Nuclear Inst. and Methods in Physics Research, A 947 (2019) 162604

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

Suppression of cosmic muon spallation backgrounds in liquid scintillator detectors using convolutional neural networks

A. Li [a,*], A. Elagin [b], S. Fraker [c], C. Grant [a], L. Winslow [c]

[a] Boston University, Department of Physics, 590 Commonwealth Avenue, Boston, MA 02215, USA
[b] Enrico Fermi Institute, University of Chicago, 5640 S. Ellis Ave, Chicago, IL, 60637, USA
[c] Laboratory for Nuclear science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Keywords:
Liquid scintillator
Neutrino detector
Deep learning
Spallation
Neural network
Background rejection

ABSTRACT

Cosmic muon spallation backgrounds are ubiquitous in low-background experiments. For liquid scintillator-based experiments searching for neutrinoless double-beta decay, the spallation product ^{10}C is an important background in the region of interest between 2–3 MeV and determines the depth requirement for the experiment. We have developed an algorithm based on a convolutional neural network (CNN) that uses the temporal and spatial correlations in light emissions to identify ^{10}C background events. Using a simple Monte Carlo simulation of a monolithic liquid scintillator detector like KamLAND, we find that the algorithm is capable of identifying 61.6% of the ^{10}C at 90% signal acceptance, with a total uncertainty of 2.7%. A detector with perfect light collection can identify 98.2% at 90% signal acceptance. The algorithm is independent of vertex and energy reconstruction, so it is complementary to other frequently-used methods and can be expanded to other background sources. This work forms the foundation for more in depth studies of detector-dependent effects and more advanced CNN-based algorithms.

1. Introduction

Neutrinoless double-beta ($0\nu\beta\beta$) decay is a hypothetical decay process by which a nucleus ejects two electrons and no neutrinos, therefore violating lepton number by two units. The observation of this process would demonstrate that neutrinos are Majorana fermions, yield valuable insight into the mechanism behind neutrino mass generation, and support a theoretical framework for matter–antimatter asymmetry in the early universe. Experiments currently being planned will instrument up to a few tons of isotope aiming for $0\nu\beta\beta$-decay half-life sensitivities between 10^{27}–10^{28} years [1–3].

Large liquid scintillator detectors are attractive for these searches because they offer cost effective scaling to large volumes and effective background reduction through self-shielding, spatial and temporal coincidence analyses, and pulse shape discrimination. The KamLAND-Zen experiment has shown the effectiveness of this technique by setting the most stringent limit on the $0\nu\beta\beta$-decay half-life independent of isotope $T^{0\nu}_{1/2} > 1.07 \times 10^{26}$ for ^{136}Xe [4].

Due to the size of these detectors, the typical 20% photocathode coverage, and $\gtrsim$1 ns timing resolution, all energy depositions are assumed to originate from a single point in space and in time. However, we know that different particle species have characteristic topologies which change in time. Even with current detector performance, these could be used for particle identification and background reduction in a variety of analyses. Recent advancements in photodetector technology, with the advent of Large Area Picosecond Photodetectors (LAPPDs) [5,6], decrease the photon arrival time uncertainty to $\leq$ 100 ps. This allows for particle identification by topology [7] and may also permit the reconstruction of particle direction through the separation directional Cherenkov light from the abundant isotropic scintillation light [8]. This has been demonstrated with muons [9–11] and recently with ^{90}Sr β-decays [12].

One of the primary cost drivers of $0\nu\beta\beta$-decay experiments is the depth at which they must be located in order to minimize backgrounds due to cosmic muon spallation. In liquid scintillator detectors doped with $0\nu\beta\beta$-decay isotopes with endpoints below 3 MeV, such as ^{136}Xe (Q = 2.458 MeV), the critical long-lived light isotope is ^{10}C (Q = 3.648 MeV, $\tau_{1/2}$ = 19.29 s). In the current KamLAND-Zen result, the largest background is the two-neutrino double-beta decay ($2\nu\beta\beta$) of ^{136}Xe, which is only reducible with improved energy resolution. In the absence of scintillator or photodetector upgrades, ^{10}C is the largest reducible background [4]. Looking at future multi-ton scintillator experiments, the increased size and shallower depth make ^{10}C the largest background for a $0\nu\beta\beta$-decay search with the JUNO experiment [13,14]. The ^{10}C background will also inform the choice of the depth and location of the proposed THEIA experiment [15]. For these reasons, we concentrate this study on ^{10}C before exploring the

* Corresponding author.
E-mail address: liaobo77@bu.edu (A. Li).

https://doi.org/10.1016/j.nima.2019.162604
Received 17 December 2018; Received in revised form 13 August 2019; Accepted 16 August 2019
Available online 19 August 2019

other significant backgrounds in KamLAND-Zen, such as ^{214}Bi, or the dominant contribution to the background in SNO+ coming from ^{8}B solar neutrino–electron elastic scattering.

The production of neutrons and light isotopes in muon spallation is an active area of study [16–19]. In muon spallation, the propagation of high energy muons through both the passive and active detector components leads to a shower of secondary particles including γ's, neutrons and pions and the breaking up of nuclei into lighter nuclei. ^{10}C is a relatively common spallation product in scintillator detectors, predominantly made through (π^+, np) [16]. Results from the Borexino experiment suggest a three-fold-coincidence of muon, neutron, and ^{11}C decay can be used to tag the ^{11}C decay relative to the neutron capture vertex and muon track [20,21]. This three-fold-coincidence, accompanied by a neutron in the final state, is also used to tag ^{10}C in the KamLAND-Zen analysis [4]. However, muon spallation is inherently chaotic and the copious number of neutrons produced through this process lead to frequent periods of detector dead-time which reduce the efficiency for identifying neutron events. This is especially true for the highest energy events with many secondary particles, which produce most of the light isotopes. This problem can be addressed with improved electronics, but due to the large dynamic range between muon events (volt-level signals) and neutron events (millivolt-level signals), difficulties remain. The challenge of tagging ^{10}C is further amplified by theoretical uncertainty in the fraction of ^{10}C production with neutron final states, and the spatial distribution relative to the muon track. For these reasons, a method independent of the progenitor muon event for identifying ^{10}C is useful for both reducing the background and verifying our understanding of the spallation process.

In this work, we show that an algorithm based on a ten-layer convolutional neural network (CNN) developed for machine vision applications can effectively separate ^{10}C from $0\nu\beta\beta$-decay events in a kilo-ton scale liquid scintillator detector, like the current KamLAND detector, without relying on muon or neutron coincidences. Since the technique is spatially invariant, it goes a step farther and is also independent of vertex reconstruction. We then perform a series of studies to understand what information the CNN is using in its discrimination and how the performance of the algorithm changes as we improve the KamLAND-like detector's performance.

The paper is organized as follows. Section 2 describes the topological differences between $0\nu\beta\beta$-decay and ^{10}C events that allow them to be distinguished by the algorithm. Section 3 provides the details of the detector Monte Carlo (MC) simulation and Section 4 provides the details of the algorithm. These are followed by the results and conclusions in Sections 5 and 6.

2. Topology of $0\nu\beta\beta$-decay and ^{10}C events

In a liquid scintillator detector, charged particles deposit energy which excites organic molecules. These molecules subsequently de-excite by releasing photons which are detected by single photon detectors like photomultiplier tubes (PMTs). The intrinsic timing of these processes is on the order of ns. Neutrino detectors with large volumes (diameters >10 m) are typically instrumented with large PMTs with the capability to resolve photon arrival times on the order of 1–5 ns.

Given the similarity between the time scales of the scintillation process and PMT readout, energy deposits are usually assumed to be point-like when reconstructing energies and vertices of physics events. This assumption neglects two effects. Gamma-rays at ^{136}Xe (Q = 2.458 MeV) $0\nu\beta\beta$-energies scatter multiple times with a mean free path on the order 10 cm which leads to a smearing of the vertex in time and space. By comparison, electrons travel <1 cm at these energies, but they are above Cherenkov threshold. Therefore, the electrons produce some directional light which is not absorbed by the scintillation process. The information from both of these processes is encoded in the pattern of photons arriving at the PMTs and can be used to identify different categories of events. Furthermore, an algorithm like a CNN which is independent of the assumptions of the vertex reconstruction would avoid the issues with the topology algorithm of Ref. [7] or direction reconstruction in Ref. [8].

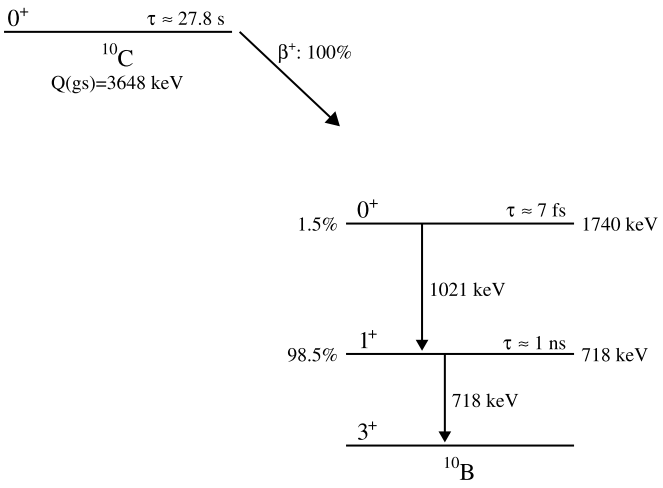

Fig. 1. Decay scheme of ^{10}C [23]. The final state of ^{10}C events consists of a positron and either one gamma with energy of 718 keV (98.5%) or two gammas with energies of 718 keV and 1021 keV (1.5%).

In a large fraction of $0\nu\beta\beta$ events, the electrons exit the nucleus at large angle with each electron getting roughly half of the available energy. We simulate the kinematics of $0\nu\beta\beta$-decay events using the same custom Monte Carlo as in Ref. [7] with momentum and angle-dependent phase factors from Ref. [22]. For ^{136}Xe, this leads to two electrons each with a kinetic energy of roughly 1.23 MeV. The Cherenkov threshold for electrons in this liquid scintillator is 0.16 MeV. Examining the path of these electrons in MC, see Section 3, we find that they travel 7.1 ± 0.9 mm with a total distance from the origin of 5.6 ± 1.0 mm in 26 ± 4 ps and drop below Cherenkov threshold after 24 ± 3 ps. We also find that the final direction of the electron before it stops does not match the initial direction, however, the total scattering angle is small while Cherenkov light is emitted.

The decay of ^{10}C is more complicated than $0\nu\beta\beta$-decay. ^{10}C is a β^+ decay that proceeds through one of two excited states as shown in Fig. 1. The event observed by the detector is a combination of a positron, the two subsequent 511 keV annihilation gammas, and either one gamma with energy of 718 keV (98.5%) or two gammas with energies of 718 keV and 1021 keV (1.5%). In addition, the 718 keV excited state is long-lived with a lifetime of 1 ns. This is significant on the time scale of events in liquid scintillator detectors.

One defining feature of $0\nu\beta\beta$-decay is the two electrons above Cherenkov threshold that algorithms, either traditional algorithms based on spherical harmonics or those based on machine learning, should pick out. Since the number of Cherenkov photons is low, the positron from ^{10}C could mimic some of these features such as one of the ring-like hit pattern. In the region of interest for ^{136}Xe, the positron spectrum produced by the ^{10}C decay therefore has a mean energy of ~0.7 MeV. Examining the path of the positron in the MC, we find that it travels 3.6 ± 0.6 mm for a total offset from the origin of 2.5 ± 0.6 mm in 13 ± 2 ps. It takes 11 ± 2 ps for the positron to fall below Cherenkov threshold. However, the ^{10}C vertex is significantly smeared in time and space by the three or more gammas in the final state and a more detailed investigation revealed that the scintillation topology and hit time distribution play an even more critical role.

3. Detector simulation

These studies are performed using a simple spherical liquid scintillator detector Monte Carlo based on GEANT4 [24,25] version 10.3.1

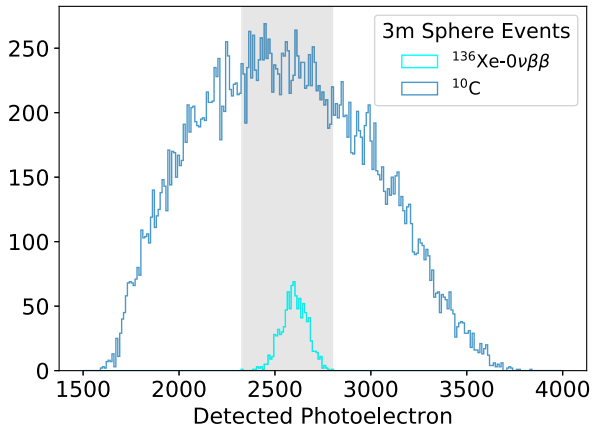

Fig. 2. The detected photoelectrons, assuming 100% photocathode coverage and QE, for ^{136}Xe $0\nu\beta\beta$-decays and ^{10}C β^+ decays generated inside a sphere with 3 m radius. The gray band indicates the energy region of interest for $0\nu\beta\beta$-decay.

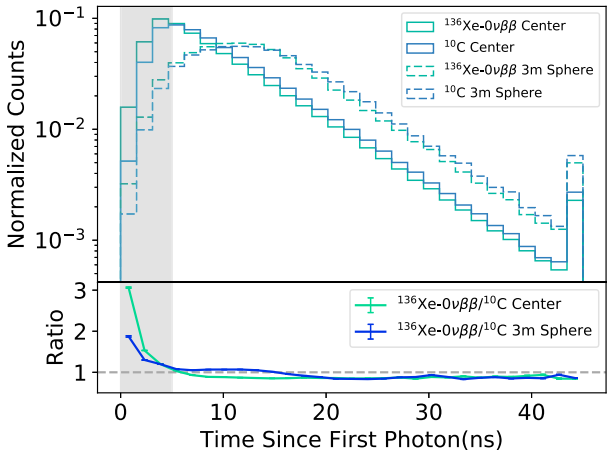

Fig. 3. TOP: Timing Profile of incoming photons for signal and background events. BOTTOM: Signal/Background ratio of timing profile histograms. Each bin represents a 1.5 ns detector snapshot, taken from the detector's sampling rate. The gray band indicates the period of major discrepancy, thus the characteristics of signal/background learned by the network. The dashed line represents a signal/background ratio equal to unity.

with a custom physics list that includes standard GEANT4 modeling of electromagnetic, optical, and radioactive decay processes. The sphere has a radius of 6.5 m and the outer surface is assumed to be 100% efficient at collecting photons. In post processing, we account for any additional propagation to the photodetectors, the details of the photodetector quantum efficiency, transit time spread and coverage, and the electronics digitization.

Although inspired by KamLAND, the goal of this MC was not to reproduce any particular detector, but to make a fast light-weight MC that is highly configurable and easily interpretable for the study of new algorithms and different detector configurations. This simulation is the same as that used in Refs. [7] and [8]. The basic scintillator properties are chosen to roughly match a KamLAND-like scintillator: 80% n-dodecane, 20% pseudocumene and 1.52 g/l PPO [26], density (ρ = 0.78 g/ml), the wavelength-dependent attenuation length [27] and refractive index [28], the scintillation emission spectrum [27], emission rise time (τ_r = 1.0 ns) and emission decay time constants (τ_{d1} = 6.9 ns and τ_{d2} = 8.8 ns with relative weights of 0.87 and 0.13) [29], scintillator light yield (9030 photons/MeV), and the Birks constant (kB ≈ 0.1 mm/MeV) [30]. The attenuation length at 400 nm, the position of the peak standard bialkali photocathode efficiency, is 25 m with a large decrease between 370 nm and 360 nm from 6.5 m to 0.65 m.

There are several effects that we are explicitly not modeling which affect mainly the late light component of the scintillator. Since the simulation is based on GEANT4, it does not simulate the re-emission of absorbed photons. The absorption is stronger at short wavelengths, so this affects Cherenkov light more than scintillation. The re-emitted photons will follow the scintillation timing, so this will appear as an additional isotropic component with a longer time constant. Similarly, GEANT4 does not model the formation of positronium in the case of β^+ emission. In vacuum, positronium decays in 0.125 ns and 142 ns for para-positronium and ortho-positronium, respectively. In scintillator, the lifetime of ortho-positronium is shortened to ~3 ns [31] leading to β^+ gaining an additional late light component. Raleigh scattering was studied in Ref. [8], and found to contribute a small late light component. Finally, there are no detector specific support structures such as balloons or acrylic vessels, so detector specific reflections are missing, which also leads to a deficit of late light.

Two types of events are generated for this study: $0\nu\beta\beta$ decay of ^{136}Xe and the β^+ decay of ^{10}C background. The kinematics of $0\nu\beta\beta$ decay events are simulated using a custom MC event generator with momentum and angle-dependent phase space factors from [22]. ^{10}C events are simulated using the default isotope decay generator in GEANT4. This correctly accounts for the long-lived first excited state of ^{10}B, but, as stated above, does not include the formation of positronium.

The energy spectrum of both event types is shown in Fig. 2 in terms of detected photo-electrons for a detector with perfect light collection. Since ^{10}C has a much broader spectrum comparing to ^{136}Xe signal, an energy cut is placed from 2.2 MeV to 2.7 MeV prior to training in order to remove background events outside region of interest.

The simulated events are positioned either at the exact center of the detector, or distributed uniformly within a 3 m-diameter spherical volume located at the center of the detector.[1] The latter matches the dimensions of the KamLAND-Zen mini-balloon, which contains the ^{136}Xe-doped scintillator. Fig. 3 compares PE arrival times between $0\nu\beta\beta$-decay and ^{10}C events. The time smearing has been included but the PMT quantum efficiency is added in later steps. In the events at the center, the excess early light from $0\nu\beta\beta$-decay and late light from ^{10}C is evident. This pattern is repeated in the 3 m sphere events but the shift in arrival time due to the vertex positions washes out some of the features. We note that the inclusion of positronium in the model would further increase the late light contribution in ^{10}C decays, improving the results of this study.

3.1. Post processing and clock latching algorithm

The photon arrival time is an important parameter of the simulation. When an optical photon is produced in liquid scintillator, it first travels to the 6.5 m simulation boundary, where the MC simulation records its arrival time and wavelength. For a detector like KamLAND, the photon needs to propagate through about 1.8 m of buffer oil to reach the actual PMT. In order to take this extra flight path into account, we calculate the group velocity of this process with the following equation:

$$v_g = \frac{c}{n - \lambda \frac{dn}{d\lambda}} \tag{1}$$

The index of refraction with respect to wavelength in Eq. (1) is extrapolated by fitting a measurement from the MiniBooNE experiment [32]. This equation is used to calculate the Time of Flight (ToF) from the 6.5 m simulation boundary to each PMT. The sum of the time to reach the 6.5 m boundary and the ToF form the actual arrival time of the photon at the face of the PMT.

Next we modeled the digitization of the signal using a clock latching algorithm. When the first photon reached a PMT a clock started ticking

[1] Unless otherwise specified, the two types of event distributions will be referred as *center* and *3 m sphere* in this text.

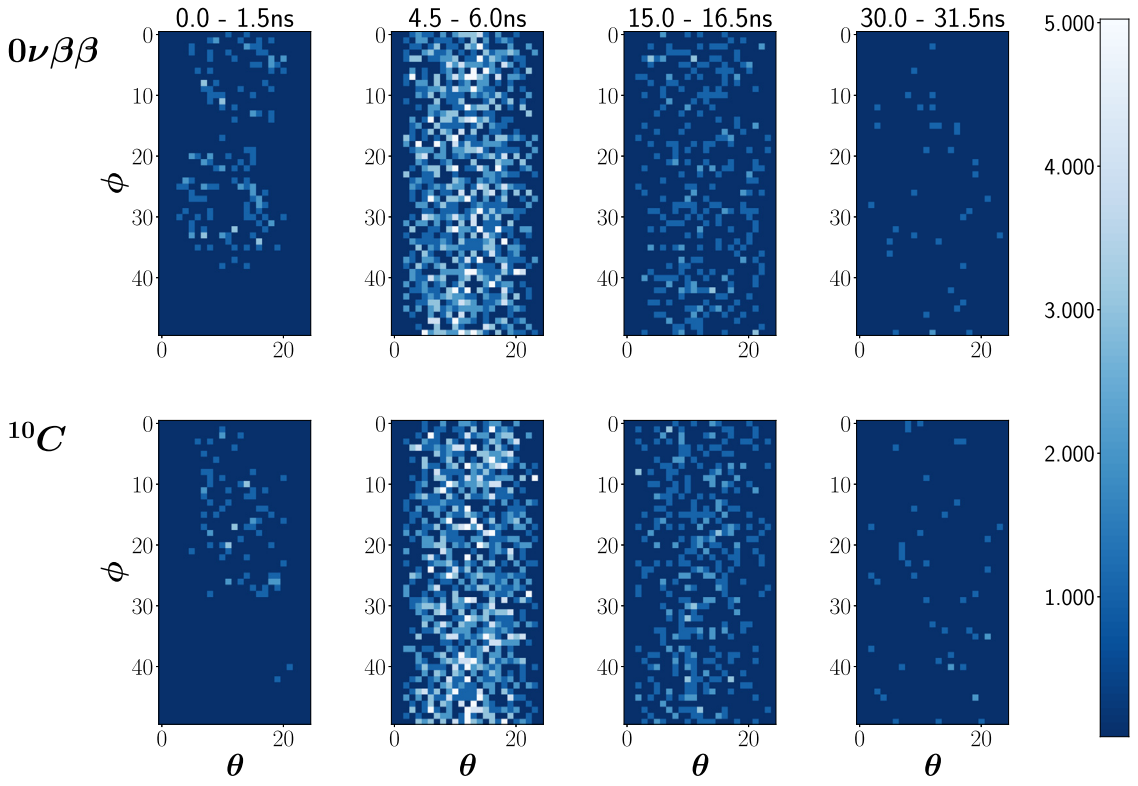

Fig. 4. Time evolution of the PMT hit-map over four example time bins for a $0\nu\beta\beta$-decay event and a ^{10}C event at the center of the detector, normalized to have the same total energy.

at a 1.5 ns interval to reproduce the sampling time of the KamLAND electronics. We observed that the choice of starting the clock on the first photon introduces a relatively small (1.9%) systematic uncertainty toward our final result. To model the time resolution of the PMTs, the arrival time of each incoming photon, including the first photon, is smeared by a Gaussian probability density function with a 1 ns standard deviation. The smeared time is latched to the ticking clock such that any photon that arriving after the current clock tick, but prior to next tick, is latched to the current tick. A total of 30 clock ticks, 45 ns after the first photon, were used to record the input event. Furthermore, four additional 1.5 ns time intervals are added prior to zero time in order to take into account that some photons are smeared backward in time. A 2D representation of the PMT hit pattern is then formed for each tick. We call each of these images a channel, analogous to the RGB channels in a photograph or a digital image.

After post-processing each event contains 34 channels, 33 from −6 ns to 45 ns, in 1.5 ns increments and one larger bin for the remaining late photons. Fig. 4 contains four example channels for $0\nu\beta\beta$-decay and ^{10}C. The distinctive double Cherenkov rings from $0\nu\beta\beta$-decay are visible in the first channel. Shortly after the first channel, the abundant scintillation light quickly dominates for the remainder of the event. The events in Fig. 4 were carefully selected to demonstrate the Cherenkov pattern. Due to limited spatial and time resolution, the Cherenkov pattern for most events is much fuzzier and more difficult to extract.

3.2. Gray disk PMT model

The photo-coverage of the detector is introduced by circumscribing a circular region around each PMT location, known as a *gray disk*. We used the KamLAND PMT locations for the 1325 17-inch PMTs as a representative PMT layout [33]. The total gray disk area was adjusted to yield the desired photo-coverage. When a photon is produced in the MC, it is associated to the closest PMT. If this photon passes through the gray disk region of the associated PMT, one photon hit is recorded for that PMT position; otherwise, the photon is rejected. This method limits the total photocathode coverage to 40%, since at that point the gray disk of PMTs start overlapping.

3.3. Quantum efficiency

The PMT quantum efficiency (QE) is wavelength dependent and is another important simulation input. The QE dependence is accounted for during the event generation. In GEANT4, a QE bit associated with each photon is introduced to indicate whether or not it was recorded by the PMT. However, in this study, we need a varying QE to act as a pressure[2] parameter in order to demonstrate the neural network's performance. A stepwise QE cut is introduced to accommodate this requirement. If the QE bit indicates that a photon was detected, then the photon always passes the QE cut. Otherwise, photons are randomly rejected based on the desired QE pressure. In the ideal situation the QE is 100% and all photons will be indiscriminately recorded. In this study, we only allow QE to vary up to a reasonable value, namingly 56% in Fig. 8.

For our baseline KamLAND-like detector model, the current model, we assume a QE of 23% and photocathode coverage of 19.6%. We define an example upgrade scenario where these parameters are roughly doubled to a QE of 36.2% and photocathode coverage of 42%.

4. Event classification algorithm

CNNs are a type of deep neural networks commonly used in the field of computer vision. To perform classification tasks using CNNs, a pixelized image is processed through several so-called *layers*, each containing a linear transformation step, followed by a non-linear activation. Among these layers, convolutional layers play the most important

[2] In this context, the term pressure refers to the level of difficulty for a neural network to classify events. See also Section 4.2.

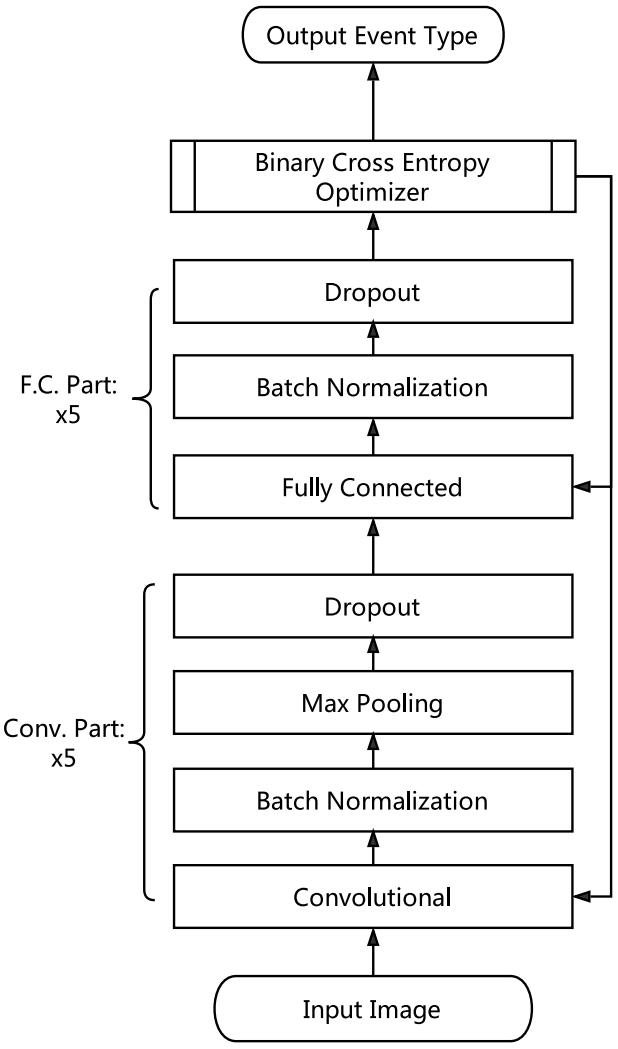

Fig. 5. Flow Diagram of the CNN. The Network is composed of Convolutional Part (bottom) and Fully Connected Part (top).

role. As the name implies they involve the application of a convolution operation between two functions. For continuous functions, it represents the Fourier transform of the products. Applying this procedure to a 2D discrete surface gives rise to the convolutional layer.

The convolutional filter, the kernel for the convolutional layer, is a fix-sized grid with specific values assigned to each block. The filter is scanned throughout the image body, and each image pixel is multiplied by a filter weight. Finally, the convolution operation is completed by summing the element-wise multiplications.

The convolutional layer is also capable of taking information from multiple channels. In this case, the convolution operation is conducted separately over each channel, and the output values are summed and fed into the next layer. For example, when a CNN is used to classify photographs, the input contains four channels: Red (R), Green (G), Blue (B), and Gray Scale. For this work, we have 34 time-based channels as described in Section 3.1.

The output of the convolutional layer contains features that are fed into the fully connected layer. This layer is where the high level decisions are made, ending with an image being classified into one of several categories with an assigned probability. Other layers in the network structure include pooling layers and dropout layers. A pooling layer reduces the image dimensionality by extracting only the maximum value from each pooling filter to represent the image. It significantly increases the processing speed with almost no sacrifice of classification accuracy. The dropout layer prevents overfitting by randomly disabling neurons in the hidden layer with a predefined dropout rate [34]. For extensive details on the functionality of CNNs, we refer the reader to Ref. [35].

4.1. Network design

The CNN used in this work is implemented in Keras [36] with a Tensorflow backend [37]. The general outline of the network is shown in Fig. 5. It can be divided into a convolutional part and a fully connected part.

The convolutional part contains 5 sets of layers where each set includes a convolutional layer, a batch normalization layer, a pooling layer, and a dropout layer. For simplicity, one such set of layers is often simply referred to as a convolutional layer. Each time a pooling layer is introduced, the image reduces to 25% of it's original size. Therefore, the convolution part is confined to five iterations.

The fully connected part also contains 5 sets of layers. Each set includes a fully connected layer, a batch normalization layer and a dropout layer. Again, this set of layers is often collectively referred to as a fully connected layer. With respect to the overall depth, the fully connected part is not limited by the image size or pooling layers. However, going too deep with the fully connected part will complicate the model and lead to overfitting. Dropout layers are inserted throughout the network in order to prevent this effect while decreasing the processing time [34].

Normalization of data plays an important role in classification tasks. Without normalization, the neural network will reveal some anomalous behavior, including non-convergence and high probability of misclassification. During a pre-training stage, each pixel is scaled to a value below unity. During the network design stage, several different normalization schemes were considered, including vector normalization, channel-wise standard scalar normalization, and batch normalization. For this work, batch normalization was chosen. This means that the normalization is performed for each incoming batch, transforming the input data, or the PMT hit patterns shown on 4, such that the mean is zero with a standard deviation of one.

Hyperparameters refer to parameters in the network that are predefined before training, and stay constant throughout the training stage. Hyperparameters of neural networks define the structure of the network, and changing these parameters turns the neural network into a different model. Typical hyperparameters include, but are not limited to, the number of layers in the network, the number of nodes in each layers, the size of the filters, and the dropout rate.

A hyperparameter search can result in a significant improvement of performance [38]. For our model, we performed the search with hyperopt [39]. Three hyperparameters are tuned to achieve the best performance, including number of nodes, number of fully connected layers, and dropout rate. The accuracy, a.k.a the percentage of events being correctly classified into its labeled class, is chosen as the criteria to evaluate network performance. We selected a preliminary training data set of 20,000 3 m sphere events with the current detector configuration. We then search across a continuous range for the dropout rate between 0 and 1 and several discrete values for the number of fully connected layers and nodes. A random search [38] of 50 attempts is executed to determine the best accuracy. The 50 trials are evaluated and compared for the best background rejection capability. After tuning, the validation accuracy increases from 49.7% to 77.3%.

4.2. Training

The algorithm is trained and validated on MC data sets, generated according to Section 3. We study two different data sets: 70,000 centered events, and 50,000 3 m sphere events. All events are stored in an 6 dimensional array. The six dimensions are correspondingly: photocoverage pressure, QE pressure, event index, time channel, polar

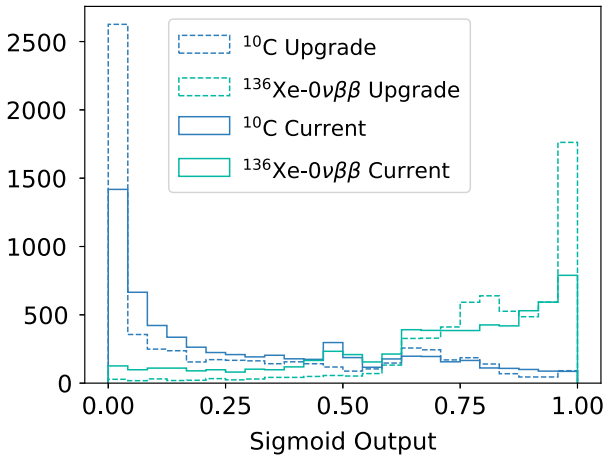

Fig. 6. Sigmoid output from simulated events isotropically distributed within a 3 m-diameter balloon. The 'Current' label indicates the current PMT configuration in KamLAND Zen (20% photocoverage and 23% QE), and the 'Upgrade' label indicates a reasonable PMT configuration for possible future liquid scintillator detector (40% photocoverage and 43% QE).

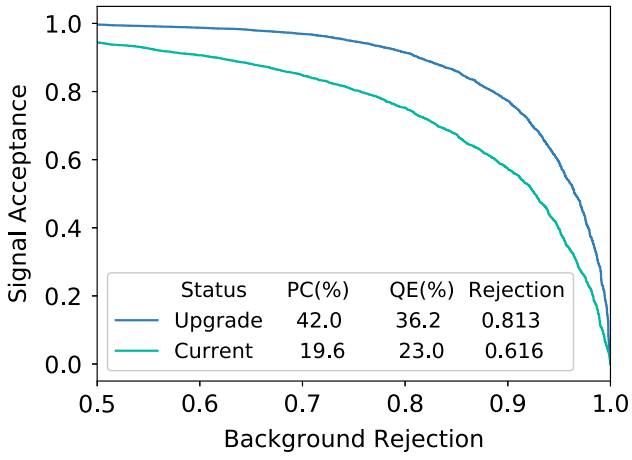

Fig. 7. ROC curves from simulated events isotropically distributed within a 3 m-diameter balloon. The quoted background rejection assumes 90% signal acceptance.

angle and azimuthal angle. The algorithm is trained and validated independently for each data set and each category contains equal amounts of $0\nu\beta\beta$-decay signal and ^{10}C background events.

The data sets are separated into training and validation subsets with a 3:1 ratio. The network is trained on batches of 10 events over 30 training cycles. An RMSProp optimizer [37] is used to apply backward propagation optimization based on binary cross entropy [40]. Due to the large data volume, sparse matrix and batch generator technology is applied to reduce memory consumption. During training stage, a learning rate decay scheduler is incorporated to reduce systematic fluctuation. After training, the CNN is applied to the validation data set to study the out-of-sample performance. A bad validation rate in the presence of good training accuracy is indicative of overfitting and we do not see this effect.

While training the CNN on each event category, the photo-coverage and PMT QE were scanned over a wide range of values to better understand the CNN performance under different levels of classification difficulty. These levels of difficulty are referred to as *pressure*. The photo-coverage was allowed to vary from 20% to 40% and QE from 23% to 56%. The lower bounds of the photocoverage and QE comes from current KamLAND Zen PMT configurations. The maximum photocoverage was calculated by expanding the Gray disk PMT radius until they start to overlap, while the PMT QE upper limit was chosen empirically as the highest possible QE in the near future. A total of 99 CNN models were trained and evaluated for the pressure maps.

5. Results

The well-trained neural network outputs a single floating point number between 0 and 1 for every input event. This number is the sigmoid output, since it comes out of a sigmoid activation function. The sigmoid output serves as the probability or metric for event classification. If the sigmoid output of a given event is close to 1, it means the event is likely to be a signal event. The value of the sigmoid output used for the classification can be more or less stringent depending on the required signal purity versus signal acceptance. Fig. 6 shows the sigmoid output for the current detector configuration and a possible detector upgrade scenario.

After the CNN is applied to the validation data set, the value of the sigmoid output cut can be varied to generate a Receiver Operating Characteristic (ROC) curve. For our application, the result is simply the signal acceptance as a function of the background rejection. Fig. 7 shows the ROC curves for the current configuration and the upgrade scenario. We find that at 90% signal acceptance we can reject 61.6% of the ^{10}C. For the scenario with increased coverage and QE this increases to 81.3%. The central events with the standard KamLAND configuration have a 97.7% rejection at 90% signal acceptance.

Fig. 8 shows the pressure maps which scan different QE and photo-coverage configurations for both the central and the 3 m sphere events. As expected, the CNN performs best for centrally located events and for higher QE and photo-coverage. The results also indicate that increasing the total light collected, whether by increasing in QE or photocathode coverage, leads to improved performance.

Within the parameters of this study, we find that it is possible to reach >99.98% discrimination for central events. This indicates that higher isotope concentrations that lead to more centrally distributed $0\nu\beta\beta$-decay events are advantageous and this motivates future design studies. We also studied the algorithm with the 3 m sphere events and perfect light collection and find 98.2% rejection at 90% signal acceptance.

We use the fluctuations observed Fig. 8 to understand the uncertainty in the algorithm. This is estimated by calculating at each point in the grid the standard deviation relative to the 4 adjacent neighbors and averaging this value over the pressure map. For central events, the uncertainty is 0.16%, while the 3 m sphere events give an 1.9%. Combined with the 1.9% systematic uncertainty from the Clock Latching Algorithm described in Section 3.1, the final uncertainty for 3 m sphere events becomes 2.7%.

In order to understand what feature the CNN is using to discriminate, we produced a data set where the Cherenkov light was removed. The results for central events with the standard KamLAND detector configuration indicated a 4% decrease in the rejection. The 3 m sphere events show a 2% increase in the rejection. This could indicate that Cherenkov light is interfering with the network's interpretation of the rise-time of the scintillation light, however this is also within our estimated uncertainty for the algorithm. As a whole, we find that the Cherenkov signal is not the dominant feature and that the scintillation light topology is driving the event separation. Within the scintillation light topology the photon timing information contains significant discriminating power. Removing spatial information and using a one-dimensional fully-connected network with timing as an input we found background rejection drops from 61.6% to 55.0% at 90% signal acceptance.

6. Conclusion

Liquid scintillator detectors have been at the heart of many of the great discoveries in neutrino physics and have been a leading

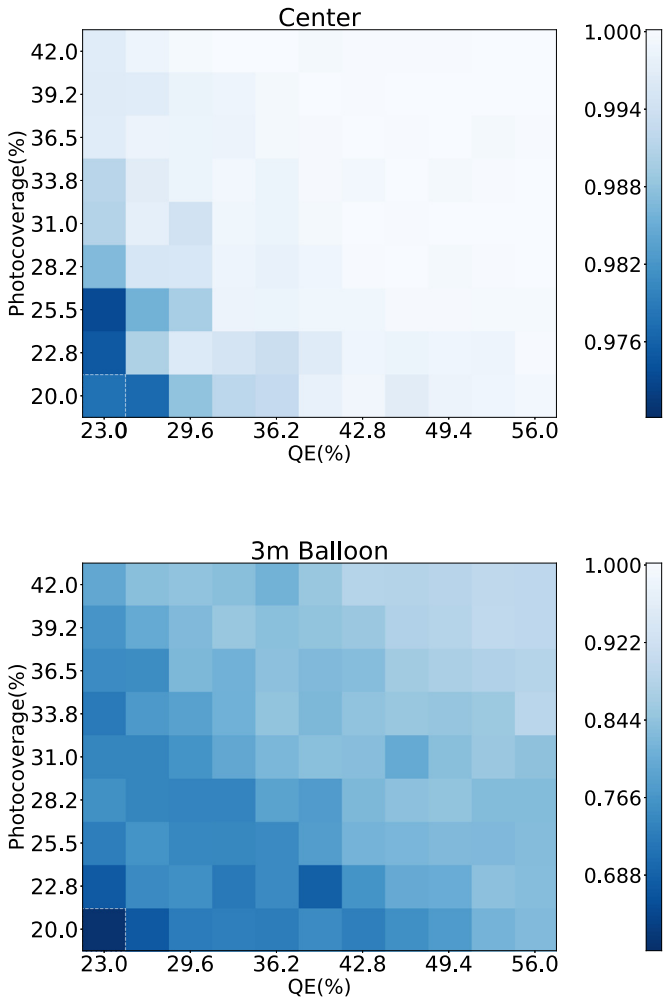

Fig. 8. Pressure map of the centrally located ^{10}C events (top) and the same for ^{10}C events within a 3 m-diameter sphere (bottom). The value in each grid is the Background Rejection Percentage assuming 90% Signal Acceptance. The dashed box indicates the current KamLAND-Zen PMT efficiency.

technology in the search for $0\nu\beta\beta$-decay. Their data is effectively a time series of images projected onto a sphere and therefore modern image processing algorithms may provide new insight into this data. In this work, we apply an algorithm from computer vision based on a CNN to extract the fundamentally different physics processes that take place in ^{10}C background events and $0\nu\beta\beta$-decay signal events. With a standard detector configuration similar to the current KamLAND detector, we find we can reject 61.6% of the ^{10}C background with 90% acceptance of the $0\nu\beta\beta$-decay signal. A detector with the same geometry and perfect light collection could achieve 98.2% rejection. We also find that the performance can be increased to better than 99.98% for centrally located events. The overall uncertainty of the algorithm is 2.7%.

These results are a basis for future studies combining machine learning techniques based on CNNs with liquid scintillator detectors. In short order, we intend to move to a spherically symmetric CNN [41] and a Bayesian classification that provides a posterior distribution for the classification. In future studies, this algorithm will be applied to other backgrounds with topologies distinct from $0\nu\beta\beta$-decay. These include ^{214}Bi decays on KamLAND-Zen's 3 m-diameter inner balloon and elastic scattering of ^{8}B solar neutrinos. Solar neutrinos are expected to be the dominant background in SNO+ [42]. We are also exploring algorithms which could move beyond simple classification to particle position and direction reconstruction. These studies are benefiting from an abundance of work being done for other applications both inside and outside of particle and nuclear physics and there are many new avenues to explore.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Numbers 1554875 and 1806440. We thank Taritree Wongjirad for contributions to earlier incarnations of this project and Jonathan Ouellet for the ^{10}C decay diagram. We also thank Kazuhiro Terao for useful discussions. This work is done in support of the NuDot, THEIA and of course KamLAND experiments and we thank them for their input. The work at the University of Chicago is supported by U.S. Department of Energy, Office of Science, Offices of High Energy Physics and Nuclear Physics under contracts DE-SC0008172 and DE-SC0015367; the National Science Foundation under grant PHY-1066014; and the Physical Sciences Division of the University of Chicago. We thank Mikhail Hushchyn, Eugeny Toropov, and Ilija Vukotic for advice on application of machine learning techniques to image classification. This research was done, in part, using resources provided by the Open Science Grid [43,44], which is supported by the National Science Foundation award 1148698, and the U.S. Department of Energy's Office of Science. This research was also performed with the help of the Shared Computing Cluster at Boston University.

References

[1] J.B. Albert, et al., nEXO Collaboration, Sensitivity and discovery potential of nEXO to neutrinoless double beta decay, Phys. Rev. 97 (6) (2018) 065503, http://dx.doi.org/10.1103/PhysRevC.97.065503.

[2] N. Abgrall, et al., LEGEND Collaboration, The large enriched germanium experiment for neutrinoless double beta decay LEGEND, in: Proceedings, Matrix Elements for the Double Beta Decay Experiments (MEDEX'17), in: AIP Conf. Proc., Prague, Czech Republic, vol. 1894, no. 1, 2017, 020027, http://dx.doi.org/10.1063/1.5007652, http://arxiv.org/abs/1709.01980.

[3] Andrea Giuliani, The mid and long term future of neutrinoless double beta decay, 2018, http://dx.doi.org/10.5281/zenodo.1286915.

[4] A. Gando, et al., KamLAND-Zen Collaboration, Search for majorana neutrinos near the inverted mass hierarchy region with KamLAND-Zen, Phys. Rev. Lett. 117 (8) (2016) 082503, http://dx.doi.org/10.1103/PhysRevLett.117.082503; Phys. Rev. Lett. 117 (10) (2016) 109903, http://dx.doi.org/10.1103/PhysRevLett.117.109903, (Addendum).

[5] Bernhard Adams, et al., Nucl. Instrum. Methods A 795 (2015) 1–11, http://dx.doi.org/10.1016/j.nima.2015.05.027.

[6] Michael Minot, et al., Pilot production & commercialization of LAPPD™, Nucl. Instrum. Methods A 787 (2015) 78–84, http://dx.doi.org/10.1016/j.nima.2014.11.025.

[7] Andrey Elagin, Henry Frisch, Brian Naranjo, Jonathan Ouellet, Lindley Winslow, Taritree Wongjirad, Separating double-beta decay events from solar neutrino interactions in a kiloton-scale liquid scintillator detector by fast timing, Nucl. Instrum. Methods A 849 (2017) 102–111, http://dx.doi.org/10.1016/j.nima.2016.12.033.

[8] C. Aberle, A. Elagin, H.J. Frisch, M. Wetstein, L. Winslow, Measuring directionality in double-beta decay and neutrino interactions with kiloton-scale scintillation detectors, J. Instrum. 9 (2014) P06012, http://dx.doi.org/10.1088/1748-0221/9/06/P06012.

[9] J. Caravaca, F.B. Descamps, B.J. Land, M. Yeh, G.D. Orebi Gann, Cherenkov and scintillation light separation in organic liquid scintillators, Eur. Phys. J. 77 (12) (2017) 811, http://dx.doi.org/10.1140/epjc/s10052-017-5380-x.

[10] J. Caravaca, F.B. Descamps, B.J. Land, J. Wallig, M. Yeh, G.D. Orebi Gann, Experiment to demonstrate separation of Cherenkov and scintillation signals, Phys. Rev. 95 (5) (2017) 055801, http://dx.doi.org/10.1103/PhysRevC.95.055801.

[11] Mohan Li, Ziyi Guo, Minfang Yeh, Zhe Wang, Shaomin Chen, Separation of scintillation and Cherenkov lights in linear alkyl benzene, Nucl. Instrum. Methods Phys. Res. A 830 (2016) 303–308, http://dx.doi.org/10.1016/j.nima.2016.05.132, URL http://www.sciencedirect.com/science/article/pii/S0168900216305411.

[12] Julieta Gruszko, Brian Naranjo, Byron Daniel, Andrey Elagin, Diana Gooding, Chris Grant, Jonathan Ouellet, Lindley Winslow, Detecting Cherenkov light from 1–2 MeV electrons in linear alkylbenzene, 2018.

[13] Shao-Feng Ge, Werner Rodejohann, JUNO and neutrinoless double beta decay, Phys. Rev. 92 (9) (2015) 093006, http://dx.doi.org/10.1103/PhysRevD.92.093006.

[14] Jie Zhao, Liang-Jian Wen, Yi-Fang Wang, Jun Cao, Physics potential of searching for $0\nu\beta\beta$ decays in JUNO, Chin. J. Phys 41 (5) (2017) 053001, http://dx.doi.org/10.1088/1674-1137/41/5/053001.
[15] J.R. Alonso, et al., Advanced scintillator detector concept (ASDC): A concept paper on the physics potential of water-based liquid scintillator, 2014.
[16] S. Abe, et al., KamLAND Collaboration, Production of radioactive isotopes through cosmic muon spallation in KamLAND, Phys. Rev. 81 (2010) 025807, http://dx.doi.org/10.1103/PhysRevC.81.025807.
[17] G. Bellini, et al., Borexino Collaboration, Cosmogenic backgrounds in borexino at 3800 m water-equivalent depth, J. Cosmol. Astropart. Phys. 1308 (2013) 049, http://dx.doi.org/10.1088/1475-7516/2013/08/049.
[18] H. de Kerret, et al., Double Chooz Collaboration, Yields and production rates of ^{9}Li and ^{8}He measured with the Double Chooz near and far detectors, 2018.
[19] Feng Peng An, et al., Daya Bay Collaboration, Cosmogenic neutron production at daya bay, Phys. Rev. 97 (5) (2018) 052009, http://dx.doi.org/10.1103/PhysRevD.97.052009.
[20] Cristiano Galbiati, Andrea Pocar, Davide Franco, Aldo Ianni, Laura Cadonati, Stefan Schonert, Cosmogenic C-11 production and sensitivity of organic scintillator detectors to pep and CNO neutrinos, Phys. Rev. 71 (2005) 055805, http://dx.doi.org/10.1103/PhysRevC.71.055805.
[21] G. Bellini, et al., Borexino Collaboration Collaboration, Final results of borexino phase-I on low-energy solar neutrino spectroscopy, Phys. Rev. D 89 (2014) 112007, URL https://link.aps.org/doi/10.1103/PhysRevD.89.112007.
[22] J. Kotila, F. Iachello, Phys. Rev. C 85 (2012) 034316, URL http://link.aps.org/doi/10.1103/PhysRevC.85.034316.
[23] ^{10}C decay scheme graphics is courtesy of J. Ouellet.
[24] S. Agostinelli, et al., GEANT4 Collaboration, Geant4: A simulation toolkit, Nucl. Instrum. Methods A 506 (2003) 250–303, http://dx.doi.org/10.1016/S0168-9002(03)01368-8.
[25] J. Allison, et al., GEANT4 Collaboration, Geant4 developments and applications, IEEE Trans. Nucl. Sci. 53 (1) (2006) 270–278, http://dx.doi.org/10.1109/TNS.2006.869826.
[26] K. Eguchi, et al., KamLAND Collaboration, First results from KamLAND: Evidence for reactor anti-neutrino disappearance, Phys. Rev. Lett. 90 (2003) 021802, http://dx.doi.org/10.1103/PhysRevLett.90.021802.
[27] O. Tajima, Development of Liquid Scintillator for a Large Size Neutrino Detector (Master's thesis), Tohoku University, 2000.
[28] Oleg Perevozchikov, Search for Electron Antineutrinos from the Sun with KamLAND Detector (Ph.D. thesis), University of Tennessee, 2009.
[29] O. Tajima, Measurement of Electron Anti-Neutrino Oscillation Parameters with a Large Volume Liquid Scintillator Detector, KamLAND (Ph.D. thesis), Tohoku University, 2003.
[30] Christopher Grant, A Monte Carlo Approach to ^{7}Be Solar Neutrino Analysis with KamLAND (Ph.D. thesis), University of Alabama, 2012.
[31] Y. Abe, et al., Double Chooz Collaboration, Ortho-positronium observation in the Double Chooz Experiment, J. High Energy Phys. 10 (2014) 32, http://dx.doi.org/10.1007/JHEP10(2014)032.
[32] J.L. Raaf, A.O. Bazarko, G.T. Garvey, E.A. Hawker, R.A. Johnson, G.B. Mills, A. Pla-Dalmau, R. Tayloe, Mineral oil tests for the MiniBooNE detector, in: 2001 IEEE Nuclear Science Symposium Conference Record, Cat. No.01CH37310, vol. 1, 2001, pp. 564–568, http://dx.doi.org/10.1109/NSSMIC.2001.1008520.
[33] KamLAND Collaboration, Private communication, 2018.
[34] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958, URL http://jmlr.org/papers/v15/srivastava14a.html.
[35] Y. LeCun, Y. Bengio L. Bottou, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, http://dx.doi.org/10.1109/5.726791, URL https://ieeexplore.ieee.org/document/726791/.
[36] François Chollet, et al., Keras, 2015, https://keras.io.
[37] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, Xiaoqiang Zheng, Tensorflow: Large-scale machine learning on heterogeneous systems, 2015, Software available from tensorflow.org. URL https://www.tensorflow.org/.
[38] Bengio Yoshua Bergstra James, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (2012) 281–305, URL http://www.jmlr.org/papers/v13/bergstra12a.html.
[39] J. Bergstra, D. Yamins, D.D. Cox, Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, in: Proceedings of the 30th International Conference on International Conference on Machine Learning - Vol. 28, ICML'13, JMLR.org, 2013, pp. I–115–I–123, URL http://dl.acm.org/citation.cfm?id=3042817.3042832.
[40] Shie Mannor, Dori Peleg, Reuven Rubinstein, The cross entropy method for classification, in: Proceedings of the 22nd International Conference on Machine Learning, ICML '05, ACM, New York, NY, USA, 2005, pp. 561–568, http://dx.doi.org/10.1145/1102351.1102422, URL http://doi.acm.org/10.1145/1102351.1102422.
[41] Taco S. Cohen, Mario Geiger, Jonas Köhler, Max Welling, Spherical CNNs, CoRR (2018) abs/1801.10130. URL http://arxiv.org/abs/1801.10130.
[42] S. Andringa, et al., Current status and future prospects of the sno+ experiment, Adv. High Energy Phys. 2016 (6194250) (2016) http://dx.doi.org/10.1155/2016/6194250.
[43] R. Pordes, et al., The open science grid, J. Phys. Conf. Ser. 78 (1) (2007) 012057, URL http://stacks.iop.org/1742-6596/78/i=1/a=012057.
[44] Igor Sfiligoi, Daniel C. Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, Frank Wurthwrin, The pilot way to grid resources using glideinWMS, WRI World Congr. 2 (2009) 428–432, http://dx.doi.org/10.1109/CSIE.2009.950.