Background rejection in NEXT using deep neural networks

NEXT collaboration

J. Renner,[a,1] **A. Farbin,**[b] **J. Muñoz Vidal,**[a] **J.M. Benlloch-Rodríguez,**[a] **A. Botas,**[a] **P. Ferrario,**[a] **J.J. Gómez-Cadenas,**[a,2] **V. Álvarez,**[a] **C.D.R. Azevedo,**[h] **F.I.G. Borges,**[c] **S. Cárcel,**[a] **J.V. Carrión,**[a] **S. Cebrián,**[d] **A. Cervera,**[a] **C.A.N. Conde,**[c] **J. Díaz,**[a] **M. Diesburg,**[p] **R. Esteve,**[f] **L.M.P. Fernandes,**[c] **A.L. Ferreira,**[h] **E.D.C. Freitas,**[c] **A. Goldschmidt,**[e] **D. González-Díaz,**[q] **R.M. Gutiérrez,**[i] **J. Hauptman,**[j] **C.A.O. Henriques,**[c] **J.A. Hernando Morata,**[k] **V. Herrero,**[f] **B. Jones,**[b] **L. Labarga,**[l] **A. Laing,**[a] **P. Lebrun,**[p] **I. Liubarsky,**[a] **N. López-March,**[a] **D. Lorca,**[a] **M. Losada,**[i] **J. Martín-Albo,**[r] **G. Martínez-Lema,**[k] **A. Martínez,**[a] **F. Monrabal,**[a] **C.M.B. Monteiro,**[c] **F.J. Mora,**[f] **L.M. Moutinho,**[h] **M. Nebot-Guinot,**[a] **P. Novella,**[a] **D. Nygren,**[b,2] **B. Palmeiro,**[a] **A. Para,**[p] **J. Pérez,**[l] **M. Querol,**[a] **L. Ripoll,**[n] **J. Rodríguez,**[a] **F.P. Santos,**[c] **J.M.F. dos Santos,**[c] **L. Serra,**[a] **D. Shuman,**[e] **A. Simón,**[a] **C. Sofka,**[o] **M. Sorel,**[a] **J.F. Toledo,**[f] **J. Torrent,**[n] **Z. Tsamalaidze,**[g] **J.F.C.A. Veloso,**[h] **J. White,**[o,3] **R. Webb,**[o] **N. Yahlali,**[a] **H. Yepes-Ramírez**[i]

[a]*Instituto de Física Corpuscular (IFIC), CSIC & Universitat de València,*
Calle Catedrático José Beltrán, 2, Paterna, Valencia, 46980 Spain

[b]*University of Texas at Arlington*
701 S. Nedderman Drive, Arlington, Texas, 76019 USA

[c]*Departamento de Fisica, Universidade de Coimbra,*
Rua Larga, Coimbra, 3004-516 Portugal

[d]*Lab. de Física Nuclear y Astropartículas, Universidad de Zaragoza,*
Calle Pedro Cerbuna, 12, Zaragoza, 50009 Spain

[e]*Lawrence Berkeley National Laboratory (LBNL),*
1 Cyclotron Road, Berkeley, California, 94720 USA

[f]*Instituto de Instrumentación para Imagen Molecular (I3M), Universitat Politècnica de València,*
Camino de Vera, s/n, Edificio 8B, Valencia, 46022 Spain

[g]*Joint Institute for Nuclear Research (JINR),*
Joliot-Curie, 6, Dubna, 141980 Russia

[h]*Institute of Nanostructures, Nanomodelling and Nanofabrication (i3N), Universidade de Aveiro,*
Campus de Santiago, Aveiro, 3810-193 Portugal

---

[1]Corresponding author.
[2]Co-spokesperson.
[3]Deceased.

ᶦCentro de Investigaciones, Universidad Antonio Nariño,
Carretera 3 este No. 47A-15, Bogotá, Colombia

ʲDepartment of Physics and Astronomy, Iowa State University,
12 Physics Hall, Ames, Iowa, 50011-3160 U.S.A.

ᵏInstituto Gallego de Física de Altas Energías (IGFAE), Univ. de Santiago de Compostela,
Campus sur, Rúa Xosé María Suárez Núñez, s/n, Santiago de Compostela, 15782 Spain

ˡDepartamento de Física Teórica, Universidad Autónoma de Madrid,
Ciudad Universitaria de Cantoblanco, Madrid, 28049 Spain

ᵐDpto. de Mecánica de Medios Continuos y Teoría de Estructuras, Univ. Politècnica de València,
Camino de Vera, s/n, Valencia, 46071 Spain

ⁿEscola Politècnica Superior, Universitat de Girona,
Av. Montilivi, s/n, Girona, 17071 Spain

ᵒDepartment of Physics and Astronomy, Texas A&M University,
College Station, Texas, 77843-4242 U.S.A.

ᵖFermi National Accelerator Laboratory,
Batavia, Illinois, 60510 U.S.A.

qCERN, European Organization for Nuclear Research,
Geneva, 1211 Switzerland

ʳDepartment of Physics, University of Oxford,
Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom

E-mail: jrenner@ific.uv.es

ABSTRACT: We investigate the potential of using deep learning techniques to reject background events in searches for neutrinoless double beta decay with high pressure xenon time projection chambers capable of detailed track reconstruction. The differences in the topological signatures of background and signal events can be learned by deep neural networks via training over many thousands of events. These networks can then be used to classify further events as signal or background, providing an additional background rejection factor at an acceptable loss of efficiency. The networks trained in this study performed better than previous methods developed based on the use of the same topological signatures by a factor of 1.2 to 1.6, and there is potential for further improvement.

Contents

1 The NEXT experiment

Double beta decay with neutrino emission ($2\nu\beta\beta$) is a process in which two simultaneous β decays occur within a nucleus,

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + 2\bar{\nu}_e. \tag{1.1}$$

This process is allowed in the Standard Model and has been observed in several isotopes. Double-beta decay has also been postulated to exist in the zero-neutrino mode, or neutrinoless double beta decay ($0\nu\beta\beta$), in which the two antineutrinos are not emitted and the total energy released in the decay, $Q_{\beta\beta}$, is carried away by the two electrons. The observation of $0\nu\beta\beta$ would imply that the neutrino is its own anti-particle, that is, a Majorana particle [1], amongst other important physical implications (see for example [2–4]).

After 75 years of experimental effort, no compelling evidence for the existence of $0\nu\beta\beta$ decay has been obtained. For a given isotope, the lifetime of $0\nu\beta\beta$ decay depends on a nuclear matrix element and a phase-space integral, both of which can be calculated to some uncertainty, and the square of the effective neutrino mass $|m_{\beta\beta}|^2 = |\sum_{i=e,\nu,\tau} U_{ei}^2 m_i|^2$ which is a combination of the neutrino masses m_i and neutrino mixing matrix elements U_{ei}. The lifetime is of the order of $10^{25} - 10^{26}$ years for a degenerate neutrino mass hierarchy ($m_1 \sim m_2 \sim m_3$), $10^{26} - 10^{27}$ years for an inverted neutrino mass hierarchy ($m_3 \ll m_1 < m_2$), and longer than 10^{27} years for a normal mass

hierarchy ($m_1 < m_2 \ll m_3$). Experiments of the current generation deploy approximately 100 kg of the candidate isotope and are subject to several tens of counts per year of background events in their region of interest (ROI) of energy selection near $Q_{\beta\beta}$ [2]. These experiments will be capable of probing only the parameter space corresponding to the degenerate mass hierarchy, perhaps pushing into the inverted hierarchy. The most sensitive lower bound to date was set by the KamLAND-Zen experiment with ^{136}Xe, at $T^{0\nu}_{1/2} > 1.06 \times 10^{26}$ years [5]. In order to completely cover the parameter space of the inverted mass hierarchy, experiments employing candidate isotope masses at the tonne-scale with background rates of (at most) a few counts per tonne-year will be required [6].

One of the technologies currently being developed is that of high pressure xenon (HPXe) Time Projection Chambers (TPCs). In particular, the NEXT collaboration is building a HPXe TPC capable of containing a total mass of 100 kg of xenon enriched at 90% in the $\beta\beta$ decaying isotope ^{136}Xe[7]. This detector, called NEXT-100, will operate at 15 bar and use electroluminescent (EL) amplification of the ionization signal to optimize energy resolution. The detection of EL light provides an energy measurement using 60 photomultipliers (PMTs) located behind the cathode (the *energy plane*) as well as tracking via a dense array of about 8,000 silicon photomultipliers (SiPMs) located behind the anode (the *tracking plane*). In addition to performing a competitive search for $0\nu\beta\beta$, NEXT-100 will explore potential techniques for operation and background rejection at the tonne-scale.

The NEXT background model predicts a background rate of 4×10^{-4} cts keV^{-1} kg^{-1} yr^{-1} in the ROI [7]. The energy resolution for NEXT-100 is assumed to be 0.7% FWHM ($\sim$ 17 keV) at $Q_{\beta\beta}$. The experiment expects, therefore, less than one count of background per 100 kg and year of exposure, and thus its sensitivity to $T^{0\nu}_{1/2}$ is not dominated by background subtraction and increases rapidly with exposure. The expected sensitivity to the $0\nu\beta\beta$ half-life is $T^{0\nu}_{1/2} > 6 \times 10^{25}$ yr for an exposure of 275 kg·yr. This translates into a $m_{\beta\beta}$ sensitivity range of [90 − 180] meV, depending on the nuclear matrix element.

The NEXT collaboration has already built and tested several kg-scale prototypes, NEXT-DBDM [8] and NEXT-DEMO [9–11], which have both demonstrated the excellent energy resolution (extrapolated to 0.5–0.7% FWHM at $Q_{\beta\beta}$) obtainable in high pressure xenon gas. NEXT-DEMO has demonstrated the feasibility of signal/background discrimination based on the topology of reconstructed tracks [12], an essential component to identifying $0\nu\beta\beta$ events and rejecting background events (see section 2). The collaboration is currently commissioning the first underground phase of the experiment, the so called NEXT-White (or NEW for short[1]). NEW deploys a mass of 10 kg of xenon at 15 bar, the energy plane hosts 12 PMTs and the tracking plane nearly 2,000 SiPMs. Operation is foreseen in 2016 and 2017, while NEXT-100 is scheduled to start operations in 2018.

A central feature of a HPXe TPC is the capability of imaging electron tracks providing a **topological signature** that can be used to separate signal events (the two electrons emitted in a $0\nu\beta\beta$ decay) from background events (mainly due to single electrons with kinetic energy comparable to the end-point of the $0\nu\beta\beta$ decay, $Q_{\beta\beta}$). In this paper, we study the performance of the topological

---

[1]The name honours the memory of the late Professor James White, whose knowledge and generosity were essential to launching the experiment.

signature, analyzing how it is affected by the various physics processes involved in the propagation of electrons in dense gas, as well as by the detector spatial resolution. We use both the conventional reconstruction of electrons in NEXT described in [12], and an alternative technique based on the use of deep neural networks (DNNs), comparing their performance.

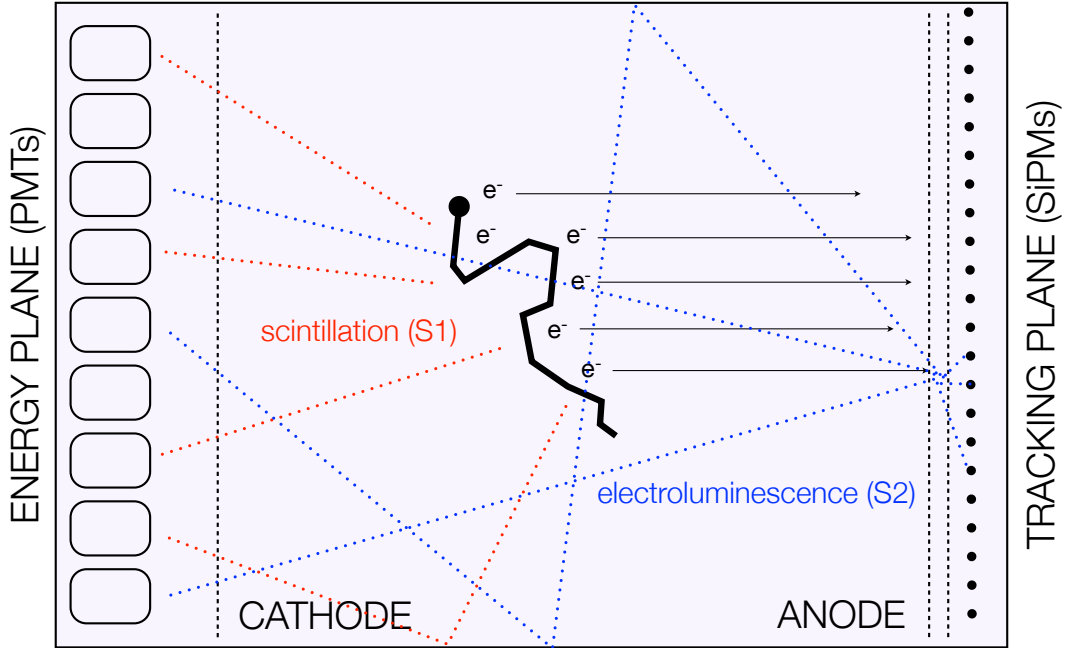## 2   Imaging tracks in a HPXe-EL TPC

Figure 1. Principle of operation of an asymmetric HPXe TPC with EL readout. (Figure from [10].)

Figure 1 shows the principle of operation of an asymmetric HPXe TPC using proportional electroluminescent (EL) amplification of the ionization signal (as is the case for NEXT-100). The detection process involves the use of the prompt scintillation light (S_1) from the gas as the start-of-event time, and the drift of the ionization charge to the anode by means of an electric field (~ 0.3 kV/cm at 15 bar) where secondary EL scintillation (S_2) is produced in a narrow region defined by a highly transparent mesh and a quartz plate coated with ITO (indium tin oxide) and TPB (tetraphenyl butadiene), called the EL gap. High voltages are applied to the two meshes to establish an electric field of ~ 20 kV/cm at 15 bar in this region. The detection of EL light provides an energy measurement using PMTs in the case of NEXT-100 located behind the cathode (the *energy plane*). The reconstruction of the track topology is carried out with a dense array of SiPMs located behind the anode (the *tracking plane*). The *x-y* coordinates are found using the information provided by the tracking plane, while *z* is determined by the drift time between the detection of S_1 and S_2. For each reconstructed spatial point, the detector also measures the energy deposited. Thus, the track is imaged as a collection of hits, and each hit is defined by a 3D space coordinate and by an associated energy deposition, as (x, y, z, E).
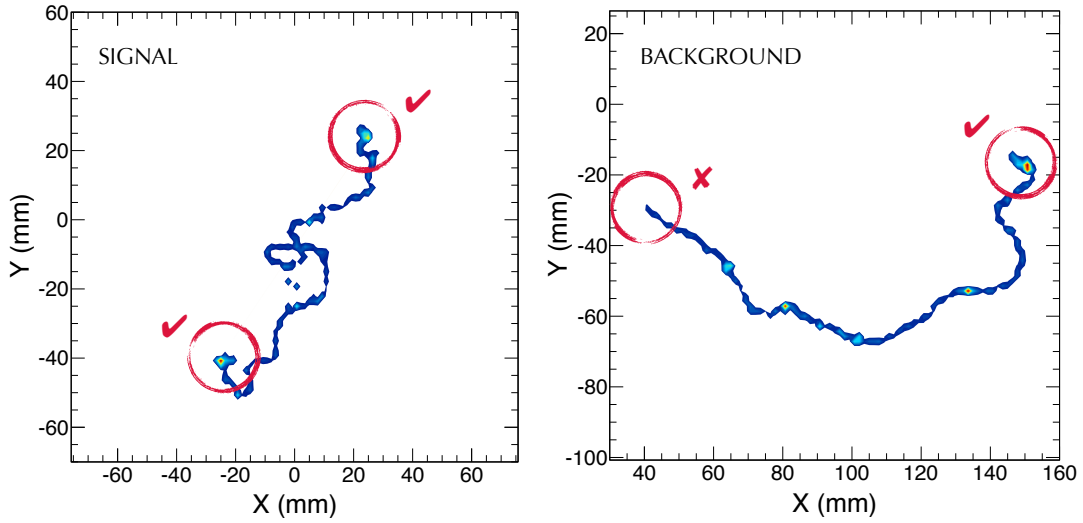
Figure 2. Monte Carlo simulation of a signal ($0\nu\beta\beta$) event (left) and a background event (right) in xenon gas at 15 bar. The color corresponds to energy deposition in the gas, red representing higher density of energy deposition and blue representing lower density. The signal consist of two electrons emitted from a common vertex, and thus it features large energy depositions (blobs) at both ends of the track. Background events are, typically, single-electron tracks (produced by photoelectric or Compton interactions of high energy gammas emitted by ^{214}Bi or ^{208}Tl isotopes), and thus feature only one blob (figure from [7]).
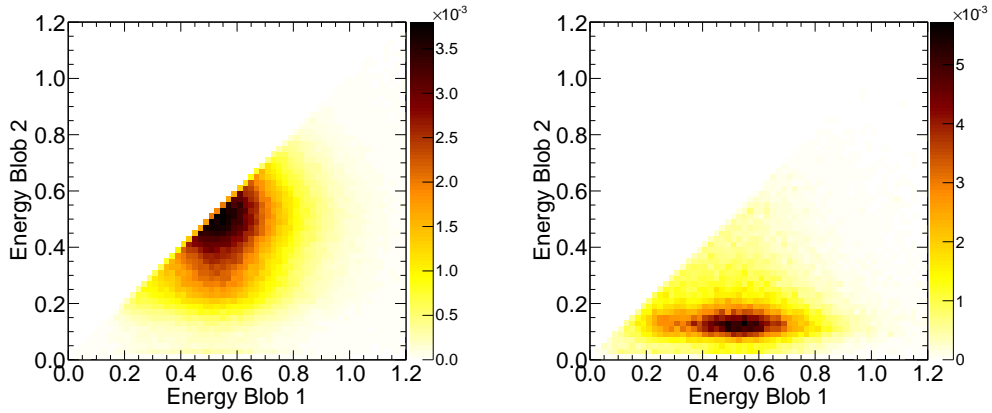

Figure 3. Probability distribution of signal (left) and background (right) events in terms of the energies of the end-of-track blob candidates. The blob candidate labelled as '1' corresponds to the more energetic one, whereas 'blob 2' corresponds to the less energetic of the two. In a signal event, the blob candidates have, on average, the same energy. In a background event, blob candidate 1 has an energy similar to that of a signal event while the energy of blob candidate 2 is very small (figure from [7]).

Electrons (and positrons) moving through xenon gas lose energy at an approximately fixed rate until they become non-relativistic. At the end of the trajectory the $1/v^2$ rise of the energy loss (where v is the speed of the particle) leads to a significant energy deposition in a compact

region, which will be referred to as a "blob". The two electrons produced in double beta decay events appear as a single continuous trajectory with a blob at each end (figure 2, left). The main background in NEXT comes from high energy gammas emitted in ^{208}Tl and ^{214}Bi decays, which occur naturally in the detector materials as part of the ^{232}Th and ^{238}U chains and enter the active volume of the detector. These gammas convert in the gas through photoelectric, Compton and pair production processes. Except in the case of pair production, these electrons typically leave a single continuous track with only one blob (figure 2-right). This topological signature was used in the Gotthard TPC to give an overall rejection rate of 96.5% for single-electron events in high-pressure xenon gas at 5 atm pressure [13]. Likewise, in NEXT, reconstruction of the signal and background topology using the tracking plane provides a powerful means of background rejection. For each track, the energy in the regions at both extremes of the track is measured and labelled as $E_{b,1}$ (the energy of the most energetic blob candidate), and $E_{b,2}$ (the energy of the least energetic blob candidate). In a signal event, $E_{b,1} \sim E_{b,2}$, while for background events $E_{b,1} >> E_{b,2}$. Figure 3 shows how this feature can be used to separate signal from background.

2.1 Reconstruction of tracks in a HPXe-EL TPC

Reconstruction of tracks in an electroluminescent HPXe TPC is complicated by the diffusion of the charge cloud during drift and also by the nature of the read-out. Scintillation light is produced over the whole width of the EL gap (5 mm in NEXT-100) spreading the signal from a single electron over a time inversely proportional to the drift velocity within the gap ($\sim$ 2 μs). Additionally, the EL light is produced isotropically and, therefore, the signal produced by the passage of an electron through the gap is expected to arrive at the tracking plane ($\sim$ several mm behind the anode) over the area defined by the intersection of the plane with the sphere of light.

In a previous paper [11], the NEXT collaboration demonstrated that a "point-like" deposition of charge due to the absorption of a point-like source (such as the xenon K$_\alpha$ X-rays) can be parameterised as a two dimensional Gaussian with a standard deviation of $\sim$ 8 mm where the spread due to EL light production is the dominant effect with subdominant contributions from transverse diffusion of the charge. As discussed below, the resolution with which we can reconstruct the centroid of this optical spread function is significantly better than this. Longitudinally, the expected spread has a noticeable dependence on the drift distance since the diffusion dominates. K$_\alpha$ events are expected to have widths in z with standard deviations of between 0.5 mm, for very short drifts, to about 5 mm for the longest drifts. In order to optimise the reconstruction of tracks these values must be taken into account by dividing the signal information into appropriate time slices and using charge information from clustered SiPM channels.

The standard NEXT algorithm searches for clusters around local maxima and then proceeds iteratively, selecting first the channel with maximum charge and forming a cluster with the first ring of sensors around it. The cluster information is then used to build a hit, whose x and y position are reconstructed as the barycentre of the charge information.

Once a set of hits is found, a connectivity criterium is defined so that the hits belonging to each separate particle can be grouped into tracks. The procedure is as follows: first, the active volume is divided into 3D pixels, known as "voxels", of fixed dimensions. Each voxel is given an energy equal to the sum of the energies of all the hits which fall within its boundaries. The collection of voxels obtained in such a way can be regarded as a graph, defined as a set of nodes and links that

connect pairs of nodes. Two voxels can then be considered connected if they share a face, an edge or a corner, with each pair of connected voxels being given a weight equal to the geometric distance between their centres. Next, the "Breadth First Search" (BFS) algorithm (see for example [14]) is used to group the voxels into tracks and to find their end-points and length. The BFS algorithm is a graph search algorithm which finds the minimum path between two connected nodes, starting from one node and exploring all its neighbours first, then the second level neighbours and so on, until it reaches the second node. The BFS algorithm divides the voxels into connected sets, known as tracks, and finds their end-points, defined as the pair of voxels with largest distance between them, where the distance of two voxels is the shortest path that connects them. The distance between the end-points is the length of the track. See [12] for a thorough discussion.

The choice of the voxel size is a compromise between a fine granularity and conservation of connectivity, which depends on the hit-finder algorithm in use. In [12], the best connectivity was found for voxels of $10 \times 10 \times 10$ mm^3. The analysis described in [7] used voxels of similar size ($10 \times 10 \times 5$ mm^3). Improvements in the hit-finder algorithm (or the use of alternative methods such as DNNs) may allow, in principle, for smaller voxels. However, the size of the voxel also reflects the effect of the spatial resolution, which in turn depends on:

1. *Tracking plane segmentation:* this includes the pitch of the SiPMs in the tracking plane as well as the SiPM response. Indeed, the use of SiPMs with very low dark current and high gain allows one to determine the location of an event by weighting the position of each sensor with the light recorded, thus improving dramatically the "digital" resolution, which goes as ~pitch/$\sqrt{12}$. The digital point resolution corresponding to a pitch of 10 mm (NEXT-100) is $10/\sqrt{12} \sim 3$ mm. Using weighted information (e.g, local barycenter algorithms), the point resolution improves to about 1 mm (see [11]).

2. *The width of the EL region:* The non-zero width of the EL gap adds an extra resolution term in the z coordinate which goes like $w/\sqrt{12}$ where w is the width of the grid. For the NEXT-100 detector, $w \sim 5$ mm, resulting also in a resolution of about 1.5 mm. At the same time, the non-negligible distance (several mm) between the gate grid (the plane that defines the beginning of the EL region) and the sensors in the tracking plane, spreads the signal of a single electron over several SiPMs. The light distribution is Gaussian and a fit to the profile recovers the position of the ionization electron.

3. *Diffusion of the drifting electron cloud:* both transverse and longitudinal diffusion are high in pure xenon (of the order of 10 and 5 mm/$\sqrt{m}$, respectively). On the other hand, work in progress within the NEXT collaboration [15] suggests that adding small amounts of cooling gases such as CH_4 or CF_4 to pure xenon (at the level of 0.1 % of CH_4 or 0.01% of CF_4) reduces both transverse and longitudinal diffusion to some 2.0 mm/$\sqrt{m}$. This is one of the most important upgrades under study for the second phase of NEXT-100.

3   Monte Carlo Simulation

NEXUS [16], the Geant4-based [17] Monte Carlo simulation of the NEXT experiment, permits an accurate modelling of the detector geometry and provides the tools to carry out both full and fast

simulations of the apparatus response. A fast simulation has been chosen for this study, given the need to generate a very large number of events for the detailed physics studies presented here, as well as for the training of the DNNs.

The simulation begins by generating a large number of signal and background events. Neutrinoless double beta events, 2 electrons with momenta generated according to a distribution calculated by the DECAY0 code [18] for $0\nu\beta\beta$ decay, are randomly created throughout the active region of the detector, while the leading background events — gamma rays of energies 2.447 MeV and 2.614 MeV corresponding to gammas emitted by daughters of ^{214}Bi and ^{208}Tl, respectively (see [7] for a thorough discussion) — are shot from the field cage of the detector geometry. The resulting locations and magnitudes of the energy depositions in the active volume are recorded as "true hits". A minimum step size of 1 mm is used in NEXUS.

The fast-simulation approach to producing reconstructed objects, starting from the true hits, takes into account the energy and spatial resolution. The former is introduced by simply smearing the total energy deposited by the event by the expected NEXT-100 resolution (we assume, conservatively, 0.7% FWHM at $Q_{\beta\beta}$); the latter, by combining the resolution associated with the pixel pitch, the EL width and the diffusion into the voxel size. Thus, to emulate the response of NEXT-100 operating with pure xenon, the true hits are replaced by voxels of $10\times10\times5$ mm^3. Comparison between fast and full simulation results in [12], with a similar voxel size of $10 \times 10 \times 10$ mm^3 as used here, showed that the efficiencies obtained for the classification cut (see section 4.2) are in agreement within 5%. These efficiency results are also in agreement with the measured value. For this reason, we believe that a voxelization of size $10 \times 10 \times 5$ mm^3 is a good proxy to capture all spatial resolution effects in a pure xenon detector with tens of centimeters of drift. In addition, we consider an optimistic scenario of $2 \times 2 \times 2$ mm^3 voxels. Spatial resolution studies using full simulation and reconstruction have also been performed with the baseline NEXT-100 EL region and tracking plane design, based on a 1 cm pitch between SiPMs and an EL gap of 5 mm, and single-point spatial resolutions of order 1 mm have been obtained in this case. This is comparable with the single-point resolution introduced by 3D voxels of size 2 mm, given by $2\sqrt{3}/\sqrt{12}$ mm = 1 mm.

4 The topological signature

4.1 Pre-selection of data

Once the list of voxels is obtained for each event, the data is processed as follows:

1. Only events near $Q_{\beta\beta}$ (in the energy window between 2.4 and 2.5 MeV) are accepted.

2. A fiducial cut is applied, ensuring that no more than 10 keV of energy was deposited within 2 cm of the edges of the active region.

3. Tracks are formed using the BFS algorithm and only events with exactly one track are accepted. This cut effectively suppresses background events which interacted by Compton scattering followed by photoelectric conversion, and also those accompanied by the emission of x-rays associated with the de-excitation of the xenon atom (e.g, after a photoelectric interaction).

Table 1 shows how the set of cuts described above reduces the signal and the primary backgrounds for Monte Carlo generated events with voxel sizes of $2 \times 2 \times 2$ mm^3 and $10 \times 10 \times 5$ mm^3 (events were generated originating from the field cage surrounding the active region of the detector). Notice that, in order to characterize the rejection power of the topological signature, a relatively large energy window of 100 keV around $Q_{\beta\beta}$ is used. The total rejection power of NEXT will be the combination of the rejection power achieved by the pre-selection (cuts 1-2 above), the topological cuts (cut 3 above and the classification cut, discussed in more detail below), and a final, stricter energy cut that accepts events in a relatively narrow ROI around $Q_{\beta\beta}$. See [7] for a detailed discussion.

Table 1. Fraction of events remaining after each analysis cut, for signal events (10^5 initial events generated within the active region of the detector) and background events from ^{208}Tl (10^9 initial events) and ^{214}Bi (10^{10} initial events) generated from the field cage surrounding the active region.

Cut	Signal Events		BG Events (^{208}Tl)		BG Events (^{214}Bi)	
	$2 \times 2 \times 2$	$10 \times 10 \times 5$	$2 \times 2 \times 2$	$10 \times 10 \times 5$	$2 \times 2 \times 2$	$10 \times 10 \times 5$
(Initial events)	1.0	1.0	1.0	1.0	1.0	1.0
Energy	7.59×10^{-1}	7.59×10^{-1}	2.27×10^{-3}	2.27×10^{-3}	1.42×10^{-4}	1.42×10^{-4}
Fiducial	6.71×10^{-1}	6.68×10^{-1}	1.19×10^{-3}	1.17×10^{-3}	8.62×10^{-5}	8.54×10^{-5}
Single-Track	3.75×10^{-1}	4.79×10^{-1}	7.90×10^{-6}	1.81×10^{-5}	3.84×10^{-6}	8.75×10^{-6}
Classification*	3.23×10^{-1}	3.67×10^{-1}	7.70×10^{-7}	2.41×10^{-6}	2.90×10^{-7}	9.59×10^{-7}

* See section 4.2

4.2 The standard NEXT classification analysis

After pre-selection and the initial topological cut eliminating events with multiple connected tracks, the events were classified as signal or background based on the presence of one or two "blobs" of energy in the reconstructed track. All possible shortest paths between two voxels were found using the BFS algorithm, and the first and last voxels of the longest of such paths[2] were considered to be the beginning and end of the track. For both the beginning and end of the track, a "blob" candidate was constructed by summing the energy of all voxels located within a given "radius" r_b of the corresponding beginning or end voxel. Note that distances between voxels were computed using the shortest path distance as determined by the BFS algorithm and not using Euclidean distance, so the quantity r_b should not be thought of literally as the radius of a sphere containing the "blob" candidate. The use of such a summation avoids, in many cases, the duplication of voxels in the two "blob" candidates. Such duplication could be present if the track wrapped around such that its end was located within a short Euclidean distance of its beginning. The summations yielded two energies, $E_{b,1}$, assigned to the greater of the two energies, and $E_{b,2}$.

The results depend on the size of the voxels, which in turn is chosen to reflect the expected performance of the detector under specific operating conditions, as discussed above. Operation

---

[2]Note that such a path may not have included, in fact most likely did not include, all voxels in the event.

with pure xenon corresponds to $10 \times 10 \times 5$ mm^3 voxels (conservative), and operation with low diffusion mixtures corresponds to voxel sizes of $2 \times 2 \times 2$ mm^3 (best expected case). Examples of events voxelized with sizes of $10 \times 10 \times 5$ mm^3 and $2 \times 2 \times 2$ mm^3 are shown in Figs. 4 and 5. The histogram of $E_{b,1}$ vs. $E_{b,2}$ is shown in figure 6 for both signal and background events analyzed with both chosen voxel sizes.
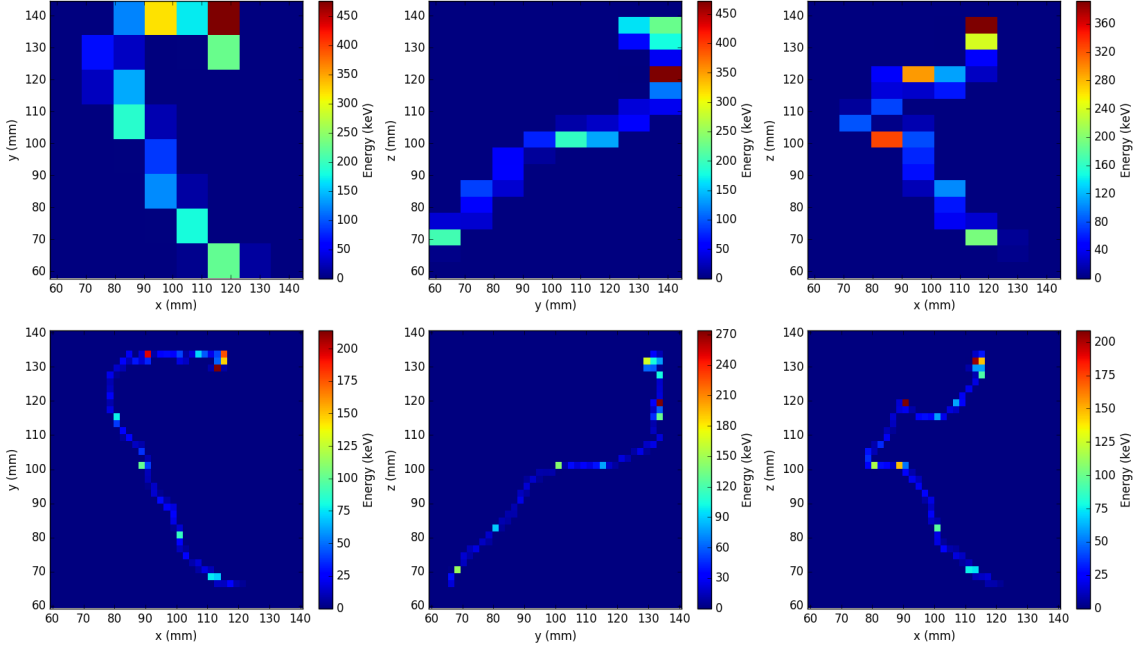

Figure 4. Projections in xy, yz, and xz for an example background event voxelized with $10 \times 10 \times 5$ mm^3 voxels (above) and with $2 \times 2 \times 2$ mm^3 voxels (below).

Finally we apply a cut designed to choose signal events with two blobs and eliminate background events with only one blob, mandating that $E_{b,1}$ and $E_{b,2}$ are both greater than a threshold energy E_{th}. This cut is applied to the events remaining after the cut requiring 1 single connected, voxelized track. For the $10 \times 10 \times 5$ mm^3 voxel size with $r_b = 18$ mm and $E_{\text{th}} = 0.35$ MeV, we eliminate all but 13.3% of remaining ^{208}Tl background events and all but 11.0% of remaining ^{214}Bi background events, and keep 76.6% of remaining signal events. For the $2 \times 2 \times 2$ mm^3 voxel size with $r_b = 15$ mm and $E_{\text{th}} = 0.3$ MeV we eliminate all but 9.74% of remaining ^{208}Tl background events and all but 7.55% of remaining ^{214}Bi background events, and keep 86.2% of remaining signal events. r_b was chosen in each case by examining the blob energy with changing r_b and selecting a value large enough to encompass the region of dense energy deposition but small enough to avoid integrating much of the less dense parts of the track. E_{th} was then varied to give a background rejection near 10%.

5 Deep Learning

The use of artificial neural networks to solve complex problems has been explored since the 1940s. In recent years, with the dramatic increase in available computing power, the use of computationally intense neural networks with many inner layers has become feasible. These neural nets that
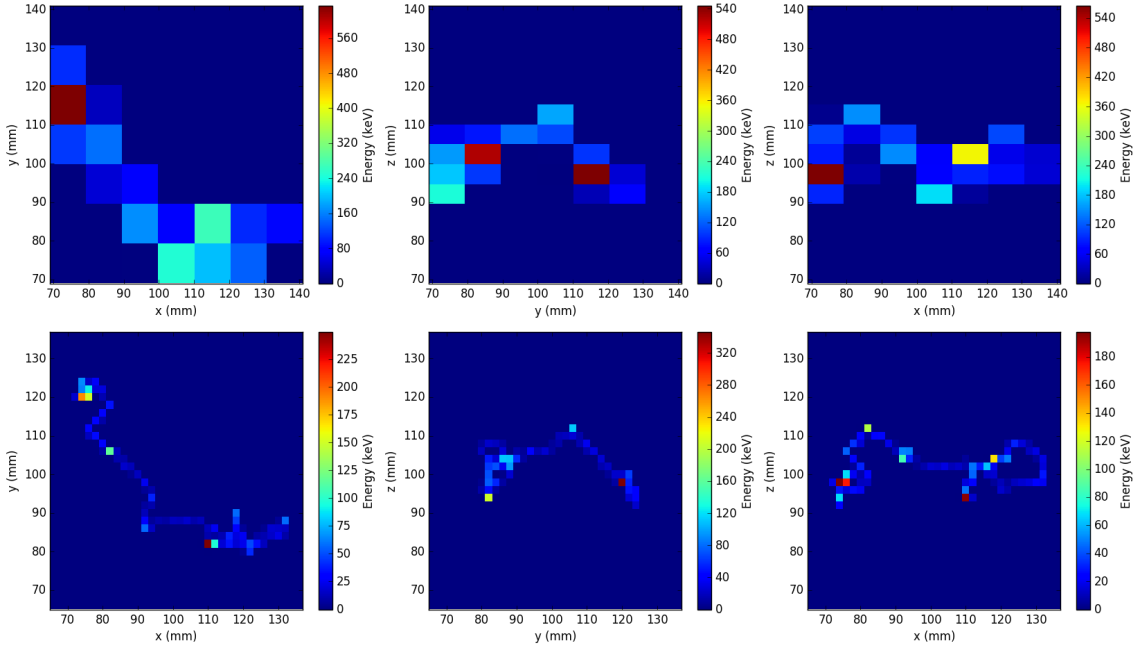
Figure 5. Projections in xy, yz, and xz for an example signal event voxelized with $10 \times 10 \times 5$ mm^3 voxels (above) and with $2 \times 2 \times 2$ mm^3 voxels (below).

are many layers deep, called deep neural networks (DNNs), are capable of learning large amounts of data exhibiting a vast array of features. This idea of "deep learning" has been applied to yield outstanding performance in solving difficult problems such as image [19] and speech [20] recognition. It has also found recent applications in physics, including event classification in high-energy and neutrino physics experiments [21–24].

Neural networks consist of layers of neurons which compute an output value based on one or several input values. The output is a function of the weighted sum of the inputs x_i plus a bias variable b, i.e. $f(\sum_i w_i x_i + b)$, where f is called the activation function and w_i are the weights of the neuron, one for each input. The idea is that with several layers of many neurons connected together, the values of the final ("output") layer of neurons will correspond to the solution of some problem given the values input to the initial layer (called the "input" layer). The weights and biases of all neurons in the network together determine the final output value, and so the network must be trained (the weights and biases must be adjusted) so that the network solves the correct problem. This is done by using a training dataset, and for each training event, presenting input data to the network, examining its resulting output, and adjusting the weights and biases of the network in a manner that minimizes the discrepancy between the output of the final layer a and the expected output y. This adjustment procedure is done by computing a cost function which depends on the actual and expected outputs and quantifies the discrepancy between them, computing the gradients of the cost function with respect to the weights and biases in all neurons, and changing the weights and biases in a manner that minimizes the cost function. After many training iterations, the weights and biases in the network will ideally have converged to values that not only yield the expected output when the network is presented with an event from the training dataset, but also yield the expected output
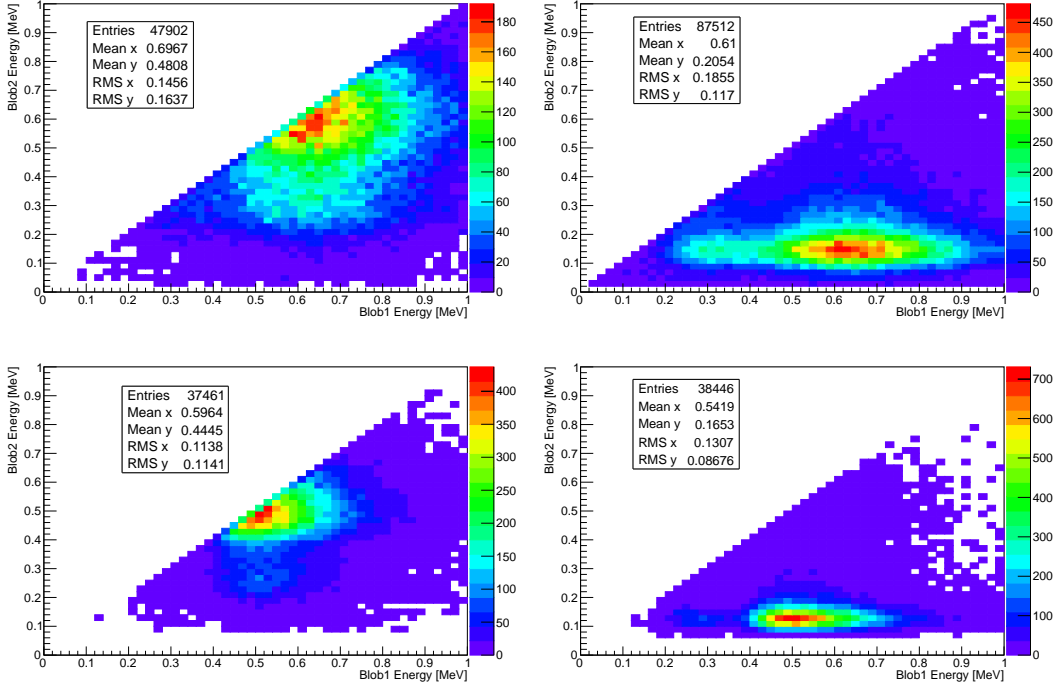
Figure 6. Computed blob candidate energies $E_{b,1}$ vs. $E_{b,2}$ for signal (left) and ^{214}Bi background (right) events with $10 \times 10 \times 5$ mm^3 voxelization (above) and $2 \times 2 \times 2$ mm^3 voxelization (below). The blob radius chosen was $r_b = 18$ mm for the $10 \times 10 \times 5$ mm^3 voxelization and $r_b = 15$ mm for the $2 \times 2 \times 2$ mm^3 voxelization.

when presented with similar events not used in training. The technical details behind implementing such a scheme mathematically will not be given here but are discussed at length in [25].

Recently, multi-layer convolutional neural networks (CNNs) have been identified as a powerful technique for image recognition problems. These neural networks consist of convolutional layers of n columns of m neurons - layers of neurons that share a common set of $m \times n$ weights and a bias. The set of weights + biases is called a filter or kernel, and this filter is combined in a multiplicative sum (a convolution) with an $m \times n$ subset of input neurons to give an output value. The filter is moved along the image, each time covering a different $m \times n$ subset of input neurons, and the set of output values corresponding to a single filter is called a feature map. With this strategy, further convolutional layers can be used to analyze the higher-level features encoded in the feature maps output from previous layers. Often to reduce the amount of computation and neurons present in deeper layers, max-pooling operations are performed, in which the neuron with maximum output value in an $m \times n$ window (or "pool") is selected, and all others in the pool are discarded. Such an operation performed on a layer of neurons leads to a new layer of reduced size. A deep CNN may be constructed from a series of several such convolutional operations and max-pooling operations, along with more conventional fully-connected layers, in which all neurons output from the previous layer are connected to the input of each neuron in the fully-connected layer, and other operations not discussed here (see figure 7 for a general schematic).

In this initial study, we make use of the GoogLeNet [19], which is a sophisticated 22-layers-
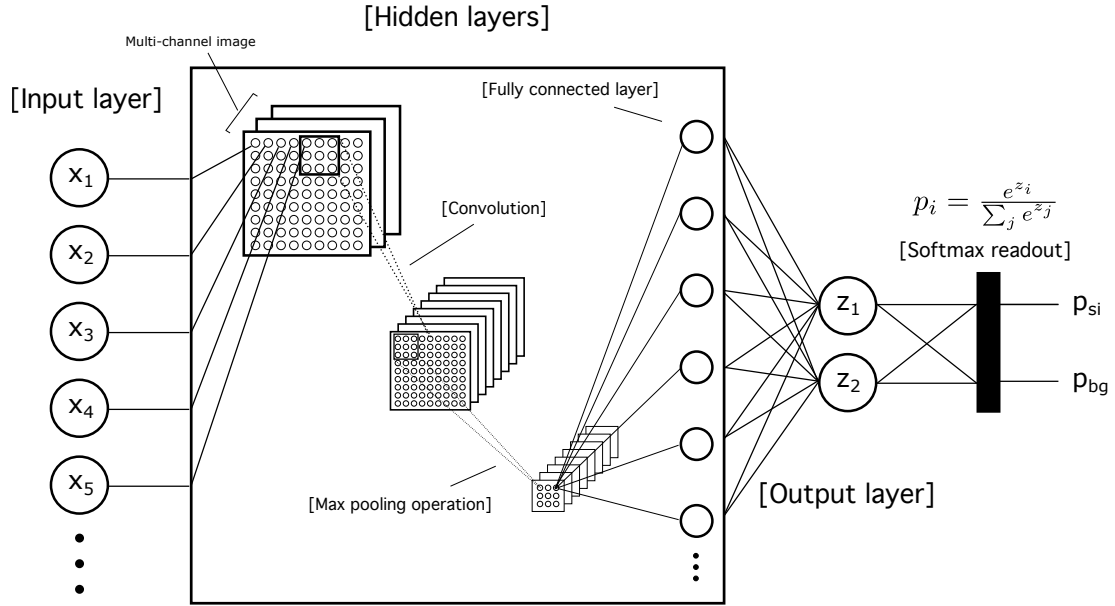
Figure 7. Schematic of a deep convolutional neural network for 2-category classification. The input layer consists of the pixel intensities of an image, possibly in multiple color channels. The hidden layers consist of several different operations performed on the input neurons - this example shows a 3×3 convolution followed by a 3×3 max-pooling operation, with the resulting neurons input to a fully-connected layer which feeds the two neurons in the output layer. The activation function of the two neurons in the final layer is such that the two outputs are exponentiated and normalized. The values in such a layer, called a "softmax" readout layer, can then be interpreted as probabilities of classification as signal or background.

deep convolutional neural network designed for image recognition. As GoogLeNet was designed to classify and identify a wide range of features in a full-color images, a more suitable network is likely to exist for our specific problem of classifying particle tracks based on topology. While further exploration of DNN architecture is essential to understanding the problem fully, our main goal in this study will be to show that DNNs can "learn" to classify NEXT events as signal or background potentially better than previously developed conventional analysis methods.

6 Event classification with a DNN

Here we investigate the performance of a DNN in classifying events into two categories, "signal" and "background," and compare the results to the conventional analysis described in section 4.2. We chose to use the GoogLeNet DNN for this initial study, as its implementation was readily available in the Caffe [26] deep learning framework along with an interface, DIGITS [27], which allows for fast creation of image datasets and facilitates their input to several DNN models. In order to generate large numbers of events with which to train the DNN, an alternate configuration of the NEXUS Monte Carlo, which we call the "xenon box" (Xe box) Monte Carlo, was run in which the NEXT-100 detector geometry was not present, and background events (single electrons) and signal events (two electrons emitted from a common vertex with a realistic $0\nu\beta\beta$ energy distribution) were

generated in a large box of pure xenon gas at 15 bar. These events were then subject to the same voxelization procedure and single-track cut as described in section 2.1.

For two different configurations of voxel size, GoogLeNet was trained on 404800 Xe box input events (50% signal, 50% background) using one or more NVidia GeForce GPUs. Each event was input to the net as a .png image consisting of three color (RGB) channels, one for each of three projections of the 3D voxelized track, $(R, G, B) \rightarrow (xy, yz, xz)$. This information for a signal event and a background event was shown earlier for different voxelizations in Fig. 4 and Fig. 5.

6.1 Analysis of NEXT-100 Monte Carlo

To compare the ability of the DNN to classify events directly with the performance of the classification analysis of section 4.2, we consider NEXT-100 Monte Carlo events that have passed the pre-selection cuts described in 4.1, with chosen voxel sizes of both $2 \times 2 \times 2$ mm^3 and $10 \times 10 \times 5$ mm^3. For each chosen voxel size, Monte Carlo events that were analyzed with the standard "blob cuts" of the classical analysis were classified by the corresponding DNN trained using Xe box events. Note that the background events used in this comparison were those produced by ^{214}Bi decay generated in the field cage surrounding the active region. The results are shown in table 2. The DNN analysis performs better than the conventional analysis, but there is still potential room for improvement.

Table 2. Comparison of conventional and DNN-based analyses. The comparison shows, for a given percentage of signal events correctly classified, the number of background (^{214}Bi) events accepted (mistakenly classified as signal).

Analysis	Signal eff. (%)	B.G. accepted (%)
DNN analysis ($2 \times 2 \times 2$ voxels)	86.2	4.7
Conventional analysis ($2 \times 2 \times 2$ voxels)	86.2	7.6
DNN analysis ($10 \times 10 \times 5$ voxels)	76.6	9.4
Conventional analysis ($10 \times 10 \times 5$ voxels)	76.6	11.0

Because the output layer of the DNN gives a probability that a given event is signal and a probability that it is background, and these probabilities add to 1, a threshold may be chosen for determining whether an event is classified as signal or background. It can be simply chosen as 50%, meaning the category with greatest probability is the classification of the event, or it can be varied to reject further background at the expense of signal efficiency. Figure 8 shows the corresponding pairs of signal efficiency and background rejection produced by variation of this threshold, while for the values reported in table 2 the threshold was chosen such that the signal efficiency matched that reported in the conventional analysis. Note that to optimize the sensitivity to $0\nu\beta\beta$ decay in the case of a non-negligible number ($\gg 1$) of background events, one would seek to maximize the ratio of signal events detected divided by the square root of background events accepted (see [7]). Thus we define a figure of merit $F = s/\sqrt{b}$, where s and b are the fractions of signal and background events accepted. This quantity is shown alongside the plot of signal efficiency vs. background rejection in Fig. 8. In table 2 we reported the values of background rejection corresponding to the

signal efficiencies studied in the classical analysis, though these did not optimize the figure of merit. For optimal figures of merit, we would have signal efficiency of 69.0% (66.7%) and background acceptance of 2.5% (6.6%) for $2 \times 2 \times 2$ mm^3 ($10 \times 10 \times 5$ mm^3) voxels.
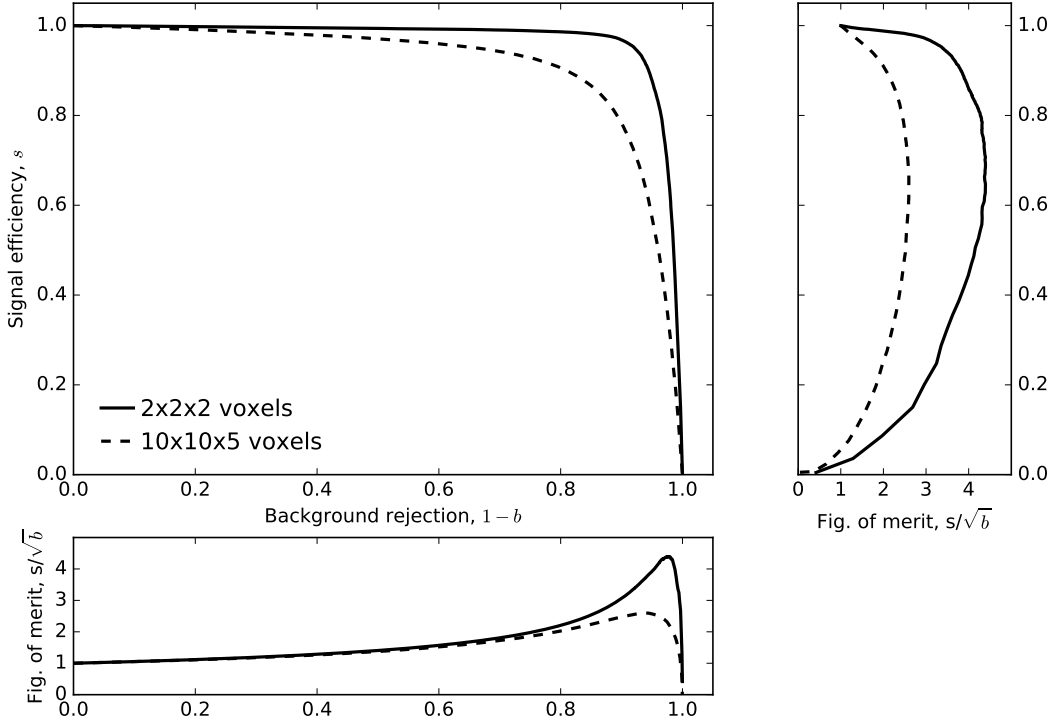

Figure 8. Signal efficiency vs. background rejection for DNN analysis of voxelized ($2\times2\times2$ and $10\times10\times5$ cubic mm), single-track NEXT-100 Monte Carlo events. The figure of merit F to be maximized in an optimal $0\nu\beta\beta$ decay search is also shown as a function of background rejection.

The improvements realized in using the DNN-based analysis combined with lower diffusion translate to significant gains in half-life sensitivity. Figure 9 shows the sensitivity at 90% confidence level calculated using the Feldman-Cousins [28] prescription as in [7] for the NEXT-100 conventional analysis and for NEXT-100 in the case of low-diffusion ($2 \times 2 \times 2$ mm^3 voxels) and using the DNN-based classification with optimal figure of merit. The substantial improvements realizable show the advantages of both an improved DNN-based analysis and achieving low diffusion in NEXT.

6.2 Evaluating the DNN analysis

We now ask what is causing some significant fraction of the events to be misclassified in the analysis described in section 6.1. To address this, a similar analysis was run on several different Monte Carlo datasets generated with differing physics effects, with the goal of developing a better understanding of where potential improvements could be made.

A simple Monte Carlo, which we call the "toy Monte Carlo" or "toy MC," was designed to produce ionization tracks of single-electron and two-electron events with a fixed energy considering minimal physical effects. Discrete energy depositions were produced with a step size less than 1
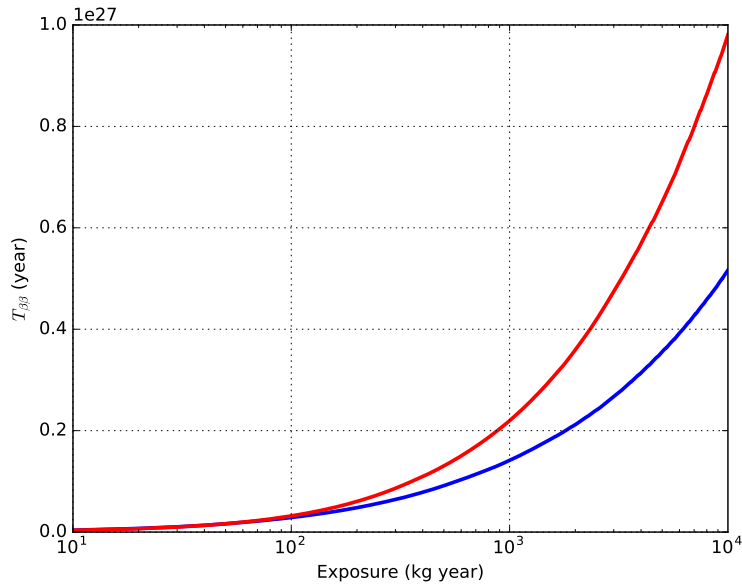
Figure 9. Sensitivity to the half-life $T^{0\nu}_{1/2}$ calculated using the Feldman-Cousins approach. The curves describe the NEXT-100 conventional analysis [7] (blue) and NEXT-100 with the improved DNN-based analysis with optimal figure of merit and low diffusion (red).

mm according to the average stopping power dE/dx as tabulated by NIST [29] for xenon at 15 atm. Electron multiple scattering was modeled by casting random Gaussian numbers to determine the angles θ_x and θ_y of deflection from the direction of travel. Assuming the particle's direction of travel is $\hat{\mathbf{z}}$, the angles θ_x and θ_y between the scattered direction and $\hat{\mathbf{z}}$ projected on the x-z and y-z planes respectively, were chosen randomly from a Gaussian distribution with sigma determined according to

$$\sigma^2(\theta_{x,y}) = \frac{13.6\text{ MeV}}{\beta p} \sqrt{dz/L_0}[1 + 0.038\ln(dz/L_0)]. \tag{6.1}$$

where dz is the thickness of xenon travelled in this step, L_0 is the radiation length in xenon, p is the electron momentum in MeV/c, and $\beta = v/c$, assuming $c = 1$.

Such tracks were generated and voxelized similar to the procedure described in section 2.1. Note that no "single-track" cut was necessary because no physics generating a secondary track was implemented. Also no energy smearing was performed. For background events, the track generation began with a single electron emitted in a random direction with energy 2.4578 MeV, while for signal events, this energy was shared equally between two electrons emitted in random directions from a single initial vertex. The DNN classified the resulting events with nearly 100% accuracy, that is, the higher of the two probabilities ($> 50\%$) of signal or background computed by the DNN corresponded to the correct classification in nearly all cases. Several modifications were then made to attempt to gain insight into the physics causing the lower classification observed in the more detailed Monte Carlo tracks. First, a realistic distribution of energies of the two electrons in signal events [18] was used, and later the magnitude of the multiple scattering was doubled (the prefactor 13.6 in equation 6.1 was increased to 27.2). The electron energy distribution caused a loss

of about 1% in average accuracy, and the increased multiple scattering an additional 1%. However, even the two effects together were not enough to account for the inaccuracy of about 8% observed in the events produced by the full Monte Carlo.

Under the controlled conditions of the Xe box simulation, many events could be generated with different aspects of the physics switched on/off. It was confirmed that with the same physics as that used in the NEXT-100 Monte Carlo, the DNN classified events with similar accuracy as before. Disabling bremsstrahlung seemed to have little effect on the accuracy. Disabling fluctuations of continuous energy losses in Geant4 had some small effect (approx. 1% increase in accuracy), and disabling the production of secondaries (disallowing the production of secondaries with a range of less than 20 cm) had a more significant effect (approx. 2.5% increase in accuracy), though still did not yield accuracy similar to that of the toy MC datasets. It was found that disabling both continuous energy fluctuations and the production of secondaries gave accuracies similar to that of the toy MC events (about 98%). A summary of the key Monte Carlo simulations run and the classification accuracies obtained is given in table 3.

There are two possible scenarios which explain why a DNN misclassifies a particular event. In the first one, the DNN is perfectly capable of taking into account all physical information available in an event but some aspects of the physics of its production have caused it to project the image of an event from the incorrect category (signal or background), given the present detector position and energy resolution. This is, for instance, the case when in a $0\nu\beta\beta$ decay event one of the electrons contains very little energy, and thus it physically resembles a single electron. Likewise, a background event in which a large secondary is produced early on in the production of the single-electron track can look like a two-electron double beta event.

However, it may well be that the present DNN is simply not capable of learning enough information to separate the two types of events, although it is physically possible. This is more difficult to understand by introducing physical effects one at a time. Rather, one should study individual events or construct more complex DNNs until no further improvement appears possible.

7   Conclusions

The NEXT topological signature of $0\nu\beta\beta$ decay events can be used to reject a significant number of background events, thus greatly increasing the sensitivity to $0\nu\beta\beta$ decay. A DNN-based analysis using GoogLeNet with just three projections seems to be capable of outperforming, by a factor of 1.2 to 1.6, depending on the resolution of reconstruction, in signal/background separation, a conventional analysis based on locating energy "blobs" at the ends of the tracks produced by energetic electrons. The production of secondaries coupled with energy fluctuations in energy deposition seems to be the principle cause of accuracy loss in the DNN analysis. Future studies geared toward developing a DNN targeted on the problem at hand, possibly exploring fully 3D convolutional networks as opposed to using 2D projections, and attempting to extract information on what characteristics of the tracks it is "learning," would lead to a more complete understanding of the possibilities and limitations of a DNN-based analysis.

Table 3. Summary of DNN analysis for different Monte Carlo datasets. The accuracy was computed assuming that the classification of the DNN corresponded to the category (signal or background) with the higher (> 50%) probability. In each case, approximately 15000 signal and 15000 background events were used in the training procedure, and between 2000-3000 signal and 2000-3000 background events independent of the training set were used to determine the accuracy.

$2 \times 2 \times 2$ **voxels**	**Run description**	**Avg. accuracy** (%)
	Toy MC, ideal	99.8
	Toy MC, realistic $0\nu\beta\beta$ distribution	98.9
	Xe box Geant4, no secondaries, no E-fluctuations	98.3
	Xe box Geant4, no secondaries	94.6
	Xe box Geant4, no E-fluctuations	93.0
	Xe box, all physics	92.1
	NEXT-100 Geant4	91.6
$10 \times 10 \times 5$ **voxels**		
	NEXT-100 Geant4	84.5

Acknowledgments

References

[1] J. Schechter and J. W. F. Valle, *Neutrinoless double-β decay in SU(2)×U(1) theories*, *Phys. Rev. D* **25** (1982) 2951.

[2] J. Gómez-Cadenas, J. Martín-Albo, J. Muñoz Vidal and C. Peña Garay, *Discovery potential of xenon-based neutrinoless double beta decay experiments in light of small angular scale CMB observations*, *JCAP* **1303** (2013) 043, [1301.2901].

[3] J. J. Gómez-Cadenas, J. Martín-Albo, M. Mezzetto, F. Monrabal and M. Sorel, *The search for neutrinoless double beta decay*, *Riv. Nuovo Cim.* **35** (2012) 29, [1109.5515].

[4] F. T. Avignone III, S. R. Elliott and J. Engel, *Double beta decay, Majorana neutrinos, and neutrino mass*, *Rev. Mod. Phys.* **80** (2008) 481, [0708.1033].

[5] KamLAND-Zen collaboration, A. Gando, Y. Gando, T. Hachiya, A. Hayashi, S. Hayashida, H. Ikeda et al., *Search for majorana neutrinos near the inverted mass hierarchy region with KamLAND-Zen*, *Phys. Rev. Lett.* **117** (2016) 082503, [1605.02889].

[6] J. Gómez-Cadenas and J. Martín-Albo, *Phenomenology of neutrinoless double beta decay*, 1502.00581.

[7] NEXT collaboration, J. Martín-Albo et al., *Sensitivity of NEXT-100 to neutrinoless double beta decay*, *Journal of High Energy Physics* **2016** (2016) 159, [1511.09246].

[8] NEXT collaboration, V. Álvarez et al., *Near-intrinsic energy resolution for 30 to 662 keV gamma rays in a high-pressure xenon electroluminescent TPC*, *Nucl. Instrum. Meth. A* **708** (2013) 101, [1211.4474].

[9] NEXT collaboration, V. Álvarez et al., *Initial results of NEXT-DEMO, a large-scale prototype of the NEXT-100 experiment*, *JINST* **8** (2013) P04002, [1211.4838].

[10] NEXT collaboration, V. Álvarez et al., *Operation and first results of the NEXT-DEMO prototype using a silicon photomultiplier tracking array*, *JINST* **8** (2013) P09011, [1306.0471].

[11] NEXT collaboration, D. Lorca et al., *Characterisation of NEXT-DEMO using xenon K_α X-rays*, *JINST* **9** (2014) P10007, [1407.3966].

[12] NEXT collaboration, P. Ferrario et al., *First proof of topological signature in the high pressure xenon gas TPC with electroluminescence amplification for the NEXT experiment*, *JHEP* **2016** (2016) 104, [1507.05902].

[13] R. Luescher, J. Farine, F. Boehm, J. Busto, K. Gabathuler, G. Gervasio et al., *Search for $\beta\beta$ decay in 136Xe: new results from the Gotthard experiment*, *Physics Letters B* **434** (1998) 407.

[14] T. H. Cormen, C. Stein, R. L. Rivest and C. E. Leiserson, *Introduction to algorithms, 2nd ed.* McGraw-Hill Higher Education, 2001.

[15] C. D. R. Azevedo, L. M. P. Fernandes, E. D. C. Freitas, D. Gonzalez-Diaz, F. Monrabal, C. M. B. Monteiro et al., *An homeopathic cure to pure Xenon large diffusion*, *JINST* **11** (2016) C02007, [1511.07189].

[16] J. Martín-Albo, *The NEXT experiment for neutrinoless double beta decay searches*. PhD thesis, University of Valencia, 2015.

[17] Geant4 collaboration, S. Agostinelli et al., *Geant4—a simulation toolkit*, *Nucl. Instrum. Methods A* **506** (2003) 250.

[18] O. A. Ponkratenko, V. I. Tretyak and Y. G. Zdesenko, *Event generator DECAY4 for simulation of double-beta processes and decays of radioactive nuclei*, *Phys. Atom. Nuclei* **63** (2000) 1282, [nucl-ex/0104018].

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov et al., *Going deeper with convolutions*, 1409.4842.

[20] G. Hinton et al., *Deep neural networks for acoustic modeling in speech recognition*, *IEEE Signal Processing Magazine* **29** (2012) 82.

[21] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. Messier, E. Niner et al., *A convolutional neural network neutrino event classifier*, *Journal of Instrumentation* **11** (2016) P09001, [1604.01444].

[22] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, *Nature Communications* **5** (2014) 1, [1402.4735].

[23] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *Journal of High Energy Physics* **2016** (2016) 69, [1511.05190].

[24] E. Racah, S. Ko, P. Sadowski, W. Bhimji, C. Tull, S.-Y. Oh et al., *Revealing fundamental physics from the daya bay neutrino experiment using deep neural networks*, 1601.07621.

[25] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2016.

[26] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick et al., *Caffe: Convolutional architecture for fast feature embedding*, `1408.5093`.

[27] NVIDIA Corporation, *NVIDIA DIGITS: Interactive Deep Learning GPU Training System*.

[28] G. D. Feldman and R. D. Cousins, *A unifed approach to the classical statistical analysis of small signals*, *Phys. Rev. D* **57** (1998) 3873.

[29] J. H. Hubbell and S. M. Seltzer, *NIST: Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients*.