_{i+1})/2, \quad \text{and} \quad \xi_l = \min(\xi_1, \xi_2), \tag{3}$$

where $\hat{n}_i$ represents the unit vector defined by the ith and ith $+$ 1 nodes of a reconstructed *top-level track*, and $i = 0$ can be identified with the first (or last) node of the *physical track*. The total number of hits, N_j, considered in each end-track, $j = 1, 2$, is relative to the total number of hits, N_{tot}, at the *top-level track*. In our analysis we have defined N_j as the 25% of N_{tot}, i.e. a track containing $N_{\text{tot}} = 28$ hits will average the 6 angles formed by the 7 hits closer to the end-tracks. The normalization factor 2 in the previous expression is introduced to assure ξ is mathematically contained in the range (0, 1), where $\xi = 0$ means that the nodes are fully aligned. As in the previous observable, we define ξ_l as the lowest value between the *twist parameter* obtained at each end-track, and obtain independent values for each projection, ξ_l^X and ξ_l^Y, for the two-dimensional case.

- *Track length.* We might also expect that two electron $0\nu\beta\beta$ tracks will produce a slightly shorter path due to the generation of two blobs, and therefore a higher dE/dx in average. Although this observable is expected to have a weak discriminating power, we still assess its performance.

The *track length*, $\mathcal{L}$, is simply calculated as the total measured distance between the track ends, following the ordered node sequence at the *top-level track* produced after *TRestTrackReconnectionProcess*. Expressed as

$$\mathcal{L} = \sum_{i=0}^{N-1} d_{i,i+1}, \tag{4}$$

where N is the total number of hits, or nodes, in the *top-level track*, and $d_{i,i+1}$ is the distance between the hit i and $i + 1$, or *connected hits*. Again, for the two-dimensional case we will obtain 2 independent length measurements for each projection, $\mathcal{L}^X$ and $\mathcal{L}^Y$.

In summary, these observables will allow us to discriminate background by differentiating the main background and signal properties. When identifying a $0\nu\beta\beta$ signal event we search for a main energetic track, with almost negligible energy deposition on secondary tracks belonging to the event. Signal events will generally lack the presence of an end-track *tail*, and therefore, the combination of low end-track values for *blob charge*, $\mathcal{Q}_l$ and *twist parameter*, ξ_l, will be decisive on the identification of such a feature. These main characteristics can be easily recognized on the background and signal events shown in figure B3, at appendix B.

Finally, it is important to remark that these parameters serve also to describe in a quantitative way each event, i.e. they can be exploited to characterize the experimental data and support the construction of a Monte Carlo background model. The nature from different background contributions due to the different contamination sources and components in the detector might be better understood by studying the behavior of these parameters on different

---

[18] A *top-level track* in our analysis identifies with the *physical track* reconstructed, see also *TRestTrackEvent* definition at appendix A,

populations of events, measured at different data taking conditions. Therefore, even if a parameter does not contribute finally on the signal-to-background significance, it may be valuable in other scenarios to help understanding the nature of the data collected with the detector.

2.2. Results

We report in this section the results obtained on background reduction considering the topological features of signal and background events. We will take into account the forecasted PandaX-III data taking conditions, i.e. a vessel filled with 10 bar of xenon+TMA 1% gas mixture, and a detailed detector response, carefully described in appendix B.

The Monte Carlo $0\nu\beta\beta$ signal events, used to define our signal efficiency have been imported in *RestG4* using a *Decay0* [20] pre-generated event population. The initial 4-momentum definition of each of the two electrons produced by the $0\nu\beta\beta$ decay is randomly positioned at the active volume of the detector, or gas volume, for each event, and then tracked using *Geant4* physics processes.

In [9], the main contamination sources contributing to the background of the experiment were analyzed, allowing to identify the most critical detector components in terms of negative impact to the overall background of the experiment. In this work we go a step further by presenting a more definite background level achievable with the PandaX-III detector. For that purpose, we have chosen representative detector components, as the high pressure copper vessel and the Micromegas readout plane, to evaluate the pattern recognition capability of our event reconstruction algorithms. We have simulated the ^{238}U and ^{232}Th decay chains at those detector components. For the copper vessel we generate our events randomly on the bulk of the material, while for the Micromegas contamination we generate surface events at the readout plane location, behind 10μm of copper. In order to guarantee enough statistics at the end of our topological analysis we have generated near 10^9 decays of ^{238}U and ^{232}Th isotopes for the vessel contribution, and 5×10^7 decays for the Micromegas contributions. Our signal efficiency has been calculated over an initial population of 10^6 generated $0\nu\beta\beta$ events.

Signal and background events have been processed following the scheme described in appendix B.1. Figure 1 shows the parameter distributions obtained during the data processing chain, and described previously on section 2.1. Except for the *track energy ratio*, it must be noted that the other distributions follow similar patterns for different simulated background components.

We apply our criteria to discriminate signal from background sequentially on these parameters, following the order given in the list provided in section 2.1. We combine the *XZ*- and *YZ*-projections into a single parameter by choosing the minimum value of each projection for the *blob charge*, $\mathcal{Q}^{\min}$, and *twist parameter*, $\xi^{\min}$, i.e. for each event we only consider the two-dimensional projection where *tail* identification is more obvious. While we have chosen the maximum *track length* of each projection, $\mathcal{L}^{\max}$, supported by the argument that $0\nu\beta\beta$ events are expected to be shorter. We have obtained our threshold values, or cuts, such that they maximize the signal-to-background significance by maximizing the figure of merit, $\epsilon_s/\sqrt{\epsilon_b}$, at each of the four topological criteria used in our analysis, i.e. *track energy ratio*, *blob charge*, *twist parameter* and *track length*. Here, ϵ_s and ϵ_b correspond to the signal and background reduction factors at each of those steps. The following topological parameter conditions were obtained after such optimization

$$\theta < 2.66 \cdot 10^{-3}, \qquad \mathcal{Q}_l^{\min} > 160\,\text{keV}$$

$$\xi_l^{\min} > 0.008 \quad \text{and} \quad \mathcal{L}^{\max} < 204\,\text{mm}.$$
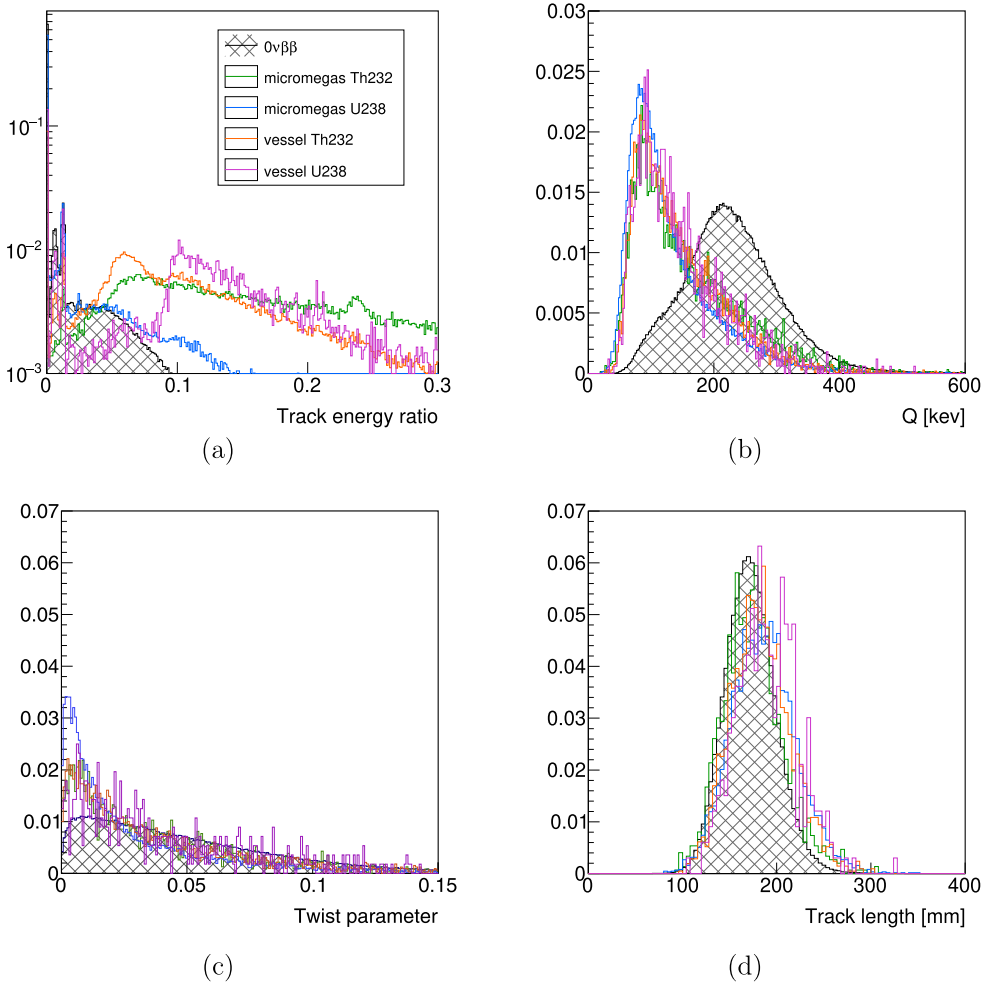
(a)　　　　　　　　　　　　　　(b)

(c)　　　　　　　　　　　　　　(d)

Figure 1. The distributions of topological parameters, normalized to unity, obtained from the *XZ*-projection of $0\nu\beta\beta$ event tracks (filled curve), compared to the equivalent distributions of the four background datasets considered in our study (colored curves). The distributions are built using only the events contained in the ROI, i.e. in the energy range (2395, 2520) keV. On top of that, each distribution accumulates the sequential topological criteria we applied in the preceding parameter. (a) The track energy ratio, or θ parameter. Different physics processes contribute to create certain structures on this distribution. Peaks found at low θ-values, $\theta \lesssim 0.01$, are related to the escape peaks of xenon, while hills appearing at higher values are related to bremsstrahlung gamma emission produced by high energy electron tracks, or Compton scattering processes that end up on a photoelectric process producing a main electron energetic track. (b) The lower track energy blob, $\mathcal{Q}_l^X$. Events satisfying $\theta < 1.33 \times 10^{-3}$. (c) The lower *twist parameter*, ξ_l^X, satisfying $\theta < 1.33 \times 10^{-3}$ and $\mathcal{Q}_l^X > 176$ keV. (d) The track length, $\mathcal{L}^X$, satisfying $\theta < 1.33 \times 10^{-3}$, $\mathcal{Q}_l^X > 176$ keV and $\xi_l^X > 0.0133$.

The final selected event population, signal or background events, will necessarily satisfy those conditions. We should mention that these values have been obtained by optimizing the dataset corresponding to the ^{238}U vessel generated events. However, it must be noted that such a choice is not critical in view of the similar behavior of the different topological
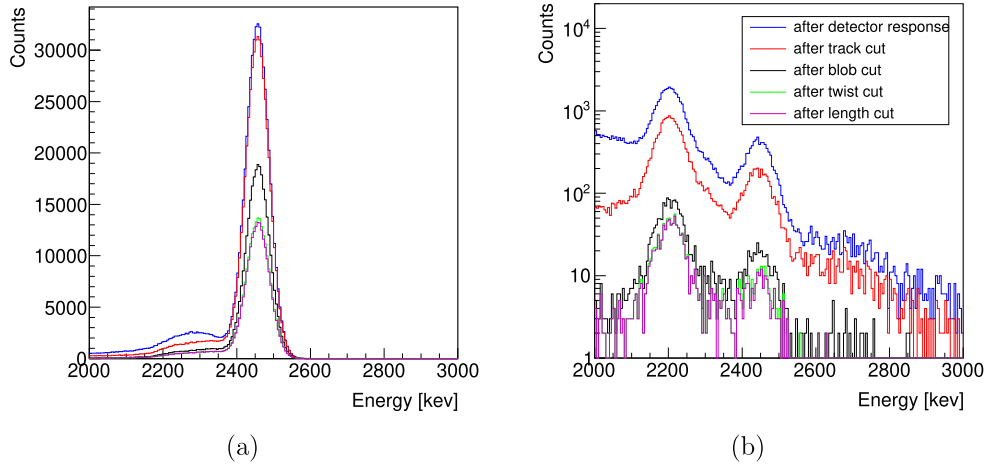
(a)   (b)

Figure 2. (a) The simulated energy spectrum response of $0\nu\beta\beta$ events around the $Q_{\beta\beta}$ value, and the resulting spectra after applying the different topological criteria sequentially, i.e. *track energy ratio*, *blob charge*, *twist parameter* and *track length*. The main peak corresponds to $0\nu\beta\beta$ events for which full decay energy was acquired by our detector, while the distribution left tail corresponds to $0\nu\beta\beta$ events with partial energy loss outside the detector readout plane boundaries, or partially contained in the detector acquisition window, limited by the electronics. (b) The equivalent background spectra resulting from ^{238}U decays generated from the copper vessel. The peaks observed correspond to two high energy gammas, 2204 and 2448 keV, from the spectrum gamma emission of ^{214}Bi present in the ^{238}U decay chain. The 3% FWHM Gaussian smearing introduced by the *TRestHitsSmearingProcess* can be recognized in both figures.

parameters produced by different background sources and components, as it can be observed in figure 1. As a consequence, the threshold values obtained are not strongly dependent on this choice. Anyhow, we must remark that the most sensitive threshold value corresponds to the *track energy ratio* parameter, as it can be induced from the distribution presented in figure 1(a). Anyhow, in our particular conditions, we observe that the low value of θ, resulting in our particular optimization, constrains our event selection practically to the population of events that contain only one single main track.

Before applying the sequential topological criteria we filter the event population to those events that are found in the ROI, defined as $Q_{\beta\beta}\pm2\sigma$, where $\sigma = 31$ keV is given by the assumed 3% FWHM energy resolution, introduced in *TRestHitsSmearingProcess* (see appendix A.2).

The energy definition considered is the value extracted from *TRestTriggerAnalysisProcess* (see appendix B.3) which contains all the effects introduced by the detector readout response. The resulting energy spectrum for $0\nu\beta\beta$, together with one of the background components studied is shown in figure 2, including the accumulated effect of sequential cuts on signal efficiency and background reduction.

Table 1 shows the percentage of accepted signal events and background reduction factors, i.e. the final number of events surviving after each cut per simulated one. The first row, *Geant4*, is the calorimetric response where we consider all the energy depositions in our active volume, here we only integrate events inside our ROI; the second row, *Response*, corresponds to the population of events that remain in the ROI after applying the detector

Table 1. Percentage of accepted ^{136}Xe signal events, or signal efficiency, together with the absolute background reduction factors at the ROI for the PandaX-III CDR baseline readout, after applying the detector response and sequential topological cuts described in the text. Each row includes the accumulative reduction of all previous cuts. For the sake of simplicity we avoid to express statistical errors in this table, although we will propagate them in table 2, where final background rates are given. Systematic errors related to the Geant4 generator were studied previously in [9] using two independent Monte Carlo generators and geometry implementations. Any other source of uncertainty will be discussed in our conclusions.

Origin	Gas volume ^{136}Xe	Lateral vessel		Micromegas	
Isotope	(%)	^{238}U	^{232}Th	^{238}U	^{232}Th
Geant4	71.1	2.28×10^{-5}	2.10×10^{-4}	2.69×10^{-3}	3.38×10^{-3}
Response	52.6	7.40×10^{-6}	8.87×10^{-5}	1.65×10^{-3}	1.56×10^{-3}
Track	37.7	1.71×10^{-6}	1.13×10^{-5}	9.60×10^{-4}	1.14×10^{-4}
Blob	26.4	2.60×10^{-7}	1.60×10^{-6}	9.28×10^{-5}	2.71×10^{-5}
Twist	22.0	1.64×10^{-7}	1.03×10^{-6}	4.87×10^{-5}	1.97×10^{-5}
Length	17.5	7.85×10^{-8}	6.07×10^{-7}	3.06×10^{-5}	1.43×10^{-5}

response, i.e. after applying readout fiducialization and boundaries imposed by the electronics acquisition window. We should remark, that when considering background events with energies above the ROI, they may finally get inside the ROI definition due to the partial event energy registered by our detector. The subsequent rows in the table show the accumulative effect of the topological cuts applied there after. The resulting topological selection contains only contributions from the ^{208}Tl isotope decay produced inside the ^{232}Th decay chain, and ^{214}Bi produced inside the ^{238}U chain.

We discover looking into table 1 that the topological background reduction is about a factor 100, and it is comparable for all the background components except for the ^{238}U Micromegas dataset, which is worse by about a factor 2. This result is explained mainly due to the failure of our topological cuts to discriminate surface events, as revealed by figure 3(a). The surface background contamination originated from the Micromegas readout is related to ^{214}Bi decays producing β-tracks emanating from the readout plane. The ^{214}Bi decay is the only isotope, inside the ^{238}U decay chain, producing a β-decay with Q_β end-point energy above our ROI. The surface contamination coming from the ^{232}Th Micromegas decays is also present, but as it is observed in figure 3(b), there is also an important volume component which is due to a 2614 keV gamma which is produced at practically each ^{208}Tl decay in the ^{232}Th chain. The volume component due to ^{214}Bi in the Micromegas is negligible, compared to its surface contribution, related to the fact that only about 1.5% of the decays produce a 2447 keV gamma.

Table 2 shows the final background contribution of the detector components studied, after renormalization with the corresponding material activities of ^{238}U and ^{232}Th for Micromegas and vessel components. We use the same values as reported in [9]. The surface contamination upper limit of Micromegas being <45 nBq cm^{-2} and <14 nBq cm^{-2}, and the radiopure copper vessel bulk contamination being 0.75 μBq kg^{-1} and 0.2 μBq kg^{-1} [21], for ^{238}U and ^{232}Th, respectively.

We conclude that the vessel contamination contribution is negligible compared to the contribution of the Micromegas readout plane, being its estimated overall contribution at the level of 1 cpy. Of course, the final value will ultimately depend on the real surface
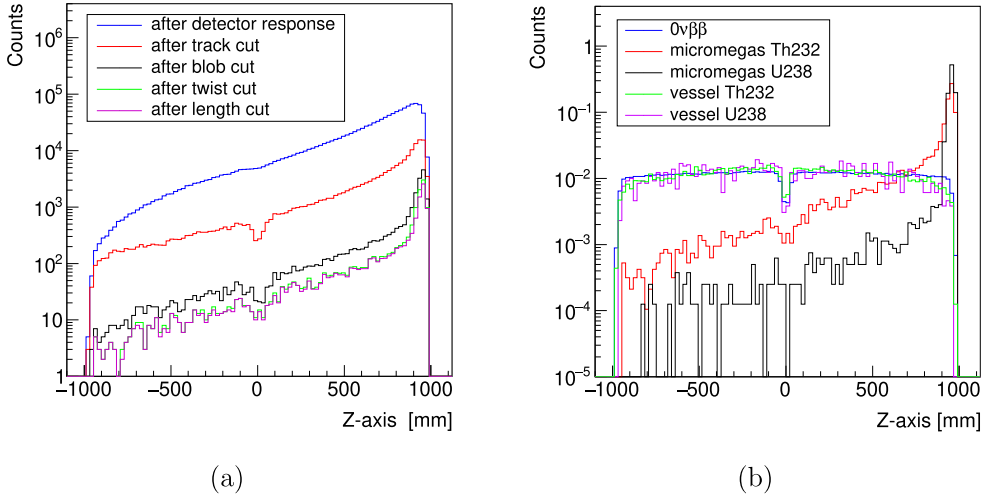
(a)   (b)

Figure 3. (a) The distribution of the average z-position for tracks from ^{232}Th generated events originated from the Micromegas readout plane, placed at $z = 1000$ mm, after different steps of the topological cuts. (b) The distribution of the average z-position for signal and the four background contributions studied, after all the topological cuts have been applied.

Table 2. Final background contribution after applying all the topological cuts, i.e. renormalizing the results of the last row presented on table 1 using the expected ^{232}Th and ^{238}U isotope contamination for the different detector components. The results are presented in the usual units, counts per year (top row) and keV^{-1} kg^{-1} yr^{-1} (bottom row). Measurements of microbulk Micromegas surface contamination provide results within the sensitivity limit of the measuring device, and therefore, our result for this component is given as an upper limit of the final background contribution to the experiment.

Lateral vessel		Micromegas		Units
^{238}U	^{232}Th	^{238}U	^{232}Th	
6 ± 1	13 ± 1	< 1534	< 223	10^{-3} cpy
0.365 ± 0.042	0.752 ± 0.032	< 310	< 45.0	10^{-6} keV^{-1} kg^{-1} yr^{-1}

contamination present at the Micromegas readout. Considering that the contamination levels used in our Monte Carlo are limits obtained experimentally we may still remain optimistic about this contribution. In section 4 we will show how we can achieve an additional reduction on the surface contamination near the readout planes by exploiting the effect of electron diffusion on the time signals registered by our TPC [13]. Finally, we must also consider our result as conservative from the point of view that there will be probably room for improvement by enhancing the existing parameter correlations, e.g. exploiting more sophisticated multivariate methods.

We reserve any discussion on the discrimination potential of the topological parameters used in our analysis for section 3, where we will also argue on the benefit of introducing the *twist parameter* for our particular readout granularity, and we will assess the appropriateness of the two-dimensional 3 mm pitch detector choice as a baseline for the PandaX-III experiment.

3. Readout topology and granularity studies

This section extends our study on background discrimination to different readout topologies and granularities. We will evaluate our new event reconstruction algorithms—detailed in appendix A—using two different readout layouts, a two-dimensional stripped layout and a three-dimensional pixel layout, to assess the potential gain on signal efficiency and/or background reduction. The results presented in this section will provide insight on the optimum readout scheme to be used in the PandaX-III data taking conditions, i.e. a vessel filled with 10 bar of xenon+TMA 1% gas mixture. We will discuss then the appropriateness of the PandaX-III baseline readout choice—detailed in appendix B.2—considering the advantages and disadvantages of two-dimensional versus three-dimensional event reconstruction, for different sizes of the readout channels, or detector granularity.

We have re-processed the same datasets analyzed in section 2. However, this time we focus only on the ^{238}U and ^{232}Th background contributions originated in the vessel. The same processing chain -as described in appendix B.1—has been applied, being the readout used for event reconstruction the only modified element. We have systematically produced different readout structures in 2D and 3D at different pitch values. For the *two-dimensional case*, the 3 mm pitch PandaX-III baseline readout module design has been generalized to allow the definition of readout channels of any pitch value. When defining a different pitch value we keep the size of the module unchanged, e.g. when we define a 1 mm pitch readout we increase the number of readout channels from 64 to 192. For the *three-dimensional case* we have simply designed a readout module using squared pixels, and for practical reasons we adapt the number of pixels to keep constant the size of the readout module as a function of the pixel size. We must remark anyhow that minor adjustments are required at the processing chain when dealing with different pitch sizes, e.g. the *cluster distance* used to identify tracks (see *TRestHitsToTrackProcess* description at appendix A.2) was set to 2.5× the pitch size, and the radius of the *blob charge* definition was fine tuned, being as low as $R = 1$ cm for 1 mm pitch, and as high as $R = 2$ cm for 1 cm pitch.

In this particular study we seek to compare the rejection potential using different readout topologies. We must remark that the fiducialization response is not relevant for such goal, and we are interested to determine the topological rejection potential without those constraints. Therefore, an important difference with respect to the analysis of the previous section 2 is that we consider the full event interacting in the gas volume. In addition, our readout plane will fully cover the gas medium, or active area [19], so that the event reconstruction limits are just bound by the field cage wall defined in the *Geant4* geometry.

Tables 3 and 4 show the effects of topological criteria, similar to those applied on section 2, for the simulated signal and background populations with different readout layouts and granularities. For each criterion (*track energy ratio*, *blob charge* and *twist parameter*) we provide the relative reduction, in percentage, with respect to the surviving event population from the previous criterion, or cut. We maximize our signal significance at each step by maximizing the quantity, $\epsilon_s / \sqrt{\epsilon_b}$. We observe that the reduction potential of the *track energy ratio* criterion is slightly better compared to the version where we included the fiducial detector response in section 2, i.e. by comparing the 3 mm pitch stripped readout results shown in table 3 with the values that can be deduced from table 1. This difference is mostly due to the fact that multi-track events have left the ROI after applying the detector response,

---

[19] Instead of using the 41-modules readout plane scheme described in appendix B.2, we extend the readout definition limits by adding readout modules as needed to cover the full active area.

Table 3. Relative signal efficiency and background reduction factors, in percentage, after each topological criterion, at different pitch values (1, 2, 3, 4 and 6 mm) for the two-dimensional stripped readout. The initial population of events considered at each step are the events in the ROI surviving from the previous criteria.

Origin Isotope	1 mm			2 mm			3 mm		
	^{136}Xe	^{238}U	^{232}Th	^{136}Xe	^{238}U	^{232}Th	^{136}Xe	^{238}U	^{232}Th
Track	69.6	9.92	2.68	63.9	8.91	2.19	69.2	11.2	2.79
Blob	58.5	9.13	9.40	59.4	10.2	10.7	65.4	12.9	14.0
Twist	85.1	60.6	65.0	93.9	78.7	80.0	94.2	80.6	81.1
Total	34.6	0.55	0.16	35.6	0.71	0.19	42.6	1.16	0.32

Origin Isotope	4 mm			6 mm		
	^{136}Xe	^{238}U	^{232}Th	^{136}Xe	^{238}U	^{232}Th
Track	73.1	13.2	3.33	79.2	16.7	4.37
Blob	64.2	14.1	14.5	54.2	12.3	12.2
Twist	94.0	83.2	81.6	90.2	82.5	74.4
Total	44.1	1.55	0.39	38.7	1.69	0.40

Table 4. Relative signal efficiency and background reduction factors, in percentage, after each topological criterion, at different pitch values (2, 4, 6, 8 and 10 mm) for the three-dimensional pixel readout. The initial population of events considered at each step are the events in the ROI surviving from the previous criteria.

Origin Isotope	2 mm			4 mm			6 mm		
	^{136}Xe	^{238}U	^{232}Th	^{136}Xe	^{238}U	^{232}Th	^{136}Xe	^{238}U	^{232}Th
Track	69.6	5.84	1.60	74.0	8.20	2.20	76.8	10.3	2.73
Blob	79.1	13.4	13.0	74.8	13.2	12.9	65.2	10.4	10.6
Twist	83.4	48.7	46.5	77.3	53.8	49.0	74.4	58.6	52.6
Total	45.9	0.38	0.097	42.8	0.58	0.14	37.3	0.63	0.15

Origin Isotope	8 mm			10 mm		
	^{136}Xe	^{238}U	^{232}Th	^{136}Xe	^{238}U	^{232}Th
Track	78.8	12.0	3.3	80.3	13.6	3.75
Blob	68.0	10.6	10.7	64.1	10.8	9.72
Twist	—	—	—	—	—	—
Total	53.6	1.20	0.35	51.5	1.47	0.36

and a subset of the background event population loses its multi-track features when we apply the detector response.

One interesting finding is that the newly introduced *twist parameter* starts to be efficient below 6 mm pitch, for both the two-dimensional and the three-dimensional readout versions. As observed in table 3, the 8 and 10 mm pitch pixel readouts studied did not improve the significance of the signal when using the *twist parameter* criterion. Figure 4 shows the evolution of the *twist parameter* and it confirms the advantage to differentiate signal and
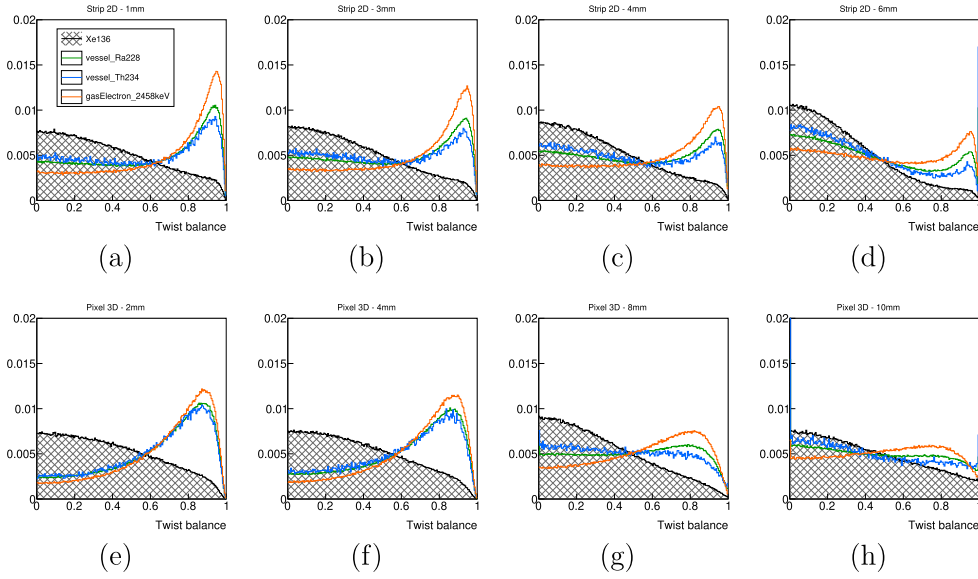
J. Phys. G: Nucl. Part. Phys. **47** (2020) 045108

J Galan *et al*

Figure 4. The distribution of the twist balance, ξ_b, parameter, a variant of the *twist parameter* calculated as $\xi_b = (\xi_h - \xi_l)/(\xi_h + \xi_l)$, being ξ_h the higher end-track *twist parameter* value, and ξ_l the lower end-track *twist parameter* value. The different plots show the evolution of this parameter for different detector granularities, and different readout topologies, 2D strips (top figures), and 3D pixels (bottom figures). The distribution from $0\nu\beta\beta$ signal events (filled curve) is compared to the background distribution of different contaminations as described in the text. For the 2D strips distributions only one of the projections is represented.

background distributions as the pitch value is reduced. Furthermore, we have added to this figure a third dataset where we generate single electrons with energy equal to the $Q_{\beta\beta}$ of ^{136}Xe, demonstrating a small dependency on the quality of the topological observables related to the nature of the background source.

The *twist parameter* provides the less powerful topological criterion (as it is deduced from tables 3 and 4). However, we should take into account that this criterion operates after the event selection of *track energy ratio* and *blob charge* criteria. It is remarkable then that the *twist parameter* criterion is still capable to improve the signal significance, meaning that the *twist parameter* definition is an additional track feature that can be explored, exploited, and optimized in a future multivariate analysis that combines the track observables more efficiently.

We summarize our results in figure 5, where we present the total signal significance, measured as $\epsilon_s/\sqrt{\epsilon_b}$, obtained with our topological criteria for different readouts. This figure allows us to assess now the gain on experimental sensitivity when reducing the detector granularity, and quantify the impact of using a two-dimensional readout compared to a three-dimensional one. We conclude from these results that the negative impact of using a stripped two-dimensional readout is counterbalanced by using a lower detector granularity, i.e. a 3 mm pitch two-dimensional readout is equivalent to a 1 cm pitch three-dimensional readout, with the advantage that the number of channels is reduced considerably in the two-dimensional version.
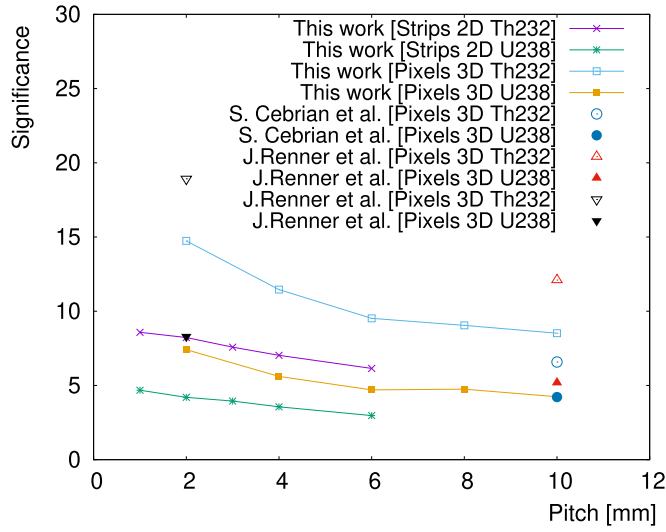
Figure 5. Total significance of topological criteria for the different detector granularities and topologies studied. The lines interconnecting data points show our results for the two different contaminations studied (^{232}Th and ^{238}U), and the two different readout topologies (Strips 2D and Pixels 3D). We add to this figure the significances obtained in previous studies, calculated from the background and signal efficiencies reported at [17, 19], and discussed in the text.

Moreover, we use figure 5 to compare our results to previously published studies—where they use similar conventional discrimination techniques. As it is observed, we find remarkably good agreement with [17] on the final significance at 1cm 3D pixel readout, for both vessel contaminations. While, for [19], although we find reasonably compatible results for the ^{238}U contamination chain (or ^{214}Bi), we find their result on ^{232}Th (or ^{208}Tl) surprisingly high when compared to our data, being the origin of the main discrepancy at the level of the *track energy ratio* criterion. The difference between our results and those reported in [19] are even larger when we consider that their 1cm results correspond to pure xenon, and it was claimed in [17] that background reduction due to blob identification is worse by a factor 3 due to the higher electron diffusion in pure xenon, and that finally implies almost a factor 2 worse on significance. The source of this discrepancy might be related to the non-independent treatment of electron diffusion and detector granularity in [19], and the higher gas pressure, 15bar, of their setup. On the other hand, we should be cautious, since direct comparison of different setups is not always obvious. The significance of the criteria may depend not only on the origin, or the particular geometry of the setup, but also on the nature of the background and the target density, or detector pressure. We have seen, for example in figure 4, how a single electron track artificially launched from the gas volume, as it is done e.g. in [22], slightly generates better results in terms of the topological parameters quality.

4. Fiducial background rejection in charge based TPC readouts

The results obtained in section 2 unveiled the weakness of our topological cuts to discriminate surface contamination events originated from the Micromegas readout planes, as revealed by figure 3(a). Our setup lacks an absolute reference time, t_o, from the interactions taking place in
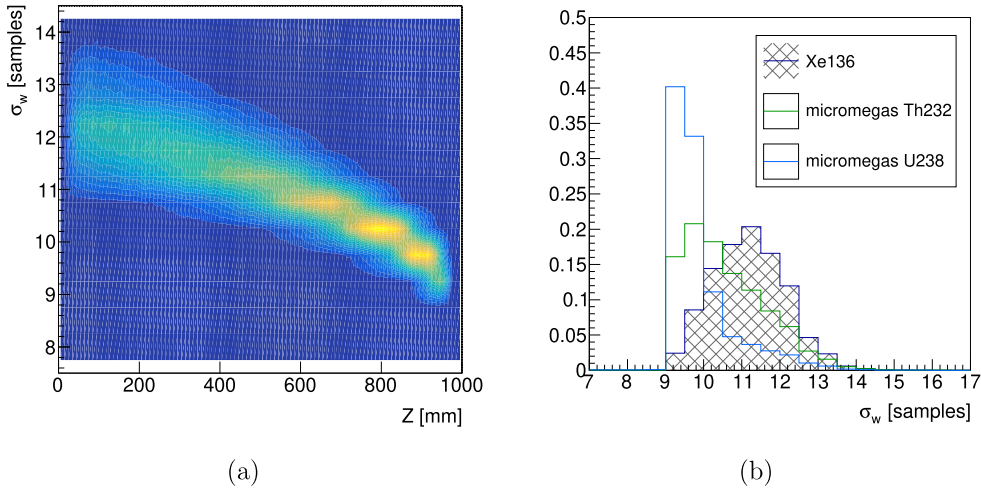
Figure 6. (a) A contour plot produced using the $0\nu\beta\beta$ signal event population, and showing the diffusion parameter, σ_w, as a function of the event position. Only one half of the symmetric TPC is shown here. Lower σ_w values correspond to events taking place near the readout plane, at $z = 1000$ mm, and larger values correspond to events taking place near the cathode, at $z = 0$ mm. (b) The distribution of the σ_w parameter for $0\nu\beta\beta$ signal events (filled curve), and the two background Micromegas components studied.

the detector volume. Without such a reference we cannot directly determine the event absolute z-position, making it hard to discriminate surface events originated on the readout planes, or the cathode.

In this section we develop a basic technique to demonstrate the remaining potential for further background reduction by attenuating the negative impact of surface contamination, even in the absence of t_o. For that we take advantage of the readout signal dependence on the diffusion of ionized electrons drifting towards the readout planes. The time signals induced on the Micromegas readout, due to a particular energy deposition on the TPC volume, will be broader when the electron cloud has drifted a longer distance towards the readout plane. However, the resulting primary electrons from β-tracks generated near the readout plane will drift a very short distance, and therefore the original charge distribution will be almost unaffected by the electron drift.

We have re-processed the $0\nu\beta\beta$ signal and Micromegas background datasets analyzed in previous sections using the same data chain parameters as defined in section 2. However, now we need to include few additional effects in order to extract a parameter, σ_w, that correlates to the diffusion of electrons in the gas. The generation of this parameter is performed under realistic conditions including the signal shaping introduced by electronics and a typical electronic noise level (further details on the extraction of this parameter are given in appendix C). In figure 6 we observe the desired dependence of σ_w with the z-position of the simulated $0\nu\beta\beta$ event, together with the σ_w distribution produced by the two Micromegas background components compared to the $0\nu\beta\beta$ signal one. We observe clearly how background surface events peak at low values of σ_w.

Therefore, we exploit the properties of σ_w to introduce an additional selection criterion in our data by accepting only the population of events that are above a certain diffusion threshold, as quantified by σ_w. Table 5 presents the results on event acceptance after applying

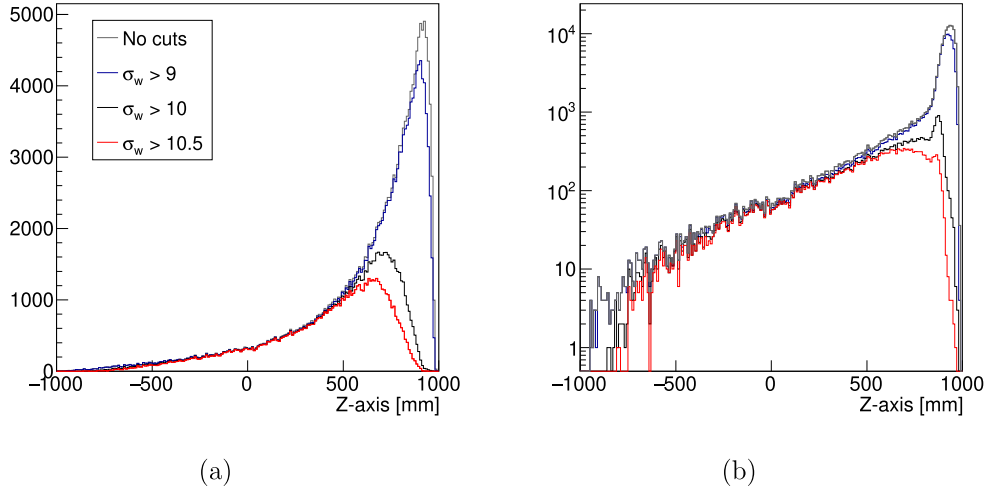(a)   (b)

Figure 7. Distribution of the average z-position for events produced by ^{232}Th (a), in linear scale, and ^{238}U (b), in logarithmic scale, generated from the Micromegas readout plane, placed at $z = 1000$ mm. Different curves represent the background reduction after applying different threshold values on the diffusion parameter, σ_w. The initial population of events, tagged as 'No cuts', are the events found in our ROI definition.

Table 5. Overall background reduction, in percentage, for events generated from the Micromegas readout plane, ^{232}Th and ^{238}U contaminations, and the corresponding signal efficiency, after applying different cuts on the diffusion parameter, σ_w.

	$\sigma_w > 9$	$\sigma_w > 10$	$\sigma_w > 10.5$
^{232}Th	93.3	53.7	44.4
^{238}U	80.9	20.1	15.1
^{136}Xe	99.4	83.0	74.5

cuts for different threshold values, where we observe that we manage to reduce by an additional factor 0.15 the contribution from the ^{238}U chain, due to ^{214}Bi decays, while keeping our signal efficiency near 75%. We must remember that the Micromegas contribution of ^{238}U was a dominant component in table 2. The background reduction on the ^{238}U chain is more significant than in the ^{232}Th chain, due to the fact that the 2614 keV gamma from ^{208}Tl is produced in almost any decay, generating finally a higher ratio of volume to surface events compared to the ^{214}Bi contribution, which produces a β decay with high energy gamma emission only at 1.5% of the decays. We should remark that the values presented on table 5 have been obtained without applying topological cuts. Still, these results are independent on previous topological cuts, since we do not observe any dependence on the z-distribution after cuts, as it would have been revealed by the $0\nu\beta\beta$ population in figure 3(b), after the topological analysis performed in section 2.

Finally, in figure 7 we show the background distribution as a function of the event position, for each of the two Micromegas isotope contaminations, after applying different σ_w threshold values. In these figures we observe how all the events from surface nature have been completely removed already for $\sigma_w > 10$, and how the remaining background event

population on table 5, i.e. 20.1% for ^{238}U and 53.7% for ^{232}Th, are only due to volume events that can only be removed by exploiting topological event features.

5. Conclusions

In this work, we have presented results on topological background discrimination considering a realistic detector response for the PandaX-III TPC baseline design. The realistic response includes an accurate description of the existing Micromegas detector layout used for the reconstruction of events in a similar way as it is done with experimental data. However, the most important element on the detector response relies on an appropriate definition of the energy of each event by considering the electronic acquisition window and the readout plane boundaries of the PandaX-III design.

We observe a negative effect on our topological background discrimination when we apply the detector response to the Monte Carlo event data, hence the importance to include those effects in our study. The results obtained show that a promising final background level is achievable. Using topological arguments, we find that for a signal reduction of $\sim 1/3$ we can achieve a reduction factor of about 100 on the background contamination. Even after including the detailed detector response, this result is still in good agreement with the preliminary estimate used to calculate the sensitivity of the first PandaX-III module at [9].

It is remarkable that the overall background contribution of the detector vessel will be negligible for the lifetime of the experiment. While an acceptable value for the contribution of one of the most uncertain sources of background in the experiment—the detector readout plane—results in a conservative contribution of 1–2 counts per year.

Furthermore, we demonstrate the capability of our setup to perform a fiducialization in the *z*-axis even in the absence of t_o, and remove a considerable amount of surface background events that cannot be eliminated by other means. By exploiting this technique we are able to reduce the Micromegas background contribution to levels well below 1 count per year, preserving 80% of the $0\nu\beta\beta$ signal.

We have also addressed the question relative to the impact of a two-dimensional readout on the pattern recognition of $0\nu\beta\beta$. It is obvious that the lack of a full three-dimensional event reconstruction leads to a reduced discrimination power. In the two-dimensional case, our event identification and classification must be based on two projections of the event, and necessarily some of those event projections hide information that can only be revealed by a three-dimensional reconstruction. Occasionally, a secondary track will be hidden by the projection and counted as energy of the main track, affecting the *track energy ratio* criterion, while in other events, the integration of the charge of an electron moving orthogonally to the projection may mimic a *blob*, reducing the rejection power of the *blob charge* and the *twist parameter* criteria. We have systematically studied the significance of our topological criteria for different detector granularities at three-dimensional pixel and two-dimensional stripped readout designs, concluding that the negative impact on two-dimensional readout pattern recognition is counterbalanced by a reduced detector granularity, i.e. a 1 cm pixel readout reaches an event topological rejection power comparable to a 3 mm stripped readout. Consequently, the 3 mm pitch stripped baseline readout choice for PandaX-III detector is much more convenient, not only in terms of detector design complexity due to the reduced number of channels required, but also minimizing the impact of typically non-radiopure electronics equipment near the detector readout planes.

We have introduced novel track parameters, as the *end-track twist* and the *track length*, never exploited before on the pattern recognition of $0\nu\beta\beta$ using conventional techniques. We

have assessed their discrimination potential, and observed how the *end-track twist* feature is still contributing to the signal-to-background significance, even after filtering the event population previously with other topological criteria, as the *track energy ratio* and the *blob charge*. Furthermore, we have found that this parameter is relevant for pattern recognition when the detector granularity is below 6 mm, concluding that such granularity or better is necessary to reveal the aforementioned track features.

We believe our results are conservative, and that there exists room for improvement and optimization through multivariate methods that better exploit the different parameter correlations. Additionally, exploring the use of novel track parameters, or improving the existing ones, can push the conventional background discrimination further. Additional parameters, as the dE/dx along the track, previously studied at [23], and its peculiarities at the end-tracks could allow to improve these results by properly combining them with the parameters studied in this work.

Recently, we have seen how machine learning methods provide enhanced capabilities for pattern recognition, as reported in [19, 22, 24]. These techniques, which provide excellent signal-to-background discrimination ratios, could never replace a conventional analysis. In a conventional analysis the different topological parameters serve as a mechanism to control the goodness of the Monte Carlo event reconstruction. In other words, we will be able to validate the different event parameters obtained by comparing with the experimental data, allowing us to be confident on our event reconstruction and the effectiveness of our topological criteria. The use of machine learning techniques, without further or complementary analysis, will not be sufficient by itself to prove the background of the detector.

The present work served to demonstrate the existing software infrastructure implemented in REST for event reconstruction, and the flexibility to add a detailed detector response at different levels, as necessary. REST has been conceived to provide a high degree of modularity allowing to connect or disconnect different *event processes* at any place of the *event data* processing chain. We have designed dedicated *event processes* and *metadata* structures to construct the PandaX-III Monte Carlo processing chain. Still, additional details can be included in future studies, that may implement even more realistic event reconstruction. A future Monte Carlo event processing may consider further details ignored in this work, as e.g. the gap inter-distance between neighbor *readout modules*, or the necessary deconvolution of the electronic signal shaping.

Finally, when comparing to real experimental conditions we will need to consider different systematic effects due to different parameters that may fluctuate, or change, during the data taking periods of the experiment, such as the electronic noise, the drift field intensity or irregularities, the concentration of TMA and its effect on the gas properties, the gain inhomogeneities on each detector module, the effect of readout module gaps on the event reconstruction, possible dead channels in the detectors, etc. All those studies must be addressed in the near future in order to have full control on the detector performance. The present work represents a necessary step towards those studies.

Acknowledgments

Appendix A. REST as a data analysis framework

Rare Event Searches with TPCs Software (RESTSoft) is a collaborative software effort providing a common framework for acquisition, simulation, and data analysis for gaseous TPCs [25]. REST is composed of a set of libraries written in C++ and is fully integrated in ROOT [26], i.e. all REST classes inherit from TObject and can be read/accessed/written using the ROOT I/O interface. The only structural dependence is related to ROOT libraries, while other packages, as *Geant4* [27] or *Garfield++* [28], can be optionally integrated and used within the REST framework when generating or processing Monte Carlo data.

During the last years, major upgrades took place on the REST core libraries [29]. An important feature introduced in REST is that *metadata* and *event data* are stored together in a unique file. We understand by *metadata* any information required to give meaning to the data registered in the *event data*, as it can be the initial run data taking conditions, the geometry of the detector, the gas properties, or the detector readout topology. Additionally, any input or output parameters, required during the processing or transformation of *event data*, using *event processes*, will be stored as *metadata*. Any previous existing *metadata* structures inside the REST input file will be transferred to any future output, assuring full traceability, as well as reproducibility of results obtained with a particular dataset.

An ambitious feature of REST is its capability to analyze Monte Carlo and experimental data using common *event processes*. This is possible by using existing REST *event processes* to condition the input data generated, for example, by a *Geant4* Monte Carlo simulation. After an appropriate *event data* conditioning, our Monte Carlo generated event will reproduce the *rawdata* of the detector (as it is shown in appendix C). Once we are at that stage, we can benefit from using the same *event processes* to analyze Monte Carlo and real experimental data. A realistic Monte Carlo *rawdata* reconstruction will allow us to assess, validate and optimize the processes that will be involved in the real event reconstruction and analysis even before the start of the physics run of the experiment.

In the following subsections, we recall the definitions of the different components of REST, *viz event types*, *event processes*, and *analysis tree*. These definitions will serve as a reference for the article. Note that we do this having in mind the case where the physics variables of interest are local energy deposits, called hits, and the signal is digitized by a sampling ADC The REST software is versatile enough, though, to handle many other cases. We include here only those components of REST that are relevant to our study.

A.1. Event types

REST defines basic *event types*, or data structures, commonly used to store *event data* generated during the data acquisition (DAQ), production and/or *event data* processing. The definition of a reduced set of basic *event types* allows for better inter-process connectivity. The following data structures defined in REST are used in our study.

- *TRestHitsEvent:* This structure contains an arbitrary number of elements or hits used to store a physical variable, ϕ_i, defined in a three-dimensional coordinate system (x_i, y_i, z_i). In our study, we use this *event type* to describe the energy depositions on the volume of the detector.
- *TRestG4Event:* It is a natural extension of *TRestHitsEvent* containing additional information gathered during the *Geant4* simulation, such as the physical interaction associated to the hit, and the volume of the simulation geometry where the interaction took place.

- *TRestTrackEvent:* A more sophisticated structure where hits are grouped into tracks, following e.g. a proximity criterion. Tracks can, in turn, be grouped into higher level structures. An *identification number* and a *parent number* are assigned to each object and allow to keep the record of its genealogy. A track that is not associated to a parent track will be denominated as an *origin track*, while a track with no daughter tracks will be denominated as a *top-level track*. Therefore, in this scheme we may distinguish different track levels related to the number of track generations.
- *TRestRawSignalEvent:* Stores the digitized signal samples, in the shape of fixed size arrays. Each array corresponds to one front-end electronics channel, and is assigned a logical *signalId*. The set of *signalId*'s is mapped to the geometry of the detection with help of a *readout metadata* table, typically known as *decoding*. The arrays generally describe the time evolution of the signal.
- *TRestTimeSignalEvent:* It contains an arbitrary number of non-fixed size arrays that define a physical quantity, ϕ_i, at arbitrary times, t_i, expressed in physical time units. Each array dataset may contain a different number of data points. This kind of *event type* can be exploited, e.g. for experimental data reduction or storage of timed signals produced in Monte Carlo generated data. As in the case of a *TRestRawSignalEvent*, each array is also identified using a *signalId* associated to the electronic channels described inside the *readout metadata* description.

We stress again that the *event types* defined in REST aspire to be as generic and abstract as possible. It is the role of *processes*, and other *metadata* information used during the data processing, to provide a physical meaning, or final interpretation, to the information that is contained in our *event type*. In the scenario where certain specialization is required, as it is the case of a *TRestG4Event*, an existing REST *event process* will be capable to transform this specialized, or dedicated, *event type* into an existing and more basic one, as it is a *TRestHitsEvent*.

The *event types* described are not only containers to store *event data* in REST, but they implement methods associated to the nature of each data structure, as e.g. spatial rotation or translation of coordinates in a *TRestHitsEvent* or time signal processing methods in a *TRestRawSignalEvent*, as e.g. differentiation, smoothing or Fast Fourier Transform (FFT) methods, together with prototyped methods, i.e. common to all the *event types*, for visualizing and printing the contents of a particular stored event.

A.2. Event processes

REST offers a large variety of *event processes* operating on the basic *event types* in order to manipulate the *event data*, transform it and extract relevant information in the intermediate steps taking place during a particular *event data* processing chain. An *event process* in REST is modular, and it can be integrated in REST by fixing only its input and output *event types*. In other words, any *event process* that participates in a data processing chain must satisfy that its input *event type* corresponds with the output *event type* of the previous *event process*, although there are exceptions as will be described in appendix A.3.

Therefore, *event processes* can be classified as processes that transform an *event type* into another, i.e. *conversion processes*, and those which operate in a particular *event type*, i.e. *hit, raw signal* or *track processes*, in which the output *event type* remains unchanged (although its content—the *event data*—will usually be transformed). In the following sub-sections we provide a description of the different processes involved in our data processing. A full,

detailed and up-to-date list of documented processes will be available at the REST API Documentation[20] for further reference.

A.2.1. Conversion processes.

- *TRestG4ToHitsProcess:* A simple process to transform a *TRestG4Event* into a *TRestHitsEvent*, visualized in figure A1(a). Any specific hits information related to *Geant4* will be lost after this step. Then, the only way in REST to have this information available at the end of the data processing chain will be through the *analysis tree*, described in appendix A.3.

- *TRestHitsToSignalProcess:* This process produces a time projection of the primary electron positions in the TPC volume into the readout plane of the detector, producing the result shown in figure A1(c). The coordinates of hits (x_i, y_i, z_i) stored in a *TRestHitsEvent* are transformed into a *TRestTimeSignalEvent*. This process uses the REST *readout metadata* structure (described in detail in appendix B.2) to associate each hit coordinates (x_i, y_i) with a detector *readout channel*[21], and perform a spatial to time conversion of the z_i coordinate using the readout plane position and the drift velocity of electrons in the gas. The gas properties can be introduced directly as an input metadata parameter, or extracted from a specific metadata class, named *TRestGas*, that interfaces with *Garfield++* to obtain the properties of any gas mixture as calculated by *Magboltz*. A sampling time, δt, can be provided optionally to discretize the resulting time values at the output *TRestTimeSignalEvent*.

- *TRestSignalToHitsProcess:* The inverse process of a *TRestHitsToSignalProcess*, producing the result shown in figure A1(d). We recover, or reconstruct, the hit coordinates using the time information inside a *TRestTimeSignalEvent* and the gas properties. The reconstructed coordinates are obtained using the *readout metadata* description and the corresponding *readout channel* number, associated to each *signalId*.

- *TRestSignalToRawSignalProcess:* This process takes as input a *TRestTimeSignalEvent* type and samples its contents into a fixed data array compatible with the *TRestRawSignalEvent* output type, i.e. fixing the number of data points and sampling time, δt, which are provided as an input parameter to this process. A reference time, or *trigger definition*, is used to define the physical time corresponding to the first sample of the resulting data array. Figure A2 describes one of the trigger methods available on this process which we use later in this study.

- *TRestHitsToTrackProcess:* This process takes as input a *TRestHitsEvent* type and transforms it into a *TRestTrackEvent*. The hit inter-distances inside the input event are evaluated, and those which fall within a certain distance will be assigned to the same identifiable *track*. This distance is introduced as a metadata input parameter to the process, named *cluster distance*. The algorithm starts creating a first track by adding an arbitrary and unassociated hit[22]. The *track* definition will only end when no more unassociated hits are found to satisfy the inter-distance relation with—and this is important—*any* of the hits already added to the *track*, i.e. the inter-distances are evaluated recursively. If after ending the definition of the first track there are still unassociated hits, the process continues to define a second track, and successively. The process ends when no more unassociated hits are found inside the *TRestHitsEvent*.

---

[20] See https://sultan.unizar.es/rest/ for further reference.
[21] As an argument to support this description we assume here that the field drifting the electrons is defined along the z-axis, although the readout plane can be placed in any arbitrary orientation.
[22] A hit that does not belong yet to any *track*.
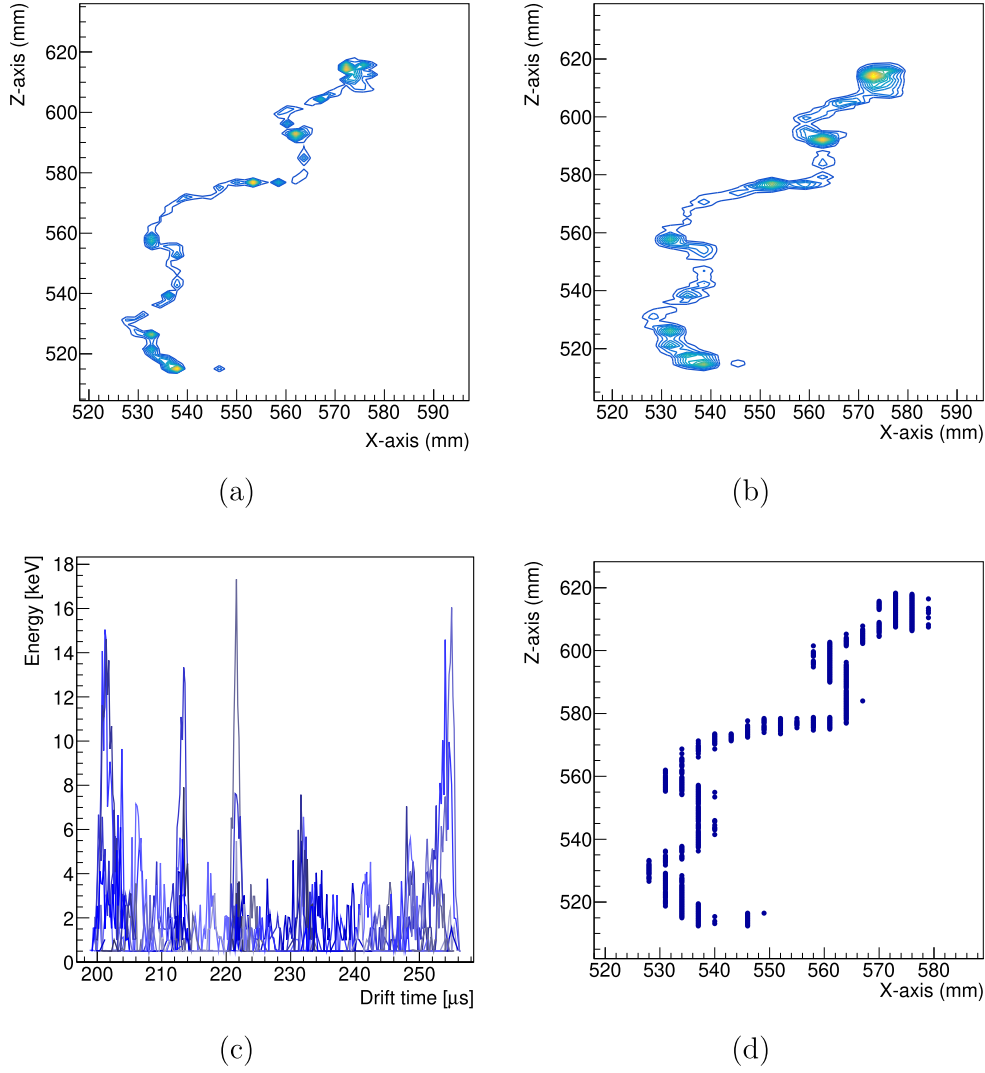
(a)

(b)

(c)

(d)

Figure A1. Representation of different event type outputs after different *event processes* on the simulation of a ^{136}Xe $0\nu\beta\beta$ event. (a) Top-left panel visualizes a *TRestG4Event* type, and represents the *XZ*-plane projection of the charge density of the three-dimensional event. (b) Top-right panel shows a *TRestHitsEvent* event type after applying the *TRestElectronDiffusionProcess* where we can appreciate the effect of electron diffusion in a xenon + 1%TMA gas mixture at 10 bar, the readout plane is placed at $z = 990$ mm. (c) Bottom-left panel shows a *TRestTimeSignalEvent* type after the conversion by a *TRestHitsToSignalProcess*. (d) Bottom-right panel shows a *TRestHitsEvent* type after the reconstruction using *TRestSignalToHitsProcess*. A scatter plot is used in this case to emphasize the effect introduced by the 3 mm pitch detector readout that can be observed along the *x*-axis, and the 200 ns sampling rate of the electronics on the *z*-axis. At this stage the *TRestHitsEvent* is not anymore a three-dimensional event, since we used here the PandaX-III stripped readout described in appendix B.2.
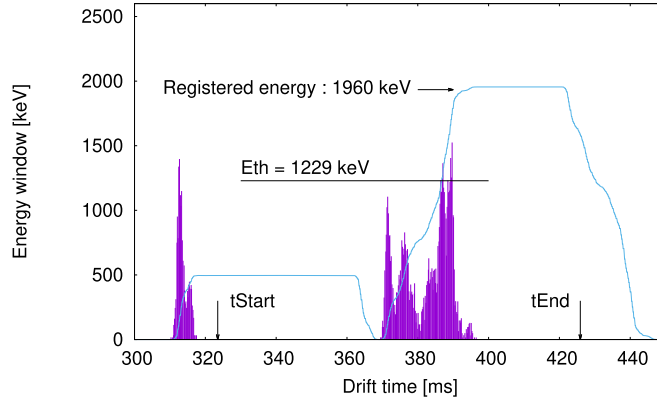
Figure A2. A figure illustrating the trigger definition using the threshold method implemented in *TRestSignalToRawSignalProcess* for a ^{136}Xe $0\nu\beta\beta$ event. The filled curve, in red, represents the charge distribution (in arbitrary units not represented on this plot) as a function of the electron drift time. The blue curve is constructed with the integral of the charges (with magnitude represented on the *y*-axis of this plot) in a fixed time width, here of about $\sim$50 μs, and corresponding to the half size of the acquisition window of the electronics, of about $\sim$100 μs. When the integration exceeds a certain threshold E_{th}, in this case equal to $Q_{\beta\beta}/2$ of ^{136}Xe, the center of the acquisition window is fixed. The resulting acquisition window is represented by t_{Start} and t_{End}. An additional offset is introduced on the definition of t_{Start} to assure few time bins will be available for *baseline* definition during the *raw signal* event processing.

A.2.2. Hit processes.

- *TRestElectronDiffusionProcess:* This process uses the longitudinal and transverse coefficients of a particular gas mixture to emulate the relative deviation of electrons from their original positions in *TRestHitsEvent*, presumably produced by primary ionization. The energy of each hit found inside the input *TRestHitsEvent* is converted to the corresponding number of primary electrons that would be produced in the ionization process. Each electron will be a new hit in the output *TRestHitsEvent* structure, and their coordinates will be randomly deviated following a Gaussian distribution related to the gas parameters, longitudinal and transverse diffusion coefficients, and the distance to the readout plane, which is given as metadata input. The result of applying this process is shown in figure A1(b). This process may optionally add the possibility to include the effect of electron attachment.

- *TRestHitsSmearingProcess:* This process smears the energy of the input *TRestHitsEvent* according to a Gaussian distribution described by parameters given as metadata input.

A.2.3. Track processes for $0\nu\beta\beta$ topology.

- *TRestTrackReductionProcess:* This process reduces the number of hits in each of the tracks stored in a *TRestTrackEvent*, as seen in figure A3(a). An input metadata parameter, N_{max}, specifies the maximum number of hits on each of the reduced tracks at the output *TRestTrackEvent*. The closest hits are merged iteratively till the N_{max} condition is satisfied. When *two* hits are merged, the resulting hit position is calculated weighting the energy of each merged hit. The input *parent tracks* are also stored in the output *TRestTrackEvent*, together with the reduced tracks which acquire a relation of inheritance.
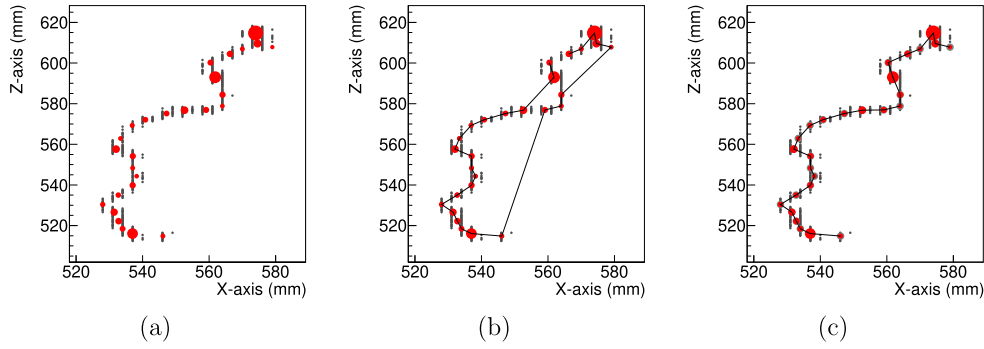
(a)   (b)   (c)

Figure A3. A *TRestTrackEvent* representation of a ^{136}Xe $0\nu\beta\beta$ event after the different track processes used for physical track identification. This event corresponds to the same event used in figure A1. (a) An image of the hit reduction produced by *TRestTrackReductionProcess*. The red circles represent the final position of reduced hits, which size corresponds with their energy value. The small gray circles on the background represent the hits of the *parent track* used as input. (b) A polyline is added to this representation to visualize the hits inter-connectivity after the *TRestTrackPath-MinimizationProcess*. If path minimization works on the whole, it produces at times obviously unphysical connections, as our example illustrates. (c) The unphysical connections are corrected using *TRestTrackReconnectionProcess*.

- *TRestTrackPathMinimizationProcess:* This process operates only on each of the *top-level tracks* found in *TRestTrackEvent* and—using graph theory—finds the shortest path that connects all the hits within each track, as seen in figure A3(b). We have integrated in this process a *Held-Karp* algorithm [30] optimized for problems that contain between 25 and 35 nodes. This algorithm was extracted from the *Concorde* travel sales problem libraries [31]. It is important to notice that the minimization is performed in a closed loop, i.e. the extremes of the physical track are not properly identified at the end of this process. From now on, we will denominate two consecutive hits as *connected hits*.
- *TRestTrackReconnectionProcess:* This process will be commonly used in combination with *TRestTrackPathMinimizationProcess*. This process will detect unphysical path connections between the hits, or nodes. We understand by unphysical connection a distance between nodes that is larger than the average by a certain number of sigmas, or where no hits from the *origin track* are found in between, as seen in figure A3(c). The track is split at those unphysical connections and reconnected with the closest hit. Using this technique, the extremes of the physical track are found naturally. Although it was not used in our analysis, this process allows to identify secondary track branches of complex event tracks when reconnection is not possible.

A.2.4. Raw signal processes.

- *TRestRawSignalShapingProcess:* This process realizes the simulation of the acquisition electronics signal shaper in a *TRestRawSignalEvent*. It implements the convolution of an analytical response function with the input *TRestRawSignalEvent*, presumably containing the original charge distribution produced in the detector. The analytical response function requires a single parameter, which is the *shaping time*. However, this process offers the possibility to introduce any arbitrary wavefunction as response function, adding realism to the generated output.

- *TRestRawSignalAddNoiseProcess:* It is used to include the effect of electronic noise and emulate the fluctuations on the acquired time signal. A value is introduced as input metadata to define the amplitude of the noise. We use a basic noise method implemented in this process that assigns an independent random value, Gaussian distributed, to each of the bins inside a *TRestRawSignalEvent*.

A.3. The analysis tree

The different *event processes* can be combined in sequence to build a more complex processing chain. In REST, only the *event data* produced in the last *event process* will be saved to the output file[23]. There are two main reasons why it is not desirable to stock intermediate *event data* into a single data file[24]. First, we risk to pile-up, or replicate unnecessarily, data that might require to allocate a non-negligible amount of disk space. Second, the traceability of changes introduced in the data might be lost by storing data at different *levels* of processing inside a unique file. This possibility would add an undesirable degree of complexity to the future tracking of the data processing, if for example, we decide to continue processing data using a previously processed file with *event data* at different processing levels but we start our *event data* processing at an intermediate state.

However, the inconvenience of this scheme resides on the fact that, during the processing, the *event data* is transformed in a way that information might be lost, as when we transform a *TRestGeant4Event* into a *TRestHitsEvent*. Therefore, it will be not available at the end of the *event data* processing. The *analysis tree* is a REST component that emerges to solve this problem. It is based on a ROOT TTree structure, and it is accessible at any stage of the *event data* processing by any *event process*. Any process in REST is allowed to create a new entry, or variable, inside the *analysis tree* which will be always available in any REST file. Future processing may add new variables to the *analysis tree* preserving the existing ones from previous processing levels. Therefore, the *analysis tree* is used to gather relevant information along the processing chain that might not be anymore available at the *event data* output of the last process.

In this context, and just for classification purposes, we can distinguish two new types of *event processes*, the *analysis processes* and the *pure analysis processes*. Both type of processes have in common that they do not modify or transform the input *event data*, and are just dedicated to add new entries to the *analysis tree*. Furthermore, the *pure analysis processes* will not even require access to the *event data*, being these processes the only ones that can be connected at any place of the processing chain without input/output *event type* restrictions. For example, a *pure analysis process* might be a process that reads a variable (or branch) in the *analysis tree* to perform a multi-peak fit, and write the results of the fit in a *metadata* structure.

Appendix B. PandaX-III Monte Carlo event simulation in REST

The first step of Monte Carlo simulation is to consider PandaX-III as a simple calorimeter and simulate the energy deposition of particles interacting in the gaseous medium, or active

---

[23] However, other *metadata* information used in the processing chain, e.g. process parameters, simulation conditions, readout definition, or gas properties, will be stored without exception, including previous historical *metadata* information.
[24] Although no restrictions apply to produce intermediate files with a snapshot of the *event data* at a particular step of the processing chain.

volume of the TPC. The calorimetric response of the conceptual PandaX-III design was studied extensively in [9] using the *Geant4*-based simulation packages *BambooMC* and *RestG4*. In *Geant4*, detailed tracks of the ionizing particles in the TPC active volume are generated. As a particle travels in the gas medium, it deposits energy along the trajectory via multiple scattering and other physical processes. *Geant4* tracks the trajectory in pre-defined steps. In each step, information such as timestamp, particle type, momentum, energy deposition, position, and physical process involved in the interaction are registered.

In our study, we use the *RestG4* package as an interface to define the simulation conditions through a REST metadata structure named *TRestG4Metadata*, and to store all event information in a *TRestG4Event* type that can be further processed inside REST, i.e. *RestG4* serves to generate a first event dataset. Event tracks can be reconstructed using this original dataset, and we call those *MC-true tracks*. The *MC-true tracks* have been generated in *Geant4* with a spatial precision of 0.2 mm, implying that every step of the simulation of the movement of a particle is no more than 0.2 mm, with particle information being recorded after each step. The *G4EmLivermorePhysics* physics list—describing the electromagnetic processes in *Geant4*—was used in the event generation.

The main features of the PandaX-III TPC are faithfully represented in our detector geometry. The cylindrical TPC is 1.5 m in diameter and 2 m in length. The cathode in the middle divides the TPC into two symmetrical drift volumes, in which electrons are drifted away from the cathode and collected by the charge readout planes located at both ends. When running at 10 bar pressure the TPC can hold 200 kg xenon gas. The active volume is contained by a 3 cm thick copper vessel, and placed in a water shielding tank in the laboratory. The geometry description, written in GDML [32], is the same as the one used at [9]. Therefore, we refer to that publication for further details on the different detector components implemented, materials used, and drawings.

B.1. Data processing chain for PandaX-III

The generated *MC-true tracks* are introduced in the REST processing scheme to achieve a realistic detector response, including electron drift diffusion, energy resolution, charge readout segmentation, and signal sampling.

Figure B1 shows the complete data chain, or *event data* flow, used to process *Geant4* Monte Carlo generated data, and includes different effects related to the detector response. Obvious *event type* conversion processes detailed in appendix A.2.1, and analysis processes described later on in this chapter, have been omitted in this drawing in order to facilitate focusing on the key aspects of the chain. In our data chain we can differentiate up to three different phases, or stages. In the *first phase* the *event data* is conditioned to take into account the physical response of the electrons drifting in the gas medium, in the *second phase* the readout topology and electronics sampling is considered, and finally in the *third phase* a physical track[25] reconstruction takes place to condition the data and make it suitable for a topological track study. The description of *event processes* required for a full understanding of the processing chain is detailed in appendix A.2.

The *first phase* of the processing takes place exclusively at the *TRestHitsEvent* level. The input generated by *RestG4*, encapsulated inside a *TRestG4Event* type, is transformed into a *TRestHitsEvent* type using *TRestG4ToHitsProcess*. Two processes are responsible to include the physical detector response, we use *TRestElectronDiffusionProcess* to emulate the electron

---

[25] We will use the term *physical track* to refer to the final state of the event reconstruction, where the reconstructed track has physical meaning.
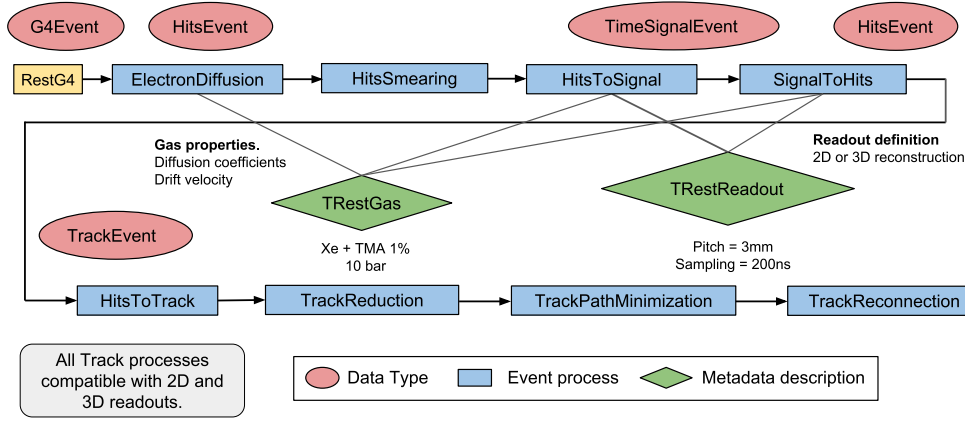
Figure B1. Data processing chain used to manipulate the Monte Carlo events as described in the text. This figure gives an overview of the data flow through the successive *event types* (red ovals). Few *event processes* (blue rectangles) have been omitted, for simplicity's sake. The chain starts with *Geant4* generated data (yellow rectangle at top-left). The flow is from left to right. We can distinguish two rows of *event processes*. The top row (*TRestElectronDiffusionProcess*, *TRestHitsToSignalProcess*, etc) represent processes used to include the detector response (or *first phase*) and event reconstruction (or *second phase*), while the bottom row represents the track processes used for *physical track* reconstruction (or *third phase*). Just in this figure, the *TRest* prefix and *Process* termination has been removed from the REST process and event names.

diffusion in the gaseous medium, and *TRestHitsSmearingProcess* to include the stochastic effect of the detector energy resolution on each independent event. The gas properties used in *TRestElectronDiffusionProcess* were obtained from *Magboltz* [33] through the *Garfield++*[28] interface integrated in *TRestGas*. In our study we used a gas mixture of xenon+1% TMA at 10bar pressure, setting the TPC drift field to $1\,\mathrm{kV\,cm^{-1}}$ leads to an electron drift velocity of $1.86\,\mathrm{mm\,\mu s^{-1}}$, and $1.46 \times 10^{-2}\,\mathrm{cm^{1/2}}$ and $1.01 \times 10^{-2}\,\mathrm{cm^{1/2}}$, for the longitudinal and transverse electron diffusion coefficients, respectively. The number of primary electrons yielded at each deposition is correlated with the energy deposited according to the work-function, or *W*-value, i.e. the energy required to extract an electron from an atom. We use a *W*-value of 21.9 eV [34] for our gas mixture. The detector energy resolution introduced in *TRestHitsSmearingProcess* was defined as 3%-FWHM at $Q_{\beta\beta} = 2\,457.83\,\mathrm{keV}$.

During the *second phase* the conversion processes *TRestHitsToSignalProcess* and *TRestSignalToHitsProcess* are responsible to introduce the detector granularity and electronics sampling of the time signal induced on the detector readout. These processes are strongly dependent on the *TRestReadout* definition, detailed in appendix B.2. For now, it is enough to mention that *TRestHitsToSignalProcess* discretizes our *event data* into time signals in steps of 200 ns, using a two-dimensional readout of 3 mm pitch. At this stage, other details related to time signal conditioning could be added, despite the fact that it is not the aim of this work to include those effects, and any future improvement will be later discussed in our final conclusions. Then, *TRestSignalToHitsProcess* reconstructs the event recovering the *TRestHitsEvent* structure including the effects introduced by the previous process, as it is shown in figure A1(d).

In the final *third phase*, the reconstructed *TRestHitsEvent* is classified into tracks through the *TRestHitsToTrackProcess* using a *cluster distance* value of 7.5 mm, i.e. two tracks are considered to be independent if the closest hits distance from one track to another is above 7.5 mm. In this phase the processes described in appendix A.2.3 are used to determine the *physical track*, evaluating the hit distances and connectivity. Figure A3 shows the result of the three *track processes* implemented in this phase of the processing chain; *TRestTrackReductionProcess* (where we use $N_{\max} = 35$), *TRestTrackPathMinimizationProcess* and *TRestTrackReconnectionProcess*. At the end of this processing chain we end up with a reconstructed *physical track* where track ends can be recognized naturally. It is the responsibility of the *analysis processes* to extract information, or observables, along the data processing chain, to be used in a final pattern recognition analysis.

It is important to remark that the *event data* processing chain discussed here is fully compatible with two and three-dimensional event representation, i.e. a *TRestHitsEvent* structure is able to identify if it contains 2D or 3D hits, and the *hit* and *track processes* will act on the *event data* accordingly, e.g. *TRestTrackPathMinimizationProcess* will independently minimize the paths of different projections and tracks produced by a two-dimensional readout. Ultimately, it is the sole responsibility of the *TRestReadout* metadata structure to define the topological nature of the *event data*.

B.2. PandaX-III readout metadata description

For charge readout, PandaX-III baseline design relies on Microbulk Micromegas technology due to its good intrinsic radiopurity levels, good energy resolution and capability to operate at high pressure [12, 13]. A modular design has been conceived to build each detector *readout plane* in the PandaX-III TPC design due to the Microbulk size limitation imposed by its fabrication process[26]. In order to cover the full active area of the detector, each *readout plane* consists of 41 independent Micromegas modules, as shown in figure B2(a). The total number of acquisition channels is limited to a manageable level [27] by physically interconnecting pixels along the horizontal or vertical axis orientations and they are read out as a single *readout channel*, or strip, as shown in figure B2(b). Each Micromegas module contains 64 *X*-strips and 64 *Y*-strips of 3 mm pitch.

The integration of a realistic complex readout scheme in REST is done through the *TRestReadout* metadata structure. *TRestReadout* is a sophisticated structure that defines an arbitrary number of *readout planes*, containing itself any number of *readout modules* composed of *readout channels* that are identified with the DAQ channels registered by the electronics. A *readout channel* itself is built of one or more *readout pixels*. A *readout pixel* is the most elementary component of the readout description in REST, and it defines its position relative to the module coordinates, its size, orientation, and shape. The combination of *readout pixels* with different sizes, orientations, and shapes at different positions allows to construct any desired readout topology in REST.

Different methods are available at *TRestReadout* and related classes to access the readout description, and determine for a given hit coordinates which is the corresponding *readout channel* in an efficient way. These methods are accessed by related *event processes* and they are exploited to translate a given *TRestHitsEvent* coordinates into *TRestTimeSignalEvent* channels, and vice versa. The *TRestReadout* definition is a crucial element in the construction

---

[26] Mainly due to the existing equipment used for production.

[27] Readout channel reduction plays an important role to simplify the design, reduce costs and minimize the impact of typically non-clean electronics from the radiopurity point of view, on a detector component that must be placed as close as possible to the readout plane to minimize electronic noise induction on the readout signal.
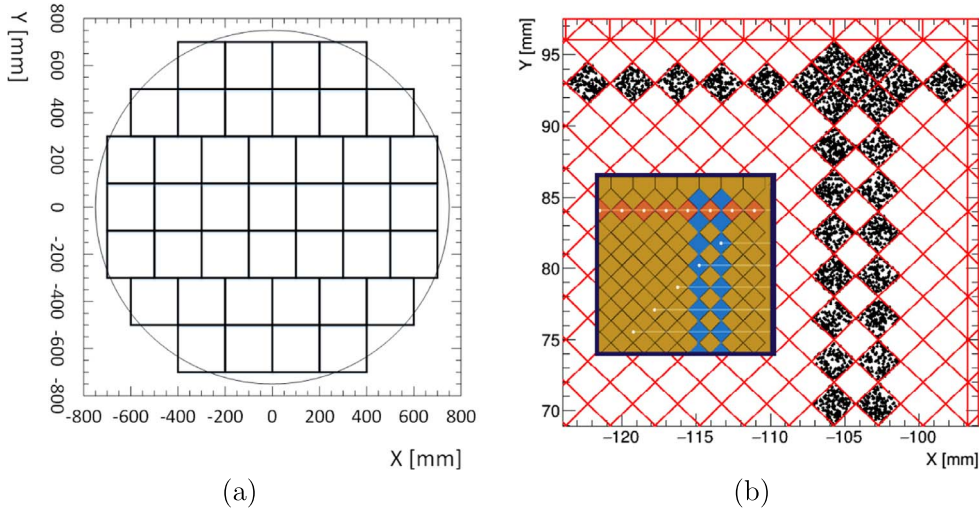
(a)   (b)

Figure B2. (a) The PandaX-III readout plane description in REST, built using 41 Micromegas modules of size 192 mm × 192 mm. The circle of size $R = 75$ cm defines the inner radius where the ^{136}Xe gas is contained. The active detector area is formed by the region of the modules found inside that circle. Small corners inside this circular region are not covered by any readout module, and therefore, they will translate into a signal efficiency loss in our final detector response. (b) The detail of a Micromegas readout module region as implemented in REST, red lines define the limits of *readout pixels*. Two *readout channels* on the *X*-axis and one *readout channel* on the *Y*-axis are drawn (black dots) using REST readout validation routines, which allow to select, or activate, a group of *readout channels* and—after generating a number of random coordinates—paint only those generated coordinates that are found inside the active channels. The small embedded image is the corresponding area extracted from the original *Gerber* design used for fabrication. The *Gerber* representation has been modified to identify the two *X*-axis (blue pixels) and one *Y*-axis (orange pixels) *readout channels* active for the readout validation.

of the data processing chain and its generic implementation provides versatility to study a common dataset with different detector readout topologies and granularities, as it is done in section 3.

In particular, the two-dimensional stripped baseline readout for PandaX-III cannot perform an unequivocal three-dimensional reconstruction of the original event topology (see discussion at [24]). In our design, a drifted electron that reaches a Micromegas module will induce a signal on either a *X*-strip or a *Y*-strip. Therefore, the *TRestHitsEvent* reconstruction after *TRestSignalToHitsProcess* is necessarily a combination of two-projections, a projection formed by *XZ*-hits (i.e. hits with valid *X* and *Z* coordinates and undefined *Y* coordinate) derived from signals corresponding to the *X*-strips, and a projection formed by *YZ*-hits resulting from the signals found at *Y*-strips. Figure B3 shows the resulting projections for a signal and a background event. End-track identification using the *analysis processes* described in the following subsection have been also represented in this figure. The main characteristics of signal and background events is discussed in section 2.1.
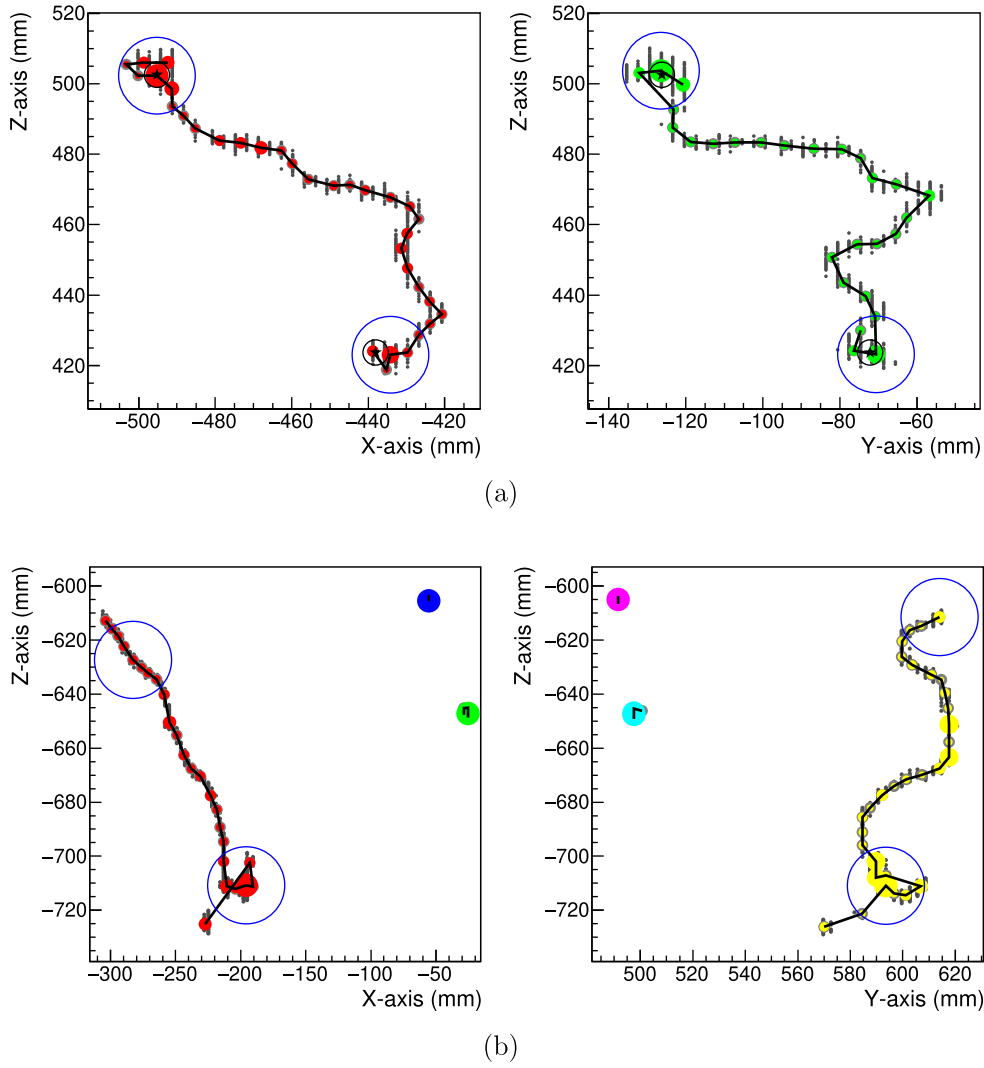
(a)

(b)

Figure B3. A *TRestTrackEvent* representation of the two-dimensional projected electron tracks for a $0\nu\beta\beta$ (a) and a background event produced in the ^{238}U decay chain (b) at the end of the data processing chain. The colored filled circles correspond to the *top-level track* of a *TRestTrackEvent*, while the gray scattered points in the background —where the 3 mm pitch granularity can be discerned—represent the hits in the *origin track*. The size of the circle in the *top-level tracks* is proportional to the hit energy, and different colors serve to identify independent tracks classified after the *TRestHitsTo-TrackProcess*. A black polyline shows the *top-level track* hits interconnectivity resulting after the *TRestTrackReconnectionProcess*, and helps to visualize the *physical track*. Large *blue circles* correspond with our *blob charge* definition ($R = 12$ mm) obtained from *TRestFindTrackBlobsProcess*, described in appendix B.3. The *starred black markers* found in the $0\nu\beta\beta$ event correspond with the positions obtained from the *TRestFindG4BlobProcess* representing the last electron energy deposition for each of the primary simulated electron tracks.

B.3. Analysis event processes

Additional *event processes* are used during the data processing chain in order to extract required event information in our posterior analysis. The *analysis processes* used, the place of the processing chain where those processes are used, the particular information extracted and/ or other parameters required are detailed in the following list.

- *TRestFindG4BlobProcess:* It is used at the beginning of the *first phase* of the data processing chain. It operates in a *TRestG4Event* and it is used to extract the *MC-true track* coordinates of each electron track end in a $0\nu\beta\beta$ event. These values are used to assess the goodness of our algorithms to identify the track ends after the physical event reconstruction.
- *TRestG4AnalysisProcess:* It is also used at the beginning of the data chain, after the previous process. This process is used to add different observables related with the *TRestG4Event* to the *analysis tree*, as e.g. the interaction types involved, or the number and type of particles involved in the event. In our particular case we use this process to extract the total energy deposited in the active volume of the detector, without any fiducialization.
- *TRestTriggerAnalysisProcess:* It is used at the *second phase* of the processing chain, at the *TRestTimeSignalEvent* level. This process uses the same trigger definition implemented in *TRestSignalToRawSignalProcess* shown previously in figure A2. We apply a sampling rate of 5 MHz with a total number of 512 sampled points, as fixed by the PandaX-III electronics system, and a trigger energy threshold, E_{th}, of 1229 keV corresponding to $Q_{\beta\beta}/2$. This process generates a new observable in the *analysis tree*, defining the energy contained in the virtual acquisition window defined. It must be noted that this process, as any other *analysis process*, does not modify the *event data*.
- *TRestTrackAnalysisProcess:* It is used at the end of the *third phase* of the processing chain. It extracts parameters used for pattern recognition of $0\nu\beta\beta$ events, such as the number of tracks, the track energy ratio, or the twist parameter described in section 2.1.
- *TRestFindTrackBlobsProcess:* It is also used at the end of the processing chain. The tracks ends have been naturally identified after the *TRestTrackReconnectionProcess*, and this process registers the coordinates or positions of the end-tracks in the *analysis tree*. This process searches for the highest density region within a 20% of the track length at the track end, in order to allow an additional degree of freedom identifying the final blob position. This process is used to generate the observables related to the *blobs charge* parameter, described in section 2.1.

Appendix C. Full Monte Carlo detector response

In order to test the detector's capability to distinguish surface events we need to extract a new parameter connected with the diffusion of electrons, from the *event data* processing chain presented in appendix B.1. We could extract such a parameter at different stages in that chain, however, a straightforward parameter characterizing the spread of the charge is the width of the pulses as recorded by the readout electronics. We require then additional event processing to accommodate our Monte Carlo generated events to a time signal similar to the recorded with the detector acquisition. Figure C1 shows the additional *event processes*, detailed also in appendix A.2, that we have introduced after *TRestHitsToSignalProcess*, at the *second phase* of the data chain described in appendix B.1.
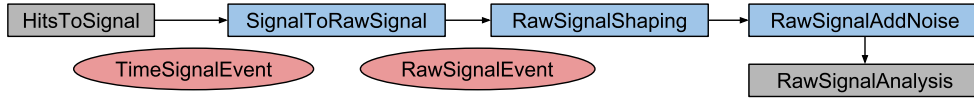
J. Phys. G: Nucl. Part. Phys. **47** (2020) 045108

J Galan *et al*

Figure C1. Extension of the event data chain to include appropriate detector readout signal conditioning, *trigger definition*, *electronics shaping* and *electronic noise*. The first process *TRestHitsToSignalProcess*—inside the first gray box—corresponds to the process shown previously in figure B1. The event processing ends with the *TRestRawSignalAnalysisProcess* after we extract the parameters of interest for the fiducial analysis carried out in section 4. The *TRest* prefix and *Process* termination has been removed from the REST process and event names, keeping the same nomenclature as in figure B1.
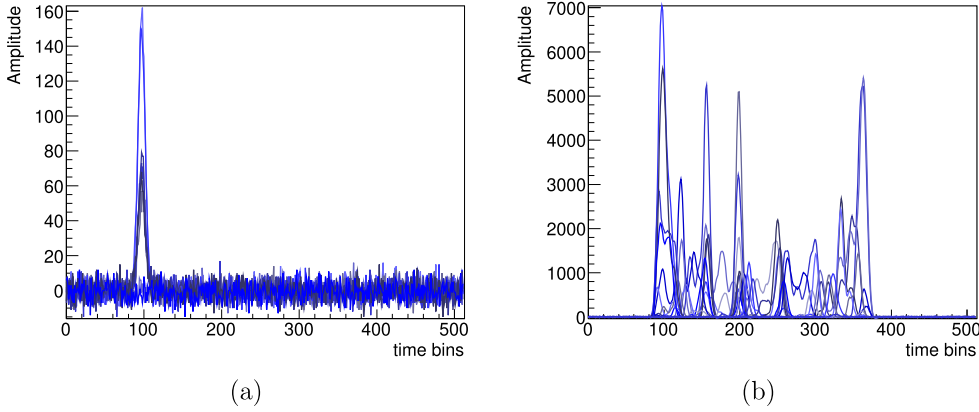

Figure C2. A *TRestRawSignalEvent* representation after applying the *event processes* present on the signal conditioning data chain, including trigger definition, electronics shaping and electronic noise contributions. (a) The readout signals for a 10 keV electron generated in the middle of our TPC volume, at $z = 500$ mm. (b) The readout signals for a $0\nu\beta\beta$ signal event.

In this extension of the processing chain we transform our *event data* into a *TRestRawSignalEvent* where we introduce different effects on the detector response, as a semi-Gaussian electronic signal shaper with 1 μs peaking time and a Gaussian noise level, at each channel. In order to adjust the noise level we have simulated a small energetic deposition, produced by a 10 keV electron, and we have chosen the noise level that approximately reproduces our signal-to-noise values when taking data in normal laboratory conditions. The unit values at this stage are completely arbitrary. Figure C2 shows the aspect of the final processed event after including all these effects.

These additional effects are included at the *TRestRawSignalEvent* level to gain realism, but at the same time to exploit the REST analysis process, *TRestRawSignalAnalysisProcess*, which is also used to analyze the real data of the detector. This analysis process is used right after *TRestRawSignalAddNoiseProcess* to extract the diffusion parameter that we are interested in. In order to define our diffusion parameter, σ_w, we pre-select those channels exceeding a certain energy threshold measured in signal amplitude at the *TRestRawSignalEvent*. The threshold value is defined as the 10% of the maximum amplitude found inside the *TRestRawSignalEvent*. Then, we obtain the FWHM of the maximum peak at each channel, and average the 9 channels which

produce the lowest FWHM values[28]. The value of σ_w is measured in samples, or bins, of 200 ns width which is the sampling value we used in the detector response.

ORCID iDs

J Galan ⓘ https://orcid.org/0000-0001-7529-9834
C Fu ⓘ https://orcid.org/0000-0002-0031-2077

References

[1] Fukuda Y *et al* 1998 Evidence for oscillation of atmospheric neutrinos *Phys. Rev. Lett.* **81** 1562–7
[2] Ahmad Q R *et al* 2002 Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury neutrino observatory *Phys. Rev. Lett.* **89** 011301
[3] Majorana E and Majorana E 1937 *Nuovo Cimento* **14** 171
[4] Barabash A S 2010 75 years of double beta decay: yesterday, today and tomorrow *DAE Symp. Nucl. Phys.* **55** I13
[5] Dell'Oro S, Marcocci S, Viel M and Vissani F 2016 Neutrinoless double beta decay: 2015 review *Adv. High Energy Phys.* **2016** 2162659
[6] Schechter J and Valle J W F 1982 Neutrinoless double beta decay in SU(2) x U(1) theories *Phys. Rev.* D **25** 2951
Schechter J and Valle J W F 1981 *Phys. Rev.* D **25** 289
[7] Avignone F T III, Elliott S R and Engel J 2008 Double beta decay, Majorana neutrinos, and neutrino mass *Rev. Mod. Phys.* **80** 481
[8] Agostini M, Benato G and Detwiler J 2017 Discovery probability of next-generation neutrinoless double-beta decay experiments *Phys. Rev.* D **96** 053001
[9] Chen X *et al* 2017 PandaX-III: searching for neutrinoless double beta decay with high pressure 136 Xe gas time projection chambers *Sci. China Phys. Mech. Astron.* **60** 061011
[10] Li J, Ji X, Haxton W and Wang J S Y 2015 The second-phase development of the China JinPing underground laboratory *13th Int. Conf. on Topics in Astroparticle and Underground Physics, TAUP 2013.; Phys. Proc.* **61** 576–85
[11] Redshaw M, Wingfield E, McDaniel J and Myers E G 2007 Mass and double-beta-decay Q value of ^{136}Xe *Phys. Rev. Lett.* **98** 053003
[12] González-Díaz D *et al* 2015 Accurate γ and MeV-electron track reconstruction with an ultra-low diffusion Xenon/TMA TPC at 10 atm *Nucl. Instrum. Methods Phys. Res.* A **804** 8–24
[13] Lin H *et al* 2018 Design and commissioning of a 600 L time projection chamber with Microbulk Micromegas *J. Instrum.* **13** P06012
[14] Andriamonje S *et al* 2010 Development and performance of Microbulk Micromegas detectors *J. Instrum.* **5** P02001
[15] Vuilleumier J C *et al* 1993 Search for neutrinoless double-β decay in ^{136}Xe with a time projection chamber *Phys. Rev.* D **48** 1009–20
[16] Galan J 2016 Microbulk micromegas for the search of $0\nu\beta\beta$ of ^{136}Xe in the PandaX-III experiment *J. Instrum.* **11** P04024
[17] Cebrián S, Dafni T, Gomez H, Herrera D C, Iguaz F J, Irastorza I G, Luzón G, Segui L and Tomás A 2013 Pattern recognition of ^{136}Xe double beta decay events and background discrimination in a high pressure Xenon TPC *J. Phys. G: Nucl. Part. Phys.* **40** 125203
[18] Ferrario P(The NEXT collaboration) *et al* 2016 First proof of topological signature in the high pressure xenon gas TPC with electroluminescence amplification for the next experiment *J. High Energy Phys.* JHEP01(2016)104
[19] Renner J *et al* 2017 Background rejection in NEXT using deep neural networks *J. Instrum.* **12** T01004
[20] Ponkratenko O A, Tretyak V I and Zdesenko Y G 2000 The Event generator DECAY4 for simulation of double beta processes and decay of radioactive nuclei *Phys. Atom. Nucl.* **63** 1282–7

---

[28] We describe in detail our particular definition of σ_w, although alternative definitions will certainly lead to similar results.

J. Phys. G: Nucl. Part. Phys. **47** (2020) 045108

J Galan *et al*

Ponkratenko O A, Tretyak V I and Zdesenko Y G 2000 The Event generator DECAY4 for simulation of double beta processes and decay of radioactive nuclei *Yad. Fiz.* **63** 1355

[21] Abgrall N *et al* 2016 The Majorana Demonstrator radioassay program *Nucl. Instrum. Methods* A **828** 22–36

[22] Ai P, Wang D, Huang G and Sun X 2018 Three-dimensional convolutional neural networks for neutrinoless double-beta decay signal/background discrimination in high-pressure gaseous time projection chamber *J. Instrum.* **13** P08015

[23] Zeng M, Li T, Cang J, Zeng Z, Fu J, Cheng J, Ma H, Liu Y and Liu Y-N 2017 3-D topological signatures and a new discrimination method for single-electron events and $0\nu\beta\beta$ events in CdZnTe: a Monte Carlo simulation study *Nucl. Instrum. Methods* A **858** 44–52

[24] Qiao H, Lu C Y, Chen X, Han K, Ji X D and Wang S G 2018 Signal-background discrimination with convolutional neural networks in the PandaX-III experiment using mc simulation *Sci. China Phys., Mech. Astron.* **61** 101007

[25] Tomas A 2013 Development of time projection chambers with MICROMEGAS for rare event searches *PhD Thesis* Zaragoza U.

[26] Brun R and Rademakers F 1997 ROOT—an object oriented data analysis framework *Nucl. Instrum. Methods Phys. Res.* A **389** 81–6

[27] Agostinelli S *et al* 2003 GEANT4: a simulation toolkit *Nucl. Instrum. Methods* A **506** 250–303

[28] Pfeiffer D *et al* 2019 Interfacing Geant4, Garfield++ and Degrad for the simulation of gaseous detectors *Nucl. Instrum. Meth.* A **935** 121–34

[29] Galan J 2016 REST v2.0: a data analysis and simulation framework for micro-patterned readout detectors *8th Sym. on Large TPCs for Low-Energy Rare Event Detection*

[30] Held M and Karp R M 1962 A dynamic programming approach to sequencing problems *J. Soc. Ind. Appl. Math.* **10** 196–210

[31] Applegate D L, Bixby R E, Chvatal V and Cook W J 2007 *The Traveling Salesman Problem: A Computational Study* (*Princeton Series in Applied Mathematics*) (Princeton, NJ: Princeton University Press)

[32] Chytracek R, McCormick J, Pokorski W and Santin G 2006 Geometry description markup language for physics simulation and analysis applications *IEEE Trans. Nucl. Sci.* **53** 2892

[33] Biagi S F 1999 *Nucl. Instrum. Meth.* A **421** 234–40

[34] Borges F I G M and Conde C A N 1996 Experimental W-values in gaseous Xe, Kr and Ar for low energy x-rays *Nucl. Instrum. Methods Phys. Res.* A **381** 91–6