A Search for Double Beta Decay of Xenon to Excited States of Barium with EXO-200

Sereres Johnston

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Elementary Particles and Fields and String Theory Commons, and the Nuclear Commons

Recommended Citation

https://scholarworks.umass.edu/dissertations_2/881

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
A SEARCH FOR DOUBLE BETA DECAY
OF XENON TO EXCITED STATES
OF BARIUM WITH EXO-200

A Dissertation Presented
by
SERERES JOHNSTON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2017

Physics
A SEARCH FOR DOUBLE BETA DECAY
OF XENON TO EXCITED STATES
OF BARIUM WITH EXO-200

A Dissertation Presented
by
SERERES JOHNSTON

Approved as to style and content by:

________________________
Andrea Pocar, Chair

________________________
Michael Ramsey-Musolf, Member

________________________
Krishna Kumar, Member

________________________
Stella Offner, Member

________________________
Rory Miskimen, Department Chair
Physics
ACKNOWLEDGMENTS

I’m so grateful to so many to finally be here today. However some require special mention, those who were instrumental in my ability to complete this work, and those who were personally close to me in emotional support. Firstly, I want to thank the two research advisers I have had during my years in grad school—Andrea Pocar who invited me into EXO-200 work and who was always there with guidance and support when a project wasn’t going how I expected. I cannot thank you enough. Also Krishna Kumar, my adviser through the first two years. It is hard to balance interesting research projects and national laboratory trips with early graduate responsibilities such as passing qualifying exams and teaching, but you did it wonderfully.

I also want to thank my fellow University of Massachusetts colleagues. Tim Daniels, without whom I would be hopeless with any number of skills such as welding, cryogenics, and gas systems. Sara Feyzbakhsh, who worked late into the night with me during collaboration meetings creating beautiful slides, and who encouraged me when tasks felt impossible. My two undergraduate students, Beryl Bell and Javier King who diligently produced far more than I expected undergraduate students would have time for, and whose infections enthusiasm always lifted my spirits.

Thirdly, I want to thank all my EXO-200 collaborators, particularly Scott Kravitz, Yung-Ruey Yen, and Dave Moore who as members of the Excited States group. It was a consistent source of motivation seeing an “excited states” email come in— it said excited, I better be excited! Without you all, I would never be here now. The expertise displayed in machine learning, fitting, and other techniques is extraordinary and the kindness with which you responded to queries is truly amazing. There are so many people I will remember forever in EXO-200, particularly everyone who I shifted
with at WIPP. What a way to turn work into a summer-camp feel at times, with barbecues and zoo trips mixed with hard work.

Lastly, but not least, I want to thank those closest to me. Andrew Hashem, who encouraged me whenever everything was broken and seemed impossible, who lived with the mess the apartment became when I was thinking of nothing but writing, who believed in me so much I just had to keep going. Dana Johnston, who was always just a phone call away to give the best mom’s advice. Sarah and Helen Johnston, who helped me with keep revising, and revising, until I was finally done. Thank you also for your graphic design ideas! Many diagrams I have made would not have been nearly as good without your help.

I am so glad to have shared this time with all of you.
ABSTRACT

A SEARCH FOR DOUBLE BETA DECAY
OF XENON TO EXCITED STATES
OF BARIUM WITH EXO-200

FEBRUARY 2017

SERERES JOHNSTON
B.Sc., ANDREWS UNIVERSITY
M.Sc., UNIVERSITY OF MASSACHUSETTS
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Andrea Pocar

This thesis presents searches for several modes of double beta decays of $^{136}\text{Xe}$ to two low-lying excited levels of $^{136}\text{Ba}$. For each final level, both $2\nu$ and $0\nu$ double beta decay processes are considered. The data and general techniques developed by the EXO-200 collaboration are used, including a Machine Learning process to improve sensitivity in multivariate space. EXO-200 is an experimental program searching for $0\nu\beta\beta$ decay in a time projection chamber filled with 175 kg of liquid Xenon enriched to 80% $^{136}\text{Xe}$, functioning as both source and detector. Experimental searches of double beta decay with $^{136}\text{Xe}$ and other isotopes are motivated by non Standard Model neutrinoless decay ($0\nu\beta\beta$) possibilities. If found this would imply the Majorana nature of neutrinos, provide access to the currently unknown absolute mass scale of the neutrino, and violate lepton conservation by 2 units. Nuclear matrix elements necessary to connecting observed neutrinoless half-life to the neutrino mass scale
depend on nuclear models. Observation of both the $2\nu$ 0+ excited and 0+ ground state in a single element may constrain these models. Sensitivity studies of search performance are described, and final sensitivities from $7.86 \cdot 10^{23}$ years to $2.00 \cdot 10^{25}$ years are reported.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. NEUTRINOS</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Neutrino History</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Neutrino Chirality</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Neutrino Oscillations</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Absolute Neutrino Mass Limits</td>
<td>7</td>
</tr>
<tr>
<td>1.5 Motivation for Majorana Neutrinos</td>
<td>12</td>
</tr>
<tr>
<td>1.6 Brief History of 0νββ Experimental Search</td>
<td>14</td>
</tr>
<tr>
<td>1.7 Predicted Rates of Decay to Excited States</td>
<td>16</td>
</tr>
<tr>
<td>2. THE EXO-200 EXPERIMENT</td>
<td>20</td>
</tr>
<tr>
<td>2.1 Time Projection Chamber</td>
<td>21</td>
</tr>
<tr>
<td>2.1.1 Detector Regions</td>
<td>22</td>
</tr>
<tr>
<td>2.1.2 Energy Deposition</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Xenon Purification and Recirculation</td>
<td>27</td>
</tr>
<tr>
<td>2.3 Background Control</td>
<td>29</td>
</tr>
<tr>
<td>2.3.1 Muon Veto System</td>
<td>29</td>
</tr>
<tr>
<td>2.3.2 Waste Isolation Pilot Plant</td>
<td>31</td>
</tr>
<tr>
<td>2.4 Calibration System</td>
<td>32</td>
</tr>
<tr>
<td>2.5 EXO-200 Status and Results</td>
<td>34</td>
</tr>
</tbody>
</table>
2.5.1 Including Run 2d ......................................................... 37

3. SIGNAL STRUCTURE AND ANALYSIS ............................. 42
   3.1 Decay Signature of $2
   \nu\beta\beta$ to the $0^+_1$ Excited State ......................... 43
   3.2 More Decay Mode Signatures ...................................... 48
   3.3 Coding Toolkits .................................................. 50
   3.4 Boosted Decision Trees .......................................... 51
   3.5 Fitting Process .................................................. 53

4. MACHINE TRAINING .................................................... 59
   4.1 Input Variables ................................................... 62
      4.1.1 Standard Variables ......................................... 63
      4.1.2 Decay Specific Variables .................................. 64
      4.1.3 Pre-Processing ............................................. 67
   4.2 Training and Applying Machine .................................. 68
      4.2.1 Building the Training Data Set ............................ 70
      4.2.2 Training and Testing the BDT Algorithm ................ 71
      4.2.3 Applying BDT Algorithm .................................. 75
   4.3 Shape Systematic Studies ....................................... 75
      4.3.1 Agreement with Calibration Source Data ................ 76
      4.3.2 Unskewing PDFs ............................................ 79
   4.4 Sensitivity Studies ............................................... 82
      4.4.1 Dominant Fit Contributions ................................ 83

5. EFFICIENCY AND SENSITIVITY .................................... 85
   5.1 Reconstruction Efficiency ....................................... 87
   5.2 Sensitivity ..................................................... 91
   5.3 Data Selection .................................................. 92
   5.4 Subtracted Low Background Data Agreement ................... 93

6. CONCLUSIONS ......................................................... 97

APPENDIX: ADDITIONAL PLOTS ....................................... 99
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>List of isotopes with high Q value which $\beta\beta$ decay. The first reported error is statistical and the second is systematic. The first column of results is the two neutrino decay to ground state half-life, the second is the limit on neutrinoless decay to ground state. The last column is the two neutrino decay to the first $0^+$ excited state level. $^{136}\text{Xe}$ is both the longest lived, and most precisely measured.</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>Start and end dates for each of the four periods making up the Run2 data. Run2 contains a total of 585.8 days livetime collected between October 5, 2011 and February 5, 2014.</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>List of background components to all double beta decay modes studied by the EXO-200 collaboration. Background locations generating these components are also listed. These backgrounds are extensively simulated, and these simulated backgrounds used to train a BDT machine learning algorithm to distinguish background from signal in each of the decay modes ending in an excited state of $^{136}\text{Ba}$.</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>List of the standard event variables used as BDT algorithm input. Multiplicity, energy, and standoff are used for all decay modes searches.</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>List of the event variables used as BDT algorithm input for all four decay modes. Target energy shown in first column, with the search function for each listing denoted by the particle type in the labels. The 1639.3 keV energy target is used both for the beta clusters produced in the $0\nu$ to $2_1^+$ decay and the $\beta$ plus $\gamma$ combination clusters produced by the $0\nu\nu$ to $0_1^+$ decay.</td>
<td>66</td>
</tr>
<tr>
<td>4.3</td>
<td>Ranking of input variables in the final boosted decision tree built for the $2\nu\beta\beta$ to $0_1^+$ excited state search with partial 3D-reconstruction (at least 75% event energy 3D reconstructed required).</td>
<td>74</td>
</tr>
</tbody>
</table>
4.4 List of shared constraints used to account for systematic uncertainties in the final fit regardless of decay mode. The values for the Excited State Normalization, which varies for each search, are shown in Table 4.6.

4.5 Unskewing Study Summary. 200 signal events of each signal type are artificially added to 1,000 simulated Low Background data sets. Histograms of best fit values plotted in appendix. Means presented here. Deviation from 200 shows magnitude of systematic error on the Excited Signal Normalization.

4.6 Excited state normalizations calculated via source agreement and unskewing for each decay mode studied.

4.7 Sensitivity Study Summary. 1,000 simulated Low Background data sets with no signal are produced for discriminator variables drawn from each decay search mode. A Profile Likelihood scan is done over the number of signal events accommodated in the fit within the 90% confidence level. Histograms of the limit are found plotted in appendix. Medians presented here. Lower values correspond to a stronger limit.

5.1 Total signal efficiency for search modes by data selection cut. Significant efficiency improvement from relaxed data cut regardless of decay mode, with particular increases seen for the $0\nu$ to $0^+$ case.

5.2 Fractional increase in sensitivity by relaxing reconstruction requirement. Definite improvement seen for each mode. The $0\nu$ decay modes show nearly three times the sensitivity found when requiring 100% reconstruction, and in the $2\nu$ modes sensitivity roughly doubles.

5.3 Table of data cuts applied to Low Background data. A subset of these apply to Source Data and MC generated simulation, as described in text.

6.1 Half Life Sensitivity in context. Published values from EXO-200 [39] and Kamland-Zen [63] compared with sensitivity values from this analysis.

A.1 Ranking of input variables in the final boosted decision tree built for the $2\nu\beta\beta$ to $0^+_1$ excited state search with full 3D-reconstruction (100% event energy 3D reconstructed required).
A.2 Ranking of input variables in the final boosted decision tree built for
the $2\nu\beta\beta$ to $2^{+}_{1}$ excited state search with full 3D-reconstruction
(100% event energy 3D reconstructed required). .......................... 110

A.3 Ranking of input variables in the final boosted decision tree built for
the $0\nu\beta\beta$ to $0^{+}_{1}$ excited state search with full 3D-reconstruction
(100% event energy 3D reconstructed required). .......................... 111

A.4 Ranking of input variables in the final boosted decision tree built for
the $0\nu\beta\beta$ to $2^{+}_{1}$ excited state search with full 3D-reconstruction
(100% event energy 3D reconstructed required). .......................... 112

A.5 Ranking of input variables in the final boosted decision tree built for
the $2\nu\beta\beta$ to $0^{+}_{1}$ excited state search with partial 3D-reconstruction
(at least 75% event energy 3D reconstructed required). .................... 113

A.6 Ranking of input variables in the final boosted decision tree built for
the $2\nu\beta\beta$ to $2^{+}_{1}$ excited state search with partial 3D-reconstruction
(at least 75% event energy 3D reconstructed required). .................... 114

A.7 Ranking of input variables in the final boosted decision tree built for
the $0\nu\beta\beta$ to $0^{+}_{1}$ excited state search with partial 3D-reconstruction
(at least 75% event energy 3D reconstructed required). .................... 115

A.8 Ranking of input variables in the final boosted decision tree built for
the $0\nu\beta\beta$ to $2^{+}_{1}$ excited state search with partial 3D-reconstruction
(at least 75% event energy 3D reconstructed required). .................... 116
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Current and projected sensitivity of EXO-200. Shaded band width driven by uncertainty in NME calculations. Allowed values of $m_{\text{min}}$ given $\langle m_\nu \rangle$ shown in colored lines on left and right halves. Normal hierarchy is shown in red on the left, with the inverted hierarchy on right in blue. From [20].</td>
<td>8</td>
</tr>
<tr>
<td>1.2 Double beta decay candidates $^{76}\text{Ge}$ and $^{136}\text{Xe}$ connected via nuclear matrix elements (NMEs) calculated by several different models. Sensitivities and limits from recent experimental publications shown. From [20].</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Plotted is the nuclear binding energy of atomic number 136 isotopes. The lowest energy isotope, $^{136}\text{Ba}$ has been set to zero to highlight the energy differences between the isotopes. It is not energetically allowed for $^{136}\text{Xe}$ to singly decay to $^{136}\text{Cs}$, but it can decay to $^{136}\text{Ba}$, with a Q value (energy difference) for the double beta decay of 2457.83 keV [73].</td>
<td>17</td>
</tr>
<tr>
<td>2.1 Diagram of EXO-200 Time Projection Chamber (TPC). The planes of circles on either end are the Avalanche Photo-Diodes (APDs) at the endcaps and the cross hatching are the wire planes. Also shown is a depiction of energy deposition, including ionization and scintillation produced via ion recombination. The central dotted line denotes the position of the cathode, held at a high negative voltage of 8kV. From [25]</td>
<td>21</td>
</tr>
<tr>
<td>2.2 Electron lifetimes for TPC1 (a) and TPC2 (b) during the first portion of Run2. Black points indicate measured values from a single source calibration run, while the blue bands show piece-wise polynomial fits to these values. Vertical dashed lines indicate discontinuities in xenon recirculation, which reduces lifetimes due to introduction of stored xenon until purity recovers. The shaded regions, where purity is increasing after an interruption of normal operation, are excluded from the final dataset. From [30].</td>
<td>25</td>
</tr>
</tbody>
</table>
2.3 Plotted is the charge vs. light energy for a calibration run with a Thorium source. The 2.615 MeV gamma line used for energy calibration shows up as the oval in the upper right quadrant. The energy resolution is markedly improved along the rotated energy axis, corresponding to the narrowest one-dimensional projection of the oval. From [77].

2.4 Simplified diagram of xenon plumbing, including the detector vessel, the recirculation system, and the storage bottles. From [25].

2.5 Simplified drawing of EXO-200 detector. The copper vessel is central, surrounded by an HFE-containing cryostat. Lead bricks provide further shielding. The vessel, cryostat, and lead bricks are contained within a clean room. Around the clean room are the muon veto panels. From [27].

2.6 Photo of author in front of EXO-200 equipment underground at WIPP. The large tank system in the center-right of the photo is the thermal-store system, which provides longer thermal control in the event of power outages in the mine. Past the container in the background on the left is the entrance to the cleanrooms.

2.7 Shown here is the relative energy resolution ($\sigma/E$, percent) vs energy. Multi Site (MS) and Single Site (SS) data is plotted separately. Improved energy resolution after the de-noising process is demonstrated. From [38].

2.8 Cartoon of two cluster photon event, with cone drawn around an axis defined via Compton scattering angle.

2.9 Image produced by resolving Compton scatters during a calibration run on Oct 3, 2011. There was a contradiction between the position described by a note left by shifters on duty that day, and the assigned position tag. The Compton imaging method described in the text easily resolves the conflict and confirms the true position of this run.

2.10 Run2abc data accumulation as a function of time. Blue is golden, red includes non-golden data taking, including calibration and noisy or low purity runs, black is real time. This is the data used in the $0\nu\beta\beta$ analysis [31].
2.11 Black points raw data, fitted lines are PDFs of backgrounds generated via MC, inset is the region of interest. Shaded grey is the 2$\nu$ββ $^{136}$Xe background to the magenta 0$\nu$ decay signal. A two dimensional fit over energy, shown here, and standoff distance, a measure of centrality of events, is performed. SS and MS spectra are simultaneously fit, with the ratio constrained via simulation. Only the SS fit is shown here. A limit of $1.1 \cdot 10^{25}$ yr. at a 90% CL is found. From [31].

2.12 Profile scan over number of counts attributed to the 0$\nu$ββ decay in the fit over the Run2abc dataset. The minimum is at the best-fit 0$\nu$ββ value of 9.9 counts. The 1$\sigma$ and 90% confidence level lines are shown. The best-fit value is consistent with zero at 1.2$\sigma$. From [31].

3.1 Double beta decay of $^{136}$Xe to $^{136}$Ba excited state, and relevant energy levels of $^{136}$Ba. ββ decay to the excited 0$^+$ state of Ba is energetically allowed, with a Q value of 879 keV. Two de-excitation γs, of 760.5 keV and 818.5 keV are emitted promptly, and cannot be resolved in time in EXO-200. Decays to the two 2$^+$ levels are also energetically allowed, though only the search for the decay to the lower 2$^+$ level is described in this thesis.

3.2 Number of clusters per event containing any energy attributable to each of the two de-excitation γs involved in the 2$\nu$ββ to the 0$^+_1$ excited state decays. Data from a small, 5,000 event simulation of this decay process with tracing from the simulated γ depositions identifying the contributions to each cluster in these events.

3.3 Angular correlation function, W(θ), between emitted de-excitation photons. Photons are preferentially emitted either directly opposite each other, or in the same direction.

3.4 Five thousand 2$\nu$ββ decays to the first 0$^+$ excited state are simulated at a central point in the EXO-200 detector. The de-excitation photons are forced to emerge anti-parallel to each other, with the direction of the larger 818 keV γ in the positive z direction and the smaller 760 γ in the opposite direction. Photon clusters are relatively simply separated by position.

3.5 Five thousand 0+ decay events with forced parallel de-excitation gamma correlation. Photon clusters not simply separated by position.
3.6 Ten million simulated $2\nu\beta\beta$ to $2^+_1$ excited state decays. Total event energy plotted, smeared with realistic detector resolution derived from EXO-200 calibration data. Beta decay shape and photon peak visible. .................................................. 49

3.7 Ten million simulated $2\nu\beta\beta$ to $0^+_1$ excited state decays. Total event energy plotted, smeared with realistic detector resolution derived from EXO-200 calibration data. Beta decay shape and photon peaks visible. .................................................. 50

3.8 Illustration of decision tree. Increasing signal proportion denoted by pinker hues, as well as the values printed in each circle. Node level denoted by the cut number, branch by cut letter. At each decision node, the cut that maximally separates signal and background among input variables is selected. ........................................... 52

3.9 Normalized PDFs for two data selection cuts for the $0\nu\beta\beta$ to $0^+_1$ excited state decay mode. Multi-site energy spectrum. Full reconstruction requirement plotted in red, with 75% reconstruction requirement in blue. Relaxing the spatial reconstruction requirement of MS events allows proportionally more events in the full Q value peak to be accepted compared to the strict full reconstruction requirement. This can be seen by the higher blue relaxed PDF line at the 2458 keV Q value as compared to the red, fully reconstructed PDF. ......................... 56

3.10 Daily activity for each background component in the golden quality EXO-200 data. Single-site spectrum. Two distinct fits are displayed. In blue, is the final fit from the published search for Majoron-emitting double-beta decay modes [32], which covers Run2abc. In black, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $0\nu\beta\beta$ to $0^+_1$ search. These rates come from spectral fits over the entire data set, with the $2\nu\beta\beta$ component listed referring to $^{136}\text{Xe}$ decays to the ground state only. .............................. 57

3.11 Daily activity for each background component in the golden quality EXO-200 data. Multi-site spectrum. Two distinct fits are displayed. In blue, is the final fit from the published search for Majoron-emitting double-beta decay modes [32], which covers Run2abc. In black, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $0\nu\beta\beta$ to $0^+_1$ search. These rates come from spectral fits over the entire data set, with the $2\nu\beta\beta$ component listed referring to $^{136}\text{Xe}$ decays to the ground state only. .............................. 58
4.1 Complete Machine Learning Pipeline. Parallel processes operate from top to bottom on data (left) and MC (right). Weighted decision trees, the output of the BDT machine learning algorithm, are produced from training on a MC-generated Toy Dataset and are applied to both data and MC to fill the discriminator variable. This process is shown in magenta with a dashed-arrow web. Source agreement, including the production of the $2\nu\beta\beta$ subtracted dataset used as a source comparison to MC, is cyan with a dash-dot arrow web. The final fit process is yellow with a solid-arrow web. Data and MC files are filled with discriminator variables from the BDT. Probability distribution functions (PDFs) of the fitting variables are then built from MC while an EXO workspace is made to hold all data events. The final fit applies the PDFs to the dataset, while the sensitivity process first generates MC simulations of a dataset to measure typical fit performance.

4.2 Distribution of all variables used for the $0\nu\beta\beta$ to $0^{+}_1$ state decay search imposing that every cluster in every event is fully reconstructed in space. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the collected data set is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $0^{+}_1$ decay-specific variables shown in the second row. Similar plots for the other search modes are found in the Appendix.

4.3 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $0^{+}_1$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. Training and testing data are more similar than expected from poissonian statistics because the fully-trained algorithm uses weights from the testing data while combining individual decision trees. Similar plots for other search modes are shown in the Appendix Figures 7.10-7.15.
4.4 “Source agreement” plot for the subtracted $2\nu\beta\beta$ decay to $^{136}\text{Ba}$
ground state source constructed as described in the text.
Distribution of discriminator variable, constructed from machine
learning algorithm, shown. Discriminator variable has been
trained for the $0\nu\beta\beta$ to $2^+_1$ excited state decay mode search.
Multi-Site component. .............................................. 78

4.5 An example profile scan over the number of signal events in the final
fit is shown on the left. This profile is one out of a thousand
simulated low background dataset built with the full
reconstruction requirement and fit on the discriminator variable
trained for maximum sensitivity to the $2\nu\beta\beta$ to $0^+_1$ decay mode,
and was selected at random. No signal events were used in the
creation of this dataset. The number of counts at the 90% confidence level for 1,000 of these simulated datasets is combined
in the histogram on the right to determine sensitivity. .............. 83

5.1 Normalized multiplicity by decay mode, compared to Run2 Low
Background data. Color coding in legend. Low background data
clearly shows lower average multiplicity than any signal. The
dominant structure for the decays to excited states is driven by
the final state level, with the two de-excitation $\gamma$'s emitted by the
$0^+_1$ decay mode pushing both of these peaks to three as compared
with the single de-excitation $\gamma$ from the $2^+_1$ decay. A smaller effect
is shown between zero and two neutrino modes, with the higher
average $\beta\beta$ cluster for both $0\nu$ modes causing a slight shift to
higher multiplicity as compared with the $2\nu$ case. ............... 86

5.2 Simulated $2\nu\beta\beta$ to $2^+_1$ excited decay mode events by efficiency of 3D
reconstructed energy. Multiplicity 2 through 5 are color coded
and shown on legend. Efficiency falls off monotonically with
increasing multiplicity. Events with multiplicity greater than five
are grouped together and shown as one additional line. ............. 88

5.3 Simulated $2\nu\beta\beta$ to $0^+_1$ excited decay mode events by efficiency of 3D
reconstructed energy. Multiplicity 2 through 5 are color coded
and shown on legend. Efficiency falls off monotonically with
increasing multiplicity. Events with multiplicity greater than five
are grouped together and shown as one additional line. ............. 89

5.4 Simulated $0\nu\beta\beta$ to $2^+_1$ excited decay mode events by efficiency of 3D
reconstructed energy. Multiplicity 2 through 5 are color coded
and shown on legend. Efficiency falls off monotonically with
increasing multiplicity. Events with multiplicity greater than five
are grouped together and shown as one additional line. ............. 90
5.5 Simulated $0\nu\beta\beta$ to $0^+_1$ excited decay mode events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line. 

5.6 Run2 Low Background events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line.

5.7 Effect of varying fraction of 3D reconstructed energy requirement per event in all excited decay modes and the low background data. Only multiplicity 2 events are shown.

5.8 Effect of varying fraction of 3D reconstructed energy required per event in all excited decay modes and the low background data. Only multiplicity greater than 5 events are shown.

5.9 $2\nu\beta\beta$ “calibration source” dataset created by subtracting fitted backgrounds from Low Background Data. Part (a) includes only Run2abc data, while part (b) includes all Run2 data. Shown is a projection of a 2D PDF onto the energy variable. Multi-site events only. Visible discrepancy near $^{40}\text{K}$ peak.

A.1 Daily activity for each background spectrum in the golden quality EXO-200 data. Single-site spectrum. Two distinct fits are displayed. In blue, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $\beta\beta0\nu$ to $0^+_1$ search with full 3D reconstruction requirement. In black is a fit over the whole Run2abcd dataset using fitting variables designed for the same mode, but at least 75\% 3D reconstruction. Absolute daily activity for numbers in fit, not normalized for different efficiency.

A.2 Daily activity for each background spectrum in the golden quality EXO-200 data. Multi-site spectrum. Two distinct fits are displayed. In blue, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $\beta\beta0\nu$ to $0^+_1$ search with full 3D reconstruction requirement. In black is a fit over the whole Run2abcd dataset using fitting variables designed for the same mode, but at least 75\% 3D reconstruction. Absolute daily activity for numbers in fit, not normalized for different efficiency.
A.3 Distribution of all variables used for the $2\nu\beta\beta$ to $0_1^+$ state decay search using a strict event cut requiring 100% of the event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $\beta\beta 2\nu$ to $0_1^+$ decay-specific variables shown in the second row.

A.4 Distribution of all variables used for the $2\nu\beta\beta$ to $2_1^+$ state decay search using a strict event cut requiring 100% of the event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $2\nu\beta\beta$ to $2_1^+$ decay-specific variables shown in the second row.

A.5 Distribution of all variables used for the $0\nu\beta\beta$ to $0_1^+$ state decay search using a strict event cut requiring 100% of event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $0_1^+$ decay-specific variables shown in the second and third rows.
A.6 Distribution of all variables used for the $0\nu\beta\beta$ to $2^+_1$ state decay search using a strict event cut requiring 100% of the event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $2^+_1$ decay-specific variables shown in the second row.

A.7 Distribution of all variables used for the $2\nu\beta\beta$ to $0^+_1$ state decay search using the relaxed event cut allowing up to 25% of the event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $2\nu\beta\beta$ to $0^+_1$ decay-specific variables shown in the second row. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.

A.8 Distribution of all variables used for the $2\nu\beta\beta$ to $2^+_1$ state decay search using the relaxed event cut allowing up to 25% of the event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $2\nu\beta\beta$ to $2^+_1$ decay-specific variables shown in the second row. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.
A.9 Distribution of all variables used for the $0\nu\beta\beta$ to $0^+_1$ state decay search using the relaxed event cut allowing 25% of event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $0^+_1$ decay-specific variables shown in the second and third rows. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.

A.10 Distribution of all variables used for the $0\nu\beta\beta$ to $2^+_1$ state decay search using the relaxed event cut allowing up to 25% of the event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $2^+_1$ decay-specific variables shown in the second row. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.

A.11 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $0^+_1$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

A.12 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $2^+_1$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

A.13 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $0\nu\beta\beta$ to $0^+_1$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.
A.14 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $2_1^+$ excited state search with at full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. ............... 112

A.15 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $0_1^+$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. .................................................. 113

A.16 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $2_1^+$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. .................................................. 114

A.17 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $0\nu\beta\beta$ to $0_1^+$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. .................................................. 115

A.18 Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $0\nu\beta\beta$ to $2_1^+$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. .................................................. 116

A.19 100% reconstructed data selection. Excited decay normalization study– MC PDF’s “unskewed” to artificially agree with calibration source data. Subtracted $2\nu\beta\beta$ to ground state source is not “unskewed.” 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0_1^+$ daughter, right column $2_1^+$ daughter. .................................................. 117
A.20 100% reconstructed data selection. Excited decay normalization study– MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2\nu\beta\beta$ to ground state source agreement are used. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^+_1$ daughter, right column $2^+_1$ daughter.  

A.21 100% reconstructed data selection. Excited decay normalization study– MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2\nu\beta\beta$ to ground state source agreement used. $2\nu\beta\beta$ subtraction altered to remove 140 fewer K40 MS events. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^+_1$ daughter, right column $2^+_1$ daughter. 

A.22 100% reconstructed data selection. Excited decay normalization study– MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2\nu\beta\beta$ to ground state source agreement used. $2\nu\beta\beta$ subtraction altered to remove 140 more K40 MS events. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^+_1$ daughter, right column $2^+_1$ daughter. 

A.23 75% reconstructed data selection. Excited decay normalization study– MC PDF’s “unskewed” to artificially agree with calibration source data. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^+_1$ daughter, right column $2^+_1$ daughter.
A.24 75% reconstructed data selection. Excited decay normalization study– MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted 2νββ to ground state source agreement used. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row 2ν decays, lower row 0ν decays. Left column 0† daughter, right column 2† daughter. 

A.25 Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. No systematic errors included, simulated dataset statistics control. 100% reconstruction selection requirement.

A.26 Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All historic systematic errors included. Excited state normalization error not included. 100% reconstruction selection requirement.

A.27 Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All systematic errors included. Final sensitivity for mode. 100% reconstruction selection requirement.

A.28 Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. No systematic errors included, simulated dataset statistics control. 75% reconstruction selection requirement.

A.29 Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All historic systematic errors included. Excited state normalization error not included. 75% reconstruction selection requirement.

A.30 Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All systematic errors included. Final sensitivity for mode. 75% reconstruction selection requirement.
Neutrinos are extremely low mass electrically neutral elementary particles with \( \frac{1}{2} \) integer spin [60]. Neutrinos are weakly-interacting leptons like their electrically charged and much more massive cousins the electrons, muons [46] and taus [22]. Because of their minimal mass, their lack of color charge, and electric neutrality, they essentially only interact via the weak nuclear force—gravity having very little effect on their tiny masses—and the strong nuclear and electromagnetic forces requiring charges neutrinos do not carry [6]. However, due to the observed conservation of lepton number, neutrinos are involved in many weak force processes such as $\beta$ decays and muon and tau decays.

Once generated, neutrinos have a very small cross section and interact rarely with even the densest materials [56]. For reactor neutrinos of energies 3-11 MeV the measured neutrino cross section is only $6.3 \cdot 10^{-44} \text{cm}^2$, nineteen orders of magnitude below the standard “barn” cross section of neutrons on uranium [13]. This small cross section corresponds to a mean free path in lead of 2.5 light years. The neutrino cross section does generally increase with energy, but even at its highest, the cross section for neutrino scattering off of free electrons is only $10^{-4}$ barn. This peak occurs around $6 \cdot 10^{15} \text{eV}$ neutrinos, and is due to the $W^-$ resonance [5]. Only extra-galactic sources can generate such tremendous energies for a single neutrino [92, 71].

Though neutrinos’ low interaction rate introduces some experimental difficulties, it also provides unique opportunities. Observing neutrinos from solar, astrophysical, and geological sources provides access to study distant or inaccessible processes. Solar
neutrinos are produced by fusion reactions which generate tremendous amounts of energy in the core of the sun [8]. Photons radiating from the surface of the sun are not directly produced by these reactions, instead coming from energy that was absorbed and re-emitted for a hundred thousand years as it slowly worked its way out of the interior [11]. In contrast, the neutrinos detected from the sun come directly from fusion events, providing direct access to nuclear processes in the core [11].

Geological neutrinos are produced by decays of Uranium and Thorium deep within the earth’s crust [75]. Such radiological decays are modeled to provide about half the Earth’s total dissipated heat. Neutrino measurements by KamLAND in 2005 were consistent with the expected decay rates, though still fairly poorly constrained [61].

Cosmic neutrinos of vastly high energy, possibly deriving from active galactic nuclei or gamma-ray bursts, have been detected by the Ice Cube laboratory. They have detected 37 events with energies between 30 TeV and 2 PeV. One of these, at the highest end of the energy range, was consistent in time and space to possibly come from an observed gamma-ray blazar outburst. Ice cube is also prepared to detect supernova neutrinos and is actively studying other sources which provide more information on astrophysical phenomena [71, 70].

In addition to using neutrinos as a probe while studying otherwise inaccessible processes, the ability of neutrinos to pass through large distances of solid material can have current practical applications as well. One of these, long distance communication, was demonstrated by MINERvANuMI in 2012. Information was transferred via a neutrino beam through 1km of earth. Though data rates with this method are still very low at only .1 bits/sec, a neutrino beam capable of going directly through the earth would benefit from the shorter distance traveled compared with cables curving around the surface of the earth [76]. Other practical neutrino physics applications include long distance monitoring of nuclear weapon tests and nuclear reactors to confirm state adherence to non-proliferation treaties [15].
Despite what we have learned so far about neutrinos and our ability to use them to answer other physical questions or for practical applications, there is much we still don’t know about their basic nature. One of these mysteries is their mass—though we know they are massive from observing neutrino oscillations, we still do not know what their mass is. Another mystery ties in with their helicity behavior and neutrality. All observed neutrinos are left handed, with spin anti-aligned with momentum, while anti-neutrinos are right handed. Combined with their neutrality, this opens the possibility that neutrinos are Majorana particles, rather than Dirac, and thus do not have distinct particle and antiparticle states. Double beta decay study is one of the experimental ways to access these questions [81, 41].

1.1 Neutrino History

In 1911 Lise Meitner and Otto Hahn discovered that the emitted electrons in beta decay had a continuous spectrum [79]. At the time the only final state particles known in beta decays were the easily detectable electron and daughter nucleus. Conservation of energy for such a two body final state decay requires that the electron carry away all the energy difference between the parent and daughter nuclear states. However, the observed continuous electron spectrum is at odds with this assumption. This conflict hinted at some new fundamental physics, either an exception to a well-tested conservation law, or some novel process.

Decades later in 1930, as experimental evidence of continuous spectra piled up, Wolfgang Pauli postulated a “desperate remedy” to rescue conservation of energy [93]. His suggestion was that beta decay did not result in a two body final state, but an additional particle was emitted along with the electron. This particle would be a difficult to detect, electrically neutral fermion of very low mass which would carry away part of the energy such that the sum of the neutrino and the electron energies together would equal that expected from the energy of the beta decay [43]. The ex-
istence of the neutrino would await experimental verification for decades longer. In 1941, Ganchang Wang proposed using the inverse beta decay, where a neutrino is captured on a nuclear proton to generate a positron and neutron, to experimentally observe neutrinos [91]. This was the reaction observed by Cowen and Reines at the Savannah River Plant in 1956, where they used a nuclear reactor as a concentrated source of neutrinos. They observed the photons from the positron annihilation and the neutron capture in coincidence to reduce backgrounds, and this signal was not diminished by the addition of a stack of bags full of water-saturated sawdust around their detector, but backgrounds from gammas and neutrons were suppressed, providing proof that the observed signal was from penetrating neutrinos [13].

In 1962, researchers at Brookhaven [46] demonstrated that a beam of neutrinos generated from muon decay did not produce any electrons when subsequently interacting, but instead produced muons. This proved there were at least two different flavors of neutrinos, one linked to electrons, and the other with muons. If neutrinos were massless as was initially believed, neutrino flavors would remain distinct. In conflict with this expectation, the Homestake experiment detecting neutrinos generated in the sun only measured a third of the expected rate [83]. This could be explained by massive neutrinos oscillating out of the electron neutrino flavor produced in the sun, and the only flavor that Bahcall and Davis’s Homestake experiment was able to detect. Neutrino oscillations were first directly observed by Super Kamiokande in atmospherically generated neutrinos in 1998 [90]. Three years later, neutrino oscillations were definitively confirmed as the source for the solar neutrino deficit when the Sudbury Neutrino Observatory observed the total flux and the electron neutrino flux simultaneously, confirming both the expected calculated total rate and the electron neutrino deficit [88].

The distinction between neutrinos and anti-neutrinos has been demonstrated experimentally [6]. High-precision measurements of cross sections for both neutrino
and anti-neutrino interactions have been completed [74]. Prior to the demonstrated solution of flavor oscillations explaining the solar neutrino flux deficit, some theories proposed neutrino to anti-neutrino oscillations as a possible solution [85]. Though this was not the answer to the solar neutrino problem, the idea of allowed oscillations between matter and anti-matter carries fundamental interest for the ultimate nature of neutrinos, and will be addressed later in Section 1.5.

1.2 Neutrino Chirality

Observed neutrinos are always left-handed, with spin anti-aligned with momentum [60]. The distinction between neutrinos and anti-neutrinos is experimentally observed, with electron neutrinos generating electrons and electron anti-neutrinos generating positrons. However the non-zero mass of the neutrino introduces a puzzle to this observed distinction between neutrinos and anti-neutrinos, as the helicity of a massive particle can be flipped by a sufficiently boosted reference frame [41].

A neutral fermion such as the neutrino is allowed to be either Dirac, with distinct particle and antiparticle states, or Majorana, where the particle and antiparticle states are not distinct. The full Lagrangian mass term possible is thus the sum for these two allowed cases:

\[
\mathcal{L} = \mathcal{L}_D + \mathcal{L}_M
\]

\[
= -m_D[(\bar{\Psi}_L)\Psi_R + H.c] - \frac{m_L}{2}[(\bar{\Psi}_L^c)\Psi_L + H.c] - \frac{m_L}{2}[(\bar{\Psi}_R^c)\Psi_R + H.c] \tag{1.1}
\]

Here \(\mathcal{L}_D\) and \(m_D\) are Dirac terms, while \(\mathcal{L}_M\) and \(m_L\) are Majorana terms. In the case of Dirac neutrinos, there are four states of the same mass—left handed neutrinos and right handed anti-neutrinos, as observed, plus their Lorenz boosted partners—right handed neutrinos and left handed anti-neutrinos which have never
been seen [23]. If neutrinos are Majorana, however, the two observed states can be the only states existing, and neutrinos and their anti-particles are related only through a sufficiently boosted Lorentz frame. Note this is only possible for massive neutrinos; massless neutrinos move at the speed of light and it is not possible to change helicity with a simple Lorentz boost [41].

1.3 Neutrino Oscillations

The three mass eigenstates \( m_1, m_2 \) and \( m_3 \), differ from the weak force flavor eigenstates \( \nu_e, \nu_\mu, \nu_\tau \), allowing the observed neutrino flavor oscillations to occur [60]. Distinct force eigenstates are not unique to neutrinos, as weak force eigenstates of quarks differ from their mass and strong force eigenstates, allowing for decays of heavier quarks into lighter quarks [59, 42]. Massive leptons such as muons also decay, but the small mass differences between neutrino mass states allow them to oscillate between flavor states rather than simply decaying.

The neutrino flavor eigenstates can be written as:

\[
|\nu_f\rangle = \sum_i U_{fi} |\nu_i\rangle, \tag{1.2}
\]

where the neutrino on the left is that which participates in weak interactions, and the neutrinos on the right are the neutrino mass eigenstates. The precise value of the mixing between states is controlled by the differences between the neutrino masses, as well as phases. These are expressed in the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix \( U \) in the previous equation. \( U \) can be decomposed as follows:
\[ U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{i\delta} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{13} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\alpha/2} & 0 & 0 \\ 0 & e^{i\beta/2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \] (1.3)

Here \( c_{ij} = \cos(\theta_{ij}) \) and \( s_{ij} = \sin(\theta_{ij}) \), where \( \theta_{ij} \) is the mixing angle between the \( i \)'th and \( j \)'th state. Of the other three parameters, \( \delta \) is a Dirac CP violating phase, and \( \alpha \) and \( \beta \) are Majorana phases only present if neutrinos are Majorana [72]. Since these phases come in on the diagonal of \( U \), they cannot be used in oscillation experiments to determine the neutrinos Majorana mass. Other than the phase parameters, all the components of the \( U \) matrix have now been measured, most recently the \( \theta_{13} \) term at Daya Bay [21].

1.4 Absolute Neutrino Mass Limits

The observed neutrino oscillations show neutrinos do have mass, and can be used to calculate precise mass differences between the mass eigenstates. Oscillations cannot be used to access the absolute neutrino mass scale, which is not yet known, though limits from several methods do exist [49]. This section will discuss three methods and current world limits for each. The first method discussed is neutrinoless double beta decay measurements in selected candidate isotopes. Second is endpoint beta decay spectrum measurement. Lastly, astrophysical observations are discussed as a third method of accessing the absolute neutrino mass scale.

If neutrinos are Majorana fermions, as described in the section 1.5, neutrinoless double beta decay (0\( \nu \)\( \beta \beta \)) can occur. In the event this decay is mediated by light Majorana neutrino exchange, an observed neutrinoless half life is directly related to
Figure 1.1: Current and projected sensitivity of EXO-200. Shaded band width driven by uncertainty in NME calculations. Allowed values of $m_{\text{min}}$ given $\langle m_\nu \rangle$ shown in colored lines on left and right halves. Normal hierarchy is shown in red on the left, with the inverted hierarchy on right in blue. From [20].

The absolute neutrino mass. The value accessible by this method depends on terms from the PMNS matrix, including the Majorana CP phase differences within the $U$ term, as follows [81]:

$$\langle m_{\beta\beta} \rangle^2 \equiv \left| \sum_i U_{ei}^2 m_{\nu_i} \right|^2$$  (1.4)
The extraction of this mass scale from experimentally observed half life limits and neutrino mass limits requires models for nuclear matrix element calculations. The connection is controlled by the following equation:

\[
[T_{1/2}^{\beta\beta}]^{-1} = G^{\beta\nu}(E_0, Z) |M_{GT}^{\beta\nu}|^2 \frac{g_V^2}{g_A^2} M_F^{\beta\nu} |^2 (m_{\beta\beta})^2 \quad (1.5)
\]
Here $M_F$ and $M_{GT}$ are model-derived nuclear matrix elements (NMEs) of the initial and final nuclear structure and their transition. The uncertainty in the models deriving these elements means the effective mass limit from a single experimental half life limit becomes a band rather than a simple point. Bands corresponding to recent limits from the EXO-200 experiment, as well as projected future sensitivity, is shown in Figure 1.1 The most aggressive limits to the effective Majorana neutrino mass have recently been reached by KamLAND-Zen and are .06-.161 eV. [64]

Candidate isotope limits can be compared with each other via calculated NMEs, in the same way these mediate the absolute neutrino mass and the observed limit. A plot of one such comparison, between $^{136}$Xe and $^{76}$Ge is shown in Figure 1.2. It can clearly be seen that the different models add an uncertainty to the interpretation of the half life in terms of either neutrino masses, or an equivalent half life in a complementary isotope.

In the event that neutrinoless double beta decay is not mediated by light neutrino exchange but another more exotic mechanism, the straightforward method of linking an observed lifetime to the absolute neutrino mass demonstrated in Equation 1.5 is no longer valid. Furthermore, if neutrinos are Dirac particles without Majorana nature, neutrinoless double beta decay is disallowed entirely, and cannot be used to probe neutrino mass.

A second method of accessing the absolute neutrino mass scale, which is effective even in the case of a purely Dirac neutrino, requires close attention to the single beta decay spectrum. The spectrum of beta decay very near the endpoint Q-value differs depending on the neutrino mass. In the limit approaching the endpoint, the mass of the neutrino itself deforms the decay spectrum as the energies of the emitted neutrino become low enough to reach the non-relativistic region and the rest mass of the neutrino is significant when compared to its kinetic energy. The value probed by this method is [14, 49]:
Typically beta decay measurements focus on tritium due to the low value of the endpoint energy. At present, the limit with this method is held by the Mainz experiment with $m_\nu < 2.3$ eV [14].

Thirdly, mass limits can be calculated via astrophysical observations, including the Cosmic Microwave Background and galaxy distribution and lensing of galaxies. This method is not sensitive to the mixing of the neutrinos or any part of the PMNS matrix, rather the value measured is:

$$m_\nu = \sum_i^3 m_i$$

Though not reliant on any mixing parameters, the extraction of this value depends critically on the cosmological models used to transform the observed large-scale mass structure into an upper bound on the summed neutrino mass. Astrophysical limits on the neutrino mass from this approach range from 0.28 eV to 1.3 eV depending on method, with a world average of $m_{\text{astro}} = .71$ eV [53]. The observed value used for this calculation is simply the power spectrum of mass scales in the universe. This would have been affected by neutrinos from the Big Bang, as these relic neutrinos smear out dense structure at smaller scales first. The amount of smearing the neutrinos are capable of performing depends on their mass, as their free-streaming length scale depends on both their temperature, measurable with the CMB, and their mass.
The power spectrum of the smaller mass scales, which would be more affected, as compared to the larger structures above 100 MPC not being as strongly affected, can determine neutrino masses [58].

1.5 Motivation for Majorana Neutrinos

Neutrinos are electrically neutral, making them the only known fundamental particles that could possibly be Majorana fermions. Further motivation for this theory comes from the natural way Majorana neutrinos explain two mysterious neutrino properties: their extremely small mass, and the observed lack of right handed neutrinos. Lepton number conservation also naturally develops as a direct consequence of neutrino helicity and weak interaction parity-violation, with the possibility of a lepton number violating decay in rare cases [23, 41].

Briefly, a free neutrino field \( \Psi \) obeys the Dirac equations:

\[
(\gamma^\mu \frac{\partial}{\partial x^\mu} + m)\Psi = 0 \\
(\gamma^\mu \frac{\partial}{\partial x^\mu} - m)\gamma_5 \Psi = 0
\] (1.8)

Fields with definite chirality are eigenstates of \( \gamma_5 \). Particles with non-zero mass cannot achieve this, though energetic particles with extremely small masses can come close. The charge conjugate field \( \Psi^c \) is defined such that if \( \Psi \) is a chirality eigenstate, \( \Psi^c \) is as well, with opposite eigenvalue. The neutrino mass is determined by mass terms in the neutrino Lagrangian. These terms must be Lorenz invariant and hermitian, restricting the possible terms to \( \bar{\Psi} \Psi \), \( \bar{\Psi}^c \Psi^c \), and the cross terms \( \bar{\Psi}^c \Psi \) and \( \bar{\Psi} \Psi^c \). Then the Lagrangian can be written in matrix form:
$-L_M = \frac{1}{2}(\Psi, \bar{\Psi}^c) \begin{pmatrix} m_D & m_M \\ m_M^* & m_D \end{pmatrix} \begin{pmatrix} \Psi \\ \Psi^c \end{pmatrix}$ \hfill (1.9)

If extended to the chiral projections and then diagonalized, this Lagrangian ultimately leads to a natural scale of neutrino masses. The mass matrix to diagonalize becomes:

\[
\begin{pmatrix}
-\gamma & 0 & m_D & m_1 - im_2 \\
0 & -\gamma & m_1 + im_2 & m_D \\
m_D & m_1 + im_2 & -\gamma & 0 \\
m_1 - im_2 & m_D & 0 & -\gamma
\end{pmatrix}
\] \hfill (1.10)

By assigning $m_R = m_1 + |m_2|$ and $m_L = m_1 - |m_2|$ for left handed or right handed majorana masses, this can be block diagonalized and simplified to the following two dimensional mass matrix:

\[
\begin{pmatrix}
m_R & m_D \\
m_D & m_L
\end{pmatrix}
\] \hfill (1.11)

Grand unified theories assign $m_D \simeq m_{\text{fermion}}$ as the mass scale of the familiar, observed, lepton or quark masses. The $m_L$ term is assigned to near zero because of the very low observed mass of the familiar left-handed neutrino, and the $m_R$ term set very high, on the GUT scale between $10^{14} - 10^{16}$ GeV where it would be understandable.
that a right-handed Majorana neutrino would not yet have been seen. Finding the
eigenvalues of the mass matrix from Equation 1.11 with these assignments yields a
light neutrino described by:

\[ m_L \simeq \frac{m_{\text{fermion}}^2}{M_{\text{GUT}}} \]

Thus, the natural mass scales corresponding to the charged fermions, are pushed
to the extremely low observed masses, by the extremely high mass Majorana neu-
trino. This is known as the see-saw mechanism and is a compelling explanation for the
observed properties of the neutrino. In this theory lepton number is not a fundamen-
tal conservation which naturally emerges from the chirality states of the neutrinos,
because the non-zero mass of the neutrino means that such chirality states are not
themselves universal to all reference frames. One of the predictions of the see-saw
theory is the occurrence of neutrinoless double beta decay, $0\nu\beta\beta$. This decay would
break lepton number by 2, and possibly be mediated by a majorana neutrino.

The previous discussion describes the simplest version, seesaw Type I, with a
singlet massive right handed neutrino. In this case a $0\nu\beta\beta$ decay is mediated by a
light Majorana neutrino exchange with standard weak vertices. Other seesaw models
exist. Types include Type II, with a higgs triplet, and type III, with a fermion triplet
[84]. Other alternative mechanisms of $0\nu$ decay include Majoron emission, with a
recent limit on this decay process published by EXO-200 [32]. Regardless of the
device mechanism, any observed $0\nu\beta\beta$ decay implies a Majorana mass term for the
neutrino, according to the Schecter Valle theorem [55].

1.6 Brief History of $0\nu\beta\beta$ Experimental Search

The standard model two neutrino double beta decay was predicted in 1935 by
Maria Goppert Mayer [50]. Since then, and particularly since the neutrinoless double
beta decay mode was proposed [45], searches for double beta decay in various isotopes
has spawned an active field of research. While no confirmed neutrinoless decays have been observed, the two neutrino mode has been measured for ten isotopes which can undergo double beta decay and also have a high endpoint energy, or Q-value. Having a high Q-value is necessary as this can reduce the backgrounds from other processes by reducing the number of other processes that are high enough in energy to be confused with neutrinoless decay at the Q-value. Two of these ten isotopes also have measured decays to the first excited state. Now there are several very large experiments being designed which are sensitive to far higher neutrinoless half-lives.

Table 1.1 has a summary of the current experimental status. Many of the measurements were taken with NEMO-3 [78], a chambered experiment which measured many different isotopes. KamLAND-Zen [62] and EXO-200 [27] both have several hundred kilograms of Xenon. Heidelberg-Moscow, a Germanium experiment, generated controversy in 2001 when a subgroup of the collaboration reported a detection claim while the majority published a limit [51]. No further evidence of discovery in Germanium has been seen, and in 2013 GERDA published limits above the claimed detection in Germanium [47].

Future large scale experiments include the next generation of EXO, nEXO, which will use a time projection chamber with a full ton of recirculated, purified Xenon and possible single ion Barium tagging [34, 35, 29]. KamLAND-Zen is also planning a next generation experiment, with 1000 kg of Xenon dissolved in liquid scintillator inside a suspended balloon inside the KamLAND detector. Pressurized gaseous Xe designs are being developed in Spain (NeXT) and China (at Jinping). Cuore will take 760 kg of natural Tellurium oxide, containing roughly 200 kg of $^{130}$Te [44], and may possibly be upgraded to include particle identification techniques as CUPID. SNO+ in Canada is also a $^{130}$Te double beta decay experiment, and is designed to use tons of $^{130}$Te combined with liquid scintillator. GERDA and Majorana Demonstrator both focus on double beta decay of $^{76}$Ge. SuperNEMO will observe 100-200 kg of $^{82}$Se, and
Table 1.1: List of isotopes with high Q value which $\beta\beta$ decay. The first reported error is statistical and the second is systematic. The first column of results is the two neutrino decay to ground state half-life, the second is the limit on neutrinoless decay to ground state. The last column is the two neutrino decay to the first 0+ excited state level. $^{136}$Xe is both the longest lived, and most precisely measured.

<table>
<thead>
<tr>
<th>isotope</th>
<th>$2\nu$ half-life (GS)(years)</th>
<th>$0\nu$ limit (years)</th>
<th>$2\nu$ half-life (ES) (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{48}$Ca</td>
<td>$4.4_{-2.4}^{+3.4} \times 10^{19}$ [7]</td>
<td>$5.8 \times 10^{22}$ [87]</td>
<td>-</td>
</tr>
<tr>
<td>$^{76}$Ge</td>
<td>$1.55 \pm 0.1 \times 10^{21}$ [51]</td>
<td>$2.1 \times 10^{25}$ [47]</td>
<td>$&gt; 3.7 \times 10^{23}$ [48]</td>
</tr>
<tr>
<td>$^{82}$Se</td>
<td>$9.6 \pm 3 \times 10^{19}$ [7]</td>
<td>$3.6 \times 10^{23}$ [7]</td>
<td>-</td>
</tr>
<tr>
<td>$^{96}$Zr</td>
<td>$2.35 \pm 0.1 \times 10^{19}$ [7]</td>
<td>$9.2 \times 10^{21}$ [7]</td>
<td>-</td>
</tr>
<tr>
<td>$^{100}$Mo</td>
<td>$7.11 \pm 0.02 \times 10^{18}$ [7]</td>
<td>$1.1 \times 10^{24}$ [7]</td>
<td>$5.7_{-0.9}^{+1.3} \times 10^8$ [78]</td>
</tr>
<tr>
<td>$^{116}$Cd</td>
<td>$2.9 \pm 0.4 \times 10^{19}$ [19]</td>
<td>$1.7 \times 10^{23}$ [19]</td>
<td>-</td>
</tr>
<tr>
<td>$^{130}$Te</td>
<td>$8.2 \pm 0.2 \times 10^{20}$ [17]</td>
<td>$4.0 \times 10^{24}$ [18]</td>
<td>-</td>
</tr>
<tr>
<td>$^{136}$Xe</td>
<td>$2.38 \pm 0.02 \times 10^{21}$ [62]</td>
<td>$1.07 \times 10^{26}$ [64]</td>
<td>-</td>
</tr>
<tr>
<td>$^{136}$Xe</td>
<td>$2.165 \pm 0.016 \times 10^{21}$ [30]</td>
<td>$1.1 \times 10^{25}$ [31]</td>
<td>-</td>
</tr>
<tr>
<td>$^{150}$Nd</td>
<td>$9.11_{-0.22}^{+0.35} \times 10^{18}$ [7]</td>
<td>$1.8 \times 10^{22}$ [7]</td>
<td>$1.07_{-0.25}^{+0.45} \times 10^{20}$ [69]</td>
</tr>
</tbody>
</table>

1.7 Predicted Rates of Decay to Excited States

Beta decay is driven by the energy difference between two nuclear states of the same atomic weight. In certain cases, nucleon pairing coefficients cause the nuclear energy of the neighboring isotope to be unfavorable to decay [82]. Nuclei with even numbers of protons and neutrons, such as Cerium, Xenon and Barium of atomic weight 136, are have lower binding energy than the even-odd Cesium and Lanthanum nuclei. This is depicted in Figure 1.3, and the greater energy required to reach the binding energy of Cesium nuclei in particular creates an energy barrier forbidding single beta decay from occurring in $^{136}$Xe. Doubly $\beta$-decaying to Barium from $^{136}$Xe, however, is energetically possible and does occur. In some cases, $^{136}$Xe included, the daughter nucleus also has excited nuclear states which are still lower in binding energy.
Figure 1.3: Plotted is the nuclear binding energy of atomic number 136 isotopes. The lowest energy isotope, $^{136}\text{Ba}$ has been set to zero to highlight the energy differences between the isotopes. It is not energetically allowed for $^{136}\text{Xe}$ to singly decay to $^{136}\text{Cs}$, but it can decay to $^{136}\text{Ba}$, with a Q value (energy difference) for the double beta decay of 2457.83 keV [73].

than the parent nucleus, and thus are double beta decay to these levels is energetically allowed.

The half life of beta decays depends on the nuclear structure of the initial and final state, as well as the operator representing the decay mechanism and a phase space factor. For the two neutrino mode, this can be written as:

$$[T_{1/2}^{2\nu}(0^+ \to 0^+) ]^{-1} = G^{2\nu}(E_0, Z) |M_{GT}^{2\nu} - \frac{g^2}{g_A^2} M^{2\nu}_F|^2$$  (1.12)
Here $M_{GT}$ and $M_F$ are matrix elements containing the nuclear structure information, and $G^{2\nu}(E_0,Z)$ is an integral over lepton phase space. In the case of a $^{136}$Xe decay to the first excited $0^+$ state of $^{136}$Ba, the form of this equation remains the same, though since the phase space integral contains a dependence on the total energy of the beta decay the smaller decay energy suppresses the expected rate. This reduction from the phase space alone is $3,956.4 \ [54]$. The nuclear matrix elements are also expected to change for the decay to excited states, though far less so, only a factor of $1.7 \ [52]$. Because the nuclear matrix elements enter the equation as a square, their full contribution to the suppression is a factor of 2.92.

The combination of phase space and nuclear matrix suppressions result in a predicted half-life nearly 12,000 times longer to the first $0^+$ as the measured rate to the ground state of $2.165 \pm 0.016(\text{stat}) \pm 0.059(\text{sys}) \cdot 10^{21} \ \text{yr} \ [30]$. Therefore, the predicted half-life for the $0^+$ excited state decay is $2.5 \cdot 10^{25}$. Given the quantity of data collected by EXO-200, approximately 15 events are expected in dataset considered in this thesis. However, the observed lifetime for $^{100}$Mo [78] and $^{150}$Nd [69] were more than an order of magnitude less suppressed than calculations predicted, so perhaps more than 200 decays to the $0^+$ excited state of Barium are in the 1.6 yrs of data considered by this thesis.

Observation of both the ground and excited state decay can constrain models used for calculating nuclear matrix elements, since certain factors, such as the details of the intermediate nuclei between parent and daughter nucleus, are common to all decays of an isotope and will cancel out of a ratio. Nuclear matrix element calculations currently use a variety of models such as the quasiparticle random phase approximations (QRPA), interacting boson model (IBM), [52] or the interacting shell model (ISM) [53].
The decay to the $2^+$ excited state, which is lower than the lowest $0^+$ excited state and therefore energetically more favorable, is suppressed by an additional phase space factor favoring antisymmetric electron and neutrino energies [41, 54].
The EXO-200 collaboration has developed an experimental apparatus to search for double beta decay in $^{136}$Xe \cite{27}. The detector contains liquid xenon enriched to 80\% $^{136}$Xe abundance and is used as both a decay source and detection medium. Properties of xenon make it ideally suited to such a use as not only is $^{136}$Xe an candidate isotope for double beta decay, xenon is an excellent detector material. Energy deposited within xenon produces high ionization and scintillation yields, and is optically transparent to its own scintillation. The pure liquid detector allows for continuous recirculation and purification of the xenon, removing radioactive and electronegative impurities. These impurities, which would otherwise constitute backgrounds, are present at low levels during initial detector filling and introduced during any event which requires stored xenon to be added to the detector.

EXO-200 is a large experiment, with 200 kg of xenon. A ton scale cryostat containing 1-methoxyheptafluoropropane (HFE) surrounds the inner detector, providing both temperature buffering and dense hydrogen atoms for shielding. Further shielding is provided by a lead wall and placement underground with an overburden of 1585 m water equivalent. Both the detector and cryostat, along with the plumbing for xenon purification and recirculation, are housed in a class 1000 clean room, and air filtration and personal entrance requirements reduce introduction of any radioactive impurities to the area. Additional background suppression is enabled by a muon veto system covering four walls of the clean room containing the detector, allowing removal from
Figure 2.1: Diagram of EXO-200 Time Projection Chamber (TPC). The planes of circles on either end are the Avalanche Photo-Diodes (APDs) at the endcaps and the cross hatching are the wire planes. Also shown is a depiction of energy deposition, including ionization and scintillation produced via ion recombination. The central dotted line denotes the position of the cathode, held at a high negative voltage of 8kV. From [25]

the data set of any events coincident with a passing muon. A simplified diagram of these components is shown in Figure 2.5.

2.1 Time Projection Chamber

The inner detector is a thin-walled cylindrical copper vessel filled with enriched liquid xenon. The detector functions as a time projection chamber (TPC), with collection of ionization and scintillation allowing three-dimensional reconstruction of events. A simplified depiction of the TPC is shown in Figure 2.1, along with an illustration of ionization and scintillation from energy deposition within the xenon.

A high negative voltage central cathode, depicted in Figure 2.1 by the dashed line in the center, and wire planes held at low voltage on either edge drift ionization electrons towards the endcaps. There the wire planes, each side a set of two planes at 60 degrees from each other, produce the ionization signals. The inner wire planes,
called V wires, collect induction charges and shield the second set of wires from induction produced by drifting electrons in the central detector region. The outer wire planes, called U wires, are held at the furthest potential from cathode, and collect the actual electrical charge. The wires are 3 mm apart and ganged in groups of three, resulting in a readout spacing of 9 mm. Since the position of the wire gang collecting charge is known, position reconstruction is possible. The wire planes are set 60 degrees angle from each other, as this allows the wire ends to be held by a hexagonal structure that more closely follows to the cylindrical walls of the detector. Because they are not orthogonal, the collection U wire plane and the induction V wire plane do not directly correspond to an x or y coordinate plane; nevertheless positions calculated from wire position can easily be transformed into standard orthogonal coordinates.

Teflon reflectors line the inner wall of the detector vessel and reflect scintillation light to the two Avalanche Photo-Diode (APD) detection platters located behind the wire planes on the inner wall of each endcap. Scintillation is detected instantaneously after energy deposition, and three-dimensional reconstruction is possible by using the time difference between APD and wire signals, and the measured drift speeds of charges. The operating voltage during Run2, the dataset considered in this thesis, held the central cathode was at negative 8 kV while keeping the collection wires at ground. This voltage difference results in a drift velocity of $1.705^{+0.014}_{-0.010}$ mm/µs [37].

2.1.1 Detector Regions

There are several notable regions within the xenon vessel. Between the Teflon reflectors and the curved wall of the copper cylinder are field shaping rings which generate field uniformity between the cathode and wire planes. The liquid xenon between the vessel walls and the inner wall of the Teflon reflectors, including that surrounding the field shaping rings, is invisible to the detector. No ionized charge
and very little scintillation light can reach the APDs or wire planes from this region. Thus, this section of the vessel, along with the area behind the APD platter, as well as in the support struts for the vessel and the connections to the recirculation system, is designated the Inactive Xenon, and it provides a level of self-shielding to radioactive backgrounds from outside the detector.

Additionally, close to these rings near the vessel wall are areas of uneven field. Ionization from energy deposited here will not reliably drift to the collection planes. The hexagonal structure holding the wire planes contributes to this effect by creating a small space between the ends of the wires and the inner wall of the copper vessel that additionally complicates signal collection from this area. This layer is not completely shielded from detection as the Inactive Xenon is, but neither is it fully detected. This “dead layer” is cut from the region included in analysis. Both the Inactive Xenon and the dead layer contribute to a self-shielding effect for the xenon in the fiducial area, as gamma backgrounds from outside the vessel, and any impurities on the eflon reflectors, tend to cluster near the edges of the detector. The fiducial volume is chosen by a balance of maximizing fiducial source mass while minimizing backgrounds, and varies for each specific analysis.

The vessel is split in two halves by the central cathode, each half a separate drift region called TPC1 and TPC2. The total length of the encapsulating vessel for one drift region, between the cathode and the inner edge of the copper vessel, is 192 mm. The vessel radius is 183 mm. Though each half functions as a separate TPC in collecting ionization signals, scintillation energy is spread between both.

Electrons within the TPC regions may capture on electronegative impurities within the xenon before reaching the wire planes, reducing the observed signal. The electron lifetime is the inverse of the exponential decay constant describing this attenuation, and is calculated separately for each TPC. Plots of this value are shown in Figure 2.2 for the early portion of Run2 data. The electron lifetime is often in excess of 3,000
\( \mu s \), which combined with the drift velocity translates to over 25 times the length of one TPC.

### 2.1.2 Energy Deposition

Energy deposited in xenon both excites and ionizes xenon atoms [24]. The xenon ions drift toward the cathode, while the electrons drift to the wire planes on either end. Some of the Xenon ion pairs recombine, producing excited or ionized xenon atoms. These atoms, as well as those directly excited by energy deposition within the xenon, join with another ground state xenon atom to create an excited or ionized dimer. This de-excites by emitting a 178 nm photon. Since this is a molecular and not atomic transition, it is not reabsorbed by the surrounding xenon and can be collected by the APDs.

Though ionization and scintillation occur regardless of the underlying process depositing energy within the xenon, subtle differences can be used to distinguish between gamma, alpha, and beta energy depositions. Gamma rays entering the xenon often scatter through a significant distance, with the number of scatters and the typical distance between them depending on the energy of the entering gamma ray. At the Q value of the \(^{136}\text{Xe}\) double-beta decay the mean free path for a gamma passing through the EXO detector is 8.5 cm, much shorter than the size of the TPC, but much larger than the pitch of the wires within the induction and collection planes. In contrast, electrons from beta decays deposit all their energy in a small volume, which is resolved into a single deposition cluster by the wire planes. Events caused by alpha particles can be distinguished because an alpha track has a dense cylindrical core with a high rate of recombination within this area, resulting in a much higher light to charge ratio than any other process [24]. These differences allow discrimination of background via detecting the overall spatial structure of an event’s energy deposition. The position-reconstructing abilities of the detector are used to define
Figure 2.2: Electron lifetimes for TPC1 (a) and TPC2 (b) during the first portion of Run2. Black points indicate measured values from a single source calibration run, while the blue bands show piece-wise polynomial fits to these values. Vertical dashed lines indicate discontinuities in xenon recirculation, which reduces lifetimes due to introduction of stored xenon until purity recovers. The shaded regions, where purity is increasing after an interruption of normal operation, are excluded from the final dataset. From [30].
energy clusters at the limit of the detector’s spatial resolution of a few millimeters, with the multiplicity of an event being defined as the number of distinct clusters of energy it has. In the analysis, two categories are defined, one for Single Site (SS) events of multiplicity 1, and another for Multi Site (MS) events with multiple energy clusters.

![Figure 2.3](image-url)

Figure 2.3: Plotted is the charge vs. light energy for a calibration run with a Thorium source. The 2.615 MeV gamma line used for energy calibration shows up as the oval in the upper right quadrant. The energy resolution is markedly improved along the rotated energy axis, corresponding to the narrowest one-dimensional projection of the oval. From [77].

The average number, $N$, of electron-ion pairs produced by a given deposited energy is generally considered a constant of a material, and the average energy required to produce a single pair known as the W-value. For liquid Xenon, the W-value is 15.6 eV per electron-ion pair generation [24]. The variance of this quantity ultimately controls
the resolution possible in the detector, and depends on a dimensionless quantity known as the Fano factor. The Fano factor is given by:

\[ \delta^2 = F \cdot N \quad (2.1) \]

where \( N \) depends on the energy, via the \( W \)-value. Though the calculated Fano factor for liquid Xenon is .059 [24], the measured value is closer to 20 [16], which is referred to as the effective Fano factor, and which causes more variation in ion pair generation than expected. The recombination rate is a function of the applied electric field, and the high electric field in the EXO-200 detector results in less recombination than would occur at lower electric fields and a higher rate of charge collection while simultaneously reducing the drift time and corresponding electron absorption over time in the liquid xenon. However, at a constant electric field such as within the TPC there is still some fluctuation in the rate of ion pair recombination rates in an single event which broadens the resolution in the separate charge and light channels. An event with greater recombination results in less charge left to collect, but more light, and events with less than average recombination rates leave more electrons free to drift to the wire planes. This effect can be seen in Figure 2.3 where the narrowest projection of the oval corresponding to the 2.615 MeV gamma calibration line is neither the scintillation nor ionization axis, but a rotated axis. The resolution (\( \sigma/E \)) at the Q value of the ground state decay is 1.53±.06\% for Single Site, and 1.65±.05\% for Multi Site events during the 2014 0\( \nu \)\( \beta\beta \)-decay search [31].

2.2 Xenon Purification and Recirculation

The xenon in the TPC is continuously recirculated to remove electronegative impurities and thus increase the electron lifetime in the detector. This reduces the
absorption rate for charge created during ionization, and thus increases the energy resolution of the charge signals detected. First the xenon is boiled off in parts of the plumbing outside the TPC. Then a magnetically-driven piston pump [26] controls the rest of the circulation, as well as provides the ability to overcome pressure losses. It passes through two heated purifiers containing zirconium getters, then is recondensed and returned to the copper detector vessel. A simplified diagram of the purification plumbing is shown in Figure 2.4.

The resulting electron lifetime is measured by the detector itself during weekly calibration runs and monitored for slight changes over time. The lifetime is calculated separately for each TPC, and when purification is running smoothly is over 3 ms as seen in Figure 2.2.
2.3 Background Control

EXO-200 uses active and passive methods to reduce backgrounds. A simplified drawing of the clean room containing the detector is visible in Figure 2.5. In the center is the copper vessel, surrounded by a cryostat kept at liquid xenon temperature, providing a thermal buffer for the detector. The cryostat also provides at least 50 cm of shielding in all directions, and is further shielded by a layer of lead inside the clean room. Outside the clean room module, a plastic scintillator muon veto system allows active detection and removal of events in coincidence with muon events. Not shown here is the underground placement of the entire experiment. The most prominent backgrounds in the region near the Q value of double-beta $^{136}$Xe include $\gamma$s produced by decays of $^{232}$Th and $^{238}$U impurities within detector components. Single-beta decays of $^{137}$Xe, a short-lived isotope produced by muon-induced neutron captures on $^{136}$Xe within the detector, is a close third to the Uranium and Thorium components [31]. $^{222}$Rn related backgrounds, such as $^{214}$Bi and $^{214}$Po decays, have a dependence on location within the TPC, as the daughters from the initial radon decay are positive ions, and drift toward the cathode [30]. Radon also shows a time dependence, with rate spikes occurring in coincidence with interruptions in regular xenon circulation and introduction of stored xenon into the detector [33]. At energies lower than the $^{136}$Xe Q value the decay of $^{40}$K becomes important.

2.3.1 Muon Veto System

The Muon Veto consists of a panel system on four outer walls of the cleanroom module containing the TPC. These panels are plastic scintillators which light up when a muon passes through them. This data is recorded in a separate data stream which can be used to “veto” events within too close a time window to the muon passing. This veto process identifies over 95% of the muons passing through the TPC, and thus can remove from the Low Background dataset most muon generated backgrounds such
as muon Bremsstrahlung and neutron capture gammas. The measured muon flux through the EXO-200 detector is $4.07 \pm 0.14 (\text{sys}) \pm 0.03 (\text{stat}) \cdot 10^{-7} \text{cm}^{-2}\text{s}^{-1}$ [36].

The muon veto process can be reversed, and was used to create a neutron capture rich dataset to study the rate of neutron capture related backgrounds. In particular, $^{137}\text{Xe}$ is a troublesome muon-associated background, as it is produced by muon generated neutron capture on $^{136}\text{Xe}$, and the resultant $^{137}\text{Xe}$ isotope $\beta$-decays with a half-life of 3.8 minutes. This is too long to veto entirely given the muon rate, and with a Q value of four MeV some of these $\beta$ decays fall within the much lower $^{136}\text{Xe}$ decay Q value, causing a muon generated background that cannot be mitigated by the standard muon veto.

In the neutron capture rich data set, events were selected having the energy signature of a $^{136}\text{Xe}$ neutron capture event. Less than 1.23 ev/day were observed by
counting between 3.25-4.5 MeV, and a fitted search resulted in less than .47 ev/day. Still, given the Q-value of the decay, and this rate, $5.88 \pm 2.18$ $^{137}$Xe events may be in the SS ROI. This is in pretty good agreement with the 7.0 counts attributed to $^{137}$Xe in the final fit unconstrained by the neutron capture search results [31].

2.3.2 Waste Isolation Pilot Plant

EXO-200 is 2200 feet (1585 m water equivalent) underground in Carlsbad New Mexico in a salt mine, the Waste Isolation Pilot Plant (WIPP). The underground placement of the experiment reduces cosmic backgrounds to an acceptable level. WIPP is operated by the Department of Energy to permanently store nuclear waste. This waste is safely contained and located at the far end of the mine from the EXO-200 installation, so does not pose any problems of radioactive backgrounds in the
Figure 2.7: Shown here is the relative energy resolution ($\sigma/E$, percent) vs energy. Multi Site (MS) and Single Site (SS) data is plotted separately. Improved energy resolution after the de-noising process is demonstrated. From [38].

experiment. Figure 2.6 is a photo of the author underground at the experimental site.

### 2.4 Calibration System

Surrounding the TPC is a small, quarter-inch diameter, tubing system used to introduce $\gamma$-ray sources of known energy in proximity to the TPC. The source is close enough to the detector, only 7-10 cm away from the inner edge of the copper vessel, that many gammas pass within the detector volume and are entirely collected there [30]. Figure 2.3 was produced with one of these source calibration runs, and demonstrates the anti-correlation method of reducing detector resolution. Gamma lines from small $^{228}$Thorium, $^{60}$Cobalt, $^{226}$Radium and $^{137}$Cesium sources are used during calibration campaigns performed a few times yearly to map out energy resolution functions. A plot of the resulting resolution as a dependence on energy is shown in Figure 2.7. In addition to the quarterly calibration campaigns, the $^{228}$Thorium source is inserted a few times per week, and the 2.6 MeV gamma line from the $^{208}$Tl decay product used to track electron lifetime, as well as any drift in energy resolution over time.
Because these high energy gammas have a high likelihood to Compton scatter multiple times in the detector, a Compton Telescope method was developed to point back to the location of the source. The angle of a scattered photon and the energy lost by the photon in the interaction are directly related by:

\[
\frac{1}{E'} - \frac{1}{E} = \frac{1}{m_e} (1 - \cos\theta),
\]  

(2.2)

where \(E\) is the original photon energy, \(E'\) is the photon energy after scattering, and \(\theta\) is the angle through which the photon scatters.

To use this method to determine source location, events are chosen with two energy depositions within the xenon, as shown in Figure 2.8. Statistically, the more energetic deposition is more likely to be the first deposition, as the less energy the resulting scattered photon has the more likely it is to be fully captured in only one additional
photo-electric interaction event. An axis is defined through the two clusters, the original photon energy is known from the energy of the calibration gamma line, and the energy deposition of the larger cluster is used to calculate the energy of the photon after that interaction. This allows the angle of scattering to be calculated. An optimization function is called to calculate the x,y, z source position consistent with the most events. This method is only usable in the very high statistics calibration runs where the source location remains constant throughout. Source calibration runs meet these requirements, resulting in high statistics from a fixed source point. One of my earliest projects with EXO-200 was to use this technique to image individual source runs. One of these is shown in Figure 2.9, where the source shows up as a light spot on an unrolled cylinder with dimensions extending to the source tube from the center of the detector.

Significant work has been expended in further developing this method to function for diffuse, low statistics sources. Overcoming these limitations would allow the technique to isolate any slightly higher-background sources near the detector, such as the high voltage cable. More importantly, it could even work to reconstruct individual candidate double beta decays to excited state events. These candidate events would be characterized by scattered de-excitation gammas Compton-reconstructed and consistent with the location of the original double-beta decay event [10].

2.5 EXO-200 Status and Results

The latest $0\nu\beta\beta$ analysis of EXO-200 covered data collected during the first three periods of the so-called Run2. Run2 spans the time between October 4, 2011 and February 5, 2014, and is broken into four periods denoted a, b, c and d. These time periods are summarized in Table 2.1. The run demarcations denote moments of importance to run conditions, such as transient deviations from typical temperatures, pressures, and flow patterns. Recovery from these events may result in subsequent
<table>
<thead>
<tr>
<th>Run Period</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run2a</td>
<td>October 5, 2011</td>
<td>April 16, 2012</td>
</tr>
<tr>
<td>Run2b</td>
<td>April 16, 2012</td>
<td>June 24, 2013</td>
</tr>
<tr>
<td>Run2c</td>
<td>June 24, 2013</td>
<td>September 1, 2013</td>
</tr>
<tr>
<td>Run2d</td>
<td>September 1, 2013</td>
<td>February 5, 2014</td>
</tr>
</tbody>
</table>

Table 2.1: Start and end dates for each of the four periods making up the Run2 data. Run2 contains a total of 585.8 days livetime collected between October 5, 2011 and February 5, 2014.

A slight change of performance in the detector. The events defining the end of Run2 data taking are the most dramatic, resulting in over a year of restricted mine access and experiment recovery. On Feb 5, 2014 an underground salt-hauling truck caught fire and burned with significant smoke and soot production [89]. Though low background physics data taking was stopped at this point due to restricted access, the clean room protected the detector itself from soot. Before the fire damage had been mitigated and regular operation resumed, a radiation release from nuclear waste stored at the south end of the mine contaminated large segments of the WIPP underground site [12]. However the structure of the mine includes four ‘air splits” which allow air flow in the mine to be completely separated from each other. The area containing EXO-200, though containing much soot from the fire, was not directly affected by the radiation. More damaging for the experiment was the extremely limited underground access for over a year and lack of reliable power. This required required the remote-controlled warming of the entire experiment and recovery of the xenon to high pressure bottles. By January 2016 the xenon recirculation in the detector had been restored and calibration data taking quickly followed. Low Background operation resumed in May 2016.

Run2abc produced $477.6 \pm 0.01$ days of “golden” data, low background data of good quality, excluding calibration runs and poor quality runs affected by bad xenon purity or noise issues. Figure 2.10 shows the accumulation of this data over time,
with a steady increase in data with little down time. The analysis searched for a zero neutrino ground state decay, and set a limit of $1.1 \cdot 10^{25}$ yr. at a 90% CL [31]. The fit from this analysis is shown in 2.11. Monte Carlo generated spectra are used to fit using a binned maximum-likelihood fit performed simultaneously in several variables. The data is separated into two categories by the number of clusters per event, with Single Site (SS) events more likely to be double beta events and Multi Site (MS) events more likely to be gamma backgrounds. The total energy of events was fitted, as was the standoff distance, a variable defined by the shortest distance between a cluster in the event and the edge of the fiducial volume.

The SS and MS fits were constrained to have a ratio defined from Monte Carlo, allowing the higher gamma rate in the MS spectrum to partly determine the SS gamma background rate as well. The upper plot is the SS spectrum, the lower is the MS, and the inset is the region of interest around the Q-value of the SS spectrum blown up. From the best fit, backgrounds within the ROI are $31.1 \pm 1.8 (\text{stat}) \pm 3.3 (\text{sys})$ counts, with the dominant fitted backgrounds being $^{232}\text{Th}$ at 16.0 counts, $^{238}\text{U}$ at 8.1 counts, and $^{137}\text{Xe}$ at 7.0 counts. These values are consistent with previous results, and separate estimates of $^{137}\text{Xe}$ activity from studies of neutron captures on $^{136}\text{Xe}$ in muon-veto-tagged data [31].

Profile scans over the $0\nu\beta\beta$ fit value were performed to test consistency with the null hypothesis. The best fit is consistent with zero at 1.2 $\sigma$, and so a limit at the 90% confidence level is set. The profile scan over the number of $0\nu\beta\beta$ counts is shown in Figure 2.12, with the 90% confidence and 1$\sigma$ lines also plotted. The profile crosses the 90% confidence level line at 24 neutrinoless decay counts, resulting in a half-life limit of $1.1 \cdot 10^{21}$ years. This corresponds to an upper limit on the Majorana neutrino mass of .19-.45 eV. This correlates to a $m_{min}$ for the lightest neutrino mass eigenstate of less than .69-1.63 eV, assuming the most disadvantageous combination of CP phases [31], with the connection between these two masses illustrated in Figure 1.1. The
fitting and profile analysis process described here for the $0\nu\beta\beta$ search is fundamentally similar to that used in the search for decays to excited states, described in more detail in Chapter 4.

2.5.1 Including Run 2d

During the break in data taking following the fire event, some partial runs which were not previously considered “golden” low background data was subsequently salvaged and added to the set of data considered suitable for physics analysis. This represented about a 5% increase over the whole Run 2abc time frame. Together with the Run 2d data, an additional 22% Run 2 livetime not included in previous analyses was added, resulting in a total of 585.8 days livetime. The searches for decays to excited states described in the following chapters use this expanded data set.
Figure 2.9: Image produced by resolving Compton scatters during a calibration run on Oct 3, 2011. There was a contradiction between the position described by a note left by shifters on duty that day, and the assigned position tag. The Compton imaging method described in the text easily resolves the conflict and confirms the true position of this run.
Figure 2.10: Run2abc data accumulation as a function of time. Blue is golden, red includes non-golden data taking, including calibration and noisy or low purity runs, black is real time. This is the data used in the $0\nu\beta\beta$ analysis [31].
Figure 2.11: Black points raw data, fitted lines are PDFs of backgrounds generated via MC, inset is the region of interest. Shaded grey is the $2\nu\beta\beta^{136}\text{Xe}$ background to the magenta $0\nu$ decay signal. A two dimensional fit over energy, shown here, and standoff distance, a measure of centrality of events, is performed. SS and MS spectra are simultaneously fit, with the ratio constrained via simulation. Only the SS fit is shown here. A limit of $1.1 \cdot 10^{25}$ yr. at a 90% CL is found. From [31].
Figure 2.12: Profile scan over number of counts attributed to the $0\nu\beta\beta$ decay in the fit over the Run2abc dataset. The minimum is at the best-fit $0\nu\beta\beta$ value of 9.9 counts. The $1\sigma$ and 90% confidence level lines are shown. The best-fit value is consistent with zero at $1.2\sigma$. From [31].
CHAPTER 3
SIGNAL STRUCTURE AND ANALYSIS

Double beta decays of $^{136}$Xe to nuclear excited states of $^{136}$Ba can be distinguished from decays to the ground state, as well as from other backgrounds, by leveraging their typically complex topologies. The excited daughter barium nucleus promptly de-excites to the ground state emitting unique gamma signatures characteristic of these decays. This chapter will first present an overview of the basic approach detailing the most promising channel, i.e., two neutrino double beta decay ($2\nu\beta\beta$) to the $0_1^+$ excited state. Three additional decay modes will then be described, followed by a description of the data analysis process used.

Data analysis is broken down into two broad steps for all EXO-200 analyses. First is the reconstruction of the raw signals from the wires and APDs into physical energy deposition clusters in specific locations in the detector, as mentioned in Section 2.1. The second step incorporates simulated Monte Carlo signal and background probability density functions (PDFs), which are used in a maximum likelihood fitting process to convert reconstructed signals into activity rates for each component decay process.

Unique to this analysis is an additional level of signal processing incorporating a machine learning-aided step between the raw signal reconstruction and the final fitting framework. This chapter introduces the machine learning algorithm used in this analysis, which improves signal-to-background discrimination power in a highly multi-variate space. The application of the algorithm to the excited state decay search is found in Chapter 4.
3.1 Decay Signature of $2\nu\beta\beta$ to the $0^+_1$ Excited State

The decay of $^{136}$Xe to the ground state of $^{136}$Ba has a Q-value of 2457.83 keV [73]. This Q value depends on the difference between energy levels of the parent and daughter nuclear configurations. A diagram showing the first couple exited levels of the $^{136}$Ba atom is shown in Figure 3.1. The $0^+_1$ excited state is 1579 keV above the ground state of barium, leaving the $^{136}$Xe nuclear state 879 keV higher still. This remaining energy difference allows double beta decay to occur to this excited state. This allowed double beta decay results in an excited barium atom which then promptly de-excites to the ground level by emitting two photons of 761 keV and 819 keV as it cascades through the $2^+_1$ level. These photons, unlike the emitted beta particles, have fixed energies and can be used to identify these decays. This section will focus on the allowed standard model $2\nu\beta\beta$ decay process to the $0^+_1$ level, leaving discussion of decays to the $2^+_2$ state, and the $0\nu\beta\beta$ decay to the $0^+_1$ state for Section 3.2. Decays to the $2^+_2$ state are not included in this analysis.

As previously discussed in chapter 2, energy deposited within the xenon by $\beta$ particles is almost always captured within one energy cluster. On the other hand, photons from the 2.6 MeV $^{208}$Tl calibration line are likely to scatter and be captured in multiple energy depositions, have a mean free path of 8.5 cm, and even occasionally escape the detector without depositing energy. De-excitation photons, having slightly less than 1 MeV of energy, are between these extremes. Detailed simulations of $2\nu\beta\beta$ decays to the $0^+_1$ excited state, complete with the promptly emitted de-excitation $\gamma$s, were generated. These simulations include realistic detector and signal reconstruction modeling, and the simulated signals are matched with the identity of the energetic particle which deposited energy within the detector volume, whether that is one of the de-excitation $\gamma$s or the $\beta$s. In less than half of these simulated decay events did the cluster containing the $\beta\beta$ energy also contain any energy from either of the de-excitation gammas. Of all the de-excitation photons simulated, one 760.5 keV $\gamma$ and
Figure 3.1: Double beta decay of $^{136}$Xe to $^{136}$Ba excited state, and relevant energy levels of $^{136}$Ba. $\beta\beta$ decay to the excited $0^+$ state of Ba is energetically allowed, with a Q value of 879 keV. Two de-excitation $\gamma$s, of 760.5 keV and 818.5 keV are emitted promptly, and cannot be resolved in time in EXO-200. Decays to the two $2^+$ levels are also energetically allowed, though only the search for the decay to the lower $2^+$ level is described in this thesis.

one 818.5 keV $\gamma$ per event, at least 40% had scattered at least once, and spread their energy to more than one detected cluster. Histograms showing the number of clusters containing any energy from each of the de-excitation $\gamma$s in 5,000 simulated $2\nu\beta\beta$ decay to the $0^+_1$ excited state events are shown in Figure 3.2, showing the number of times each de-excitation photon scattered in these events.

The position reconstruction uncertainty for cluster position ranges from a low of .42 mm in the Z direction to a high of 2.4 mm in the U-wire (charge collection) coordinate [40]. Consequently, photon scattering events separated by more than a few mm can be distinguished as separate energy clusters. Calculating from liquid xenon density near 168 K [80] and NIST photon cross section scattering data [68],
Figure 3.2: Number of clusters per event containing any energy attributable to each of the two de-excitation $\gamma$s involved in the $2\nu\beta\beta$ to the $0^+_1$ excited state decays. Data from a small, 5,000 event simulation of this decay process with tracing from the simulated $\gamma$ depositions identifying the contributions to each cluster in these events.

the mean free path of a photon of 800 keV is 4.9 centimeters. Thus, it is unlikely for photons near that energy, either the 760 keV or 818 keV de-excitation $\gamma$s, to scatter twice in the bulk xenon so closely that the scattering energy depositions are not separately resolvable. Studies of simulated decay and de-excitation events indicate that photon interaction close to the deposition of $\beta$ energy, so close as to constitute a single combined cluster, is higher than naively expected from this calculation. More than 20% of simulated events have at least a small amount of energy from a soft scatter off one of the $\gamma$s within the distance defined by the cluster containing the $\beta\beta$ energy. This higher than expected rate of combined clusters may arise from processes included within the simulation but not accounted for in the calculation of the mean free path within bulk liquid xenon, such as the fact that the de-excitation photons are being emitted from an atom and are thus close to the atomic electrons, or are passing through the cloud of ionization electrons caused by deposition of the $\beta\beta$ energy, or some other mechanism not captured by the xenon bulk cross section.
The two $\gamma$s emitted as the $0_1^+$ excited state relaxes are angularly correlated by a function shown in Figure 3.3. These two photons are preferentially emitted anti-parallel, or parallel to each other. A study of photon cluster separability was conducted for both the parallel, and anti-parallel photon cases. The benefit of demonstrating methods that separate clusters into groups originated by the same de-excitation $\gamma$ is that the total energy deposited in the such a group of clusters would have characteristic, known energy corresponding to the de-excitation gap which could be used to discriminate signal events from background. Five thousand each of anti-parallel and parallel photon emission decays were simulated from the same point in the detector [65], with the originating cause, or “ancestor particle” for each cluster determined. The results are shown in Figures 3.4 and 3.5. Relatively simple algorithms can be developed to separate clusters from each original photon in the event they are emitted anti-parallel, as the clusters from each photon are well separated in space. One simple spatial algorithm, kmeans clustering, separates all the clusters in an event into three larger groupings in such a way as to minimize separation between

![Figure 3.3: Angular correlation function, $W(\theta)$, between emitted de-excitation photons. Photons are preferentially emitted either directly opposite each other, or in the same direction.](image)
clusters within the larger groupings. Kmeans clustering was developed and tested by a group at Stanford [66], with success roughly in proportion to the number of events expected to have anti-parallel photon emission. However, it is much more difficult to separate clusters by ancestor in the event they are emitted in the same direction. The equally preferential representation of the worst-case, parallel, and best-case, anti-parallel, of the angular correlation function has so far limited the ability to develop a search method that uses combined energy from multiple clusters as a discriminating variable in a search for decays to excited states. The goal of reconstructing individual photon scattering tracks within a single event, including that using Compton Telescope reconstruction methods of section 2.4, remains out of reach.

![Anti-Parallel Photons](image)

Figure 3.4: Five thousand $2\nu\beta\beta$ decays to the first $0^+$ excited state are simulated at a central point in the EXO-200 detector. The de-excitation photons are forced to emerge anti-parallel to each other, with the direction of the larger 818 keV $\gamma$ in the positive z direction and the smaller 760 $\gamma$ in the opposite direction. Photon clusters are relatively simply separated by position.

Even without reconstructed single photons, a great deal of topological information exists to separate excited state decays from ground state or other backgrounds, such as the number of clusters per event, and the typical energy spectra of individual clusters.
Figure 3.5: Five thousand 0+ decay events with forced parallel de-excitation gamma correlation. Photon clusters not simply separated by position.

Figure 5.1 displays the number of clusters per event for both this decay mode, and others described in the following section. Variables built to highlight these event characteristics are input into a Machine Learning algorithm. Details of the algorithm are described in section 3.4, and the application to EXO-200 event variables is found in chapter 4. This approach is also used in the 2015 EXO-200 paper reporting a limit of \( > 6.9 \times 10^{23} \) yr in the search of excited decays of this type [39].

### 3.2 More Decay Mode Signatures

Additional decay modes beyond the \( 2\nu\beta\beta \) to \( 0^+_1 \) excited state are energetically allowed, though detection of any \( 0\nu \) decay mode would require new physics beyond the Standard Model. It is possible, though suppressed through angular momentum considerations as described in Section 1.7, for direct decays to the \( 2^+_1 \) excited state to occur. For both of these, searches for the \( 0\nu\beta\beta \) decay mode is also of interest. In contrast to the phase space suppression factor of 3956 between the \( 2\nu \) to \( 0^+_1 \) excited state and ground state, the phase space suppression in the \( 0\nu \) decay is only 23.8 [54].
The difference is not made up by the nuclear matrix elements, as the suppression of $0\nu$ decay to the $0^+_1$ excited state as compared to decays to the ground state is only a factor of 1.9 [52], yielding a total suppression well over two hundred times smaller than that for $2\nu$ decays. Thus it may be possible, given a search algorithm that sufficiently discriminates backgrounds, that the $^{136}\text{Xe}$ detector which discovers $0\nu$ decay to the ground state, will also discover the $0\nu$ to $0^+_1$ excited state. Limits for all of these decay modes, plus the decay to the second $2^+_2$ level, were set in 2015 by the Kamland-Zen Collaboration [63].

![$2\nu\beta\beta$ to $2^+$ Excited State](image)

Figure 3.6: Ten million simulated $2\nu\beta\beta$ to $2^+_1$ excited state decays. Total event energy plotted, smeared with realistic detector resolution derived from EXO-200 calibration data. Beta decay shape and photon peak visible.

A GEANT4 generator for the $2^+_1$ excited state decay was developed for this analysis. The energy spectrum of simulated events using this generator are shown in Figure 3.6. This can be compared to the existing $0^+_1$ generator as seen in Figure 3.7. In both cases, the beta decay spectrum is translated up in energy, and can be seen between the photon and Q value energies. These events have complete energy collection- with both the $\beta$ and $\gamma$ contributions completely collected within the detector. The $2^+_1$
Figure 3.7: Ten million simulated $2\nu/3\nu$ to $0^+_1$ excited state decays. Total event energy plotted, smeared with realistic detector resolution derived from EXO-200 calibration data. Beta decay shape and photon peaks visible.

decay spectrum is conceptually simpler, with the photon peak from events where the $\beta$ energy was not collected clearly visible. Below this are events missing some or all photon energy. The principle is the same for the $0^+_1$ decay, though with two $\gamma$s, an additional possibility is available between these regions where one photon energy is collected, while the other one escapes the detector. Ten million each of $0\nu$ and $2\nu$ decays to the $2^+_1$ excited state of $^{136}$Ba were simulated using this new generator, and the resulting MC files used for all subsequent analysis described for this mode.

### 3.3 Coding Toolkits

ROOT [2] is an object oriented programming library developed by CERN, and is commonly used in the particle and nuclear physics community. It also has an associated, RooFit [3] toolkit used to handle probability distribution functions and final fitting processes. The Toolkit for MultiVariate Analysis (TMVA)[4] is a ROOT-integrated machine learning environment for processing multivariate classification and
regression techniques. All of these packages, as well as GEANT4, are crucial in completing the analysis described in this thesis.

The TMVA package includes object-oriented implementations in C++ for multivariate Machine Learning methods such as rectangular cut optimization, artificial neural networks, and boosted decision trees [1], among others. Brief studies were done by the group of EXO-200 analysts searching for $\beta\beta$-decays to excited states of barium comparing learning methods, and simple boosted decision tree was deemed to be better than others based on simulated discrimination power, and selected for use [86].

3.4 Boosted Decision Trees

Decision trees are a greedy classification model for data sets with many relevant variables each weakly separating signal from background. A sample decision tree is shown in Figure 3.8. Greedy in this context refers to the cut selection at each node, which maximizes the immediate discrimination gain possible from a single cut among all input variables at that node, even if that variable had previously been used as a cut in an earlier node. This process continues until the desired noise or background separation, or the nesting limit is reached. Completing the arboreal analogy, the final nodes in the decision tree are “leaves” and values of +1 imputed to events in signal leaves, with −1 assigned to all background leaf events. The purity of a leaf is the fraction of its events which are from the signal. Each decision tree, though a better discriminator than any individual input variable, is still relatively weak. The weakness of the decision tree method is that a small change of input may cause a large change in the training and resulting machine algorithm. Decision trees provide a very good initial machine algorithm—however, they are prone to overfitting and instability. Thus, Boosted Decision Tree (BDT) algorithms produce several hundred
Figure 3.8: Illustration of decision tree. Increasing signal proportion denoted by pinker hues, as well as the values printed in each circle. Node level denoted by the cut number, branch by cut letter. At each decision node, the cut that maximally separates signal and background among input variables is selected.

to a thousand decision trees which are combined for stronger separating power, and weighted to reduce overfitting.

Training sets are built with an equal number of signal and background events, and the first decision tree produced with each event weighted equally. After the first training, all the misclassified events are given boosted weight and a second training is completed, resulting in a second decision tree. Any events misclassified in that tree are again boosted, until the desired total number of trees are completed. Once the trees are finished, their performance is tested on a second set of simulated signal and background events, with the purity of the final leaves determining the weight of
that tree in the final combination. When applying this algorithm, every event is run through each of the decision trees, and the weighted average calculated. Since each decision tree produces a binary output of + or −1, the final discriminator value is a spectrum between these extremes.

3.5 Fitting Process

RooFit and associated RooMinuit classes are used during the fitting process. These tools are built to handle PDFs, many component fits, and minimization functions, and work seamlessly with output from TMVA and other ROOT packages. Two separate PDFs for each signal and background component are created, one each for two versions of data selection. More details about cut definitions is found in Chapter 5. An example of both PDFs compared against each other is shown in Figure 3.9. The PDF is actually a 2D object in both the energy and discriminator variable, but the second variable has been projected out to produce this plot.

Background PDFs are generated for the components listed in Table 3.1. Each component has variable parameters, such as the ratio between SS and MS events. For the neutron capture PDF, the relative contributions from capture on different materials is allowed to float within measured systematic error. A negative log likelihood minimization function is applied to find the best fit activity values for the components, including the floating parameters such as the SS fraction and neutron capture normalization, as well as additional parameter constraints described in Section 4.3.

Since the background components are combined to produce simulated “Low Background” data sets for use in training the BDT machine algorithm, as described in more detail in Chapter 4, it is very important to have a background model that allows for the proper ratio of each component to be determined. For the recent EXO-200 paper using Run2abc data to search for the 2νββ decay to the 01+ excited state [39],
Table 3.1: List of background components to all double beta decay modes studied by the EXO-200 collaboration. Background locations generating these components are also listed. These backgrounds are extensively simulated, and these simulated backgrounds used to train a BDT machine learning algorithm to distinguish background from signal in each of the decay modes ending in an excited state of $^{136}\text{Ba}$.

<table>
<thead>
<tr>
<th>Background Location</th>
<th>Background Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Liquid Xenon</td>
<td>$^{214}\text{Pb}$</td>
</tr>
<tr>
<td>Active Liquid Xenon</td>
<td>$^{135}\text{Xe}$</td>
</tr>
<tr>
<td>Active Liquid Xenon</td>
<td>$^{137}\text{Xe}$</td>
</tr>
<tr>
<td>Air Gap</td>
<td>$^{214}\text{Bi}_{\text{nochain}}$</td>
</tr>
<tr>
<td>All Vessel</td>
<td>$^{60}\text{Co}$</td>
</tr>
<tr>
<td>All Vessel</td>
<td>$^{54}\text{Mn}$</td>
</tr>
<tr>
<td>All Vessel</td>
<td>$^{232}\text{Th}$</td>
</tr>
<tr>
<td>All Vessel</td>
<td>$^{238}\text{U}$</td>
</tr>
<tr>
<td>All Vessel</td>
<td>$^{65}\text{Zn}$</td>
</tr>
<tr>
<td>Cathode Surface</td>
<td>$^{214}\text{Bi}_{\text{nochain}}$</td>
</tr>
<tr>
<td>Inactive Liquid Xenon</td>
<td>$^{222}\text{Rn}$</td>
</tr>
<tr>
<td>Inner Cryostat</td>
<td>$^{232}\text{Th}$</td>
</tr>
<tr>
<td>All</td>
<td>bb2n to ground</td>
</tr>
<tr>
<td>All</td>
<td>Neutron Capture</td>
</tr>
</tbody>
</table>

this background model was set by the best-fit values for background components as found by the recent analysis on the same Run2abc dataset searching for double-beta decays with the emission of Majorons ("Majoron fit" here-after) [32]. This Majoron fit is a two-dimensional maximum likelihood fit in both total event energy and the standoff distance, a measure of centrality in the detector for events. At certain points in the analysis described in Chapter 4, a background model is needed which includes the full Run2 data including the Run2d period as described in Section 2.6. Thus, a second background model is generated by performing a background-only fit over this larger dataset. This fit is also a two-dimensional maximum likelihood fit simultaneous in both MS and SS like the Majoron Fit, but over total event energy and the variable derived from the BDT machine learning algorithm rather than the standoff distance.

Figure 3.10 and 3.11 show the single-site and multi-site daily background activity rates for each of these two fits. The differences are small enough not to impact this
search, as the point where the Majoron fit is required rather than that covering the full Run2 dataset is necessary does not require very high accuracy, and in fact is artificially limited by statistics, and can be attributed to the fact that the two fits use: a) different datasets, and b) different fit variables. Both fits are used at different points of this analysis as “reference” activity rates to build simulated datasets in this analysis. The more recent fit, including full Run2 data, is used for sensitivity studies, while the Majoron fit is used to build training MC data sets.

An additional fit comparison is shown in Figures A.1 and A.2 in the Appendix. This comparison shows the differences between the fully reconstructed and partially reconstructed data sets in a background-only fit. The motivation for including two reconstruction requirements is described in more detail in Chapter 5, and relates to relaxing a fiducial requirement which has particular relevance for highly-multisite events. The partially reconstructed requirement allows some events containing clusters which have only a U-wire collection signal and no V-wire induction signal, and thus less than fully 3D position information, to be included within the data set. As the number of clusters increases, the chances that at least one cluster in an event is near or below the induction signal threshold, while still well above the collection signal threshold, increases.
Figure 3.9: Normalized PDFs for two data selection cuts for the $0\nu\beta\beta$ to $0^+_1$ excited state decay mode. Multi-site energy spectrum. Full reconstruction requirement plotted in red, with 75% reconstruction requirement in blue. Relaxing the spatial reconstruction requirement of MS events allows proportionally more events in the full Q value peak to be accepted compared to the strict full reconstruction requirement. This can be seen by the higher blue relaxed PDF line at the 2458 keV Q value as compared to the red, fully reconstructed PDF.
Figure 3.10: Daily activity for each background component in the golden quality EXO-200 data. Single-site spectrum. Two distinct fits are displayed. In blue, is the final fit from the published search for Majoron-emitting double-beta decay modes[32], which covers Run2abc. In black, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $0\nu\beta\beta$ to $0^+_1$ search. These rates come from spectral fits over the entire data set, with the $2\nu\beta\beta$ component listed referring to $^{136}$Xe decays to the ground state only.
Figure 3.11: Daily activity for each background component in the golden quality EXO-200 data. Multi-site spectrum. Two distinct fits are displayed. In blue, is the final fit from the published search for Majoron-emitting double-beta decay modes [32], which covers Run2abc. In black, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $0\nu\beta\beta$ to $0^+_1$ search. These rates come from spectral fits over the entire data set, with the $2\nu\beta\beta$ component listed referring to $^{136}$Xe decays to the ground state only.
The search for double-beta decays to nuclear excited states of the product nucleus makes use of a boosted decision tree (BDT) [1] described in more detail in Section 3.3.1, to maximize sensitivity to signal. Briefly, decision trees are quickly trained on multivariate datasets defined by the user. An individual decision tree is composed of a sequence of nodes where decisions are made. At each node, the choice of which variable to use and the position of the variable’s cut is made to maximize the separation between signal and background for the population of that node. However, even given this greedy approach to building discriminating trees, single trees are relatively weak, binary, discriminators. Boosted decision trees combine up to a thousand or more individual trees trained in a sequential fashion. This combination results in a smooth discriminator variable that improves discriminatory power and reduces overtraining to anomalies.

As discussed in more detail in Section 3.5, EXO-200 analyses have historically used two-dimensional profile likelihood fits split by single-site and multi-site events, to search for a number of different signals. These include a precision measurement of $2\nu\beta\beta$ $^{136}$Xe-to-$^{136}$Ba ground state decay [30], and limits for $0\nu\beta\beta$ $^{136}$Xe to the $^{136}$Ba ground state [31] and similar Majoron-emitting modes [32]. These fits use total event energy, and standoff distance—a variable measuring the minimum distance between any charge cluster in the event to the edge of the xenon fiducial volume on a simultaneous fit to SS and MS data. As mentioned earlier, the SS-to MS ratio of each component included in the fit is determined by MC. Separating single- and multi-site
events has been shown to be effective in isolating events with $\beta$-only physics from those including $\gamma$ processes—namely the de-excitation photons from excited $^{136}$Ba atoms.

The analysis searching for decays to excited states of $^{136}$Ba also uses a 2-dimensional profile likelihood fit split by single-site (SS) and multi-site (MS) events [39]. The discriminator variable produced by the boosted decision tree machine training algorithm replaces standoff distance as the second variable in the fit.

This chapter details the process for producing the discriminator variable, which smoothly ranges from +1 for signal-like events, to −1 for background-like events. Later sections discuss the process of quantifying the systematic uncertainties of the search by comparing calibration source data and simulated Monte Carlo (MC) data. Calculating this systematic error relies on using the trained discriminator variable in two-dimensional profile likelihood fits of simulated low background data sets with a known number of artificially-injected, simulated signal events. Lastly, the process of determining search sensitivity is explained, a procedure which relies on similar simulated low background data sets, but without added signal.

A schematic diagram of the complete analysis process is shown in Figure 4.1. Briefly, parallel processes operate on data and MC in concert, with a training dataset being built from a small subset of simulated MC events (magenta), and a discriminator variable defined by training on this dataset. This trained algorithm is applied to all events in data and MC files. Probability Distribution Functions (PDFs) of the fitting variables to be used in performing profile-likelihood fits are also created. Shown in cyan is the process of determining systematic error through source agreement, including producing the $2\nu\beta\beta$ dataset used as a comparison source. The full process takes approximately a week using batch processing queues at SLAC.
Figure 4.1: Complete Machine Learning Pipeline. Parallel processes operate from top to bottom on data (left) and MC (right). Weighted decision trees, the output of the BDT machine learning algorithm, are produced from training on a MC-generated Toy Dataset and are applied to both data and MC to fill the discriminator variable. This process is shown in magenta with a dashed-arrow web. Source agreement, including the production of the $2\nu\beta\beta$ subtracted dataset used as a source comparison to MC, is cyan with a dash-dot arrow web. The final fit process is yellow with a solid-arrow web. Data and MC files are filled with discriminator variables from the BDT. Probability distribution functions (PDFs) of the fitting variables are then built from MC while an EXO workspace is made to hold all data events. The final fit applies the PDFs to the dataset, while the sensitivity process first generates MC simulations of a dataset to measure typical fit performance.
4.1 Input Variables

As previously described, BDTs are used to search complex multivariate datasets for a particular signal signature. The initial step is to finalize a list of input variables. Though a BDT’s discrimination power is not easily compromised by including weak or useless variables, using too many similar variables or ones with limited separation capability may slow the analysis. Furthermore, the algorithm is only capable of finding relatively simple correlations, so inputting the full 3D position and energy of each cluster in every event and attempting to independently discover, for instance, the Compton Scattering formula and use it, will absolutely fail. Since the decision nodes only cut on a single variable, and there must be a relatively shallow number of decision nodes in each tree, independently discovering a new complicated relationship between the variables and cutting on that is not something a BDT can do.

To some extent, more complex information can be extracted by increasing the maximum number of decision points allowed in individual decision trees, currently set at 5. Increasing this depth hurts training time and does not provide a strong benefit, so improving final performance comes down to choosing input variables which each provide enough uncorrelated discrimination power but not so many that the training slows. Keeping the number of input variables below 10 is sufficient. Variations of these BDT definitions were studied by the Excited States work group, and described in the technical note [86].

No matter which decay mode the BDT is tuned to search for, fewer than 10 input variables are used. The final lists can be separated into two classes: standard event variables which are used in every search, and decay specific variables which are used to find unique signatures marking a particular decay mode.
4.1.1 Standard Variables

Three standard event variables are chosen as inputs, as these have well understood and very good agreement between data and MC: event multiplicity, total event energy, and standoff distance, shown listed in Table 4.1. Event multiplicity is the number of separate ionization energy clusters deposited in an event. Total event energy sums energy from every cluster and uses the light/charge anticorrelation and de-noising to define a combined energy scale with improved detector resolution. Standoff distance is the minimum distance between any cluster in an event and the edge of the fiducial volume. Note that although the standoff distance is not used on its own as a variable in the final 2D fits, it is used as one of the input variables for the discriminator variable. Since the discriminator variable is one of the fit variables, the standoff distance has indirect influence on the final fit results. This reduced importance of one of the primary variables used in previous EXO-200 searches to this analysis is justified by the highly multi-site nature of the signal. Standoff distance is most useful for separating single-site $\gamma$-producing backgrounds outside the detector from the uniform double beta decay signal within the xenon, and has less of an influence in distinguishing multi-site components, including decays to excited states.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Shorthand</th>
</tr>
</thead>
<tbody>
<tr>
<td>event multiplicity</td>
<td>total number of clusters in event</td>
<td>multiplicity</td>
</tr>
<tr>
<td>total event energy</td>
<td>summed energy from all clusters</td>
<td>energy</td>
</tr>
<tr>
<td>standoff distance</td>
<td>distance between event and fiducial volume</td>
<td>standoff</td>
</tr>
</tbody>
</table>

Table 4.1: List of the standard event variables used as BDT algorithm input. Multiplicity, energy, and standoff are used for all decay modes searches.

Using total event energy as an input variable requires some additional work. For MC events, energy is calculated with access to the actual simulated energy deposited in the detector, resulting in artificially perfect energy resolution. Measured values of detector energy resolution are imposed over this artificially perfect spectrum which
allow MC to realistically mimic data. The correction is only applied to MC in aggregate on total energy distributions. This process effectively convolves the resolution and energy distribution together when producing PDFs. This solution is not suitable for the machine learning process because the machine learning input variables must be calculated on a per-event level, rather than as a collective application to the whole MC energy spectrum.

For the machine learning process, a stochastic smearing term is added to the MC total energy in each event. This term, $\Delta E$, is randomly drawn from a Gaussian distribution with standard deviation of the detector energy resolution at that event’s energy. Because the smearing variable is added to the total energy of an event after individual cluster energies have been combined, the process for single-site and multi-site events is functionally identical. Only the resolution function determining the width of the Gaussian from which the stochastic value is drawn differs, as the multi-site energy resolution is different, and typically larger, than the single-site one. The calibration source energy distributions show no appreciable difference between the event-specific method described here and the aggregate process used in previous analyses [86].

### 4.1.2 Decay Specific Variables

In addition to the three standard variables, several decay-specific variables are used. Because each decay mode’s signature is different, the number of decay specific variables differs for each search. Decay-specific variables are designed to look for sharp energy signals possible in the event, such as those associated with de-excitation $\gamma$s emitted by the excited $^{136}\text{Ba}$ daughter as it relaxes to the ground state, or the entire Q-energy for $0\nu\beta\beta$ events. These decay-specific variables will be collectively known as $\epsilon$-search variables, since they are defined in relation to a specific energy relevant to the search.
In the $2\nu\beta\beta$ decay to the $0^+_1$ state, there are two sharp energy signals possible with $\epsilon_1 = 760.5$ keV and $\epsilon_2 = 818.5$ keV. For the $\beta\beta2\nu$ decay to the $2^+_1$ state, there is only one, the $\epsilon_2 = 818.5$ keV. When searching for $0\nu$ decays, additional sharp energy signals associated with the $\beta\beta$ energy of the decay are also expected, and so associated variables are developed to identify them. In addition to variables searching for single energy signals, combinations of expected sharp energy signals are also valuable.

The decay mode with the most distinct sharp energy signals possible is the $\beta\beta0\nu$ decay to the $0^+_1$ state. The individual energy variables are $\epsilon_{\gamma760} = 760.5$ keV, $\epsilon_{\gamma818} = 818.5$ keV, and $\epsilon_{\beta\beta} = 878.8$ keV. Combination variables included as BDT inputs are $\epsilon_{\gamma1\gamma2} = 1579$ keV for both $\gamma$ signals together, $\epsilon_{\beta\beta\gamma} = 1639.3$ keV for a $\beta\beta$ and $\gamma$ signal together, and $\epsilon_{\text{total}} = 2458$ keV. The effective multi site detector resolution at the full Q value of 2458 keV is $\sigma/E = 1.65 \pm 0.05\%$ [31], which rises at lower energies as shown in Figure 2.7. However, the absolute rather than relative resolution is still higher at larger energies, such that the difference in energy between the de-excitation $\gamma$s, 58 keV, is much more important at lower rather than higher total cluster energies. Thus while a separate variable is used to search for the 760.5 keV and 818.5 keV $\gamma$s, only one variable, at 1639.3 keV, is used to search for the combination between the $0\nu\beta\beta$ cluster plus one of the two de-excitation $\gamma$s. For the $0\nu$ decay to the $0^+_1$ excited state, six energy-signal variables are added to the three standard event variables for a total of nine input variables to the BDT. This is below the soft maximum of 10. All other decay modes use a subset of these variables. A list of the decay specific variables by decay mode is found in Table 4.2

All of the $\epsilon$-search variables are defined in the same way, with different energies, as follows:

$$\epsilon_{S_{\text{energy}}} \equiv \min_{j \in S} \{|E_j - \epsilon_{\text{target}}|\}$$

(4.1)
Table 4.2: List of the event variables used as BDT algorithm input for all four decay modes. Target energy shown in first column, with the search function for each listing denoted by the particle type in the labels. The 1639.3 keV energy target is used both for the beta clusters produced in the $0\nu \to 2^+_1$ decay and the $\beta \gamma$ combination clusters produced by the $0\nu \to 0^+_1$ decay.

Here $\epsilon_{energy}$ is the variable to be used as a machine learning input, $\epsilon_{target}$ is the value of one of the sharp energy-signals expected in the decay, and $E_j$ is the energy of the jth energy cluster in the event. As with the total event energy variable discussed in Section 4.1.1, there is an additional complication when applying this function to MC, and it is solved in virtually the same way but at the cluster rather than total event energy level. The cluster energy is calculated with access to the simulated deposited energy, with artificially perfect resolution. This is then smeared to realistic multi-site energy by adding a stochastic variable selected from a Gaussian distribution applying the multi-site resolution to an individual cluster. The resulting variable, $\epsilon_{S_{energy}}$, peaks at zero for events where clusters have the energy of one of the emitted photons or $\beta\beta0\nu$s. It does not have a sharp minimum in events where a photon’s energy is shared over more than one energy cluster.

Significant work has been done to develop a more complex $E_j$ value with the goal is of reconstructing energies of scattered photons in order to increase final fit sensitivity. One of these efforts is detailed in Beryl Bell’s undergraduate thesis [10]. There, she expands and tests a Compton-scattering based clustering algorithm the author developed in an attempt to trace back individual photon clusters adding up to the total energy of one of the original photons in the decay. This scatter-only algorithm
has only been applied to the $0^+_1$ final excited state modes, but ultimately did not help. It was ineffective due to a combination of cluster energy uncertainties and clusters containing energy from more than one original particle. These issues were addressed by correctly including all possible scatter combinations and by considering the energy of grouped clusters as well as the position-based scattering behavior. Additional remaining issues have been identified, but in the current form this grouping is still not helpful. At Stanford University, Scott Kravitz applied a k-means cluster algorithm using spatial compactness alone [66]. This successfully reconstructed photons which were emitted back to back, one of the two favored directions, but was extremely poor in reconstructing photons emitted in the same direction– which was the other favored emission direction as discussed in Section 3.1. It ultimately did not improve sensitivity, and imposed a systematic cost, so was abandoned. At Drexel University, Yung-Ruey Yen continued to develop his doctoral thesis work [94], which set a limit of $1.2 \cdot 10^{23}$ yr for the $2\nu\beta\beta$ decay to the $0^+_1$ excited state mode using a simple spectral fit analysis and also described the development of an energy-only algorithm to distinguish the $\beta\beta$ cluster from the two $\gamma$-rays. However, further work on this this energy-only method also did not show improvement.

4.1.3 Pre-Processing

The finalized BDT input variables must be calculated for every event in all data and MC files, in a step known as “pre-processing.” The files it creates are shown in the very top level of Figure 4.1, and pre-processing is assumed to have been completed when data or MC files are referred to later in this text. In pre-processing the BDT input variables are calculated and stored in a streamlined ROOT object. After this point only the input variables are accessible. Wire signals, individual cluster information, raw APD signals, etc. are no longer available.
Every relevant MC event, whether signal, background, or calibration source, is pre-processed. However, since all searches use a subset of the variables needed for the $\beta\beta_{0\nu}$ decay to the $0^+_1$ search, pre-processing is only done once. The BDT is then trained on the subset of variables relevant for each decay mode. This approach saves disk space, an important consideration for this search. Each set of full MC files uses approximately 100 GBs and separate versions of these files must exist for the fully and partially reconstructed versions of the data for each of these four search modes. The minimum disk space required just for MC used in the searches to excited state decays approaches 1 TB.

The distributions of the input variables used in all eight decay searches are shown in Figures A.7 to A.10 in the Appendix. As an example, those for the $2\nu\beta\beta$ to $0^+_1$ decay mode with full 3D reconstruction are shown in Figure 4.2. In place of plotting each background MC distribution separately, only the decay signal and combined backgrounds are shown.

### 4.2 Training and Applying Machine

Once the input variables are determined and calculated for all data and MC files, the BDT is ready to be trained. First an appropriate training dataset is generated. This must include equal quantities of signal and background events, and the background events should be as similar to the expected real data background as possible. This dataset is built by taking individual simulated events from the MC files. Then the algorithm is trained on this dataset and a large number of individual decision trees produced. Lastly, the trained machine is used to calculate a discriminator variable for every data and MC event that was previously “pre-processed.”
Figure 4.2: Distribution of all variables used for the $0\nu\beta\beta$ to $0^+\text{state}$ decay search imposing that every cluster in every event is fully reconstructed in space. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the collected data set is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $0^+_1$ decay-specific variables shown in the second row. Similar plots for the other search modes are found in the Appendix.
4.2.1 Building the Training Data Set

The training data set is built by taking individual simulated events from MC files, both background and signal, and aggregating them into one million “background” events and one million “signal” events. To match the simulated background dataset to that of the collected low background data as much as possible, the activity of backgrounds included in the training dataset come from fits to the data. Activity rates for the training set come from the final fit in the Majoron search [32]. As shown in Section 3.5, the difference between this and a later fit conducted using BDT output and the full Run2 dataset does not differ significantly, justifying this choice.

Once the background activity rates are set, the training data sets are built. All standard cuts are applied, such as fiducial volume and energy threshold conditions. These standard cuts are briefly described in Section 5.3. Two training data sets are built for each decay search mode, one requiring all events to have fully 3D-reconstructed clusters, and the other requiring only 75% of each event’s energy to come from fully 3D-reconstructed clusters. The motivation for extending the data selection to partially-reconstructed events is recovering significant signal efficiency, and is discussed in more detail in Section 5.1. Though the 75% requirement is somewhat arbitrary, it is seen in Chapter 5 that this choice keeps the majority of events by avoiding the sharpest drop off in signal efficiency, and has been previously used in an EXO-200 publication [36]. Because only one data set is required to train each decay mode, actual MC events are used instead of building PDFs of the necessary variables from the full MC statistics, and dynamically producing variables from the resulting PDF. Because of the discrete nature of single events in a large simulated data set of a million events, some very low activity backgrounds may not have the exact relative ratios they do in the background-model fit. This is shown not to harm the effectiveness of the boosted decision tree algorithm.
Following the best practices of using BDTs, each dataset is split in half, with half used to train individual decision trees, and the other half to test decision tree performance on a dataset it was not trained on. This reduces the ability of the BDT to “overtrain” to specific details of the dataset at hand.

4.2.2 Training and Testing the BDT Algorithm

Training of the BDT algorithm proceeds according to the detailed algorithm description in Section 3.3.1. Eight-hundred and fifty separate nested decision trees up to 5 nodes deep are sequentially constructed from the input variables in the training half of the generated data set, with variant BDT versions with differing numbers of trees or nodes not improving the ability to separate signal from background with minimal systematic error [86]. At each node, the algorithm is “greedy,” i.e., it finds the cut out of all of the variables which maximizes discrimination power for the events which populate that particular node. This greediness at each node keeps the algorithm robust against weak variables, as a variable which does not give any useful information will simply not be selected as a cut on most nodes. After the first decision tree is created, the events which are misclassified by that tree are given increased weight, and a second decision tree produced which better separates those events. This process is repeated sequentially, until every subsequent tree is trained with increased weight on those events which the previous tree is unable to properly identify.

The decision trees are then tested against the reserved testing half of the generated data set. Their performance at correctly classifying events in this testing half determines the weight of each tree in the final boosted decision tree. This penalizes any trees which may have found non-representative patterns in the training dataset and guarantees the final BDT will perform well at discriminating signal and background events in data sets it has not explicitly trained on, such as the collected Run2 EXO-200 data. This weighting process protects against statistical fluctuations and
rests on the assumption that there are no systematic differences between the two half-sets.

The final output from the BDT training process is a discriminator variable calculated for every event. The discriminator variable is found by passing an event’s input variables through each of the separate 850 decision trees, which classify it as either +1 for signal-like, or -1 for background-like. Then this classification is multiplied by the weight of the decision tree which produces it, and the weighted average across all trees calculated. This value can take any number between +1 to -1, with those values falling closer to the extreme values being those events where the majority of high weighted trees agree on a classification type.

Plots of the discriminator variable as it has been defined for each of the eight searches of interest are shown in Figures A.13 to A.18 in the Appendix. An example, for the $2\nu\beta\beta$ decay to the $0^+_1$ daughter excitation level with partial 3D reconstruction allowed is shown in Figure 4.3.

Each input variable can be ranked for how crucial it is to building the BDT discrimination variable. This ranking is based on the fraction of decision tree nodes where an input variable is used, weighted by the weight of the trees those nodes belong to. Rankings are shown together with the plot of discriminator variable for the corresponding search mode in the Appendix. An example table is shown in Table 4.3.

Since most of the data in EXO-200’s Run2 dataset are $2\nu\beta\beta$ decays of $^{136}\text{Xe}$ to the ground state of $^{136}\text{Ba}$, variables which discriminate well against this type of event tend to rank high regardless of decay search. This is what drives the high rankings of multiplicity. Decays to the ground state are single-site events 82.5% of the time [28], while most decays to the daughter excited states, no matter the decay mode, are not. Because the decay-specific variables listed in Table 4.2 have a similar definition they are partially correlated. This may have the effect of elevating the use of one
of these variables at the expense of the others, which could have yielded a similar discrimination.
100 % Reconstruction Required, 2νββ Decay to 0⁺ State

Figure 4.3: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the 2νββ to 0⁺ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra. Training and testing data are more similar than expected from poissonian statistics because the fully-trained algorithm uses weights from the testing data while combining individual decision trees. Similar plots for other search modes are shown in the Appendix Figures 7.10-7.15.

Table 4.3: Ranking of input variables in the final boosted decision tree built for the 2νββ to 0⁺ excited state search with partial 3D-reconstruction (at least 75% event energy 3D reconstructed required).
4.2.3 Applying BDT Algorithm

Once the machine is completely trained and tested, the set of weighted decision trees used to create the discriminator variable is applied to each event in data and MC which were previously “pre-processed.” The results of this process is represented as the second set of boxes down in Figure 4.1, directly beside the red arrows representing the calculation itself, and is carried out as follows.

After the discriminator variable is created, the ROOT-associated structures of data RooFit workspaces and MC PDFs can be built. The data RooFit workspace is used to store all of the low background data in a convenient format, and sets up all the RooFit variables that are necessary to build PDFs and ultimately fit the data. The MC PDFs translate the full statistical power of the MC datasets into simple two-dimensional histograms for each component used in fits.

4.3 Shape Systematic Studies

Many systematic uncertainties in this analysis are the same as in previous published EXO-200 work [31, 32, 39] and do not need to be revisited. These are presented in Table 4.4. The SS fraction constraint accounts for deviations in the single-site fraction between data and MC. It is set by observing the discrepancy in the proportion of single-site events between the data from calibration source runs and simulated versions of these source runs. The Common Normalization accounts for detection efficiency uncertainty, and is also derived from the calibration sources which have a known activity. The Background Normalization corrects for uncertainties in the location of degenerate backgrounds. The Neutron Capture Fractions allow the relative rates of individual neutron-capture related backgrounds to float against each other. Lastly, the radon activity within the liquid xenon obtained by measuring the $^{214}$Bi - $^{214}$Po coincidences in the detector, and the uncertainty on this measurement is translated into a constraint on the normalization of the Radon in the Liquid Xenon.
As the discriminator variable is unique to this analysis, a new systematic error arising from discrepancy between data and MC in this variable is calculated and applied to the final fit. This is the Excited State Normalization constraint in the last line of Table 4.4, and must be calculated separately for each decay mode.

The Excited State Normalization is measured via an “unskewing” process described in more detail in the following section. Briefly, it involves artificially adjusting, or “unskewing”, a set of MC PDFs by the observed ratio between source data and simulated sources. These unskewed PDFs are used to generate toy datasets with a known number of injected signal events. These are then fit with our regular MC PDFs in exactly the same way the final fit to the real data is performed. The difference between the number of injected signal events and the fit results gives a measure of the systematic error due to source disagreement from using this variable in the technique. Because the discriminator variable differs for every decay search, this process must be repeated for each, with different outcomes.

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS Fraction</td>
<td>4</td>
</tr>
<tr>
<td>Common Normalization</td>
<td>8.6</td>
</tr>
<tr>
<td>Background Normalization</td>
<td>20</td>
</tr>
<tr>
<td>Neutron Capture Fractions</td>
<td>20</td>
</tr>
<tr>
<td>Radon in Liquid Xenon</td>
<td>10</td>
</tr>
<tr>
<td>Excited State Normalization</td>
<td>varies</td>
</tr>
</tbody>
</table>

Table 4.4: List of shared constraints used to account for systematic uncertainties in the final fit regardless of decay mode. The values for the Excited State Normalization, which varies for each search, are shown in Table 4.6

4.3.1 Agreement with Calibration Source Data

Three $\gamma$-ray sources are used for calibration of EXO-200, $^{228}$Th, $^{226}$Ra, and $^{60}$Co. One set of “source agreement” ratios is produced for each one. These sources are used over several different “calibration periods,” so plenty of data is available to be compared to simulated calibration sources. The dominant background in this analysis
is the $2\nu\beta\beta$ decay to the ground state of $^{136}$Ba. Taking advantage of this fact, fitted background contributions other than this are removed from the Low Background data set via a PDF subtraction process to produce a “pure” $2\nu\beta\beta$ $^{136}$Xe decay to ground state data set. This in turn is used as a fourth comparison point to MC. The background activities are set by the same fits shown in Figures 3.10 and 3.11. Both the previously published Majoron fit and the new background-only discriminator and energy fit are used to produce a subtracted dataset to show that the resulting source agreement is not strongly affected. It was thus decided to simply use the new fit, since it includes the entirety of Run2. The new background-only fit uses energy and a discriminator, but the choice was made to use one fit, the fully-reconstructed $0\nu\beta\beta$ decay to the $0^+_1$ excited state mode, rather than produce separate background models for each decay mode in this analysis.

An example source agreement plot is shown in Figure 4.4. Plots are made for all source sets, for both MS and SS components, and for discriminator and energy variables. The simple ratios for each bin between the source data and simulated MC are stored for later PDF adjustment. This allows the MC PDFs to be forced into agreement with the observed data in order to study the effects of any data-MC shape differences on the final sensitivity. Usually this adjustment uses the simple ratio between the data and MC, however this does not work well in any bin containing very few data events. In this case, low statistics can result in very high, unrealistic ratios between source and simulation.

Several versions of gradually reducing the impact of low statistics bins were studied. The best way to reduce unwarranted large source agreement ratios caused by very few events is to use a graduated scheme that slowly reduces the impact of the calculated ratio as the number of data events falls. By inspecting the bin statistics and the calculated ratios at several threshold values, a gradual method of reducing the effect of low statistics bins was produced [86]. When a bin has over 18 events in
Figure 4.4: “Source agreement” plot for the subtracted $2\nu\beta\beta$ decay to $^{136}$Ba ground state source constructed as described in the text. Distribution of discriminator variable, constructed from machine learning algorithm, shown. Discriminator variable has been trained for the $0\nu\beta\beta$ to $2^+_1$ excited state decay mode search. Multi-Site component.
it, the data / MC ratio is saved with no modifications. Between 10 and 18 events in a single bin the calculated ratio’s deviation from 1 is reduced by a factor of 10. Points in this statistical region will still have a small influence in the ultimate systematic error as the PDF’s will still be skewed partially in these bins, but not enough to dominate the systematic error. For those bins with fewer than 10 events, the calculated data / MC ratio is not saved at all, and is simply set to 1. This modified skewing function preserves enough real information to be actually useful, and reduces the systematic error from 38% down to 12.7% for the published version of the search for 2νββ decay to 0⁺ [39]. However, the addition of Run 2d data appears to have exacerbated a discrepancy caused by imperfect ⁴⁰K MS subtraction, and the systematic error on signal shape stays relatively high. This ⁴⁰K discrepancy is discussed in more detail in Section 5.4.

4.3.2 Unskewing PDFs

After all the source agreement plots have been created and ratios saved, these are used to “unskew” the MC PDFs to behave more like the low background data. Each background or signal component is unskewed according to the source data type which it most resembles, though in some cases this is difficult to determine. If it is unclear which calibration source a background most resembles, unskewing coefficients of the ²²⁸Th source are used because they are the ones determined with higher statistics. These forced-to-empirical PDFs are used to build toy low background datasets with respective ratios determined by previous fits. Two hundred excited state signal events are then injected into each of 1,000 datasets. These are then fit with the unmodified MC PDFs, and the deviation from the fitted number to the injected number of signal events is used as the measure of systematic error for the Excited State Normalization factor listed as variable in Table 4.4, since each mode has a different value.
A test of this process is performed by unskewing one component at a time. This way, dominant contributors can be clearly identified. In the analysis for the paper describing the EXO-200 collaborations search for $2\nu\beta\beta$ decays to the $0^+_1$ excited state of $^{136}$Ba, the unskewing effect is very minimal for the calibration sources. Conversely, the effect is significant for the subtracted $2\nu\beta\beta$ component and so this component is not tested alone. Instead, for each decay mode, and also for fully and partially reconstructed data, the unskewing effect from the three calibration together sources is contrasted with the effect from all four sources, to demonstrate the separate effect of the subtracted $2\nu\beta\beta$ unskewing. The separate effect of the subtracted data set could also have be seen by unskewing the $2\nu\beta\beta$ component alone, however this additional test was unnecessary as the unskewing from all four unskewing components is required to calculate the full Excited State Normalization systematic, and the contrast between the sources and the subtracted data can be determined by comparing the effect of unskewing the three sources against the total from all four.

The reason why the majority of the unskewing systematic error comes from the $2\nu\beta\beta$ subtracted data source is not fully understood. This is the only data source which requires a subtraction process to produce, and it also has the most issues with low-statistics data bins. Furthermore, it is the largest background component to the searches for $\beta\beta$ decays to the excited states of $^{136}$Ba regardless of mode, given that it comprises the majority of the Low Background data events.

Figures A.19 through A.24 in the Appendix show the tests done to determine the contribution from the Excited State Normalization systematic, followed by the final study determining the overall systematic error. The results are summarized here in Table 4.5. Though the sensitivity studies in the following section reference medians as a more suitable statistical choice due to the skewing of the mean by the positive tail in sensitivity distributions, the shape studies done here do not show a strong effect from the choice of representative statistic.
<table>
<thead>
<tr>
<th>Study</th>
<th>$0\nu$ to $0^+_1$</th>
<th>$0\nu$ to $2^+_1$</th>
<th>$2\nu$ to $0^+_1$</th>
<th>$2\nu$ to $2^+_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unskew Source</td>
<td>204.3</td>
<td>211</td>
<td>194.5</td>
<td>178.6</td>
</tr>
<tr>
<td>Unskew All</td>
<td>239.2</td>
<td>223.6</td>
<td>284.3</td>
<td>135.7</td>
</tr>
<tr>
<td>More $^{40}$K subtracted</td>
<td>239.5</td>
<td>221.2</td>
<td>282.5</td>
<td>168</td>
</tr>
<tr>
<td>Less $^{40}$K subtracted</td>
<td>237.9</td>
<td>222.5</td>
<td>282.3</td>
<td>126.8</td>
</tr>
<tr>
<td><strong>Partial Reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unskew Source</td>
<td>193.1</td>
<td>208.1</td>
<td>156.8</td>
<td>260.9</td>
</tr>
<tr>
<td>Unskew All</td>
<td>184</td>
<td>188</td>
<td>115</td>
<td>230</td>
</tr>
</tbody>
</table>

Table 4.5: Unskewing Study Summary. 200 signal events of each signal type are artificially added to 1,000 simulated Low Background data sets. Histograms of best fit values plotted in appendix. Means presented here. Deviation from 200 shows magnitude of systematic error on the Excited Signal Normalization.

Two sets, one for each 3D-reconstruction requirement, are shown sequentially, and each figure shows the results of the same study as applied to all four possible excited state decay modes. Both fully reconstructed and 75% reconstructed selection cases test the effect of unskewing only calibration sources (most minimal), then unskewing all sources, including the subtracted $2\nu\beta\beta$ spectra. For the fully reconstructed selection case, one additional test is conducted to show the effect of the $^{40}$K peak residual in the MS spectrum. Two versions of the subtracted dataset are produced, one artificially removing an additional 140 MS K40 events than fitting suggests, and the other artificially removing 140 fewer such events. The spectrum was altered by 140 events because a similar test with 120 additional events for the Run 2abc demonstrated small but visible changes to the $^{40}$K energy peak by eye [67], and the full dataset is approximately 20% larger, so correspondingly more additional events are used. The results of this study are shown in Figures A.21 and A.22.
<table>
<thead>
<tr>
<th>Decay</th>
<th>Level</th>
<th>3D Reconstruction</th>
<th>Normalization (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta\beta 2\nu$</td>
<td>$0^+_1$</td>
<td>(100% 3D)</td>
<td>42.2</td>
</tr>
<tr>
<td>$\beta\beta 2\nu$</td>
<td>$0^+_1$</td>
<td>(&gt; 75% 3D)</td>
<td>42.5</td>
</tr>
<tr>
<td>$\beta\beta 2\nu$</td>
<td>$2^+_1$</td>
<td>(100% 3D)</td>
<td>32.2</td>
</tr>
<tr>
<td>$\beta\beta 2\nu$</td>
<td>$2^+_1$</td>
<td>(&gt; 75% 3D)</td>
<td>15</td>
</tr>
<tr>
<td>$\beta\beta 0\nu$</td>
<td>$0^+_1$</td>
<td>(100% 3D)</td>
<td>19.6</td>
</tr>
<tr>
<td>$\beta\beta 0\nu$</td>
<td>$0^+_1$</td>
<td>(&gt; 75% 3D)</td>
<td>8</td>
</tr>
<tr>
<td>$\beta\beta 0\nu$</td>
<td>$2^+_1$</td>
<td>(100% 3D)</td>
<td>11.8</td>
</tr>
<tr>
<td>$\beta\beta 0\nu$</td>
<td>$2^+_1$</td>
<td>(&gt; 75% 3D)</td>
<td>6.0</td>
</tr>
</tbody>
</table>

Table 4.6: Excited state normalizations calculated via source agreement and unskewing for each decay mode studied.

4.4 Sensitivity Studies

To properly evaluate the ability to find decays to excited states of $^{136}$Ba in the low background data requires performing a two-dimensional profile likelihood fit over the energy and discriminator variables. However, performing such a fit on the actual low background data before finalizing all fit constraints and processes may bias the search. To avoid this risk, the fitting procedure is done on toy MC datasets intended to look similar to the true low background data. One thousand individual toy MC generated datasets are produced with background component ratios given by the background-only $0\nu\beta\beta$ to $0^+_1$ mode fit results. Unlike in the systematic study, no signal events – no decays to excited states of $^{136}$Ba – are included in the simulated data sets. The events which populate these toy datasets are generated by directly producing events drawn from the PDF distribution for energy and discriminator variable. This guarantees there are no low statistics issues and each MC dataset is actually distinct from the others.

Each toy dataset is then fitted by a two dimensional profile likelihood fit, just as if it were a fit to real Low-Background data. A single example profile is included in Figure 4.5, on an example simulated low background dataset which has a non-zero best fit value. Every fit includes all Gaussian constraints tabulated in Tables 4.4 and
4.6. In addition to this fit, a profile analysis is conducted over the number of signal events accommodated by the fit. From this profile analysis, the 90% confidence limit determined for each dataset. Then, a histogram is produced of all of these limits. Since no signal events were injected into the Low Background data set, the median of this distribution represents the expected sensitivity of the search. These histograms can be found in Figures A.27 and A.30 in the Appendix. The median values are tabulated in Table 4.7.

4.4.1 Dominant Fit Contributions

To determine what drives the final fit sensitivity, a study is done varying the errors included in the fit. First, all parameters representing systematic errors are fixed without the ability to float. The sensitivity in this case is calculated as described in the previous section. Then, the standard errors, excluding the Excited State Normalization, are allowed to float within their usual constraint values as shown in Table 4.4.
Comparing these sensitivities isolates the signal normalization which depends on the “source agreement” of the discriminator variable. Finally all the errors are included in the fit by allowing their associated parameters to float (with Gaussian penalties further from central value). The median from 1,000 fits in this case, for each mode, is the final sensitivity of the search.

Table 4.7: Sensitivity Study Summary. 1,000 simulated Low Background data sets with no signal are produced for discriminator variables drawn from each decay search mode. A Profile Likelihood scan is done over the number of signal events accommodated in the fit within the 90% confidence level. Histograms of the limit are found plotted in appendix. Medians presented here. Lower values correspond to a stronger limit.

<table>
<thead>
<tr>
<th></th>
<th>$0\nu \to 0^+_\nu$</th>
<th>$0\nu \to 2^+_\nu$</th>
<th>$2\nu \to 0^+_\nu$</th>
<th>$2\nu \to 2^+_\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Systematics</td>
<td>10.00</td>
<td>15.00</td>
<td>15.00</td>
<td>25.00</td>
</tr>
<tr>
<td>No Excited State Normalization</td>
<td>25.00</td>
<td>20.08</td>
<td>51.45</td>
<td>140.31</td>
</tr>
<tr>
<td>Full Systematics</td>
<td>25.03</td>
<td>20.94</td>
<td>76.37</td>
<td>175.00</td>
</tr>
<tr>
<td><strong>Partial Reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Systematics</td>
<td>25.00</td>
<td>15.01</td>
<td>50.00</td>
<td>85.05</td>
</tr>
<tr>
<td>No Excited State Normalization</td>
<td>20.00</td>
<td>15.00</td>
<td>35.00</td>
<td>150.26</td>
</tr>
<tr>
<td>Full Systematics</td>
<td>20.00</td>
<td>15.00</td>
<td>40.00</td>
<td>157.60</td>
</tr>
</tbody>
</table>
CHAPTER 5
EFFICIENCY AND SENSITIVITY

As mentioned in Section 4.2.1, two separate versions of the data selection cuts are used in this analysis. The standard data selection removes all events which do not have complete 3D reconstruction from both the U and V wire planes described in Section 2.1. In practice, this means that any event with even one cluster below the threshold for V-wire induction signal detection, even if it is above the threshold of U-wire charge collection signal detection, is excluded from analysis. Because the decays to excited states regardless of the decay mode typically have high average multiplicities, as shown in Figure 5.1, the chances for a single cluster in a typical even to not have an induction signal and thus not be completely 3D reconstructed is relatively high. The signal efficiency, which is linearly proportional to the sensitivity of the search, is gravely affected by the loss of these events.

The multiplicities of the considered decay channels, including both 0ν and 2ν processes to both the 01+ and 21+ excited state levels, are compared against the multiplicity of Run2 Low Background events in Figure 5.1. Multiplicities have been normalized at their peaks to aid in direct shape comparison, and the primary effect arises from the daughter excited state level—as would be expected as the 01+ excited level emits twice as many de-excitation γs as the 21+ level, and these γs cause the most scatters and multiple energy depositions within the xenon in each event. Both decays ending at the 2+ state have a peak multiplicity at two and fall off rapidly, while both decays to the 0+ state peak at three and fall off more slowly. There is also a more subtle but consistent effect from the beta component process as well, with the 0νββ cases shifted
Figure 5.1: Normalized multiplicity by decay mode, compared to Run2 Low Background data. Color coding in legend. Low background data clearly shows lower average multiplicity than any signal. The dominant structure for the decays to excited states is driven by the final state level, with the two de-excitation $\gamma$s emitted by the $0^+$ decay mode pushing both of these peaks to three as compared with the single de-excitation $\gamma$ from the $2_1^+$ decay. A smaller effect is shown between zero and two neutrino modes, with the higher average $\beta\beta$ cluster for both $0\nu$ modes causing a slight shift to higher multiplicity as compared with the $2\nu$ case.

very slightly towards higher multiplicity compared with the $2\nu\beta\beta$ cases. This can be explained by the greater average energy of the $\beta\beta$ deposit creating a slightly higher chance of a larger charge cluster which is interpreted as a multiplicity two event by the detector. The Run2 Low Background data, in magenta, is overwhelmingly single site, and as compared to the others, most rapidly falls off to higher multiplicities. This data is highly dominated by the $2\nu\beta\beta$ decay to the ground state of $^{136}\text{Ba}$ which is predominantly single-site, with some additional, primarily multi-site, backgrounds included.
The efficiency loss by the standard 100% 3D reconstruction requirement motivates the study of an alternative data selection choice which retains more signal while having limited and manageable systematic effects on the analysis. The recent EXO-200 paper on muon-associated Neutron captures [36] also required high efficiency for multi-cluster events. This paper used a 75% 3D reconstruction requirement—i.e. 75% of the energy of an event must come from clusters in the event which have full 3D positions determined. After looking at the efficiency curves for decays to excited states while varying the spatial reconstruction requirement, and the systematic error from the 75% 3D reconstruction requirement, this analysis elected to also use this requirement as the definition of the relaxed data cut.

5.1 Reconstruction Efficiency

In order to keep only the events where the main portion of the deposited energy is well located in space, the relaxed 3D reconstruction requirement weights clusters within an event by energy. This allows for more flexibility in responding to individual evens than would be allowed by an alternate method allowing a set number of clusters per event to be non-reconstructed. This alternative would have a sharp binning effect corresponding to the multiplicity of the event, rather than a smoother structure depending on the energy distribution between event clusters.

Figures 5.3 to 5.6 show the effect of varying the weighted energy reconstruction requirement for all four different excited state decay modes, as well as the Low Background data set. To highlight the effect of the reconstruction cut alone, it is the only cut applied in these plots. All plots are made comparing the total summed event energy over all charge clusters, and the summed event energy of all 3D reconstruction clusters. For a given ratio of 3D reconstructed to total event energy, the cumulative fraction of events meeting or exceeding the cut requirement is shown. Thus, all start at full efficiency (1) while at the minimum cut requirement (0% reconstructed en-
Figure 5.2: Simulated $2\nu\beta\beta$ to $2_1^+$ excited decay mode events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line.

Energy). This falls off to the right as the requirement becomes more stringent. At the far right with 100% reconstruction efficiency is the cut requirement which has been used in most previous EXO-200 analyses, including the published search for the $2\nu\beta\beta$ decay to the $0_1^+$ excited state [39].

Events with only two clusters are least affected by the 3D reconstruction requirement, and are shown as the highest, black line in every plot. The efficiency reduction is similar in signal and data for two cluster events, differing only to the extent that the energy distributions of multiplicity two events from different sources cause a stronger reduction higher for two-cluster events with lower average cluster energy. These effects within a single multiplicity across all modes and the Low Background data can be seen in Figures 5.7 and 5.8. For brevity, only multiplicity two and greater than...
Figure 5.3: Simulated $2\nu\beta\beta$ to $0^+_1$ excited decay mode events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line.

five are shown. It is clear that for low multiplicity events, especially the two cluster events seen in Figure 5.7, though efficiency losses for signal are dramatic as the reconstruction requirement becomes stricter, losses in the low background data are even higher. However, the multiplicity structure of the simulated signal events and the low background data set differ dramatically, with signals being much more multi-site than the typical gamma backgrounds in the Run2 dataset as seen in Figure 5.1. Thus because the fraction of two cluster signal events is so much higher than in background, even the higher relative recovery of partially reconstructed background events to signal events does not negatively affect the increased separation power provided by a lower 3D requirement. For very high multiplicity events, greater than five clusters, the differential effect of the reconstruction requirement nearly disappears, as seen in
Figure 5.4: Simulated $0\nu\beta\beta$ to $2^+_1$ excited decay mode events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line.

Figure 5.8. Though the deficit by fraction within a multiplicity may be equivalent, the multiplicity structure of the source spectra is not, and gaining equivalent efficiency in both background and signal at these high multiplicity events improves the accepted signal to background ratio.

It is clear from all efficiency plots, that a 75% 3D reconstruction requirement is near the inflection point of the efficiency loss, and can allow maximum benefit with lower systematic effect. The gain of using a cut that has been previously tested in an EXO-200 analysis outweighs the slight differences between modes which may otherwise have generated a slightly different value for each decay search. The signal efficiencies for both selection cuts is shown in Table 5.1. Improvement is dramatic across the board.
Figure 5.5: Simulated $0\nu\beta\beta$ to $0_1^+$ excited decay mode events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line.

### 5.2 Sensitivity

By combining the efficiencies described in the Section 5.1, summarized by Table 5.1, with the number of fitted signal counts in 1,000 simulated low background datasets for each decay mode of interest, summarized in Table 4.7, the improvement in calculated limit for each decay mode can be determined. Table 5.2 presents these increases for all four decay modes of interest, including two $2\nu\beta\beta$ decay processes, one each to daughter $0_1^+$ and $2_1^+$ excited levels, and the equivalent two $0\nu\beta\beta$ processes. It is clear that substantial improvement is shown across the board by using the more relaxed reconstruction requirement, with the most dramatic improvement resulting in nearly three times the limit derived from fully-reconstructed events alone. Though
Figure 5.6: Run2 Low Background events by efficiency of 3D reconstructed energy. Multiplicity 2 through 5 are color coded and shown on legend. Efficiency falls off monotonically with increasing multiplicity. Events with multiplicity greater than five are grouped together and shown as one additional line.

this section only presents the relative improvement, the final sensitivity limits for all cases are shown in Table 6.1.

### 5.3 Data Selection

In addition to the reconstruction requirement previously described this chapter, several additional cuts are applied to ensure only good quality events are included in this analysis. A list of these cuts is included in Table 5.3. Calibration source data does not require and thus does not use the Golden and Muon Veto cuts, but maintain all others listed. MC generated data are also cut, though like the calibration data only a subset of the listed cuts apply. These are the Scintillation Number, Multiplicity,
Figure 5.7: Effect of varying fraction of 3D reconstructed energy requirement per event in all excited decay modes and the low background data. Only multiplicity 2 events are shown.

Fiducial Volume, Energy, and Missing Position cuts. A cut unique to simulated data, requiring non-zero simulated energy deposition, is roughly equivalent to the Scintillation Energy cut.

5.4 Subtracted Low Background Data Agreement

$^{40}$K comprises a significant portion of the low energy MS background in the Low Background data. When $^{40}$K undergoes $\beta^+$ decay, as it does 10.86% of the time, it ends at an excited level of the daughter argon nucleus and subsequently emits a 1.46 MeV de-excitation $\gamma$. [57]. For currently uncertain reasons, the subtraction process which removes fitted backgrounds from the Low Background dataset to produce a subtracted $2\nu\beta\beta$ “source,” does not perfectly remove this $^{40}$K peak for MS events
Figure 5.8: Effect of varying fraction of 3D reconstructed energy required per event in all excited decay modes and the low background data. Only multiplicity greater than 5 events are shown.

and a strong residual is visible in Figure 5.9. This remnant is more prominent in the full Run2 dataset than the Run2abc subset, which increases the associated excited state normalization systematic for the $2\nu\beta\beta$ to the $0_{1}^{+}$ excited state decay. It rises from the 15% used in the published search [39] to the 42.2% in Table 4.5. This explains why the final sensitivity for the full Run2 100% reconstructed data, at $1.15 \cdot 10^{24}$ in Table 6.1, is lower than the published $1.7 \cdot 10^{24}$ yr [39]. A test with artificially reduced excited state normalization over the full Run2 data reproduced the higher sensitivity.
Table 5.1: Total signal efficiency for search modes by data selection cut. Significant efficiency improvement from relaxed data cut regardless of decay mode, with particular increases seen for the $0\nu$ to $0^+_1$ case.

<table>
<thead>
<tr>
<th>Decay Mode</th>
<th>$0\nu$ to $0^+_1$</th>
<th>$0\nu$ to $2^+_1$</th>
<th>$2\nu$ to $0^+_1$</th>
<th>$2\nu$ to $2^+_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Reconstruction</td>
<td>31.2%</td>
<td>54.6%</td>
<td>25%</td>
<td>39.2%</td>
</tr>
<tr>
<td>Partial Reconstruction</td>
<td>70.1%</td>
<td>85.4%</td>
<td>38.6%</td>
<td>60.8%</td>
</tr>
</tbody>
</table>

Table 5.2: Fractional increase in sensitivity by relaxing reconstruction requirement. Definite improvement seen for each mode. The $0\nu$ decay modes show nearly three times the sensitivity found when requiring 100% reconstruction, and in the $2\nu$ modes sensitivity roughly doubles.

<table>
<thead>
<tr>
<th>Study</th>
<th>$0\nu$ to $0^+_1$</th>
<th>$0\nu$ to $2^+_1$</th>
<th>$2\nu$ to $0^+_1$</th>
<th>$2\nu$ to $2^+_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity Increase</td>
<td>2.81x</td>
<td>2.19x</td>
<td>2.95x</td>
<td>1.72x</td>
</tr>
</tbody>
</table>

Table 5.3: Table of data cuts applied to Low Background data. A subset of these apply to Source Data and MC generated simulation, as described in text.

<table>
<thead>
<tr>
<th>Cut Name</th>
<th>Cut Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise</td>
<td>Remove events tagged as noise.</td>
</tr>
<tr>
<td>Drift Time</td>
<td>Require light and charge signals within maximal drift time.</td>
</tr>
<tr>
<td>Diagonal Cut</td>
<td>Remove events ($\alpha$-decays) with a high light to charge ratio.</td>
</tr>
<tr>
<td>Scintillation Number</td>
<td>Keep only events with a single scintillation cluster.</td>
</tr>
<tr>
<td>Scintillation Energy</td>
<td>Require nonzero scintillation energy.</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>Remove multiplicity zero events.</td>
</tr>
<tr>
<td>Fiducial Volume</td>
<td>Hexagonal. 162 mm apothem cut. 10 mm and 182 mm Z cut.</td>
</tr>
<tr>
<td>Solicited Trigger</td>
<td>Remove the .1 Hz forced trigger event class.</td>
</tr>
<tr>
<td>Energy</td>
<td>Require at least 980 keV total event energy.</td>
</tr>
<tr>
<td>Muon Veto</td>
<td>Remove events up to 5 ms following hit on veto panel trigger.</td>
</tr>
<tr>
<td>Missing Position</td>
<td>Two cases. Either 100% or 75% reconstructed.</td>
</tr>
<tr>
<td>Golden</td>
<td>Event must fall within “golden” data quality periods.</td>
</tr>
</tbody>
</table>
Figure 5.9: $2\nu\beta\beta$ “calibration source” dataset created by subtracting fitted backgrounds from Low Background Data. Part (a) includes only Run2abc data, while part (b) includes all Run2 data. Shown is a projection of a 2D PDF onto the energy variable. Multi-site events only. Visible discrepancy near $^{40}\text{K}$ peak.
CHAPTER 6
CONCLUSIONS

A search for double-β decays of $^{136}$Xe to the $0^+_1$ and $2^+_1$ excited states of $^{136}$Ba is presented in this thesis. Four decay modes are of interest, including $0\nu$ and $2\nu$ decays to both $0^+_1$ and $2^+_1$ daughter excited states. A machine learning algorithm, described in more detail in Chapters 3 and 4, has been developed to boost the sensitivity of this search. Final half-life sensitivities for the four modes of interest are presented in the context of published world limits in Table 6.1. The novel partially reconstructed data selection described in Chapter 5, together with the inclusion of an additional 22% of data discussed in Section 2.5.1, provides a benefit over previous published EXO-200 results. Searches for decay modes other than the $2\nu\beta\beta$ to the $0^+_1$ state with EXO-200 data are new, and half-life sensitivities range from $7.86 \cdot 10^{23}$ years to $2.00 \cdot 10^{25}$ years.

The $2\nu\beta\beta$ decay to the $0^+_1$ excited state, which is expected to have the shortest half-life of all modes considered with $T_{1/2}^{\text{estimate}} = 2.5 \cdot 10^{25}$ yr, remains beyond the reach of this analysis. Model uncertainty while calculating this expected value, and additional data currently being collected may yet allow discovery of this decay mode with EXO-200. The extraction of additional topological information not yet fully exploited by EXO-200 analysis could also boost the search sensitivity. Possibilities include the recovery of photon scattering information within an individual event, and making explicit use of the angular correlation between emitted de-excitation $\gamma$s from the $0^+_1$ daughter state.

A discovery of this decay mode could provide useful in constraining nuclear matrix element calculations. The $0\nu\beta\beta$ decay to the $0^+_1$ excited state of $^{136}$Ba is expected
to be less suppressed than $2\nu\beta\beta$ with respect to the ground state, making their simultaneous discovery at future, larger detectors (such as nEXO) a possibility.

<table>
<thead>
<tr>
<th></th>
<th>$0\nu \text{ to } 0_1^+$</th>
<th>$0\nu \text{ to } 2_1^+$</th>
<th>$2\nu \text{ to } 0_1^+$</th>
<th>$2\nu \text{ to } 2_1^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sensitivity)</td>
<td>$4.37 \cdot 10^{24}$ yr</td>
<td>$9.15 \cdot 10^{24}$ yr</td>
<td>$1.15 \cdot 10^{24}$ yr</td>
<td>$7.86 \cdot 10^{23}$ yr</td>
</tr>
<tr>
<td><strong>Partial Reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sensitivity)</td>
<td>$1.23 \cdot 10^{25}$ yr</td>
<td>$2.00 \cdot 10^{25}$ yr</td>
<td>$3.39 \cdot 10^{24}$ yr</td>
<td>$1.35 \cdot 10^{24}$ yr</td>
</tr>
<tr>
<td><strong>Published Limits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXO-200</td>
<td>-</td>
<td>-</td>
<td>$6.9 \cdot 10^{23}$ yr</td>
<td>-</td>
</tr>
<tr>
<td>Kamland-Zen</td>
<td>$2.4 \cdot 10^{20}$ yr</td>
<td>$2.6 \cdot 10^{20}$ yr</td>
<td>$8.3 \cdot 10^{23}$ yr</td>
<td>$4.6 \cdot 10^{23}$ yr</td>
</tr>
</tbody>
</table>

Table 6.1: Half Life Sensitivity in context. Published values from EXO-200 [39] and Kamland-Zen [63] compared with sensitivity values from this analysis.
APPENDIX
ADDITIONAL PLOTS

A.1 Fit Comparison

Figure A.1: Daily activity for each background spectrum in the golden quality EXO-200 data. Single-site spectrum. Two distinct fits are displayed. In blue, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $\beta\beta_0\nu$ to $0^+_1$ search with full 3D reconstruction requirement. In black is a fit over the whole Run2abcd dataset using fitting variables designed for the same mode, but at least 75% 3D reconstruction. Absolute daily activity for numbers in fit, not normalized for different efficiency.
Figure A.2: Daily activity for each background spectrum in the golden quality EXO-200 data. Multi-site spectrum. Two distinct fits are displayed. In blue, is a background only fit on the whole Run2abcd dataset using the fitting variables designed for the $\beta\beta_0\nu$ to $0^+_1$ search with full 3D reconstruction requirement. In black is a fit over the whole Run2abcd dataset using fitting variables designed for the same mode, but at least 75% 3D reconstruction. Absolute daily activity for numbers in fit, not normalized for different efficiency.
A.2 Decay Mode Input Variables

100 % Reconstruction Required, $2\nu\beta\beta$ Decay to $0^+$ State

Figure A.3: Distribution of all variables used for the $2\nu\beta\beta$ to $0^+_1$ state decay search using a strict event cut requiring 100% of the event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $\beta\beta2\nu$ to $0^+_1$ decay-specific variables shown in the second row.
100 % Reconstruction Required, $2\nu\beta\beta$ Decay to $2^+$ State

Figure A.4: Distribution of all variables used for the $2\nu\beta\beta$ to $2_1^+$ state decay search using a strict event cut requiring 100% of the event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $2\nu\beta\beta$ to $2_1^+$ decay-specific variables shown in the second row.
Figure A.5: Distribution of all variables used for the $0\nu\beta\beta$ to $0^+_{1\nu}$ state decay search using a strict event cut requiring 100% of event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $0^+_{1\nu}$ decay-specific variables shown in the second and third rows.
Figure A.6: Distribution of all variables used for the $0\nu\beta\beta$ to $2^+_1$ state decay search using a strict event cut requiring 100% of the event energy to come from fully-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $2^+_1$ decay-specific variables shown in the second row.
Figure A.7: Distribution of all variables used for the $2\nu\beta\beta$ to $0^+_1$ state decay search using the relaxed event cut allowing up to 25% of the event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $2\nu\beta\beta$ to $0^+_1$ decay-specific variables shown in the second row. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.
75 % Reconstruction Required, $2\nu\beta\beta$ Decay to $2^+$ State

Figure A.8: Distribution of all variables used for the $2\nu\beta\beta$ to $2^+_1$ state decay search using the relaxed event cut allowing up to 25% of the event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $2\nu\beta\beta$ to $2^+_1$ decay-specific variables shown in the second row. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.
Figure A.9: Distribution of all variables used for the $0\nu\beta\beta$ to $0^{+}$ state decay search using the relaxed event cut allowing 25% of event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $0^{+}$ decay-specific variables shown in the second and third rows. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.
Figure A.10: Distribution of all variables used for the $0\nu\beta\beta$ to $2^+_{1}$ state decay search using the relaxed event cut allowing up to 25% of the event energy to come from non-3D reconstructed events. Signal distributions are plotted with solid blue. A combined background distribution, grouping all backgrounds together to mimic the real Low Background data, is plotted with dashed red. One million events, half signal, and half background, are used to create the distributions, and normalized counts plotted. The standard event variables of energy, multiplicity, and standoff distance are in the top row, with the $0\nu\beta\beta$ to $2^+_{1}$ decay-specific variables shown in the second row. Standoff distance is not well defined for events with non-3D reconstructed clusters. For these cases it is set to the non-physical value of -999.
A.3 Decay Mode Discriminator Variables

100 % Reconstruction Required, $2\nu\beta\beta$ Decay to $0^+$ State

![Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $0^+_1$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.](image)

Figure A.11: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $0^+_1$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>multiplicity</td>
<td>27.4%</td>
</tr>
<tr>
<td>2</td>
<td>energy</td>
<td>25.8%</td>
</tr>
<tr>
<td>3</td>
<td>standoff distance</td>
<td>14.4%</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_{\text{sum}}$</td>
<td>13.6%</td>
</tr>
<tr>
<td>5</td>
<td>$\gamma_{760}$</td>
<td>9.8%</td>
</tr>
<tr>
<td>6</td>
<td>$\gamma_{818}$</td>
<td>9%</td>
</tr>
</tbody>
</table>

Table A.1: Ranking of input variables in the final boosted decision tree built for the $2\nu\beta\beta$ to $0^+_1$ excited state search with full 3D-reconstruction (100% event energy 3D reconstructed required).
Figure A.12: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $2^+$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>energy</td>
<td>32%</td>
</tr>
<tr>
<td>2</td>
<td>multiplicity</td>
<td>25.9%</td>
</tr>
<tr>
<td>3</td>
<td>$\gamma_{ALL}$</td>
<td>19.3%</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_{818}$</td>
<td>13.7%</td>
</tr>
<tr>
<td>5</td>
<td>standoff distance</td>
<td>9.1%</td>
</tr>
</tbody>
</table>

Table A.2: Ranking of input variables in the final boosted decision tree built for the $2\nu\beta\beta$ to $2^+$ excited state search with full 3D-reconstruction (100% event energy 3D reconstructed required).
Figure A.13: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $0\nu\beta\beta$ to $0^+_1$ excited state search with full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

Table A.3: Ranking of input variables in the final boosted decision tree built for the $0\nu\beta\beta$ to $0^+_1$ excited state search with full 3D-reconstruction (100% event energy 3D reconstructed required).
Figure A.14: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $2^+_1$ excited state search with at full event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

Table A.4: Ranking of input variables in the final boosted decision tree built for the $0\nu\beta\beta$ to $2^+_1$ excited state search with full 3D-reconstruction (100% event energy 3D reconstructed required).
75 % Reconstruction Required, $2\nu\beta\beta$ Decay to $0^+$ State

Figure A.15: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $0^+_1$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>multiplicity</td>
<td>27.1%</td>
</tr>
<tr>
<td>2</td>
<td>energy</td>
<td>26.7%</td>
</tr>
<tr>
<td>3</td>
<td>standoff</td>
<td>14.7%</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_{\text{760}}$</td>
<td>14.2%</td>
</tr>
<tr>
<td>5</td>
<td>$\gamma_{760}$</td>
<td>9.1%</td>
</tr>
<tr>
<td>6</td>
<td>$\gamma_{818}$</td>
<td>8.1%</td>
</tr>
</tbody>
</table>

Table A.5: Ranking of input variables in the final boosted decision tree built for the $2\nu\beta\beta$ to $0^+_1$ excited state search with partial 3D-reconstruction (at least 75% event energy 3D reconstructed required).
75 % Reconstruction Required, $2\nu\beta\beta$ Decay to $2^+$ State

Figure A.16: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $2\nu\beta\beta$ to $2^+$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>energy</td>
<td>34.1%</td>
</tr>
<tr>
<td>2</td>
<td>$\gamma_{ALL}$</td>
<td>23.1%</td>
</tr>
<tr>
<td>3</td>
<td>multiplicity</td>
<td>22.8%</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_{818}$</td>
<td>12.7%</td>
</tr>
<tr>
<td>5</td>
<td>standoff distance</td>
<td>7.3%</td>
</tr>
</tbody>
</table>

Table A.6: Ranking of input variables in the final boosted decision tree built for the $2\nu\beta\beta$ to $2^+$ excited state search with partial 3D-reconstruction (at least 75% event energy 3D reconstructed required).
75 % Reconstruction Required, $0\nu\beta\beta$ Decay to $0^+_{1}$ State

Figure A.17: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the $0\nu\beta\beta$ to $0^+_{1}$ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>energy</td>
<td>23.7%</td>
</tr>
<tr>
<td>2</td>
<td>multiplicity</td>
<td>17.1%</td>
</tr>
<tr>
<td>3</td>
<td>standoff distance</td>
<td>12.6%</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_B$</td>
<td>9.0%</td>
</tr>
<tr>
<td>5</td>
<td>$\gamma_{ALL}$</td>
<td>8.9%</td>
</tr>
<tr>
<td>6</td>
<td>$\gamma_{760}$</td>
<td>7.9%</td>
</tr>
<tr>
<td>7</td>
<td>$\gamma_{818}$</td>
<td>7.4%</td>
</tr>
<tr>
<td>8</td>
<td>$\gamma_{B760}$</td>
<td>7.1%</td>
</tr>
<tr>
<td>9</td>
<td>$\gamma_{comb}$</td>
<td>6.2%</td>
</tr>
</tbody>
</table>

Table A.7: Ranking of input variables in the final boosted decision tree built for the $0\nu\beta\beta$ to $0^+_{1}$ excited state search with partial 3D-reconstruction (at least 75% event energy 3D reconstructed required).
75 % Reconstruction Required, 0νββ Decay to 2+ State

Figure A.18: Normalized discriminator variable for signal (solid blue) vs background (hatched red) for the 0νββ to 2+ excited state search with at least 75% event energy 3D-reconstructed. Discriminator variable spectra from the training half of the data set (colored points) are overlapped to the color-filled shapes of testing data spectra.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>energy</td>
<td>34.1%</td>
</tr>
<tr>
<td>2</td>
<td>standoff distance</td>
<td>18.9%</td>
</tr>
<tr>
<td>6</td>
<td>γ_s18</td>
<td>14.5%</td>
</tr>
<tr>
<td>9</td>
<td>γ_B760</td>
<td>13.3%</td>
</tr>
<tr>
<td>4</td>
<td>γ_ALL</td>
<td>11.9%</td>
</tr>
<tr>
<td>2</td>
<td>multiplicity</td>
<td>7.3%</td>
</tr>
</tbody>
</table>

Table A.8: Ranking of input variables in the final boosted decision tree built for the 0νββ to 2+ excited state search with partial 3D-reconstruction (at least 75% event energy 3D reconstructed required).
A.4 Unskewing Studies

A.4.1 Full 3D Reconstruction Requirement

100 % Reconstruction, Unskew Source Only

Figure A.19: 100% reconstructed data selection. Excited decay normalization study—
MC PDF’s “unskewed” to artificially agree with calibration source data. Subtracted
$2\nu\beta\beta$ to ground state source is not “unskewed.” 200 injected signal events into 1,000
simulated Low Background datasets, fit by standard Low Background process. Devi-
ation from 200 injected signal determines systematic error. Figure shows four decay
modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^{+}$
daughter, right column $2^{+}$ daughter.
Figure A.20: 100% reconstructed data selection. Excited decay normalization study—MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2
u\beta\beta$ to ground state source agreement are used. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^{+}_1$ daughter, right column $2^{+}_1$ daughter.
Figure A.21: 100% reconstructed data selection. Excited decay normalization study–MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2\nu\beta\beta$ to ground state source agreement used. $2\nu\beta\beta$ subtraction altered to remove 140 fewer K40 MS events. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^+_1$ daughter, right column $2^+_1$ daughter.
100 % Reconstruction, 140 more 40K counts

Figure A.22: 100% reconstructed data selection. Excited decay normalization study—MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2\nu\beta\beta$ to ground state source agreement used. $2\nu\beta\beta$ subtraction altered to remove 140 more K40 MS events. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0_{1}^{+}$ daughter, right column $2_{1}^{+}$ daughter.
A.4.2 Partial 3D Reconstruction Requirement

75 % Reconstruction, Unskew Source Only

Figure A.23: 75% reconstructed data selection. Excited decay normalization study—MC PDF’s “unskewed” to artificially agree with calibration source data. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0\nu_i^+$ daughter, right column $2\nu_i^+$ daughter.
Figure A.24: 75% reconstructed data selection. Excited decay normalization study—MC PDF’s “unskewed” to artificially agree with source data. Both calibration source data and subtracted $2
u\beta\beta$ to ground state source agreement used. 200 injected signal events into 1,000 simulated Low Background datasets, fit by standard Low Background process. Deviation from 200 injected signal determines systematic error. Figure shows four decay modes simultaneously, upper row $2\nu$ decays, lower row $0\nu$ decays. Left column $0^+_1$ daughter, right column $2^+_1$ daughter.

A.5 Sensitivity Studies

A.5.1 Full 3D Reconstruction Requirement

A.5.2 Partial 3D Reconstruction Requirement
100 % Reconstruction, No Constraints in Fits

2n, 0+ (No Constraints)

0n, 0+ (No Constraints)

2n, 2+ (No Constraints)

0n, 2+ (No Constraints)

Figure A.25: Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. No systematic errors included, simulated dataset statistics control. 100% reconstruction selection requirement.
Figure A.26: Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All historic systematic errors included. Excited state normalization error not included. 100% reconstruction selection requirement.
100 % Reconstruction, All Constraints Included

Figure A.27: Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All systematic errors included. Final sensitivity for mode. 100% reconstruction selection requirement.
75 % Reconstruction, No Constraints in Fits

Figure A.28: Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. No systematic errors included, simulated dataset statistics control. 75% reconstruction selection requirement.
75 % Reconstruction, No Excited State Constraint in Fits

Figure A.29: Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All historic systematic errors included. Excited state normalization error not included. 75% reconstruction selection requirement.
Figure A.30: Histogram of 90% confidence level limits from 1,000 individual simulated Low Background datasets. Profile around best fit value for each dataset until 90% confidence level reached. All systematic errors included. Final sensitivity for mode. 75% reconstruction selection requirement.


[34] EXO-200, Collaboration. An RF-only ion-funnel for extraction from high-pressure gasses. *Int. J. Mass Spec.* 379 (2015), 11–120.


[66] Kravitz, S. *private communication.*

[67] Kravitz, S. *private communication.*


[77] Moore, Dave. *private communication.*


