
UNIVERSIDADE D
COIMBRA

Andrey Solovov

# EXPLORATION OF MACHINE LEARNING TECHNIQUES

## FOR DISCRIMINATION OF NEUTRINOLESS DOUBLE BETA DECAY OF $^{136}$Xe

**Dissertação no âmbito do Mestrado em Astrofísica e Instrumentação para o Espaço orientada pelo Professor Doutor Francisco Filipe Bento Neves e pelo Professor Doutor Alexandre Miguel Ferreira Lindote e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia.**

Junho de 2021

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Dissertation

# Exploration of Machine Learning techniques for discrimination of neutrinoless double beta decay of $^{136}$Xe

*Author:*
Andrey Solovov

*Supervisors:*
Dr. Francisco Filipe Bento Neves
Dr. Alexandre Miguel Ferreira Lindote

*Submitted in fulfillment of the requirements for the degree of Master in Astrophysics and Space Instrumentation (MAIE)*

*in the*

Laboratório de Instrumentação e Física Experimental de Partículas (LIP)
Department of Physics

June 16, 2021

UNIVERSITY OF COIMBRA

# *Abstract*

Faculty of Sciences and Technology
Department of Physics

Master in Astrophysics and Space Instrumentation (MAIE)

**Exploration of Machine Learning techniques for discrimination of neutrinoless double beta decay of $^{136}$Xe**

by Andrey Solovov

Detection of neutrinoless double beta decay ($0\nu\beta\beta$) is a scientific goal with significant implications for neutrino physics and cosmology, and a new generation of large-scale detectors is underway for probing $0\nu\beta\beta$ half-lives in the $10^{25} - 10^{27}$ yr range. The LUX-ZEPLIN (LZ) dark matter experiment will use a liquid xenon time projection chamber (TPC) with competitive specifications for the search of $0\nu\beta\beta$ events of $^{136}$Xe. With the current background rejection analysis, the projected LZ $0\nu\beta\beta$ background rate is $6.8 \times 10^{-3}$ counts/keV/kg/year in the relevant energy region for this decay, with almost all of the background occurring due to events with a single primary electron. The two primary electrons of $0\nu\beta\beta$ events produce a different energy deposition topology than these backgrounds, motivating the investigation of topology-based discrimination techniques.

In this work, the viability of a discrimination cut based on machine learning algorithms was examined. A modular and expandable framework was developed for feature extraction and dynamic dimensionality reduction on simulated time-series data, for the purpose of binary classification. This framework was applied to discrimination of the two classes of event, with four different binary classifiers (k-nearest neighbors, support vector machines, Gaussian process, and random forests), using data simulated (with equal balancing of the two classes) for this purpose. To identify the best configuration, a fast performance comparison and assessment method was developed, capable of extrapolating to balancings other than the one employed. In effect, a versatile procedure for systematic viability assessment of binary classification approaches was conceived.

With the aid of the Instituto Nacional de Computação Distribuída, and using a custom version of the ANTS2 package with Geant4 integration, two datasets with $\sim 10^4$ datapoints each were simulated: one where all the primary electrons of the two classes had the most favorable initial emission direction for this particular analysis (vertical); and one where the initial emission direction was isotropic. A miniaturized LZ TPC was modeled, recreating topology transport physics: drift diffusion; electroluminescence; and pulse generation in the upper PMT array. The Gaussian process classifier was seen to perform best out of the four selected. For a target $0\nu\beta\beta$ sensitivity of 80%, it reduced the single electron background to $\sim 22\%$ of its initial rate for vertical emission, and to $\sim 37\%$ of the original rate for isotropic emission. However, for ISO there was no predicted improvement to LZ sensitivity. The main cause of the difficulty in discrimination was determined to be drift diffusion, with a promising avenue of investigation being Gaussian deconvolution methods.

UNIVERSIDADE DE COIMBRA

# *Resumo*

Faculdade de Ciências e Tecnologia
Departamento de Física

Mestrado em Astrofísica e Instrumentação para o Espaço (MAIE)

## Exploração de técnicas de aprendizado de máquina para discriminação do decaimento beta duplo, sem neutrinos, do $^{136}$Xe

por Andrey Solovov

A detecção do decaimento beta duplo sem neutrinos ($0\nu\beta\beta$) é um objectivo científico com implicações significativas nas áreas da física de neutrinos e cosmologia, e uma nova geração de detectores de grande escala está em desenvolvimento para sondar meias-vidas de $0\nu\beta\beta$ na gama dos $10^{25} - 10^{27}$ anos. A experiência de matéria escura LUX-ZEPLIN (LZ) usará uma câmara de projecção temporal (TPC) de xénon líquido com especificações competitivas para a procura de eventos de $0\nu\beta\beta$ do $^{136}$Xe. Com os cortes correntemente aplicados, a taxa de eventos de fundo projectada para o LZ é de $6.8 \times 10^{-3}$ contagens/keV/kg/ano na região de energia de interesse para este decaimento, quase toda devido a eventos com a emissão de um só electrão primário. Os dois electrões primários dos eventos $0\nu\beta\beta$ produzem uma topologia de deposição de energia diferente da dos fundos, motivando o desenvolvimento de um corte baseado em algoritmos de auto-aprendizagem que explore essa diferença topológica.

Neste trabalho foi examinada a viabilidade e o potencial de diferentes algoritmos de auto-aprendizagem aplicados a este problema. Foi desenvolvida uma infraestrutura expansível e modular para a extracção de propriedades e redução dinâmica de dimensionalidade em sinais temporais simulados deste tipo de detectores. Esta infraestrutura foi aplicada à discriminação das duas classes de evento, com quatro classificadores binários diferentes (*k-nearest neighbors*, *support vector machines*, *Gaussian process* e *random forests*), usando dados simulados (com balanço igual entre as duas classes) para este propósito. Para identificar a melhor configuração, foi desenvolvido um método rápido de comparação e avaliação de desempenho, capaz de extrapolar para balanços diferentes dos utilizados. Em suma, foi concebido um procedimento de avaliação de viabilidade sistemático para abordagens de classificação binária.

Com o auxílio do Instituto Nacional de Computação Distribuída, e usando uma versão customizada do pacote ANTS2 com integração do Geant4, foram simulados dois conjuntos de dados com $\sim 10^4$ pontos cada: um em que todos os electrões primários de cada classe tinham a direcção inicial de emissão mais favorável (vertical); e um em que a direcção de emissão inicial era isotrópica. Foi modelada uma TPC de LZ miniaturizada, simulando a física relevante para a topologia: deriva e difusão dos electrões de ionização; electroluminescência; e geração dos pulsos na matriz de PMTs superior. Verficou-se que o classificador por *Gaussian process* tinha o melhor desempeho dos quatro seleccionados. Dada uma sensibilidade-alvo a $0\nu\beta\beta$ de 80%, o classificador reduziu o fundo devido a electrões singulares a $\sim 22\%$ da sua taxa inicial para emissões verticais, e a $\sim 37\%$ da sua taxa inicial para emissões isotrópicas. No entanto, para emissões isotrópicas não foi prevista uma melhoria na sensibilidade do LZ. A principal causa de dificuldade na discriminação determinou-se ser a difusão durante a deriva, com os métodos de desconvolução Gaussiana a serem uma via promissora para investigação futura.

# *Acknowledgements*

This work would not have been possible without so many people.

My supervisors, Dr. Francisco Neves and Dr. Alexandre Lindote, with their unwavering support, their insights and advice, have taken a vague ambition and mapped the way to achieving it. I thank them not only for that, but also for tirelessly driving me to be better, for their perseverance when things got out of hand, and for not letting anyone throw in the towel in the many times that things looked grim. I could not have been luckier than to have them as mentors.

Professor Maria Isabel Silva Ferreira Lopes put this wonderful opportunity and life-changing endeavor underway. I thank her, in addition to that, for her advice and support, and for welcoming me into the LIP Dark Matter group, where I congregated with people that inspire me, and learned things that have molded me into who I am.

Dr. Andrey Morozov, Dr. Paulo Braz, Cédric Pereira and Antonio Gurgel have been instrumental in shaping up this work into its current form. I thank Andrey for his interest, openness, guidance, novel ideas and for the opportunity to work with ANTS2 more deeply as a member of its development team. I thank Paulo for sharing his experience and insights, for his constant help and support, and for pointing the way forward in so many respects. Of course I thank Cédric for the simulation data that got the ball rolling, but more important than that were all the discussions, wisecracks, fun times and great moments shared. I thank him for that and for being a bigger inspiration than he'll likely ever know. I thank Antonio for all the help with chiseling the Docker image and repo into shape, for all the wonderful conversations about music, amateur radio, tech, among other things, and for always pushing me to improve.

I thank the LIP computing teams in Lisbon and Minho for their invaluable help and support getting the distributed simulations up and running, with special mentions of Dr. João Pina and Professor António Pina, without whom there would be nothing to conclude. Also a big thank you to the LZ collaboration for their openness, approachability, friendly disposition and inspiration to strive. I also thank the LIP Competence Center on Simulation and Big Data for many things, but most of all for organizing the 2019 Data Science School and Symposium at UMinho, where I learned the methods employed here and to which can be traced my current embedding.

A very special thank you to Professor José Lopes Pinto da Cunha for his boundless investment, concern, understanding and patience. I cannot express how grateful I am for all the help he has provided over the years, and for how he achieves the impossible for care of his students whenever necessary. Also an equally special thank you to Professor João Manuel de Morais Barros Fernandes and Professor Vitor Hugo da Rosa Bonifácio for their kind help, insights, fruitful conversations, hospitality, inspiration and tireless support in my research endeavors at OGAUC and UA.

A special thank you to all the friends I made, and all the wonderful people I have crossed paths with, at LIP, at the university, at the observatory and SAC. Drs. Filipe Veloso, Hugo Simões and Luís Pereira, as well as Ana Galvão, Ana Vasconcelos, Bruno, Guilherme, Giorgio, Henrique, Jacinto, João Jordão, Marcela, Marlene, Miguel, Ricardo, Sérgio... life would not be the same without you. A very warm thank you to David Mascarenhas, for the great and impactful moments shared, for the inspiring conversations, the sage advice, for expanding my worldview and for helping me grow as a person. I am incredibly lucky to be your friend.

And how can I ever thank enough the three people who have never ceased helping and enduring since the start? All the toil, the turbulence, frustration and scares, all the small victories, reliefs, surprises and successes... I am lucky to be able to share them with you. There are so many things for which I want to thank you, and hopefully in time I will. I would not have had this journey in any other way. Olga Solovova and Vladimir Solovov, you are my first and continued inspiration, thank you. Inês Dinis, I am forever indebted to you for all the help and motivation, and for making every moment valuable. Love of my life and best friend: thank you so much.

Finally, protracted as this process was, this is in order:
Thank you to *literally everyone* for your patience.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **$0\nu\beta\beta$** | Neutrinoless Double Beta Decay |
| **$1e$** | Single Electron |
| **$2\nu\beta\beta$** | Ordinary Double Beta Decay |
| **AUC** | **A**rea **U**nder the **C**urve |
| **b2b** | **B**ack-**t**o-**B**ack Emission *(both primaries oriented vertically)* |
| **BAU** | **B**aryon **A**symmetry in the **U**niverse |
| **CBR** | **C**osmic **B**ackground **R**adiation |
| **CDG** | **C**osmic **D**iffuse **G**amma ray background |
| **CL** | **C**onfidence **L**evel |
| **CMB** | **C**osmic **M**icrowave **B**ackground |
| **DEC0** | **DEC**AY**0** Primary Energy Distributions |
| **DT** | **D**ecision **T**ree |
| **HPXe** | **H**igh-**P**ressure **Xe**non |
| **EL** | **E**lectro**l**uminescence |
| **FV** | **F**iducial **V**olume |
| **FWHM** | **F**ull **W**idth at **H**alf **M**aximum |
| **GP** | **G**aussian **P**rocess |
| **GXe** | **G**aseous **Xe**non |
| **ISO** | **Iso**tropic Emission *(but in $0\nu\beta\beta$ events, both primaries still have the same energy)* |
| **kNN** | **k**-**N**earest **N**eighbors |
| **LXe** | **L**iquid **Xe**non |
| **LZ** | **L**UX-**Z**EPLIN |
| **N.A.** | **N**atural **A**bundance |
| **OD** | **O**uter **D**etector |
| **P1A** | **P**erformance Convergence Comparative Test |
| **P1B** | Cost Curve Comparative **P**erformance Test |
| **P2** | ROC Curve Quantitative **P**erformance Test |
| **PDE** | **P**hoton **D**etection **E**fficiency |
| **pdf** | **P**robability **D**istribution **F**unction |
| **PMT** | **P**hoto**m**ul**t**iplier |
| **PTFE** | **P**oly**t**etra**f**luoro**e**thylene |
| **RBF** | **R**adial **B**asis **F**unction |
| **RF** | **R**andom **F**orests |
| **ROI** | Energy **R**egion **o**f **I**nterest |
| **ROC** | **R**eceiver **O**perating **C**haracteristic |
| **S1** | Primary **S**cintillation |
| **S2** | Secondary **S**cintillation |
| **SiPM** | **Si**licon **P**hoto**m**ultiplier |
| **SS** | **S**ingle **S**catter |
| **SVM** | **S**upport **V**ector **M**achine |
| **TPC** | **T**ime **P**rojection **C**hamber |
| **WIMP** | **W**eakly **I**nteracting **M**assive **P**article |
| **XFV** | E**x**tended **F**iducial **V**olume |

# Chapter 1

# Introduction

The current surplus of matter over antimatter, commonly referred to as the baryon asymmetry in the universe (BAU), remains a pressing issue in cosmology. It is only known that BAU was generated in the early universe, and that the generating mechanism would require extensions to the Standard Model (SM), several candidates having been suggested [1]. One attractive option, because it requires only minor SM extensions, is the generation of baryon asymmetry via the decay of ultraheavy sterile neutrinos in the early universe: these particles are predicted to decay more often into leptons than antileptons (*leptogenesis*), and the resulting asymmetry can be transferred into baryons (*baryogenesis*) in such a way that it becomes permanent. The see-saw mechanism in Majorana neutrino masses allows for the existence of sufficiently heavy sterile neutrinos, but is only applicable for the case that neutrinos and antineutrinos be the same particle [2] [3].

If neutrinos are in fact Majorana, then one interaction that becomes possible is neutrinoless double beta decay ($0\nu\beta\beta$). Observation of $0\nu\beta\beta$ would therefore be strong evidence for the sterile neutrino decay model of BAU generation. The search for $0\nu\beta\beta$ spans over three decades, with modern approaches emphasizing technologies allowing for large masses of sensitive volume and low background rates. Current half-life constraints exceed $10^{19}$ years for all allowed isotopes. Despite its main goal being weakly interacting massive particle (WIMP) search, the LUX-ZEPLIN (LZ) detector displays very promising characteristics for $0\nu\beta\beta$ search as well. With 623 kg of $^{136}$Xe it has the second-largest $0\nu\beta\beta$-active mass of all current-generation $0\nu\beta\beta$ experiments, and its background rate is projected to be competitive.

The majority of the projected $0\nu\beta\beta$ background in the LZ detector is caused by the same type of event: a photo-electric effect interaction of a gamma-ray photon (originating either from the surrounding rock or from the detector materials) with the active volume of the detector. The sensitivity of the LZ detector for $0\nu\beta\beta$ search may thus be improved significantly if this type of event were successfully discriminated. One possible approach is to discriminate between the topological signatures of the $0\nu\beta\beta$ decays and the single electron recoil interactions. In this work, a machine learning signal processing procedure was developed to test the viability of this approach, with the possibility of implementation into the LZ background analysis.

The principal contribution in this work is a parameterization, feature extraction and data optimization procedure that prepares pulse-like time-series (*waveform*) data for use on binary classification algorithms. It is modular and expandable, allowing for the easy inclusion and deletion of extracted parameters and features without side-effects, and includes an automatic feature selection and dimensionality reduction procedure to ensure that feature correlation does not negatively affect classifier training. In effect, multiple configurations of viable training data can be generated quickly, without the need for extensive fine-tuning of the extracted feature set.

Binary classification was performed using four algorithms: k-nearest neighbors; support vector

machines with radial basis function kernel; Gaussian process; and random forests. Waveform data generated by $0\nu\beta\beta$ and gamma-ray background event topologies in an LZ-like TPC was simulated using the ANTS2 simulation and data processing package [4], with Geant4 physics lists and NEST v2 data [5] used for topology generation. With the aid of distributed computing due to the Lisbon and Minho nodes of the Infraestrutura Nacional de Computação Distribuída [6], two binary classification waveform datasets were produced, roughly equally balanced between the two classes, with $\sim 10^4$ datapoints per dataset.

For each waveform dataset and classification algorithm, a set of different parameterization, feature extraction and data optimization approaches was tested. In total, 24 different configurations were produced for each of the datasets, with the goal of obtaining the optimal performance. A two-step performance comparison approach was developed for this optimization. It uses performance convergence and cost curves to quickly compare the performance of all configurations, with the added ability to extrapolate to different class balancings than the one used. In effect, the developed machine learning signal processing procedure allows for the quick preparation and comparison of a large set of binary classification algorithms and data processing methods, thus serving as a tool with applications not only in testing the viability of topology discrimination in LZ, but also with potential applicability for other viability assessment problems of a similar nature.

## 1.1 Outline of the dissertation

Chapter 2 describes the parameters of the baryon asymmetry problem. It first begins by presenting the available evidence for BAU, narrowing down the timeframe allowable for baryogenesis, and presenting the conditions necessary for it to take place. Then it shows that baryogenesis is not achievable within the SM, and then focuses on the sterile neutrino decay model.

Chapter 3 consists of a description of the physics of $0\nu\beta\beta$ decay with an emphasis on observables, followed by a review of the state of experimental $0\nu\beta\beta$ decay search. This chapter also describes the functioning principle of a time projection chamber (TPC), their performance when applied to $0\nu\beta\beta$ search and their background environment.

Chapter 4 details the characteristics of the LZ detector applied to $0\nu\beta\beta$ search, namely its construction parameters, background and how the obtained data is considered, highlighting aspects to consider during simulation and the possibility of using binary classification for a background cut.

Chapter 5 motivates the use of binary classification for background rejection and describes the selection method for the four chosen binary classifiers. It also describes the functioning principle and of each chosen classifiers, as well as detailing some of their applications, advantages and disadvantages. Special focus is given to the Gaussian process classifier, which proved to perform best during the tests. Classifier performance assessment tools are also introduced here.

Chapter 6 begins by describing and motivating the simulation approach taken in this work. It follows with a technical description of the performed simulations, the developed data processing framework, and the developed approach to configuration selection and performance assessment.

Chapter 7 presents the obtained performance results, as well as a discussion on the viability of binary classification for the given problem, on the adopted methodology and the applicability of the developed tools for other purposes.

Chapter 8 presents conclusions and discusses future work.

# Chapter 2

# Baryon Asymmetry in the Universe and Neutrinoless Double Beta Decay

The goal of this chapter is to describe how detecting neutrinoless double beta decay ($0\nu\beta\beta$) would be strong evidence for the most attractive model of *baryogenesis* **{fn2a}**: lepton number violation, termed *leptogenesis* (Chapter 16 of [8]). To do this, some background information must be provided first. The chapter is structured so as to frame the argument in the following way:

1. Show evidence against baryon symmetry, making BAU an attractive alternative;
2. Discuss the time-frame when baryon asymmetry must have originated;
3. Explain the necessary conditions for producing the asymmetry;
4. Show that it is not explained by the Standard Model;
5. Show how it is explained by leptogenesis, and its connection to $0\nu\beta\beta$ decay.

## 2.1 Evidence against baryon-symmetric matter-antimatter spatial distribution models

Antibaryon content can be measured using different observables, their reliability depending on the observed scale. Within the Solar System or its vicinity, antibaryon content can be gauged by the flux of heavy antibaryon particles (*e.g.* antihelium), or from interactions of matter with antimatter. Outside of the Solar System, indirect methods of measuring antibaryon content have to be employed, which rely on identifying the spectral signature of matter-antimatter annihilation: Doppler-broadened $\gamma$-ray and X-ray peaks around specific energies [9].

Evidence against baryon symmetry in the universe is extensive: the box in the next page presents different possible matter-antimatter spatial distribution models for a baryon-symmetric universe, as well as evidence against them, spanning scales from the neighborhood of the Solar System to the observable universe.

**{fn2a}**— The name used to refer to the dynamical generation of a baryon asymmetry in quantum field theory (Chapter 1 of [7]).

### Antimatter bodies within the Solar System

Evidently presolar nebulae cannot have a homogenous mix of matter and antimatter: for a presolar nebula of number density $n$, the collapse time is $\sim n^{-1/2}$, whereas its annihilation time in case it were also homogeneously baryon-symmetric would be only $\sim n^{-1}$ [10]. A presolar nebula would have annihilated away upon collapse, leaving insufficient material with which to construct a planetary system.

However it can be supposed that different presolar nebulae could be predominantly matter or antimatter, while baryon symnmetry in the universe is maintained. It would then be possible for the Solar System to capture an ejected antimatter body formed by a predominantly antimatter presolar nebula. The spectra of antimatter bodies in the Solar System would display an annihilation signature when interacting with baryon gas, namely with solar wind [10]. There are no currently known bodies of this type in the Solar System.

### Predominantly baryonic or antibaryonic patches in the Milky Way

The antibaryon content in the Milky Way can be measured with cosmic rays. The detection of a significant flux of antinucleids heavier than antiprotons would be a good indicator of the presence of antimatter in the galaxy. The Alpha Magnetic Spectrometer experiment determined that the flux ratio of antihelium to helium cosmic rays is only $F_{\overline{\text{He}}}/F_{\text{He}} < 1.1 \times 10^{-6}$ [11].

Furthermore, the theoretical limit of the fraction of antimatter in the interstellar medium (ISM) is below $10^{-15}$, due to the annihilation time in the ISM being only $\sim 300$ years [10]. In fact, the presence of large collections of antimatter in the form of stars and gas clouds would imply the existence of sources of annihilation signature within the ISM, which are seen to be absent [12].

### Predominantly baryonic or antibaryonic patches in galactic clusters

The ratio of baryons to antibaryons in the intergalactic medium (IGM) of galactic clusters can be measured by comparing the fluxes of bremsstrahlung X-rays and annihilation $\gamma$-rays. These two fluxes depend on the same observables except for one: the fraction of antimatter to matter (which the bremsstrahlung flux does not depend on) [10]. The X-ray and $\gamma$-ray fluxes were measured for 55 galaxy clusters by Edge *et al.* [13] and by the EGRET experiment [14], respectively. The upper limits of the fraction of antimatter for each cluster in the survey ranges between a minimum of $< 5 \times 10^{-9}$ for the Virgo cluster and a maximum of just $< 1 \times 10^{-6}$ for the Bullet cluster.

### Predominantly baryonic or antibaryonic patches in the universe

The previous points imply that the matter and antimatter should be organized in large discrete regions. By comparing the observed cosmic diffuse $\gamma$-ray background (CDG) spectrum with that of model spectra for discrete regions of different sizes, Cohen, De Rújula and Glashow demonstrate in [15] that the size of these discrete matter-antimatter regions would be $\gg 1$ Gpc **{fn2.1a}** . Due to the temperature uniformity of the cosmic background radiation (CBR), it can be shown that any existent baryonic and antibaryonic patches in the universe will have been adjacent to each other since before the recombination epoch, with no voids separating them. The borders between the two types of regions would then have produced a significant amount of annihilation $\gamma$-rays, which would contribute to the cosmic CDG, so that the CDG amplitude is inversely correlated with the size of the regions. Figure 2.1 shows that the observed amplitude is consistently at least 1 standard deviation below a conservative estimate of the amplitude for regions of size 1 Gpc, for all $\gamma$-ray energies above $\sim 4$ MeV.

**{fn2.1a}** — For comparison, the observable universe is $\sim 29$ Gpc [16]

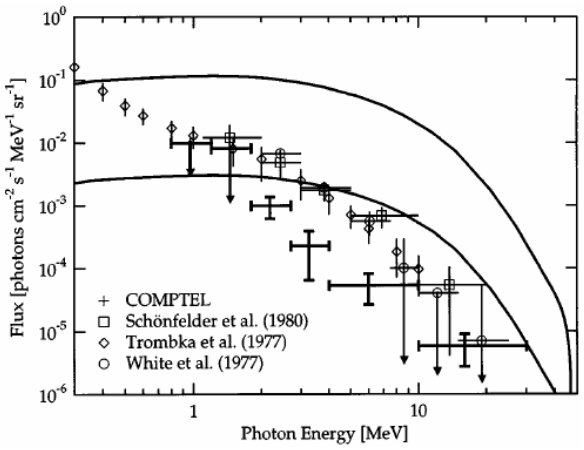


FIGURE 2.1: CDG spectrum data from COMPTEL, Fichtel et al. (1975), Mazets et al. (1975), Schönfelder et al. (1980), Trombka et al. (1977) and White et al. (1977), versus the expected spectra for the case of discrete matter-antimatter regions of sizes 0.02 Gpc (continuous curve above) and 1000 Mpc (continuous curve below). It can be seen that the observed flux is consistently at least one error bar below a conservative estimate of the amplitude for regions of size 1 Gpc, for all energies above about 4 MeV.
Source: [15]

The data listed in the preceding box presents comprehensive evidence against baryon symmetry in the universe. The antibaryon content in the universe was shown to be almost negligible by measures across all observable scales, making BAU an attractive alternative.

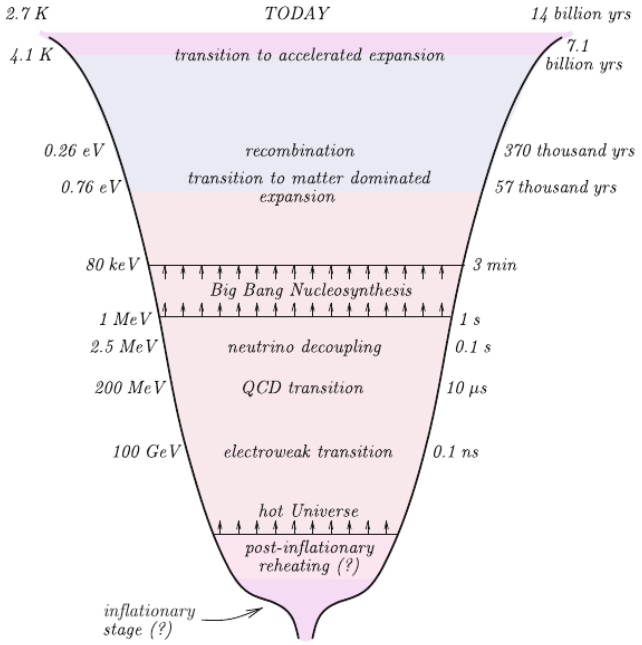


FIGURE 2.2: Stages of universe evolution starting from inflation, with relation between temperature $T$ (left) and age of the universe (right).
Source: [7]

## 2.2 Baryon asymmetry in the early universe

Beyond effectively proving that BAU exists, the argument in [15] also implies that the asymmetry already existed prior to the recombination epoch (Figure 2.2). This section will now examine the baryon content prior to the recombination epoch, spanning until the quark-hadron phase transition, or quantum chromodynamic (QCD) phase transition , when the quarks and gluons clustered to form color-singlet hadrons, namely the first protons and neutrons.

The following discussion will have three reasonable assumptions:

**Assumption 1.** For practical purposes, during the QCD transition, baryons can already be considererd nonrelativistic particles.

*Relativistic effects are conventionally negligible for speeds $v \lesssim 0.1c$, corresponding to energy $\lesssim$ 1 GeV, while the QCD transition occurs when the temperature of the universe, $T$, is around $T \sim 0.2$ GeV (see Figure 2.2).*

**Assumption 2.** At sufficiently large length scales $\mathcal{O}(0.1\,\text{Gpc})$ [17], the observable universe is spatially homogeneous and isotropic.

*Evidenced by the uniformity of the temperature of the cosmic microwave background (CMB) to the $\sim 10^{-5}$ level ( [18] and Section 1.3 of [19]).*

**Assumption 3.** At some point post-inflation and before the QCD transition, the universe had symmetric baryon-antibaryon content.

*Even supposing that at the Big Bang the universe had already been baryon-asymmetric, this asymmetry would have not survived inflation (Section 1.2 of [7]), so the BAU today is not due to initial conditions of the universe. Some process has to have caused baryogenesis in a baryon-symmetric universe.*

The evolution of the abundance of some arbitrary particle species $\psi$, $Y_\psi \equiv n_\psi/s$ (where $n_\psi$ is its number density and $s$ is the entropy density **{fn2.2a}**), is strongly linked to its *mass* $m_\psi$ and to the *temperature of the universe*. $T$ is related to the size of the universe, which expands at a rate given by the *Hubble parameter* $H \equiv \dot{R}/R$, where $R$ is the *cosmic scale factor* **{fn2.2b}**. These factors in turn affect the interaction rates $\Gamma_\psi \equiv n \langle \sigma |v| \rangle$ of the species (where $n$ is the number density of target particles for the case of a stable particle species, like the proton; $\sigma$ is some cross section; and $|v|$ is the relative velocity).

For the early universe, its small size and high temperature meant that $\Gamma_\psi \gtrsim H$: interactions occured faster than the particles moved away from each other, and hence particle species were in thermal equilibrium. For the case of fermions, their abundances therefore followed Fermi-Dirac statistics. However, as the temperature decreased and the universe expanded, for certain species the expression above flipped, $\Gamma_\psi \lesssim H$, and the species fell out of thermal equilibrium: a process termed *decoupling* or *freeze out*. The interactions freeze out and the abundance is said to *freeze in* to a certain value. Figure 2.3 shows the effect of decoupling on the relative abundance

**{fn2.2a}** — The entropy density is defined as $s \equiv S/V$, with entropy per comoving volume $S = R^3(\rho + p)/T$ (where $\rho$ and $p$ are the thermal equilibrium energy density and pressure) being conserved in thermal equilibrium. For the early universe, the entropy density can be approximated as $s \approx (2\pi^2/45) \times g_* T^3 = 1.80 \times g_* n_\gamma$, where $g_*$ is the total number of effectively massless degrees of freedom, $n_\gamma$ is the photon density, and the second equality is due to the fact that relativistic particles in thermal equilibrium have $n \sim T^3$. The current entropy density is $s = 7.04 \times n_\gamma$ (for more information, see Section 3.5 of [19]).

**{fn2.2b}** — A dimensionless factor representing the size of the universe relative to its present size: $R(t) \in [0, 1]$, such that $R(t = \textit{current age of universe}) = 1$.

of a massive particle, for decreasing temperature. The current baryon abundance is in the order of $10^{-10}$ **{fn2.2c}**.

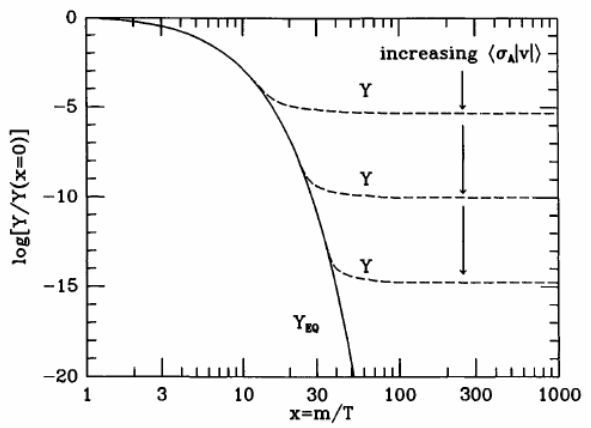


FIGURE 2.3: The effect of decoupling on the relative abundance of a massive, stable particle, for decreasing temperature, and hence increasing time elapsed since the Big Bang. $Y_{\mathrm{EQ}}$ is the abundance of the species had it always remained in thermal equilibrium. $\sigma_A$ is the annihilation cross-section. We can see that the abundance appears to "freeze in" to a certain value once the species decouples. Without decoupling, it would continue to decrease.
Source: Section 5.2, pg. 126 of [19]

Once the species decouples, its abundance no longer behaves according to Fermi-Dirac statistics, and instead begins conforming to the Boltzmann transport equation (BTE): an equation that describes the phase space occupation of a thermodynamic system out of equilibrium. Supposing that $\psi$ is a stable massive nonrelativistic particle (**Assumption 1.**) in an isotropic and homogeneous universe (**Assumption 2.**), and supposing that at some point in the very early universe there existed no asymmetry between particles $\psi$ and antiparticles $\bar{\psi}$ of the species (*so* $Y_\psi = Y_{\bar{\psi}}$) (**Assumption 3.**), then the BTE can be expressed in an intuitive manner (see Section 5.1 of [19] for a more detailed explanation):

$$\frac{dn_\psi}{dt} = -3Hn_\psi - \langle\sigma|v|\rangle\,[n_\psi^2 - (n_\psi^{\mathrm{Eq.}})^2], \tag{2.1}$$

where $n_\psi^{\mathrm{Eq.}} \propto T^{2/3} e^{-m_\psi/T}$ is the number density of the species if it remained in thermal equilibrium, and thus preserved Fermi-Dirac statistics. Here we can see that $-3Hn_\psi$ accounts for the dilution caused by the expansion of the universe. The positive term $-\langle\sigma|v|\rangle\,[n_\psi^2 - (n_\psi^{\mathrm{Eq.}})^2]$ accounts for interactions that change the number of existing particles of the species: its simple shape arises from the assumption that there exists no asymmetry between the particles and antiparticles of the species. Refactoring (Eq. 2.1) in terms of the abundance gives:

$$\frac{x_\psi}{Y_\psi^{\mathrm{Eq.}}}\frac{dY_\psi}{dx_\psi} = -\frac{\Gamma^{\mathrm{Eq.}}}{H}\left[\left(\frac{Y_\psi}{Y_\psi^{\mathrm{Eq.}}}\right)^2 - 1\right], \tag{2.2}$$

**{fn2.2c}** — The baryon number abundance is given as $B \equiv (n_b - n_{\bar{b}})/s = 3.81 \times 10^{-9}(\Omega_B h^2)$, where $\Omega_B h^2 \approx 0.0223^{+0.0007}_{-0.0009}$ is the physical baryon density, resulting in $B = (0.48 - 0.98) \times 10^{-10}$( [20] and Section 3.5, pg. 81 of [19])

where $x_\psi \equiv m_\psi/T$ exceeds 3 for the nonrelativistic case; $\Gamma^{\text{Eq.}} = n^{\text{Eq.}} \langle \sigma_A |v| \rangle$ is what the interaction rate would have been had the species remained in thermal equilibrium; $\sigma_A$ is the annihilation cross-section; and $Y_\psi^{\text{Eq.}} \equiv n^{\text{Eq.}}/s$. Solving (Eq. 2.2) gives the abundance of the species today, $Y_\psi^{\text{Now}}$. which **{fn2.2d}** is seen to behave roughly as:

$$Y_\psi^{\text{Now}} \propto \frac{1}{T_{\text{F}} \langle \sigma_A |v| \rangle},$$

where $T_{\text{F}}$ is the freeze-out temperature. In this model, still maintaining the assumption that $Y_\psi = Y_{\bar{\psi}}$, $T_{\text{F}}$ is seen to only depend on $\langle \sigma_{\text{A}} |v| \rangle$, being almost directly proportional to it (Section 5.2, pg.125 of [19]). This way, by specifying $\langle \sigma_{\text{A}} |v| \rangle$, both the freeze-out temperature and the current abundance can be obtained.

This can be applied to baryons by inputting the appropriate interactions. By factoring into $\langle \sigma_{\text{A}} |v| \rangle$ only the most common nucleon-antinucleon annihilation modes (those being the annihilation of the nucleon-antinucleon pair into a group of mesons containing a number of pions of order unity [21]), Section 5.2, pg. 127 of [19] gives a $Y_\psi^{\text{Now}}$ value in the order of $10^{-20}$, for a $T_{\text{F}}$ of 22 MeV. This optimistic $Y_\psi^{\text{Now}}$ value is $\sim$ 9 - 10 orders of magnitude smaller than the observed current baryon abundance in the universe. Additionally, in Section 6.2, pg. 159 of [19] it is determined that forcing $Y_\psi^{\text{Now}}$ to be the current baryon abundance returns a $T_{\text{F}}$ value of $\sim$ 40 MeV.

If the baryon content of the universe had been symmetric after $T \sim 40$ MeV, then the matter and antimatter would have annihilated to the point that the baryon abundance today would have been much smaller than it actually is, as evidenced by the argument above.

## 2.3 Sakharov Conditions for Baryogenesis

Some process had to transition the universe from baryon symmetry post-inflation to baryon asymmetry at $T \gtrsim 40$ MeV, less than a tenth of a second after the Big Bang. This section describes the three conditions that this process would have to meet to produce baryogenesis, as originally proposed by Andrei Sakharov [22].

Intuitively, there cannot be the same number of baryons and antibaryons at the start and at the end of the process: the baryon number has to be violated. It also has to favor the production of particles over antiparticles. Baryogenesis will therefore require *C violation*: a violation of charge conjugation symmetry. However, this condition is not sufficient, as it does not account for quantum field theory (QFT) effects **{fn2.3a}**, which further impose that a simultaneous violation of symmetry toward charge conjugation and parity transformation, *CP violation*, must occur in tandem with the C violation.

**{fn2.2d}** Aside from some constants, and a factor distinguishing s-wave from p-wave annihilations (Section 5.2, pg.124 of [19]).

**{fn2.3a}** — Massive fermions, like quarks or neutrinos, do not exhibit chiral symmetry (see [2]), so it could occur, for example, that the production rate of left-handed quarks will be different from the production rate of right-handed quarks. Under CP conservation, the production rate of left-handed quarks will be equal to the production of rate of right-handed *antiquarks*, and that of right-handed quarks will equal that of left-handed antiquarks. Under CP conservation, therefore, the rates of production of quarks and antiquarks of both chiralities would still remain the same, even though C-symmetry had been violated (the antiparticle of a left-handed quark is a left-handed antiquark, and their rates of production were different in the above example [23]).

Finally, if a stable species (namely protons **{fn2.3b}** and neutrons) is in thermal equilibrium, then the annihilation rate for the particles will be equal to the rate at which they are produced. Therefore, as long as thermal equilibrium is maintained, no BAU can be generated.

With this, the three *Sakharov conditions* can now be listed:

**Condition 1.** Baryon number violation;

**Condition 2.** C and CP violation;

**Condition 3.** Departure from thermal equilibrium.

These conditions are sufficient, as ensured by the CPT theorem **{fn2.3c}**.

## 2.4 Baryogenesis within the Standard Model

To achieve baryogenesis in the Standard Model (SM), it is necessary to find some process within it that would violate the three Sakharov conditions. In perturbative processes — processes that can be represented by perturbative methods, these being the kind of processes observed below temperatures of the quark epoch ($T \gtrsim 150\,\text{MeV}$) — the baryon number is always conserved. However, the temperatures beyond $T \gtrsim 100\,\text{GeV}$ of the electroweak epoch — when the electromagnetic and weak interactions were still unified — allowed for non-perturbative processes — processes that cannot be represented by perturbative methods — whose conserved quantity was $B - L$ (where $B$ and $L$ are the baryon and lepton number) instead of $B$ and $L$ individually [1].

This is due to the structure of the vacuum in the electroweak gauge theory. In this theory, the vacuum gauge field configurations are degenerate, with different configurations resulting in the same zero field strength and energy, but with varying values of the so-called Chern-Simons number, $N_{CS}$, a topological quantum number involving the winding of the weak isospin [25]. The vacuum degeneracy is pictured in Figure 2.4, where we can see that between each vacuum — with zero energy and integer $N_{CS}$ — there exists an energy barrier, whose value has been minimized to $E_{\text{sph}}$ — also called the *sphaleron barrier* — corresponding to non-vacuum field configurations. A transition from one vacuum configuration to another would have a nonzero probability via a tunneling process, called an *instanton process*, and an integer difference between the initial state and the final state in $N_{CS}$ would violate $B$ and $L$ by 3 [23].

In the electroweak gauge theory itself, the tunnelling rate between different vacuum configurations is high enough to serve as a thermal equilibrium: the resulting baryon number violation from some change in configuration is quickly cancelled out by another change in configuration, and cannot be preserved. Meanwhile, once the electroweak symmetry breaks, the tunnelling rate in the resulting gauge group is greatly suppressed, and so the baryon number is "frozen" in place: the first Sakharov condition is not met in that broken symmetry vacuum.

**{fn2.3b}** — [22] proposes the possibility that protons decay with a large half-life ($\gtrsim 10^{50}$ yrs.). The current most sensitive results (Super-Kamiokande) place a lower bound on the proton half-life at $> 1.67 \times 10^{34}$ yrs. [24]. This work, however, supposes that protons are stable.

**{fn2.3c}** — Any self-consistent Lorentz invariant quantum field theory is symmetric under simultaneous charge conjugation, parity transformation and time reversal [2].

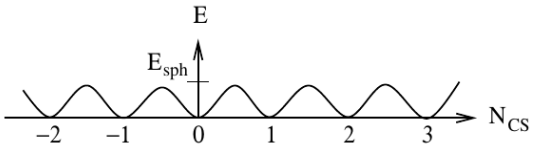


FIGURE 2.4: Gauge field configurations of the electroweak theory, as they relate to the Chern-Simons number, $N_{CS}$. Vacuum gauge field configurations correspond to integer $N_{CS}$.
Source: [23]

However, if the electroweak phase transition occurs discontinuously (i.e. it is a 1st order phase transition), then there will exist bubbles of gauge field where the symmetry has already been broken, submerged in an electroweak gauge field, with discrete domain walls separating the bubbles from their surroundings (see Figure 2.5). This environment would serve to escape from thermal equilibrium, as the rate of the $B$-violating process would slow down faster than it would be able to compensate for excess baryons. We thus see that the electroweak phase transition obeys the **1st** and **3rd** Sakharov conditions. Finally, the **2nd** condition — C and CP violation — can be met by having CP-violating interactions occuring at or near the domain walls [25]. CP violating processes are known to occur: they have first been experimentally observed in 1964, by way of $K_0 - \bar{K}_0$ mixing, in the Fitch-Cronin experiment [26].

Unfortunately, in order for the electroweak phase transition to be discontinuous, the Higgs mass must be below about 70 GeV [27]It is seen instead to be 125.38 $\pm$ 0.14 GeV [28]. Under these conditions, the phase transition is smooth, and hence the Sakharov conditions can no longer be met, thus ruling out the electroweak phase transition as an avenue for explaining the BAU. Explanations for the baryon asymmetry thus have to be found beyond the SM.

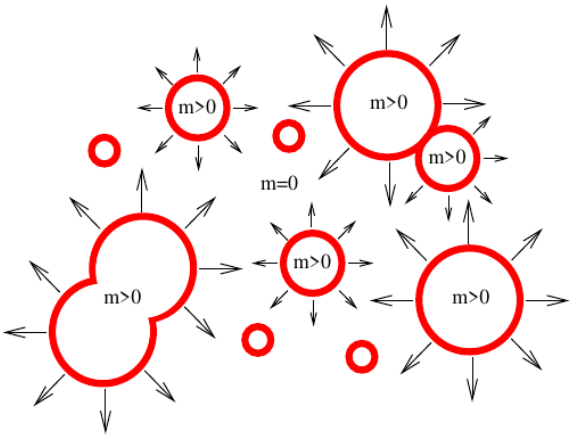


FIGURE 2.5: Representation of the case of a discontinuous electroweak phase transition. Bubbles of broken symmetry (with the EM and weak interactions already separated, and the Higgs particle mass $m$ nonzero) are surrounded by the electroweak gauge field, separated by discrete domain walls. The rate of baryogenesis is zero in the bubbles, and nonzero outside of them. The occurence of CP-violating processes at the domain walls would meet all three Sakharov conditions, thus making this process a candidate for baryogenesis.
Source: Section 5, pg. 83 of [29]

## 2.5 Baryogenesis through Majorana neutrinos

There are several proposed mechanisms for baryogenesis arising from extensions to the SM **{fn2.5a}**, a very attractive one being baryogenesis through leptogenesis caused by the decay of extremely heavy neutrino species before electroweak symmetry breaking [30]. The only two extensions to the SM required by this mechanism (aside from nonzero neutrino masses) are for neutrinos to be *Majorana particles* (i.e that neutrinos be their own antiparticles) and for *sterile neutrinos* (i.e. those that do not couple to the weak force) resulting from this extension to have a mass in the $10^9 - 10^{14}$ GeV range, as motivated by Grand Unified Theories (GUTs) [30]. The details of this mechanism for baryogenesis are described below.

Neutrinos emerge in QFT as quanta of a Dirac field: a solution to the Dirac equation [2]. In this field, a massive particle is represented as a quantum of a 4-component spinor. If this particle is charged, then its spinor has to be complex so as to not violate charge conservation: it has to be a so-called *Dirac spinor* [2]. The particle is then termed a *Dirac particle*. For a Dirac particle, there exists a corresponding antiparticle with opposite charge. However if the massive particle is neutral, then charge conservation cannot constrain the spinor, so that it can either be complex or real. A real spinor of the Dirac equation is called a *Majorana spinor* and then its quanta are named accordingly [2]. Neutrinos are neutral particles, so if we suppose that they are massive they can be of either type.

Any Dirac field spinor can be separated into left- and right-chiral projections, $\Psi = \Psi_L + \Psi_R$, and hence will be composed out of some combination of two irreducible field spinors, called *Weyl spinors*, one for each chirality. The right-chiral Weyl spinor will be a doublet under the $SU(2)_R$ group and a singlet under $SU(2)_L$, and vice-versa. Weak interactions in the SM are $SU(2)_L$, and hence only left-chiral neutrinos $\nu_L$ and their CP-conjugates, the right-chiral antineutrinos $(\nu_L)^c \equiv \bar{\nu}_R$ appear in nature today. A Dirac spinor will include both Weyl spinors, while a Majorana spinor only includes either the left-chiral Weyl spinor, or the right-chiral one. This means that there are three distinct ways to append a neutrino mass term to the SM Lagrangian: one for the Dirac spinor; and two for both opposite chirality Majorana spinors. While Dirac neutrinos cannot obtain contributions from the two Majorana mass terms, nothing prevents Majorana neutrinos from obtaining contributions from both the Majorana mass terms as well as the Dirac mass term [3].

For a single neutrino generation, the Majorana mass can therefore appear as shown below:

$$-\mathcal{L}^{\nu}_{\text{mass}} = \overline{\left(\boldsymbol{\nu}_L^{\text{Maj.}}\right)^c}\,\boldsymbol{\mathcal{M}}^{\text{Maj.}}\boldsymbol{\nu}_L^{\text{Maj.}} \quad , \quad \boldsymbol{\mathcal{M}}^{\text{Maj.}} = \begin{bmatrix} m_L & m_{\text{D}} \\ m_{\text{D}} & m_R \end{bmatrix}, \tag{2.3}$$

where $(\boldsymbol{\nu}_L^{\text{Maj.}})^T = \begin{bmatrix} \nu_L & (\nu_R)^c \end{bmatrix}$, with its elements being the two possible left-chiral neutrino states; $m_{\text{D}}$ comes from the Dirac mass term; and $m_L$ and $m_R$ come from the two opposite chirality Majorana mass terms. One can now predict that $m_L \ll m_{\text{D}} \ll m_R$ [31], with $m_{\text{D}}$ of magnitude comparable to other fermion masses. The mass contributions can then be diagonalized to give two eigenvalues: a very low mass $m_\nu \approx m_{\text{D}}^2/m_R$, corresponding to a left-chiral active neutrino $\nu$; and a very high mass $M_N \approx m_R$, corresponding to a right-chiral, sterile neutrino $N$. This is known as the *see-saw mechanism* [3]. This argument can be easily extended to include three neutrino generations, thus returning three small mass eigenvalues,

**{fn2.5a}** — Most of them are beyond the scope of this thesis. For more information on baryogenesis mechanisms other than neutrino leptogenesis, see [1].

corresponding to the active neutrino mass eigenstates $\nu_1, \nu_2, \nu_3$ **{fn2.5b}**; and three very large mass eigenvalues, corresponding to three sterile neutrinos $N_1, N_2, N_3$ [3].

The sterile neutrinos have several different decay modes [32]. Among them, there are decays into a lepton and a Higgs boson, $N \to LH, N \to \bar{L}\bar{H}$ **{fn2.5c}**, where the interference between tree-level and loop-level processes (Figure 2.6) will result in CP violation (the **2nd Sakharov condition**) and a nonzero $L$. The high sterile neutrino masses imposed by GUTs guarantee that the decay already happens out of thermal equilibrium (the **3rd Sakharov condition**) when $T \sim M_1$, meaning that the lepton asymmetry is preserved to a sufficient degree. Finally, like in Section 2.4, this temperature is high enough for non-perturbative processes to take place, which by $B - L$ conservation will then transfer the asymmetry from leptons into quarks (the **1st Sakharov condition**) [32]. Thus leptogenesis meets the conditions for baryogenesis. The baryon asymmetry will then freeze to its current value at $T \sim 40$ MeV, and will later be transferred into nuclides during the Big Bang nucleosynthesis epoch (see Figure 2.2), resulting in the BAU seen today.

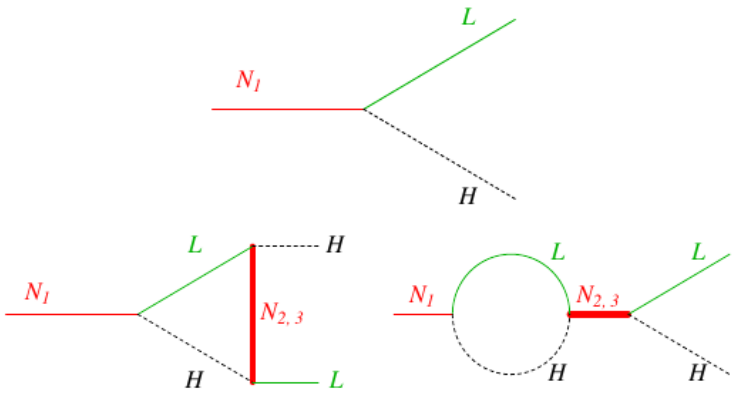


FIGURE 2.6: Diagrams of the $N_1 \to LH, N_1 \to \bar{L}\bar{H}$ processes, $N_1$ being the lightest sterile neutrino, and $N_{2,3}$ being one of the other two sterile neutrinos. The top diagram is tree-level, coupling $N_1$ to the $L$ and $H$ fields with a strength $\lambda_1$; and the rest are loop-level, the $N_{2,3}$ coupling as $\lambda_{2,3}$. The decay width of the $N_1 \to LH$ process will have the form $\Gamma(N_1 \to LH) \propto |\lambda_1 + A\lambda_1^*\lambda_{23}^2|^2$, with $A$ a complex CP-conserving loop factor, while $\Gamma(N_1 \to \bar{L}\bar{H}) \propto |\lambda_1^* + A\lambda_1\lambda_{23}^{2*}|^2$, so $\Gamma(N_1 \to LH) \neq \Gamma(N_1 \to \bar{L}\bar{H})$.
Source: [32]

Again, however, this mechanism requires that neutrinos be Majorana particles. It is not yet known if they are. The next Chapter discusses the physics and detection of neutrinoless double beta decay ($0\nu\beta\beta$) , a process that, if seen to occur, would prove it.

**{fn2.5b}** — Not to be confused with the neutrino flavor eigenstates, $\nu_e, \nu_\mu, \nu_\tau$. The mixing of the mass eigenstates results in the neutrino flavor oscillations.

**{fn2.5c}** — The second decay in reality is $\bar{N} \to \bar{L}\bar{H}$, but the case of a Majorana neutrino allows for this simplification.

# Chapter 3

# $0\nu\beta\beta$ Decay Morphology and Detection Technology

This chapter summarizes the characteristics of $0\nu\beta\beta$ decay relevant to this work and the state of $0\nu\beta\beta$ search. Section 3.1 discusses the physics of the decay, with focus on geometry and time characteristics, whereas Section 3.2 gives an overview of the basic detection approach, technologies, challenges and current state of the discipline.

## 3.1 Physics and morphology of double beta decay

Double beta decay is a rare process during which a nuclide with atomic number $Z$ and mass number $A$ decays into a nuclide with atomic number $Z+2$ and equal mass number by way of two simultaneous $\beta^-$ decays. It only occurs if simple beta decay is not energetically possible, when the nuclear binding energies $B$ have the following configuration:

$$B(Z+2,A) < B(Z,A) < B(Z+1,A), \tag{3.1}$$

as exemplified in Figure 3.1 for $A = 136$, $^{136}$Xe being the nuclide of interest there.

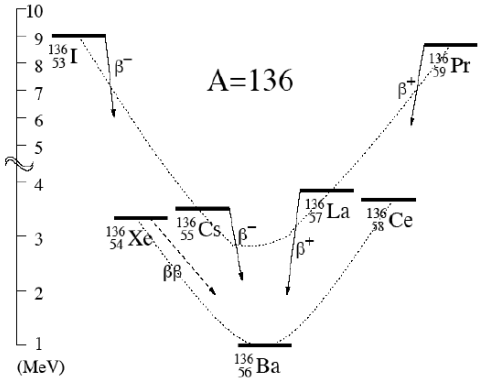


FIGURE 3.1: Nuclear binding energies for nuclei with $A = 136$.
Source: [33]

Each of the $\beta^-$ decays releases an electron and an electron antineutrino, so that *ordinary double beta decay* ($2\nu\beta\beta$) can be described as:

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

This type of decay is allowed in the SM and has already been directly observed in nature for different nuclides by several experiments since 1987 (see [34]).

If neutrinos are Majorana particles, $\nu = \bar{\nu}$, then there will also be a nonzero probability for *neutrinoless double beta decay* ($0\nu\beta\beta$) :

$$(Z, A) \rightarrow (Z + 2, A) + 2e^{-}.$$

This process has not been observed yet. Note that unlike ordinary double beta decay, the neutrinoless variant requires extensions to the SM, as the lepton number is not conserved: $\Delta L = 2$.

The standard mechanism for $0\nu\beta\beta$ decay involves the exchange of a light Majorana neutrino between two virtual $W$ bosons emitted by parent nucleons [3] (Figure. 3.2). The same mechanism allows for three other $L$-violating processes: the double positron emission process ($(Z, A) \rightarrow (Z - 2, A) + 2e^{+}$); as well as single positron emission plus single electron capture ($e^{-} + (Z, A) \rightarrow (Z - 2, A) + e^{+}$); and double electron capture ($2e^{-} + (Z, A) \rightarrow (Z - 2, A)$). Additionally, there are possible non-standard mechanisms, via exchange of particles other than light neutrinos (namely a sterile neutrino or a heavy supersymmetric particle, like a neutralino or gluino) or with the emission of one or two [35] Majorons **{fn3.1a}**. The mechanisms and processes referred in this paragraph are outside the scope of this work, however. For more information, see [3] and [36].

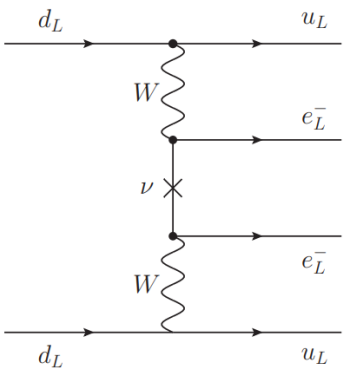


FIGURE 3.2: $0\nu\beta\beta$ decay mechanism via a Majorana neutrino exchange.
Source: [3]

The two identifying measurable features of a double beta decay are the energies and paths taken by the two released electrons. For a given nuclide, both $2\nu\beta\beta$ and $0\nu\beta\beta$ release the same amount of energy: the Q-value $Q_{\beta\beta}$, which is the difference between the binding energies of the initial and the final nuclide.

$$Q_{\beta\beta} \equiv B(Z, A) - B(Z + 2, A)$$

For the case of $2\nu\beta\beta$, this energy is distributed over two electrons and two neutrinos, while for $0\nu\beta\beta$ it is distributed over only the two electrons, as the recoil energy of the nucleus is negligible [3]. In the rest frame, the sum of the energy of the two $0\nu\beta\beta$ electrons will be $Q_{\beta\beta}$.

**{fn3.1a}** — Majorons are spinless, light or massless bosons that, depending on the model, may or may not be Goldstone bosons associated with spontaneous lepton number symmetry breaking, and may or may not carry a lepton charge. For more information, see [35].

Figure 3.3 displays the distributions for the energies of one of the electrons and of both electrons together for $2\nu\beta\beta$ and $0\nu\beta\beta$, according to the model used by the `DECAY0` event generator [37] [38], employed for simulating initial electron kinematics by the LZ collaboration for sensitivity projection [39]. The top panel shows that the two $0\nu\beta\beta$ electrons will most likely have the same energy, $Q_{\beta\beta}/2$. In that case, to preserve linear momentum of the system, they will escape the nucleus in opposite directions, *back-to-back* (b2b). If the energies of the two electrons are different, then the angle formed by the two escape directions for either type of decay will follow a distribution that is approximated in `DECAY0` as [37]:

$$F(\cos\theta) = 1 - \beta_1\beta_2\cos\theta,$$

where $\theta$ is the angle between initial electron directions and $\beta_i$ is the velocity of electron $i$. In 2012, Kotila and Iachello provided a theoretical model using exact Dirac wave functions with finite nuclear size and electron screening [40], with which the semiempirical `DECAY0` model is in good agreement [41].

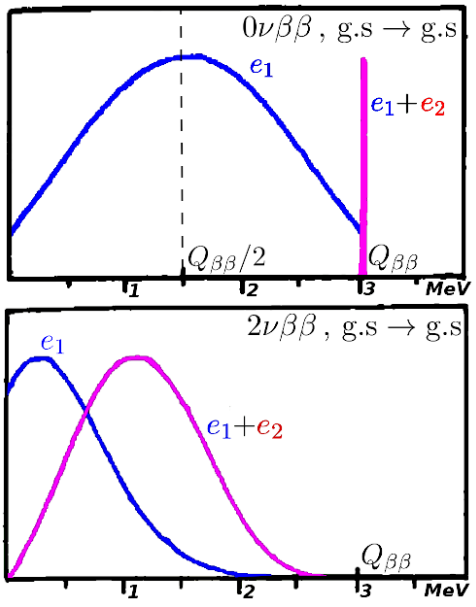


FIGURE 3.3: Theoretical energy distributions for a single electron *(blue)* and for both electrons together *(magenta)* in the case of $0\nu\beta\beta$ *(top)* and $2\nu\beta\beta$ *(bottom)*, for the double beta decay from a ground-state $^{100}$Mo nucleus to a ground-state $^{100}$Ru nucleus ($Q_{\beta\beta} = 3.034$ MeV). The graphs chosen are for the most common mechanism of the two distinct decay modes. The amplitude on the two graphs is in arbitrary units.
Source: [37]

After escape, the $0\nu\beta\beta$ *(primary)* electrons interact with the surrounding medium, losing energy to ionization and excitation of its atoms. As the primary electron energy decreases, the rate of energy loss $dE/dx$ rapidly increases, resulting in a *topology* where a track ends at two *blobs* on either side (Figure 3.4). The trajectory of the electrons is altered by multiple scattering on the nuclei of the medium. Additionally, the track can suffer branching due to $\delta$-electrons, or become disjoint due to bremsstrahlung, although these occurences are relatively rare.

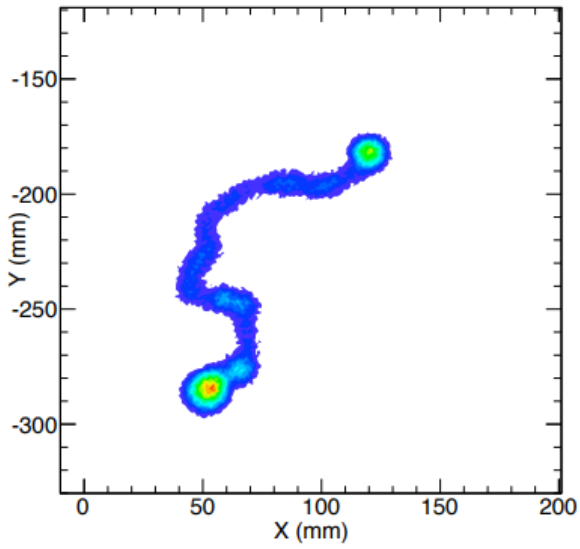


FIGURE 3.4: $0\nu\beta\beta$ track simulation in 10 bar xenon gas, distinctly showing the two blobs on either end.
Source: Section 6.2.1 of [42].

The half-lives of ordinary double beta decay, $T_{1/2}^{2\nu\beta\beta}$, and of its light neutrino-exchanging neutrinoless variant, $T_{1/2}^{0\nu\beta\beta}$, can be factorized as [40] [43] [3]:

$$T_{1/2}^{2\nu\beta\beta} = \left(G^{2\nu}\left|M^{2\nu}\right|^2\right)^{-1} \quad , \quad T_{1/2}^{0\nu\beta\beta} = \left(G^{0\nu}\left|M^{0\nu}\right|^2 \langle m_{\beta\beta}\rangle^2\right)^{-1}, \tag{3.2}$$

where $G^{2\nu} \propto (Q_{\beta\beta})^{11}$ and $G^{0\nu} \propto (Q_{\beta\beta})^5$ [44] [3] are phase-space factors; $M^{2\nu} \sim 0.06$ [34] and $M^{0\nu} \sim 4$ [3] are nuclear matrix elements; and $\langle m_{\beta\beta}\rangle$ is the effective Majorana neutrino mass. The $T_{1/2}^{2\nu\beta\beta}$ values and measured $T_{1/2}^{0\nu\beta\beta}$ lower bounds for 11 isotopes of experimental interest are listed in Table 3.1. The shortest one is for $^{100}$Mo, with $T_{1/2}^{2\nu\beta\beta} = (7.11 \pm 0.56) \times 10^{18}$ years [45].

The effective Majorana neutrino mass is given by:

$$\langle m_{\beta\beta}\rangle = \left|\sum_{i=1}^{3} U_{ei}^2 m_i\right|,$$

where $m_1, m_2$ and $m_3$ are the eigenvalues of the three light neutrino mass eigenstates, and $U_{e1}, U_{e2}$ and $U_{e3}$ are components of the neutrino mixing matrix:

$$\begin{bmatrix}\nu_e\\ \nu_\mu\\ \nu_\tau\end{bmatrix} = \begin{bmatrix}U_{e1} & U_{e2} & U_{e3}\\ U_{\mu1} & U_{\mu2} & U_{\mu3}\\ U_{\tau1} & U_{\tau2} & U_{\tau3}\end{bmatrix}\begin{bmatrix}\nu_1\\ \nu_2\\ \nu_3\end{bmatrix}.$$

The neutrino mixing matrix can be parameterized in terms of three angles $\theta_{12}, \theta_{23}, \theta_{13}$, a Dirac phase $\delta_{CP}$ and two Majorana phases $\alpha_1, \alpha_2$. The angles and Dirac phase have been estimated using solar, atmospheric, reactor and accelerator neutrino experiments [46], but the Majorana phases are uncoupled from these phenomena. With the detection of $0\nu\beta\beta$, one of the two Majorana phases can be constrained [47].

The neutrino mass eigenvalues $m_1, m_2$ and $m_3$ are not known completely. Two mass-squared differences $\Delta m^2_{ij} = m^2_i - m^2_j$ have been estimated using solar and atmospheric experiments: $\Delta m^2_{sol} \equiv \Delta m^2_{21}$, and $\Delta m^2_{atm} \gg \Delta m^2_{sol}$ is approximately the mass splitting between $m_3$ and whichever eigenvalue is closest. From the study of the propagation of neutrinos through solar matter, it is also known that $m_1 < m_2$ [48]. Two pieces of information are missing. First, it is not known whether the ordering of the masses is $m_1 < m_2 < m_3$ (normal ordering) or $m_3 < m_1 < m_2$ (inverted ordering). Second, the lightest neutrino mass is still unknown. Together with neutrino oscillation experiments and cosmological probes, the calculation of $\langle m_{\beta\beta} \rangle^2$ via the measurement of $T^{0\nu\beta\beta}_{1/2}$ would help to constrain the mass ordering and the value of the lightest neutrino mass [46]. Figure 3.5 shows the $3\sigma$ allowed bands for the effective Majorana neutrino mass against the lightest neutrino mass eigenvalue ($m_1$ in the normal ordering, $m_3$ in the inverted ordering).

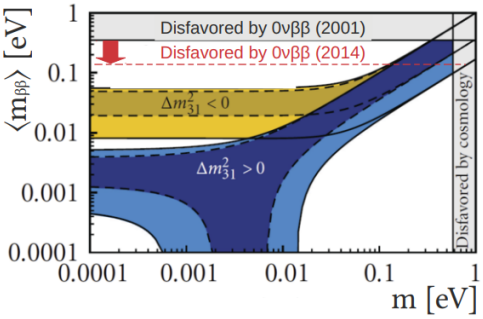


FIGURE 3.5: $3\sigma$ allowed bands for the effective Majorana neutrino mass against the lightest neutrino mass eigenvalue, for the inverted ordering (yellow) and the normal ordering (blue). The bands are mapped out according to constraints given by neutrino oscillation parameters and cosmological oscillations. Upper limits on the effective mass are given by $0\nu\beta\beta$ experiments until 2014. Source: [49].

## 3.2 $0\nu\beta\beta$ search

This Section describes $0\nu\beta\beta$ detection methods, background sources and the current state of the discipline. Section 3.2.1 describes the challenges of $0\nu\beta\beta$ search, Section 3.2.2 discusses detection methods and technologies, Section 3.2.3 covers background sources and reduction methods, and Section 3.2.4 describes the current generation of $0\nu\beta\beta$ detectors.

### 3.2.1 Challenges for $0\nu\beta\beta$ search

The challenges of $0\nu\beta\beta$ search are its rarity in nature and its similarity to the more common $2\nu\beta\beta$. The binding energy configuration referred in Equation 3.1 is only known to occur in 35 isotopes [43], and due to the $(Q_{\beta\beta})^5$ dependency of the neutrinoless phase space factor $G^{0\nu}$, only the 11 isotopes with $Q_{\beta\beta} > 2$ MeV are typically considered viable for $0\nu\beta\beta$ detection [3]. The very long half-life means that even other rare events, like neutrino interactions with matter, can constitute significant backgrounds in a $0\nu\beta\beta$ experiment **{fn3.2a}**. As referred in Section 3.1, $2\nu\beta\beta$ events display very similar features to those of $0\nu\beta\beta$, the only distinctions being the presence of emitted neutrinos and the constant sum energy of the two emitted electrons. The measured $T^{0\nu\beta\beta}_{1/2}$ lower limits listed in Table 3.1 show that the half-lives of $0\nu\beta\beta$ are larger than those of $2\nu\beta\beta$, typically by at least a factor of $10^3$.

**{fn3.2a}** — As an example, in the SNO+ experiment, $^8$B solar neutrinos constitute $\sim 34\%$ of the expected background counts [50].

| Isotope | N.A. (%) | $Q_{\beta\beta}$ (MeV) | $T_{1/2}^{2\nu\beta\beta}$ (yr) | $T_{1/2}^{0\nu\beta\beta}$ (yr) |
|---|---|---|---|---|
| $^{48}$Ca | 0.187 | 4.263 | $5.3 \times 10^{19}$ | $> 5.8 \times 10^{22}$ |
| $^{76}$Ge | 7.8 | 2.039 | $1.88 \times 10^{21}$ | $> 8.0 \times 10^{25}$ |
| $^{82}$Se | 8.7 | 2.998 | $0.87 \times 10^{20}$ | $> 3.6 \times 10^{23}$ |
| $^{96}$Zr | 2.8 | 3.348 | $2.3 \times 10^{19}$ | $> 9.2 \times 10^{21}$ |
| $^{100}$Mo | 9.8 | 3.035 | $7.06 \times 10^{18}$ | $> 1.1 \times 10^{24}$ |
| $^{110}$Pd | 11.7 | 2.018 | $4.40 \times 10^{19}$ | $> 6.8 \times 10^{23}$ |
| $^{116}$Cd | 7.5 | 2.813 | $2.69 \times 10^{19}$ | $> 2.2 \times 10^{23}$ |
| $^{124}$Sn | 5.8 | 2.293 | $0.78 \times 10^{20}$ | $> 2.0 \times 10^{19}$ |
| $^{130}$Te | 34.08 | 2.527 | $7.91 \times 10^{20}$ | $> 1.5 \times 10^{25}$ |
| $^{136}$Xe | 8.9 | 2.459 | $2.18 \times 10^{21}$ | $> 1.07 \times 10^{26}$ |
| $^{150}$Nd | 5.6 | 3.371 | $9.34 \times 10^{18}$ | $> 2.0 \times 10^{22}$ |

TABLE 3.1: Percentage natural isotopic abundances, $Q_{\beta\beta}$ and $T_{1/2}^{2\nu\beta\beta}$ values for 11 isotopes of experimental interest. All natural abundances, $Q_{\beta\beta}$ and $T_{1/2}^{0\nu\beta\beta}$ values except for $^{110}$Pd and $^{124}$Sn are from [43]. The $T_{1/2}^{2\nu\beta\beta}$ except for $^{110}$Pd and $^{124}$Sn are from a systematical review of experimental results [51]. $^{110}$Pd natural abundance and $Q_{\beta\beta}$ are from [52], the $T_{1/2}^{2\nu\beta\beta}$ is from a systematical review [53] and the $T_{1/2}^{0\nu\beta\beta}$ is was calculated in [54]. $^{124}$Sn natural abundance and $Q_{\beta\beta}$ are from [55] and the $T_{1/2}^{2\nu\beta\beta}$ was calculated using a Woods-Saxon single-particle basis [56] and the $T_{1/2}^{0\nu\beta\beta}$ is from a 5285 h 1.1$\ell$ liquid scintillator run [57].

### 3.2.2 Detection methods and technologies

The clearest signature of a $0\nu\beta\beta$ event is its electron sum energy of $Q_{\beta\beta}$: all $0\nu\beta\beta$ detectors built up to date are some form of calorimeter with a target sensitive volume where the energy deposited is screened for a $Q_{\beta\beta}$ value [43]. Many current-generation experiments have also used event topology reconstruction to improve background discrimination.

$0\nu\beta\beta$ detectors are designed to optimize *sensitivity*, defined as the 90% or 95% confidence level (CL) upper limit of the number of counts expected to be obtained from a detector's experimental run when no actual $0\nu\beta\beta$ events occurred, with an expected background rate $B$ [58]. Sensitivity is usually expressed in terms of the achieved corresponding lower limit of $T_{1/2}^{0\nu\beta\beta}$ or upper limit of $\langle m_{\beta\beta} \rangle$ (Eq. 3.2), calculated as:

$$T_{1/2}^{0\nu\beta\beta} = k_{\beta\beta}^{T} \times \varepsilon \sqrt{\frac{M_{\beta\beta}\, t}{B\, \Delta E_{\beta\beta}}} \quad , \quad \langle m_{\beta\beta} \rangle^2 = k_{\beta\beta}^{m} \times \left( T_{1/2}^{0\nu\beta\beta} / k_{\beta\beta}^{T} \right)^{-1} , \tag{3.3}$$

where $\varepsilon$ is the $0\nu\beta\beta$ event *detection efficiency*, $t$ is the experimental run-time, $M_{\beta\beta}$ is the mass of $0\nu\beta\beta$-active material, $\Delta E_{\beta\beta}$ is the detector's energy resolution at energy $\sim Q_{\beta\beta}$, and $k_{\beta\beta}^{T}$ and $k_{\beta\beta}^{m}$ are two constants that depend on the chosen isotope and confidence level [43] [58]. A background count occurs whenever a non-$0\nu\beta\beta$ event interacts with the detector in a way that is indistinguishable from $0\nu\beta\beta$, and the rate $B$ is commonly expressed in units of counts/(keV kg yr). The factor $M_{\beta\beta}\, t$ is referred to as the *exposure* of the experiment.

Summarily, the design of a $0\nu\beta\beta$ detector optimizes sensitivity by accomplishing the following:

**Improve detection efficiency**

Depends on the detection technology and on the signal features captured by the detector construction.

**Increase exposure**

Depends on the choice of isotope and detector construction. Decisive factors include the costs of procurement / enrichment, and the scalability of the chosen detection technology (currently preferred isotopes described in Section 3.2.4).

**Improve energy resolution near $Q_{\beta\beta}$**

Allows to define a narrower energy region-of-interest (ROI), reducing the counted $0\nu\beta\beta$ candidates due to background events. Depends on technology and construction.

**Reduce background count rate**

Depends on the choice of isotope, detection technology, detector construction and location. A higher $Q_{\beta\beta}$ reduces $T_{1/2}^{0\nu\beta\beta}$ while also having the ROI in an energy region with less background activity as fewer background events are capable of depositing such high energies. Beyond the isotope choice, background reduction is achieved via background shielding, vetoing and rejection (Section 3.2.3).

Next follows a description of the operating principle and performance characteristics of the five base detection technologies in $0\nu\beta\beta$ search, as well as their variants. Given the focus of this work, the description of time projection chambers is more detailed, while that of the other technologies is short. Detector sensitivities will be addressed in Section 3.2.4.

### Cryogenic bolometer

Bolometers are detectors that can perform calorimetric measurements via the heating of a superconducting solid absorber, typically a crystal when applied to $0\nu\beta\beta$ search [43]. A schematic of the operating principle for the case of $0\nu\beta\beta$ search is shown in Figure 3.6. The target (absorber) is coupled to a heat sink with temperature $T_0$ (typically $\sim 10 - 20$ mK [59]) via a weak thermal link of thermal conductivity $K$. After an atom in the absorber suffers a $0\nu\beta\beta$ event, the primary electrons interact with the crystal lattice and a large fraction of their energy is converted into phonons [60], causing a peak in the temperature of the absorber (typical values for $0\nu\beta\beta$ search are in the order of $\sim 0.1$ mK/MeV [59]) which is then measured by a highly sensitive thermometer.

The energy resolution of absorbers can reach as low as $\sim 1$ eV due to the low thermal capacity of the superconducting crystal. The resolution of bolometers is instead limited by the noise due to the cryogenics and read-out electronics [60], however modern $0\nu\beta\beta$ bolometer experiments still manage to attain resolutions of $\sim 4 - 10$ keV near $Q_{\beta\beta}$ [61] [62]. The heat capacity is proportional to the mass of the absorber [60], and so $0\nu\beta\beta$-detecting bolometers generally use sets of absorber elements, each with mass $\lesssim 1$ kg and volume $\lesssim 5\,\mathrm{cm}^3$, typically cubic in shape and arranged in an array of towers. Although the fact that the sensitive volume is itself the active source ensures high detection efficiency, the small size of the absorber elements means that some $0\nu\beta\beta$ events near the absorber surface only partially deposit their energy, and otherwise escape detection by the entire absorber array, reducing the detection efficiency to $\sim 80 - 90\%$ [60] [63]. Currently the largest $0\nu\beta\beta$ bolometer detector is CUORE, using 741 kg

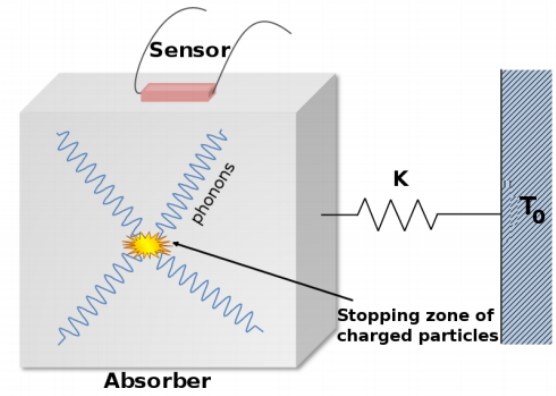


FIGURE 3.6: Schematic of the operating principle of a bolometer element in the $0\nu\beta\beta$ case.
Source: [60]

of unenriched $TeO_2$ [62]. The scalability of bolometer detectors is limited by the difficulty of working at extremely low temperatures [43].

The most suitable $\beta\beta$-decaying isotopes for absorber crystal growth are $^{100}$Mo, $^{116}$Cd, $^{82}$Se and $^{130}$Te [59]. For certain crystals, a *scintillating bolometer* design is also possible, where crystal scintillation is also read-out, in order to reduce the background rate due to $\alpha$ particles by way of differences in their photon/phonon yield ratios [59].

### Semiconductor

Semiconductor detectors in $0\nu\beta\beta$ search typically use fully depleted $^{76}$Ge-enriched high-purity germanium (HPGe) crystals as the sensitive volume, operated at cryogenic temperature. The crystal is placed in a reverse-biased high-voltage circuit with a read-out, maintained at a temperature of typically $\gtrsim 80$ K by a cryostat, and after an atom in the crystal suffers a $0\nu\beta\beta$ event, the primary electrons interact with the sensitive volume to release a number of electron-hole pairs proportional to the deposited energy, which then drift to the contacts and produce an electric pulse [64].

Figure 3.7 shows the two crystal configurations typically used in $0\nu\beta\beta$ search. The closed-ended coaxial configuration maximizes the attainable volume of fully depleted crystal (volumes of $\gtrsim 700\,\text{cm}^3$ have been achieved [65]), whereas the more recently adopted p-type point contact (PPC) or broad-energy (BEGe) configurations ($0\nu\beta\beta$ searches typically use PPCs and BEGes with volumes of $\gtrsim 150\,\text{cm}^3$ [66] [64]) allow for improved pulse shape discrimination [67]. In order to increase exposure, semiconductor experiments typically use arrays of several crystals: unlike in the case of bolometers, the cryostat temperatures are sufficiently high to not significantly hinder scalability. Currently the largest $0\nu\beta\beta$ detectors are GERDA-II (35.6 kg of Ge with 86% enrichment [68]) and the MAJORANA Demonstrator (29.7 kg of Ge with 88% enrichment + 14.4 kg natural Ge [69]), and the LEGEND collaboration is projected to begin data collection in 2021 with 200 kg of Ge with 88% enrichment, as the first phase of a future tonne-scale experiment [70].

The high yield of charge carriers due to the narrow band gap in germanium (0.67 eV [64]) allows for excellent energy resolution in the crystal. Semiconductor $0\nu\beta\beta$ experiments commonly attain energy resolutions of $\sim 2$ - 3 keV at $Q_{\beta\beta}$ [72] [73] [74]. Detection efficiency after

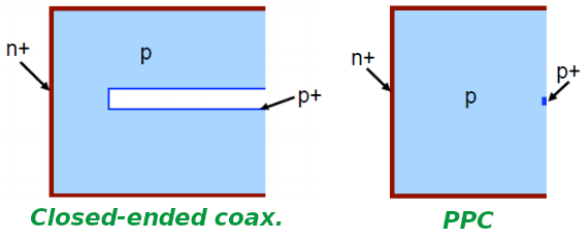


FIGURE 3.7: Cross section of the two crystal configurations typically used in semiconductor $0\nu\beta\beta$ search, closed-ended coaxial on the left schematic and p-type point contact on the right schematic. Both configuratios have a cylindrical design. The bulk of the crystals (light blue) is HPGe lightly doped with acceptor impurities. The $p^+$ contacts are HPGe heavily doped with acceptor impurities and the $n^+$ contacts are HPGe heavily doped with donor impurities. Typically a high positive voltage is applied to the $n^+$ contact, while the $p^+$ is grounded. Broad-energy crystals are similar to PPCs, except the $p^+$ contact is wider. Source: [71]

pulse shape discrimination is usually $\sim 70\%$ for coaxial crystals and $\sim 85\%$ for PPC and BEGe crystals [75] [72]. Beyond the use of $^{76}$Ge, $^{116}$Cd has also been studied using CdZnTe $0\nu\beta\beta$ semiconductors, namely in the COBRA collaboration ($\sim 0.38$ kg of unenriched CdZnTe), although currently the technology is not yet competitive [76].

### Scintillator

Scintillation detectors can perform calorimetric measurements by using de-excitation photons as information carriers. After an atom in the sensitive volume of a scintillator suffers a $0\nu\beta\beta$ event, the primary electrons excite the medium, which then releases fluorescence and phosphorescence photons that travel freely through the material until being captured by photodetectors, with which the scintillation — and therefore the incident energy — can be measured (Chapter 8 of [65]). Despite providing inferior energy resolutions, this family of detectors provides certain advantages, namely excellent background rates and good scalability. It is typically divided into two subfamilies: *organic scintillators*; and *inorganic scintillators*. In $0\nu\beta\beta$ search, organic scintillators are generally a solution of liquid scintillator with some $0\nu\beta\beta$-decaying material, while inorganic scintillation detectors usually use solid crystal sensitive volumes grown with a $0\nu\beta\beta$-decaying component **{fn3.2c}**.

Figure 3.8 shows the typical configurations used for organic and inorganic scintillators in $0\nu\beta\beta$ search. Most detectors of both subfamilies converge on the same three basic parts: an *inner detector* where the $0\nu\beta\beta$-decaying material is located, usually adjacent to a liquid scintillator layer for background vetoing **{fn3.2d}**; an *outer detector* surrounding the inner detector, serving as a background shield; and a photomultiplier (PMT) array around the outer detector. Organic inner detectors typically use a large, (mostly) spherical liquid sensitive volume, whereas inorganic scintillators use arrays of small crystals, separated by gaps to aid with background rejection [80].

**{fn3.2c}** — An exception to this is the XMASS experiment, which uses ultrapure liquid xenon (LXe) as a liquid scintillator [77]. The first generation of the experiment (XMASS-I) has completed data taking in 2019 [78], and now a >20 tonne upgrade (XMASS-II) is being planned, with a 5 tonne intermediate upgrade (XMASS-1.5) currently in the design stage [79].

**{fn3.2d}** — The terms "background vetoing", "background shielding" and "background rejection" are defined in Section 3.2.3.

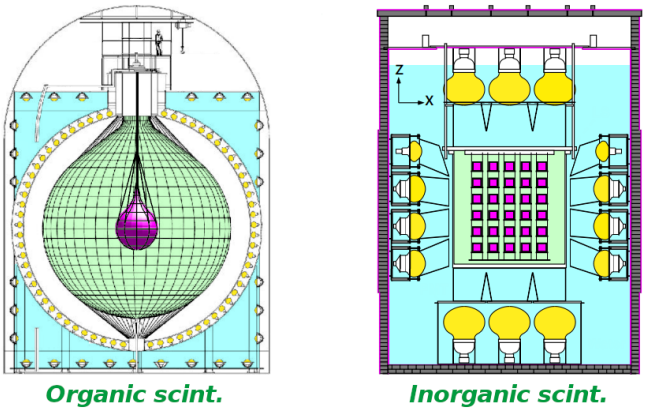


FIGURE 3.8: Cross sections of a typical organic scintillator (KamLAND-Zen 800, left) and typical inorganic scintillator (CANDLES-III, right). The $0\nu\beta\beta$-decaying scintillators / solution are highlighted in magenta. In both cases they are surrounded by liquid scintillator, highlighted in green. Photodectors (photomultipliers in this case) are highlighted in yellow, arranged into arrays. The material highlighted in blue serves to shield the detector from cosmic and earth-originating radiation.
Sources: [81] and [82]

Neutrino detectors can often be easily repurposed for $0\nu\beta\beta$ search as organic scintillators. This approach provides some advantages, namely the possibility of improved background self-shielding (see Section 3.2.3) with the large detector bulk, ease of purification due to the liquid medium, and excellent scalability due to the large detector size requirements for neutrino search. Two of the three most prominent organic scintillator $0\nu\beta\beta$ experiments used this approach: KamLAND-Zen [83], currently using 750 kg of 90% $^{136}$Xe-enriched xenon in a gas-liquid solution with liquid scintillator [43]; and SNO+ [50], with ~1330 kg of natural tellurium [84].

The largest prominent inorganic scintillator $0\nu\beta\beta$ experiment is CANDLES-III, with 305 kg of natural $CaF_2$ crystals [43] **{fn3.2e}**. Calcium crystal scintillators are the most common, because the high $^{48}$Ca decay Q-value allows for a significantly more background-clean measurement than other isotopes: the ELEGANT VI experiment, for example, recorded no background events in either of its two runs, with exposures of ~4.3 and ~9.3 kg × yr, respectively [86]. Large $^{48}$Ca exposures prove challenging, however, due to the low natural isotopic abundance and high cost of enrichment. Beyond the use of $^{48}$Ca, $^{116}$Cd has also been studied using $CdWO_4$ scintillator crystals, namely in the Aurora experiment (1.162 kg 82% $^{116}$Cd-enriched $CdWO_4$), although currently the technology is not yet competitive [87].

### Tracking Calorimeter

Tracking calorimeters search for $0\nu\beta\beta$ events by identifying the energy, momentum and trajectory of the two primary electrons and the opening angle between them. It is the only currently pursued detection technology where the sensitive volume is not itself the $0\nu\beta\beta$ event source. Instead, very thin source foils (~40 mg/cm$^2$ [88]) are placed in a medium that the primaries can traverse freely, providing 3D position information via sparse ionizations of the medium

**{fn3.2e}** — The XMASS experiment used 800 kg of pure liquid xenon [79], however it is primarily adapted for dark matter search, and correspondingly has comparatively low $0\nu\beta\beta$ sensitivity ($T^{0\nu\beta\beta}_{1/2} \sim 10^{21}$yr [85] compared to $T^{0\nu\beta\beta}_{1/2} \sim 10^{26}$yr obtained in KamLAND-Zen [83]).

or via interaction with drift cells within it. The primaries are then captured by scintillators, yielding calorimetry data. The two currently active tracking calorimeter $0\nu\beta\beta$ search experiments, DCBA-T3 [89] and SuperNEMO [90], also apply a magnetic field to curve the electron trajectories, for improved background rejection (Section 3.2.3). Figures 3.9 and 3.10 show the operating principle of the two detectors.

These detectors provide unique advantages compared to other detection technologies. Most important is that the design with disparate source and sensitive volume allows for the use of almost any of the available isotopes as a $0\nu\beta\beta$ source, only necessitating that the production of a source foil be possible. Another advantage is that the excellent topology reconstruction, together with charged particle identification, allows for very efficient background rejection (Section 3.2.3), as well as making this technology capable of probing physics that the other technologies are unable to, namely the theoretically proposed neutrinoless quadruple beta decay ($0\nu4\beta$) [89]. There are also considerable limitations, however. In order to ensure that the primary electrons escape the source foil, the foils have to be very thin, so a sufficient exposure would require a very large experiment, meaning that scalability is poor. The disparate source and sensitive volume design also severely reduces detection efficiency, and the interactions with the tracking medium cause the energy resolution to be poor compared to that of the other technologies.

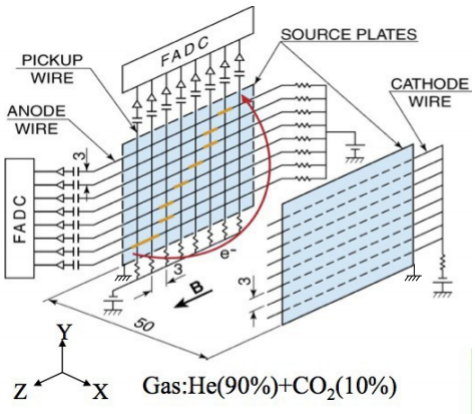


FIGURE 3.9: Operating principle of the DCBA-T3 detector. As an atom in the source plate suffers a $0\nu\beta\beta$ event, the primary electrons are released into 90% He + 10% $CO_2$ gas chambers at atmospheric pressure on either side of the plate, where they are subjected to a 2.4 kG magnetic field [89]. The primary electrons ionize the gas along their trajectories, and the ionization electrons drift to anode wires along a uniform electric field. The pickup and anode wires are then used for calorimetry and $yz$ track position reconstruction. The $x$ track positions are reconstructed via the electron drift time.
Image source: [91]

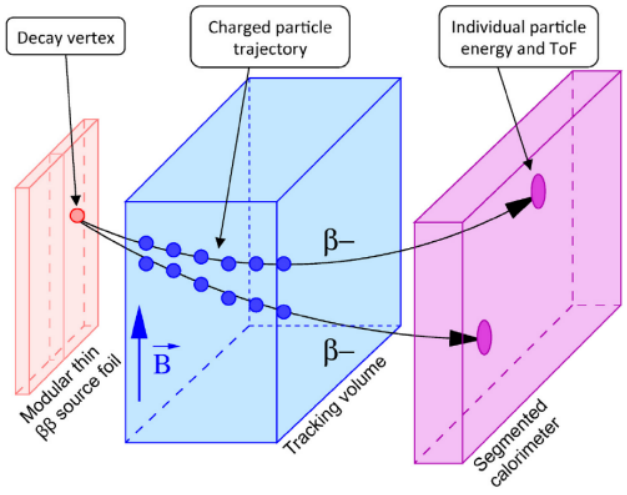


FIGURE 3.10: Operating principle of the SuperNEMO detector. As an atom in the source foil suffers a $0\nu\beta\beta$ event, the primary electrons are released into 95% He +4% ethanol +1% Ar gas chambers at 7 mbar above atmospheric pressure on either side of the foil [92] [88], where they are subjected to a 25 G magnetic field [93]. The tracking volumes contain ∼2000 Geiger-mode wire drift cells with which the electron trajectories are reconstructed. Surrounding the tracking volume are scintillators coupled to PMTs, which are used for calorimetry and timing.
Image source: [93]

### Time Projection Chamber

Originally proposed by David R. Nygren in 1974 [94] for particle identification using spatial and calorimetric data retrieved from gas ionization electrons, the time projection chamber (TPC) design has since been refined and adopted for applications in, among other disciplines, particle accelerators [95], gamma-ray astronomy [96], and dark matter search [97]. In $0\nu\beta\beta$ search, it is implemented as a detector where the sensitive volume is itself the active source, with a xenon medium, either in the form of high pressure gas (HPXe, density $\lesssim 0.1\text{g/cm}^3$), of liquid (LXe, density $\sim 2.95$ g/cm$^3$), or of a combination of both (2-phase TPC). These three variants are described further on, following a primer on xenon as a detection medium.

#### *Xenon as a detection medium — particles released upon deposition*

After an atom in the sensitive volume of a TPC suffers a $0\nu\beta\beta$ event, the primary electrons interact with the medium, depositing an energy $E_0$ into three channels: heat, excitation and ionization. This can be expressed as [98]:

$$E_0 = N_\text{i}\varepsilon + N_\text{ex}E_\text{ex} + N_\text{i}E_\text{i}, \tag{3.4}$$

where $N_\text{i}$ and $N_\text{ex}$ are the average number of ionized and of excited atoms; $E_\text{i}$ and $E_\text{ex}$ are the average primary electron energy loss per ionization and per excitation; and the $N_\text{i}\varepsilon$ component represents the energy transferred to the medium as heat, $\varepsilon$ being the average kinetic energy of the electrons with energy below the excitation potential immediately after the last collision that yielded an excitation or ionization [99]. It is observed that $N_\text{i}$ is virtually proportional to $E_0$ [100]. The values of $\varepsilon$, $E_\text{i}$ and $E_\text{ex}$ do not have a significant dependence on the energy of the incident particle [101], and they are all roughly in the same order of magnitude as the ionization

potential, $I$, which is 12.13 eV in xenon gas and 11.67 eV in liquid xenon [101]. The value of $\varepsilon$ is $\sim 0.3\,I$ [101]. $E_{\rm i}$ is slightly larger than $I$, accounting for the small probability of multiple ionization and ion excitation [99]. $E_{\rm ex}$ is slightly smaller than $I$, because the atomic levels in xenon are close to the ionization limit [101]. The ratio $N_{\rm ex}/N_{\rm i}$ has also been determined empirically [99]. It depends on the incident particle, and for electrons it is about 0.4 - 0.6 for xenon gas and 0.1 - 0.2 for liquid xenon [101].

The three deposition channels yield corresponding particles, mainly by the sequence of processes pictured in Figure 3.11, namely phonons (heat), 178 nm photons **{fn3.2f}**(scintillation), and electron-ion pairs. The comparatively high temperatures of gas and liquid Xe (triple point at 0.81600 bar and 161.38 K [102] [103], far above cryogenic temperatures) make bolometry unfeasible, so the phonons are lost, but the photons and electrons are typically retrieved in $0\nu\beta\beta$ search. In Xe, the average energy required to produce an electron-ion pair, $W_{\rm i}$, is virtually independent from the type and energy of the incident particle, and displays a virtually linear dependence on the density **{fn3.2g}**. It can be obtained from 3.4 as [101]:

$$W_{\rm i} \equiv \frac{E_0}{N_{\rm i}} = \varepsilon + \frac{N_{\rm ex}}{N_{\rm i}} E_{\rm ex} + E_{\rm i} = \begin{cases} 22.0 \pm 0.2\,\text{eV},\ \sim 1\,\text{atm Xe gas}\ [105]\,[106] \\ 15.6 \pm 0.3\,\text{eV, liquid Xe}\ [107] \end{cases} . \tag{3.5}$$

The average energy required to produce a scintillation photon, $W_{\rm sc}$, is defined in a similar manner, as $W_{\rm sc} \equiv E_0/N_{\rm sc}$, where $N_{\rm sc}$ is the total number of scintillation photons produced. As shown in Figure 3.11, both the excitation and ionization channels yield scintillation photons, with the ionization channel yielding them via recombination at the $0\nu\beta\beta$ event site [101]:

$$N_{\rm sc} = rN_{\rm i} + N_{\rm ex}, \tag{3.6}$$

where $r$ is the fraction of electrons that recombine. By applying a strong electric field, $r$ can be reduced to $r \sim 0$. If no electric field is applied, most released electrons recombine: $r \sim 1$. Substituting $W_{\rm sc} \equiv E_0/N_{\rm sc}$ and $W_{\rm i} \equiv E_0/N_{\rm i}$ into Eq. 3.6, $W_{\rm sc}$ can be expressed as [101]:

$$W_{\rm sc} = W_{\rm i} \times \frac{1}{r + N_{\rm ex}/N_{\rm i}} \approx \begin{cases} (13.6\text{ - }55.5)\ \text{eV},\ \sim 1\,\text{atm Xe gas} \\ (12.8\text{ - }159)\ \text{eV}\,,\ \text{liquid Xe} \end{cases} , \tag{3.7}$$

where the energy ranges were calculated using the upper and lower bounds of the empirical values of $W_i$ listed in Eq. 3.5, and of the ranges for $r$ and $N_{\rm ex}/N_{\rm i}$.

It is expected that a $^{136}$Xe $0\nu\beta\beta$ event would therefore yield as electrons and photons a minimum of $N_{\rm i} \sim 1.1 \times 10^5$ and $N_{\rm sc} \sim 0.4 \times 10^5$ in low pressure Xe gas, or $N_{\rm i} \sim 1.5 \times 10^5$ and $N_{\rm sc} \sim 0.2 \times 10^5$ in liquid Xe. A large proportion of the electrons and photons can be made to remain free in the medium provided that the concentration of impurities is small: Xe is transparent to the scintillation photons from the de-excitation of $Xe_2^*$, so their absorption length is $>$100 cm in pure Xe [108]; and forcing the electrons to drift away from the energy deposition site by a sufficiently strong electric field suppresses most recombination, preserving a large proportion of the free electrons **{fn3.2h}**, which then remain free unless they suffer

**{fn3.2f}** — The scintillation spectrum of xenon is a near-Gaussian rough continuum with a peak scintillation wavelength of 178 nm, and a spectral width (FWHM) of ∼14 nm (see Fig. 5 of [104]).

**{fn3.2g}** — Low pressure gaseous Xe $W_{\rm i}$ values for incident $^{35}$S $\beta$ electrons and $^{210}$Po $\alpha$ particles ware measured and compared in [105]: the two $W_i$ values differed by $<1\%$. The dependence of $W_i$ for Xe to the density and the incident particle energy was studied in [106], and it was determined that $W_i$ is virtually inversely proportional to the density, ranging from low pressure Xe to LXe, and that for two different densities there was no perceptible variation in $W_i$ at varying incident particle energies.

attachment to electronegative impurities (such as oxygen and water), though current Xe purification technologies allow the electron mean lifetime to significantly exceed the average time taken by a free electron to traverse the entire length of a TPC (Section 1.5 of [109]).

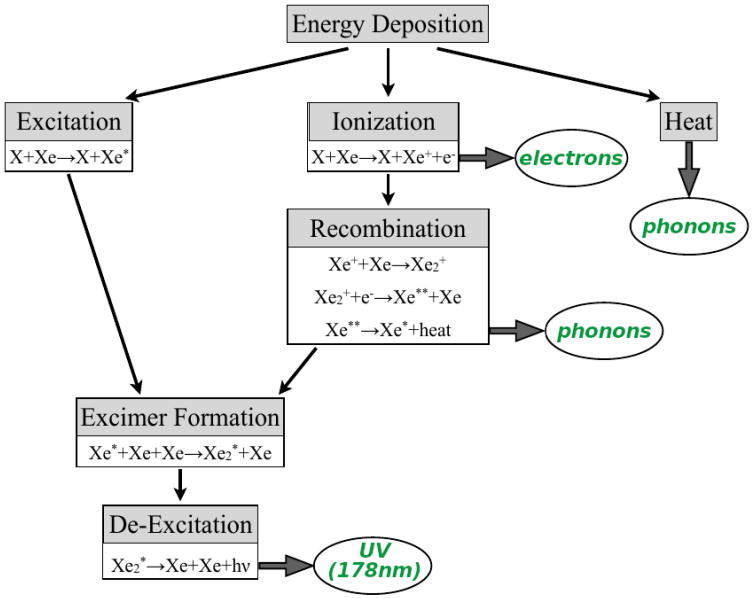


FIGURE 3.11: Main processes producing the released phonons, photons and electron-ion pairs. The symbol "X" represents ionizing radiation. "UV (178 nm)" corresponds to scintllation photons. As depicted, besides the excitation channel, the ionization channel also yields scintillation photons via prompt recombination.
Image source: [101]

***Xenon as a detection medium — calorimetry***

The preceding discussion shows that the Xe ionization energy deposition channel displays attractive properties for calorimetry via free electron counting: $W_{\rm i}$ has virtually no dependence either on the type or on the energy of the incident particle, and $N_{\rm i}$ for $\sim$ MeV range particles is fairly large since $W_{\rm i}$ is in the $\sim$ eV range. Excluding all subsequent effects, the energy resolution of the medium cannot be better than the limit imposed by the statistical fluctuations in the number of ionizations. Given $W_{\rm i} \equiv E_0/N_{\rm i}$, the best possible, *intrinsic Xe ionization energy resolution*, $\delta E_{\rm i}$, is defined as [101]:

$$\delta E_{\rm i} = 2.355 \times W_i\, \delta N_{\rm i} \ \ \text{(FWHM)}, \tag{3.8}$$

where $\delta N_{\rm i}$ is the standard deviation in the number of ionizations.

The distribution of the number of ionizations is typically modeled as sub-Poissonian, with a dependency factor (the *Fano factor*) $F$, giving a standard deviation [101]:

$$\delta N_{\rm i} = \sqrt{F N_{\rm i}}. \tag{3.9}$$

**{fn3.2h}** — A usable parameterization of the dependence of the number of deposited free electrons , $N$, on the applied electric field, $\mathcal{E}$, is $N(\mathcal{E}) = N_i \times (1 + k/\mathcal{E})^{-1}$, where $k$ characterizes the recombination strength and has the value $k \approx 2.4$ kV/cm for LXe with $\sim$ 0.1 MeV incident electrons [99].

The $F$ value for low-pressure gaseous xenon was measured to be 0.13 - 0.17 [101], with theoretical values in close agreement [110], and corresponds to $\delta E_{\mathrm{i}} \approx 0.27\,\%Q_{\beta\beta}$. For LXe, $F$ is theoretically estimated to be $\sim 0.059$ [111], however the experimental value is unclear. Conti *et al.* gives a commonly cited $F$ figure of $\gtrsim 20$ in [112], based on LXe resolution measurements. The ionization energy resolution in LXe is inferior to the resolution in low pressure xenon gas by a factor of $\sim 11$, due to fluctuations in the partitioning of the energy between free electrons and scintillation that become prominent as the density is increased [113]. Martín-Albo suggests that the fluctuations are due to the increased effect of recombination on the obtained number of free electrons [101]. The intrinsic ionization resolution depends on the density of the medium and also on the strength of the applied electric field (Figure 3.12). The ionization energy resolution obtained in practice for either gas or liquid xenon should be expected to be worse than the intrinsic resolution, as beyond the statistical fluctuations it is also affected by recombination and attachment to impurities.

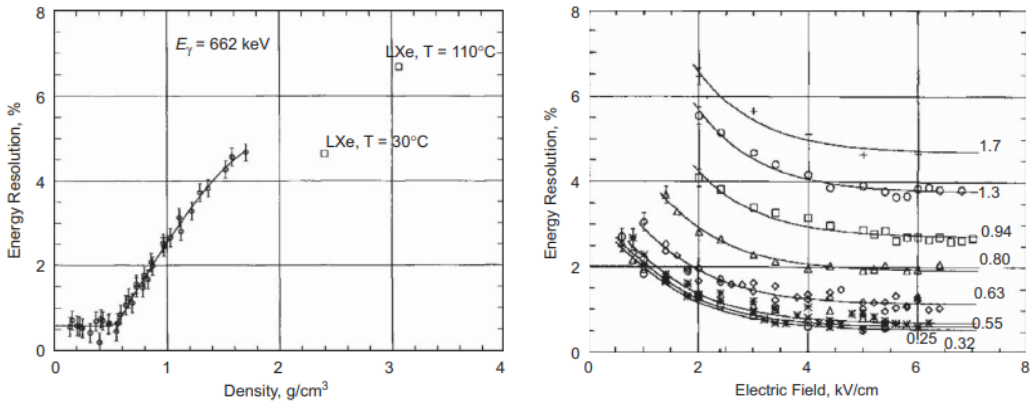


FIGURE 3.12: FWHM intrinsic ionization energy resolution measured for 662 keV $\gamma$-rays at varying xenon densities (left panel) and electric field strengths (right panel). The numbers on the right of the right panel indicate the densities of the xenon used.
Image source: [114] version, originally from [106]

There is less material characterizing the intrinsic scintillation resolution: it requires a more complex model and is inferior to the ionization resolution in typical applications. It is advantageous, however, to use both photon and electron counting together for calorimetry, particularly for LXe. In that case, a significant fraction of scintillation photons are expected to be from recombination of free electrons, and so there is a visible anti-correlation between the number of counted photons and electrons [115]. By using a linear combination of the two measurements, improved energy resolutions can be achieved (Figure 3.13).

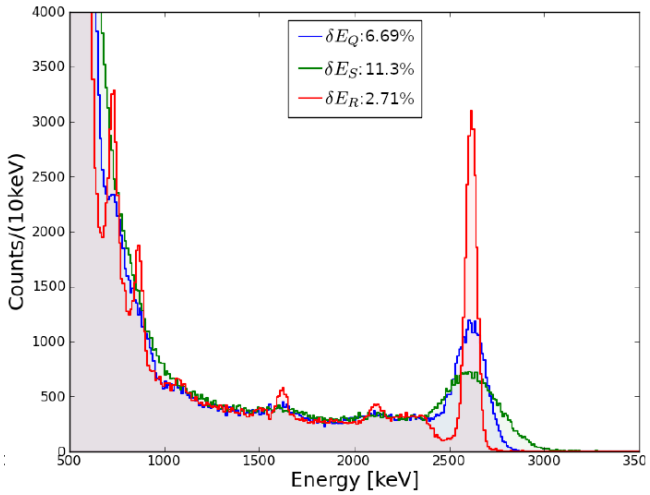


FIGURE 3.13: Electron, scintillation and joint response for EXO-200 during a $^{228}$Th calibration run. $\delta E_Q$ is the FWHM energy resolution from electron counting, $\delta E_S$ is the FWHM energy resolution from photon counting and $\delta E_R$ is the FWHM joint energy resolution. The joint energy is calculated as $E_R = E_Q \cos\theta + E_S \sin\theta$, where $\theta$ is calculated according to the measured anti-correlation of $E_Q$ with $E_S$.
Image source: [115]

### *Xenon as a detection medium — topology discrimination*

The path taken by the $0\nu\beta\beta$ event primary electrons results in a track of free electrons and scintillations: an event topology. The interaction cross-sections of the primary electrons with the Xe are largely independent of the density of the medium, so the size of the topology, namely the distance between blobs (see Section 3.1), can be expected to be roughly inversely-proportional to the density of the medium. There is little available literature characterizing the dimensions of the topologies. The NEXT-100 conceptual design report refers a track length in the order of 30 cm in 10 bar HPXe [42]. Monte Carlo simulations performed for this work (see Chapter 6 for implementation details) suggest an average blob-to-blob distance in LXe of $\sim$ 4 mm.

The scintillation photons immediately escape the $0\nu\beta\beta$ event site, but the electrons initially remain near the site. Therefore by applying an electric field and causing the free electrons to collectively *drift* at a constant average *drift velocity* toward some readout plane, the arrival time of individual electrons at the readout and their 2D position along the plane can be used to retrieve 3D topological information of the event.

Unfortunately, as the group of electrons traverses the medium, it becomes increasingly difficult to discern the original event topology from it. The electrons experience Brownian motion inside the medium: they suffer many collisions with its atoms, which besides ensuring that the group of free electrons maintains an average drift speed instead of accelerating as the electric field is applied, also prevents the electrons from traveling in a straight line. Instead the electrons perform a "random walk" in the $x$, $y$ and $z$ coordinates [101] (typically defined so that the electric field is applied along the $z$ direction). Along the direction of the electric field, the drifting electrons experience *longitudinal diffusion*, while along the other two coordinates they experience *transverse diffusion*. Supposing that an electron originates at position $(0, 0, 0)$, the probability of finding it in a certain position $(x, y, z)$ at time $t$, $n(x, y, z, t)$, can be described

by the following distribution [116]:

$$n(x,y,z,t) = \frac{1}{4\pi D_T t \times \sqrt{4\pi D_L t}} \times \exp\left[-\frac{1}{4t}\left(\frac{x^2+y^2}{D_T} + \frac{(z-v_d t)^2}{D_L}\right)\right], \tag{3.10}$$

where $v_d$ is the drift velocity, $D_L$ is the longitudinal diffusion coefficient and $D_T$ is the transverse diffusion coefficient. These quantities in turn can be expressed as [116]:

$$v_d = \mu\mathcal{E} \quad , \quad D_L = \frac{k_B T}{e}\left(\mu + \mathcal{E}\frac{\partial\mu}{\partial\mathcal{E}}\right) \quad , \quad D_T = \frac{k_B T}{e}\mu, \tag{3.11}$$

where $\mu$ is the electron mobility, $\mathcal{E}$ is the electric field strength, $e$ is the electron charge, $k_B$ is the Boltzmann constant and $T$ is the temperature. The expression for $D_T$ is just the Einstein-Smoluchowski relation, and the expression for $D_L$ an extension of it [116]. Figure 3.15 shows the measured dependence of $v_d$ with $\mathcal{E}$, and Figure 3.16 shows the measured dependence of $\mu$ on $\mathcal{E}$ and on the density of the medium. With typical drift field strengths, $\partial\mu/\partial\mathcal{E} \lesssim 0$, and so $D_L \lesssim D_T$. The reviewed LXe diffusion coefficient measurement data in Figure 3.14 is in agreement with this for $\mathcal{E} \gtrsim 100$ V/cm. Given that $D_L$ and $D_T$ are (partly) proportional to $\mu$, the right panel in Figure 3.16 suggests that the diffusion coefficients are significantly larger in LXe than in gaseous xenon.

In effect, the original topology becomes increasingly "smeared" as it drifts through the medium, but it experiences less "smearing" along the direction of the electric field than perpendicular to it. This effect is more detrimental in liquid xenon than in gaseous xenon, as the topologies in liquid have a smaller size, and they experience stronger diffusion than in gas. Topological reconstruction in gaseous xenon has been shown to be viable in HPXe TPC experiments, while the viability of topological reconstruction in LXe is the subject of this work.

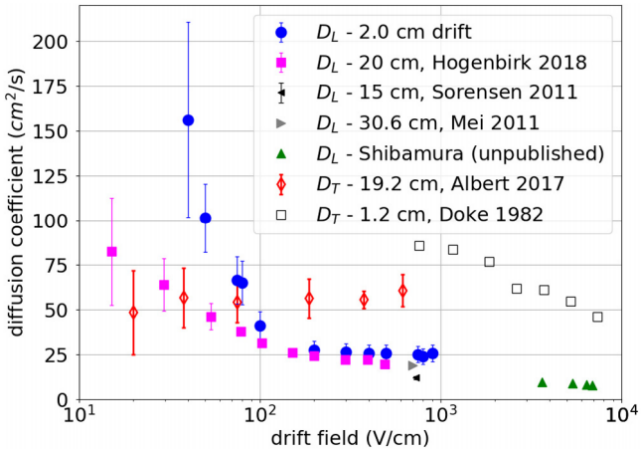


FIGURE 3.14: Review of measurements of $D_L$ and $D_T$ for varying $\mathcal{E}$, for LXe. Image source: [117]

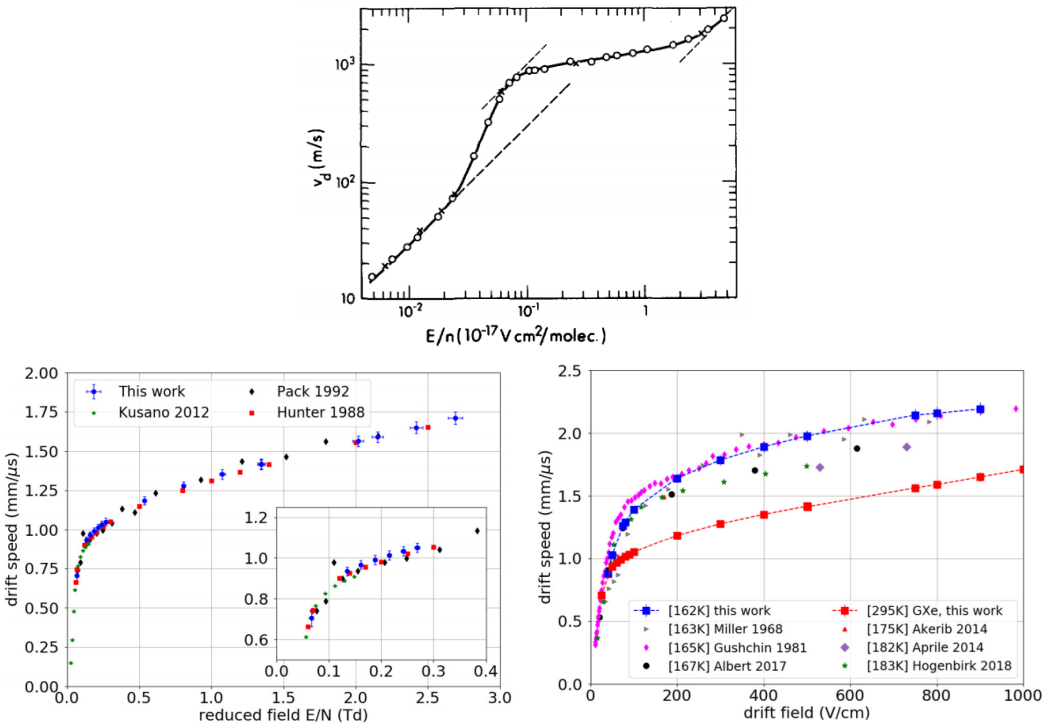


FIGURE 3.15: *Top panel*: Measurement of $v_d$ for varying $\mathcal{E}$ in 236 K, $2.95 \times 10^{10}$ molecule/cm$^3$ xenon gas; *Bottom panels*: Review of measurements of $v_d$ for varying $\mathcal{E}$, for gaseous xenon (left panel) and for different temperatures of LXe (right panel). The reduced field referred in the left panel is $\mathcal{E}$ scaled by the number density of the gaseous xenon. E/n in the top panel is also the reduced field.
Image source: [118] (top) [117] (bottom)

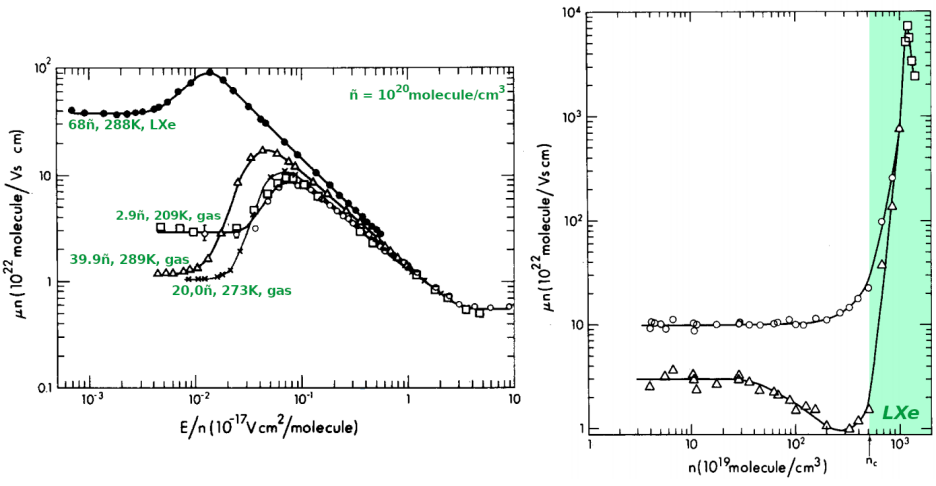


FIGURE 3.16: *Left panel:* Normalized measurements of electron mobility for varying electric field strength, for different temperatures and densities of the medium. *Right panel:* Low field strength electron mobility (△) and peak electron mobility (○) for varying density of the medium.
Image source: [118]

***Xenon as a detection medium — additional benefits***

The preceding paragraphs show that xenon is a medium with favorable characteristics for particle detection. The ionization channel provides a competitive energy resolution near $Q_{\beta\beta}$, which can be further aided by the scintillation signal. The presence of two detectable signal channels is useful for background rejection (namely for distinguishing heavy incident particles like $\alpha$ radiation or neutrons from light particles like electrons or $\gamma$ radiation, see Section 3.2.3), as are the topology reconstruction capabilities provided by the free electrons. There are some other benefits to using a xenon detection medium for $0\nu\beta\beta$ search, aside from the energy deposition.

Xenon in itself is very radiopure. Aside from $^{136}$Xe, the only radioactive isotopes in natural xenon are $^{124}$Xe (N.A. 0.095%, $T_{1/2} = (1.8 \pm 0.6) \times 10^{22}$ yr [119]) and $^{126}$Xe (N.A 0.089%, $T_{1/2} > 1.9 \times 10^{21}$ yr 90% CL [120]) which can decay by double electron capture, both depositing $\sim 64$ keV of de-excitation energy [120], of little concern for $0\nu\beta\beta$ search. It should be noted, however, that several radioactive isotopes of xenon can be produced from natural xenon via neutron capture. Of these, $^{137}$Xe is concerning: it decays via a $\beta^-$ event with $T_{1/2} \approx 230$ s, and has a Q-value of 4.173 MeV, meaning that it is capable of emitting electrons with energies overlapping the $^{136}$Xe $Q_{\beta\beta}$ [39]. The generation of $^{137}$Xe is typically suppressed by shielding the medium from external neutron sources and by preventing its contamination with neutrogenic nuclides. This point is discussed further in Section 3.2.3.

Although xenon is a rare gas, procurement is not prohibitively expensive (In 2019, the projected Xe market price was $\sim 3.70$ \$/g [121] [122]), allowing for tonne-scale experiments. It also allows for efficient enrichment: the world Xe enrichment capacity is of a few tonnes per year [123], with 90% $^{136}$Xe-enrichment commonly attained. The use of a liquid / gaseous detection medium also allows for efficient purification from both electronegative impurities (which reduce the energy resolution) and radioactive impurities (which increase the background rate, discussed further in Section 3.2.3). In effect, the radioactive background rate from sources within the detection medium can be made almost negligible (for an example, see Section 4.2).

In addition to excellent natural radiopurity, LXe provides good self-shielding from external backgrounds, $\gamma$ radiation in particular. Figure 3.17 shows the mass attenuation coefficient for xenon at varying energies of incident $\gamma$ radiation. The minimum mass attenuation coefficient is for an incident energy of $\sim Q_{\beta\beta}$, with a value of $\sim 3.5 \times 10^{-2}$ cm$^2$/g. A beam of $\gamma$ particles with energy $\sim Q_{\beta\beta}$ would be reduced in intensity by half after traversing $\lesssim 10$ cm of LXe. With a large volume of pure LXe medium, this allows for a very quiet inner volume (termed the *fiducial volume*), protected by a surrounding xenon vetoing region (see Section 3.2.3).

TPCs are designed to take advantage of the favorable properties of xenon.

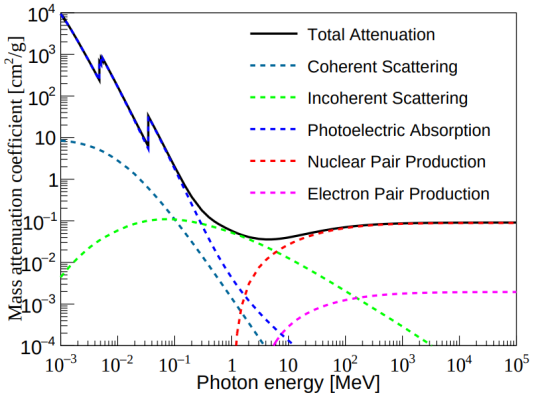


FIGURE 3.17: Mass attenuation coefficient for xenon at varying energies of incident $\gamma$ radiation.
Image source: [124]

### *Xenon TPCs — basic design*

The TPC is a detector technology designed to obtain both a calorimetric measurement and a position or topology reconstruction of an event occurring in the medium. A cylindrical tank with a 2D readout plane on at least one of the faces is filled with the detection medium (Xe in the case of $0\nu\beta\beta$ search), and a uniform electric field is applied to the medium using electrodes. Figure 3.18 shows the basic design and working principle.

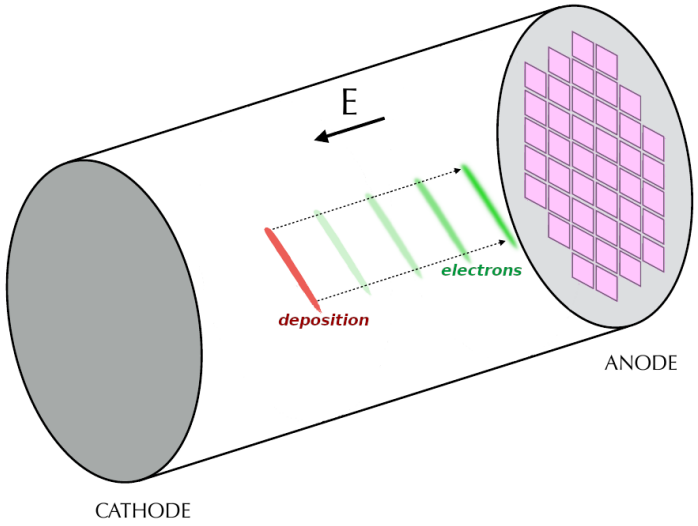


FIGURE 3.18: Diagram of a basic TPC design. The cylinder is filled with detection medium. An energy deposition occurs along a track (highlighted in red). The spacial arrangement of the free electrons (highlighted in green) produced during the energy deposition matches the topology of the track. The electrons are then drifted by the electric field E toward the 2D readout plane (highlighted in magenta). The drift field is generated by two electrodes on either side of the TPC: the cathode and the anode.
Image source: [101]

The large number of free electrons produced in the case of $0\nu\beta\beta$ events makes viable the use of readout planes that count the electrons directly (as is done in single-phase LXe TPCs, namely by the EXO collaboration [123]), via charge collection wires (a Frisch grid) or tiles. This approach should minimize the electron and photon yield variances of the TPC, but in most implementations it also significantly increases the signal-to-noise ratio in the readout electronics [101], making it impractical. Even in $0\nu\beta\beta$ search it is more common to apply signal amplification than to eschew it. Amplification is typically achieved by dividing the detector volume with a *gate* electrode into two connected cylindrical regions: a large *drift region* between the cathode and the gate, where the electric field $E_{\text{drift}} \sim 10^2$ V/cm does not impart enough energy to the electrons for them to excite the medium; and a thin *amplification gap* between the gate and the anode, where the electric field $E_{\text{amp}} \gtrsim 10^3$ V/cm imparts enough energy to the electrons for them to excite the medium, also referred to as producing *electroluminescence* (EL). Electroluminescence allows for better energy resolution than a typical avalanche, and so electroluminescence is preferred for $0\nu\beta\beta$ search [101].

Electroluminescence produces a shower of scintillation photons, which are then captured by the readout plane. In effect, there are two scintillation signals per event: a primary scintillation signal when the energy deposition initially occurs, referred to as *S1*; and a secondary scintillation signal produced by the free electrons exciting the medium at the electroluminescence gap, referred to as *S2*. The readout plane is typically an array of low-background PMTs. They capture both the S2 and the S1 signal, and the time difference between the signals can serve to reconstruct the position of the deposition site within the TPC.

Regarding the choice of phase of the drift region, selecting either LXe or HPXe has benefits and drawbacks, as shown in the preceding paragraphs. Summarily, HPXe has superior energy resolution and topology discrimination capabilities, while LXe is significantly denser and has excellent self-shielding. Applying an amplification gap to LXe reduces the signal-to-noise ratio, but if an amplification gap is used, it has to be inside a layer of gaseous xenon. Hence, in $0\nu\beta\beta$ search, the TPC family of detectors can be divided into three subfamilies: (single-phase) HPXe TPCs; (single-phase) LXe TPCs; and two-phase TPCs. Each has its own particularities in the design and performance characteristics, as described next.

#### *Xenon TPCs — LXe TPCs*

Currently there is one prominent projected single-phase LXe TPC experiment in $0\nu\beta\beta$ search: nEXO (Figure 3.19). It is designed to optimize the use of the good self-shielding, signal channel background discrimination, and energy resolution characteristics of pure LXe. Its design is conservative compared to that of other TPCs (capture of both charge and scintillation channels, single charge capture plane, bottom-positioned cathode, no signal amplification), the focus being on minimizing background sources and improving signal fidelity. The use of a large volume of LXe (∼5 tonnes, 90% enrichment) results in a near-ideal detection efficiency [123], as well as a large fiducial volume (∼4 tonnes) at the center of the vessel, where the only remaining background source is the LXe medium itself. Like in its predecessor, EXO-200, the calorimetric measurement in nEXO uses both the free electron and the scintillation signals. The energy resolution is estimated as 2.36 $\%Q_{\beta\beta}$ (FWHM) [125]. Care in the design was taken to ensure that the event topology was not distorted by the instrumentation, however there is little information on topology reconstruction development.

The nEXO design is also informed by the challenges revealed in EXO-200. Aside from a larger bulk (110 kg of 80.6% $^{136}$Xe-enriched LXe in EXO-200 [43]) and the change from a horizontal orientation of the vessel axis to a vertical orientation, there are two significant differences in the

detector designs: the EXO-200 cathode was mounted at the center of the vessel, not at the bottom like in nEXO; and the charge collection planes in EXO-200 used charge collection wires (Frisch grids), whereas the charge collection plane in nEXO uses metalized fused silica tiles. The cathode can be a significant origin of background (Section 3.2.3), and so a central cathode mounting bisects the fiducial volume with a background source, so for nEXO the cathode was placed at the bottom of the detector, where most of the fiducial volume is shielded from its radiation. The change from charge collecting wires to tiles was due to the increased susceptibility of the wires to reconstruction ambiguity and to microphonic pickup as the scale of the detector increased.

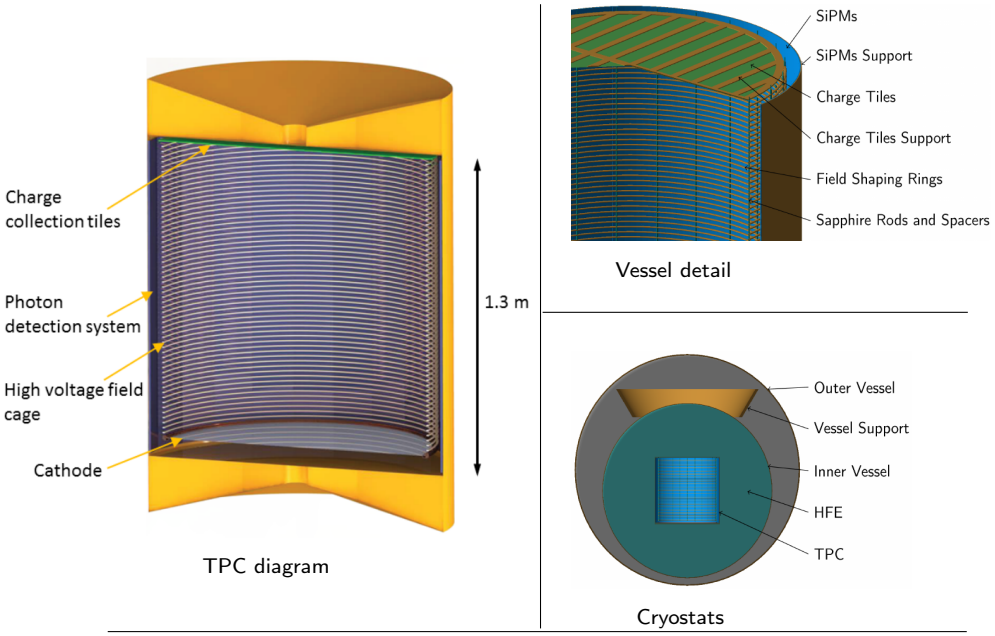


FIGURE 3.19: Diagrams of the nEXO design concept. *Left panel:* Diagram of the nEXO TPC. The detection medium is ∼5 tonnes of 90% $^{136}$Xe-enriched LXe at 165 K (∼4 tonnes of fiducial volume), placed inside a vertically mounted cylindrical copper vessel, and the projected electric field is 400 V/cm, generated by a cathode at the bottom of the detector, an anode at the top and 58 field shaping rings. The detector reads both an ionization and a scintillation signal, using a metalized fused silica tiled anode at the top of the detector for charge collection, and using side-mounted SiPMs for light collection. To improve resolution near $Q_{\beta\beta}$, no amplification was applied to the charge signal. *Top-right panel:* detail of the top of the xenon vessel interior, showing sensor placements and supports, as well as the field shaping rings. *Bottom-right panel:* nEXO cryostat / shielding layers. Two cryostats surround the TPC, separated by a vacuum insulation layer (located between the inner and outer vessels). The inner cryostat is the interior of the inner vessel, and it is filled with HFE-7000 refrigerant fluid. The outer cryostat (not pictured) is the exterior of the outer vessel, and it is composed of a large water tank (projected as having a 13 m diameter and 13.3 m height). Both cryostats double as shielding from external backgrounds.
Image source: [123]

***Xenon TPCs — HPXe TPCs***

Figures 3.20 and 3.21 show conceptual schematics of the two most prominent current HPXe TPC $0\nu\beta\beta$ experiments: NEXT-100 and PandaX-III. The two detectors share a similar basic design: both mount the axis horizontally and focus primarily on background rejection methods. But the approach taken in the two experiments is different. While the NEXT collaboration emphasizes the energy resolution and complements its background-rejecting capability with topology reconstruction, using for that purpose an asymmetric layout of photon readout planes (one energy readout plane and one tracking readout plane), PandaX-III emphasizes topology reconstruction at the cost of some energy resolution, using for that purpose a diffusion-reducing additive in the medium (trimethylamine, which has the side-effect of worsening the primary scintillation signal [126]) and using charge readout planes with superior $xy$ resolution than the tracking plane in NEXT-100. In effect, NEXT-100 has a projected energy resolution of 0.7% (FWHM) at $Q_{\beta\beta}$ [127] and an $xy$ imaging resolution of $\sim$10 mm [128], while PandaX-III has a projected energy resolution of 3% (FWHM) and an $xy$ imaging resolution of $\sim$3 mm [129]. In turn, the estimated background rates are $< 4 \times 10^{-4}$counts/(keV kg yr) for NEXT-100 [130] and $\sim 10^{-4}$counts/(keV kg yr) for PandaX-III [43]. Due to the low density of HPXe, however, detection efficiency is projected to be quite low: 28% in the case of NEXT [130] and 54.2% for PandaX-III [131].

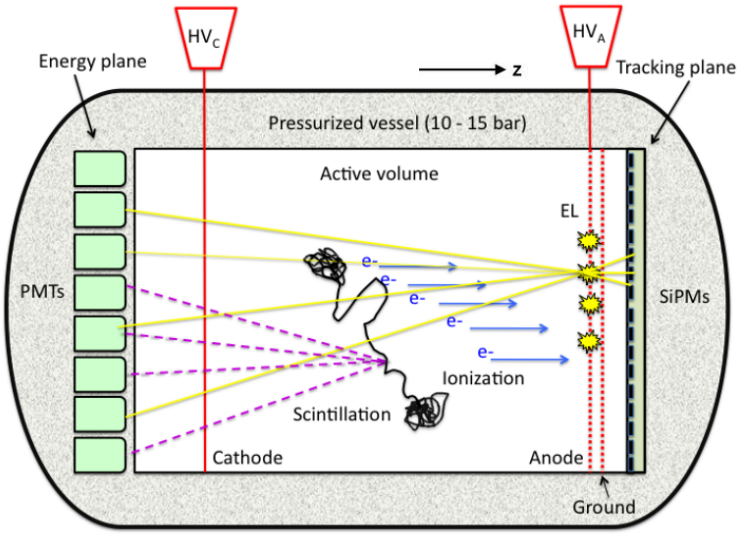


FIGURE 3.20: Diagram of the NEXT base design concept: the Separate, Optimized Functions TPC (SOFT). The sensitive medium is 100 kg of 91% $^{136}$Xe-enriched 15 bar xenon [101] [43]. The charges are converted into an S2 signal once they drift into the electroluminescence gap (marked "EL"). Good calorimetry and track reconstruction have conflicting instrumental requirements [101], and so an asymmetric design is used, where the opposite faces have readout planes with different functions: the face nearest the cathode performs calorimetry with low-background PMTs using the S1 and S2 signals, while the face nearest the EL gap performs $xy$ reconstruction with the more compact SiPMs using the S2 signal. The $z$ coordinate is reconstructed using the time difference between the S1 and S2 signals.
Image source: [132]

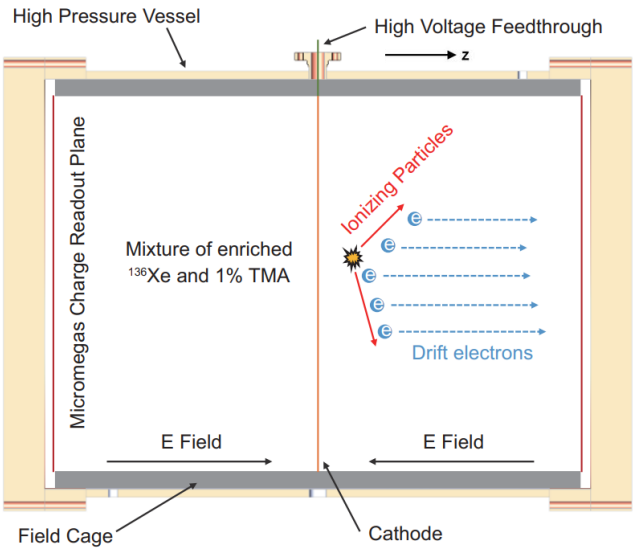


FIGURE 3.21: Diagram of the PandaX-III design concept. The sensitive medium is 200 kg of a 99% / 1% mixture of 90% $^{136}$Xe-enriched 10 bar xenon together with trimethylamine (TMA) quencher. The event energy and $xy$ position are retrieved from the ionization channel using Microbulk Micromegas (which also provide amplification, see [133]). The electric field is produced in two opposite directions from the center of the detector by a single high voltage (HV) electrode at ∼100 kV, and the drift region on either side of the electrode is ∼1 m long. TMA is used to improve the quality of the ionization signal and to suppress the diffusion of drift electrons. In the current design, the scintillation signal is not measured, and so the $z$ position of the events is not retrieved, however timing is still used for 3D reconstruction.
Image source: [131]

### *Xenon TPCs — Two-phase TPCs*

Two-phase TPCs are generally designed for multiple scientific goals at once. There are two current prominent projected two-phase Xe TPC experiments in $0\nu\beta\beta$ search, LZ and XENONnT, and both of them are also dark matter search experiments. The description of the LZ detector is the focus of Chapter 4, so this paragraph will focus on XENONnT, though the core design principles are similar. Figure 3.22 shows a conceptual schematic of its operating principle. XENONnT is projected to use ∼8 tonnes of natural LXe, which, like in LXe TPCs, provide a near-ideal detection efficiency and a large central fiducial volume (∼4 tonnes). The S1 and S2 readouts are performed with top and bottom arrays of 253 and 241 PMTs, respectively. Free electron and scintillation signal anti-correlation is used to improve the energy resolution, which is projected to be $\sim 2.4\,\%Q_{\beta\beta}$ [134]. There is little information on topology reconstruction development for XENONnT.

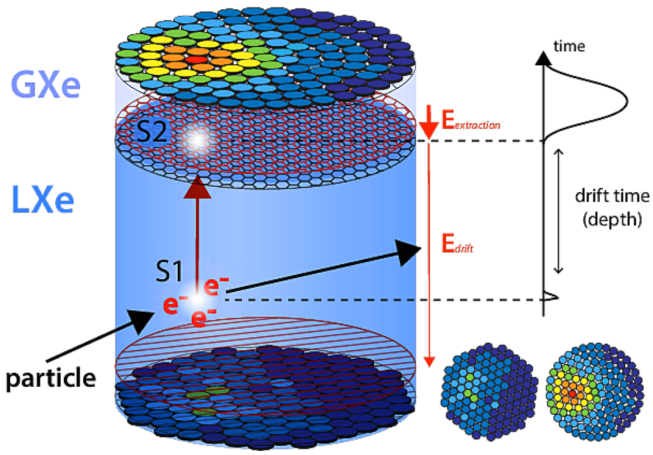


FIGURE 3.22: Diagram of the XENON design concept, used by XENONnT. The diagram shows an incident particle, but otherwise applies for $0\nu\beta\beta$ search. The sensitive medium is ∼8 tonnes of natural LXe. The S1 signal is measured by the top and bottom PMT arrays, and the free ionization electrons drift up to the EL gap due to $E_{drift} \approx 80$ V/cm, produced by a bottom-mounted cathode grid and gate grid inside the LXe, near the surface. The S2 signal is then produced by a stronger electric field, $E_{extraction}$, of $\mathcal{O}(10\,\text{kV/cm})$. The $z$ position of the event is calculated by the time difference between S2 and S1.
Image source: [134]

Table 3.2 summarizes the performance characteristics of each detector technology, and Table 3.3 lists the current and recent experiments using each of the referred technologies. As can be seen, each detector technology is only compatible with a subset of the 11 isotopes. This point is further discussed in Section 3.2.4.

| Technology | Efficiency | ΔE | BG | Scalability | Topology |
|---|---|---|---|---|---|
| **Bolometer** | 80 - 90% | ∼0.2% | $\sim 10^{-2}$ | Poor | None |
| **Scint. Bol.** | 80 - 90% | ∼0.2% | $\sim 10^{-4}$ | Poor | None |
| **Semiconductor** | 70 - 85% | $\lesssim$0.2% | $\sim 10^{-3}$ | Good | Limited |
| **Org. scint.** | >90% | ∼4% | $\sim 10^{-5}$ | Excellent | None |
| **Inorg. scint.** | >90% | ∼3% | Near zero | Good | None |
| **Track. Cal.** | 10 - 30% | ∼5 - 10% | Near zero | Poor | Excellent |
| **HPXe TPC** | 25 - 55% | ∼1 - 3% | $\sim 10^{-4}$ | Good | Excellent |
| **LXe TPC** | >90% | ∼2.5% | $\sim 10^{-3}$ | Excellent | Limited |
| **2-Phase TPC** | >90% | ∼2% | $\sim 10^{-4}$ | Excellent | Possible |

TABLE 3.2: Table comparing the typical performance characteristics of the different detector technologies. "Efficiency" refers to the detection efficiency. $\Delta E$ is the energy resolution (FWHM), in units of percentage of $Q_{\beta\beta}$. "BG" is the background rate, in units of counts/(keV kg yr). "Topology" refers to the detector's capability at topology discrimination.

| | $^{48}$Ca | $^{76}$Ge | $^{82}$Se | $^{96}$Zr | $^{100}$Mo | $^{110}$Pd | $^{116}$Cd | $^{124}$Sn | $^{130}$Te | $^{136}$Xe | $^{150}$Nd |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bolometer** | | | | | *AMoRE-Pilot* | | | TIN.TIN | CUORE | | |
| **Scintillating Bolometer** | | | | *CUPID-0* | *CUPID-Mo* *LUMINEU* | | | | | | |
| **Semicond.** | | GERDA-II *LEGEND-200* MJD | | | | | *COBRA* | | | | |
| **Organic Scintillator** | | | | ZICOS | | | | | KamLAND-Zen 800 SNO+ | | |
| **Inorg. Scint.** | CANDLES-III | | | | | | *AURORA* | | | *XMASS-1.5* | |
| **Track Cal.** | | | | | SuperNEMO DCBA-T3 | | | | | | |
| **HPXe TPC** | | | | | | | | | | NEXT-100 *PandaX-III* | |
| **LXe TPC** | | | | | | | | | | nEXO | |
| **2-phase TPC** | | | | | | | | | | LZ XENONnT | |

TABLE 3.3: Detection technologies and isotopes used by recently completed (until 2017), on-going and planned $0\nu\beta\beta$ experiments. The SuperNEMO and DCBA-T3 tracking calorimeters can be used with any isotope, so the entire row was marked in orange. Prototype and proof-of-concept experiments in italics.

### 3.2.3 Background sources and reduction methods

Background counts occur in a $0\nu\beta\beta$ detector because there are non-$0\nu\beta\beta$ events that can occasionally interact with its sensitive medium in a manner that is indistinguishable from $0\nu\beta\beta$. Background contributions can originate from outside the detector (*external background sources*), or from within the detector itself (*internal background sources*), either from the sensitive medium or from the detector components. Background sources typically undergo some event that can release one or two particles that together deposit $\sim Q_{\beta\beta}$ of energy into the sensitive medium, and so background reduction typically consists of either preventing the event from occurring; keeping the particle from interacting with the sensitive medium; or using characteristics of the produced signal to discard the interaction. This Section first describes the physics of how particles can generate background counts, then which events can produce background count-generating particles, followed by a descripiton of the background sources, and finally background reduction methods are addressed. Because of the scope of this work, the focus of this Section is on the case of two-phase Xe TPCs, though a significant part of the presented information is applicable to the other detector technologies as well. For a more general description, see [43].

#### Background count-generating particles

The short time resolution of typical $0\nu\beta\beta$ detector electronics (Table 4.1) makes it unlikely for an accumulation of low Q-value events to be read as a single $\sim Q_{\beta\beta}$ energy pulse, so typically only the single events with Q-value $\gtrsim Q_{\beta\beta}$ can directly cause background counts.

Interactions of light particles ($e^-, \gamma, \nu$) with the sensitive volume constitute a significantly larger contribution to the background rate than interactions with heavy particles ($n$, $\alpha$). This is because heavy particles depositing sufficient energy will almost always interact with the Xe nucleus, causing a *nuclear recoil* (NR), whereas light particles interact with an electron of the Xe atom, causing an *electron recoil* (ER), and these two types of recoil can be discriminated with high efficiency due to their difference in $N_{\text{ex}}/N_{\text{i}}$ values (from Eq. 3.4; in LXe, $N_{\text{ex}}/N_{\text{i}} \sim 1$ for NR [99], and $N_{\text{ex}}/N_{\text{i}} \sim 0.1$ - 0.2 for ER [101]), which in two-phase Xe TPCs results in distinct

S1/S2 ratios. Figure 3.23 shows the difference in S1 and S2 signal amplitudes for recoils due to $\gamma$-rays and due to neutrons: it can be seen that the recoils due to the two types of particle occupy distinct regions of S1/S2 ratios, with negligible overlap.

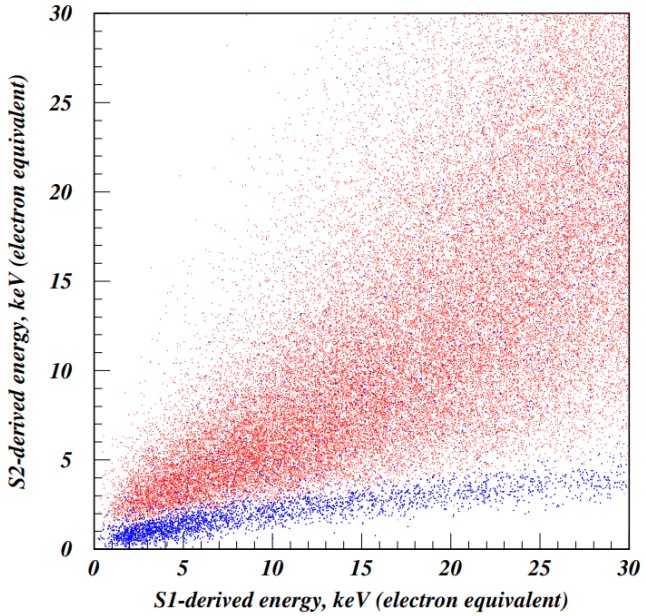


FIGURE 3.23: Discrimination of light and heavy particles in a LXe TPC with 3.9 kV/cm drift field. Scatter plot of S1 and S2 signal amplitudes for Compton-scattered $^{137}$Cs $\gamma$-rays (red dots) and for recoils due to elastic scattering with Am-Be neutrons (blue dots).
Image source: [99]

Background count-generating particles are therefore typically singular, light and very energetic: either single electrons with energy $\sim Q_{\beta\beta}$; single $\gamma$ particles with energy $\gtrsim Q_{\beta\beta}$ **{fn3.2i}**; or single neutrinos with energy $\gtrsim Q_{\beta\beta}$. Exceptions include groups of promptly released light particles with sum energy $\gtrsim Q_{\beta\beta}$, namely $2\nu\beta\beta$ events, which release two electrons similarly to $0\nu\beta\beta$; $^{60}$Co decay into $^{60}$Ni, where a 0.31 MeV electron and two $\gamma$ particles, of energies 1.17 MeV and a 1.33 MeV, are released in immediate succession (branching ratio of the decay chain is 99.88%, and the combined half-life of the two excited states is ~4.4 ps [135], whereas readout times are typically in the ns — $\mu$s scale); and neutrino capture on $^{136}$Xe (the charged-current interaction $\nu + {}^{136}\text{Xe} \rightarrow e^- + {}^{136}\text{Cs}^*$), with the release of an electron and of usually 3 de-excitation $\gamma$ particles [123].

### *Single electrons*

The $\sim Q_{\beta\beta}$ energy single electrons interact with the medium similarly to the $0\nu\beta\beta$ electrons, though the higher electron energy implies that the mean interaction length for ionization is larger and the mean interaction length for bremsstrahlung is smaller, so that the topology produced by a $\sim Q_{\beta\beta}$ single electron deposition in the medium typically has a longer track than the topology produced by a $0\nu\beta\beta$ event, with a higher likelihood of a bremsstrahlung event (for more information, see Section 4.2 and Chapter 6).

**{fn3.2i}** — In LXe, for an electron to be released with $Q_{\beta\beta}$ by Compton scattering, the incident $\gamma$ must have an energy of at least $1.09 \times Q_{\beta\beta}$ =2.692 MeV.

### $\gamma$ particles

The dominant photon interaction in LXe, for energies $\mathcal{O}(Q_{\beta\beta})$, is Compton scattering, with a mean interaction length of $\sim$10 cm (see Figure 3.24). By the Klein-Nishina formula, photons with energy $\gtrsim Q_{\beta\beta}$ are $\gtrsim$ 20 times more likely to Compton-scatter with a negligible change in direction (imparting a small energy on the target electron) than to reverse direction (imparting an energy $\sim Q_{\beta\beta}$ on the target electron). This means that most usually $\gamma$ particles can be rejected by the presence of several faraway deposition sites, each with deposited energy $\ll Q_{\beta\beta}$. However a small fraction of incident $\gamma$ particles will instead interact with the sensitive medium only a single time via photoelectric effect (interaction length $\gtrsim$ 100 cm), releasing a $\sim Q_{\beta\beta}$ energy single electron into the sensitive medium, which may then be mistaken for a $0\nu\beta\beta$ event.

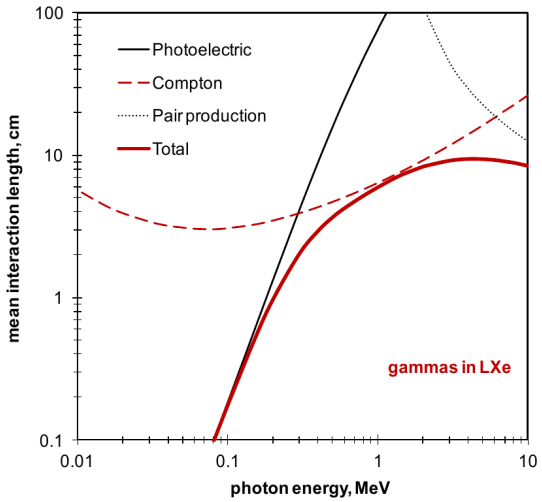


FIGURE 3.24: Mean interaction lengths for photons of varying energies in LXe due to photoelectric effect, Compton scattering and pair production.
Image source: [109]

### $^8B$ solar neutrinos

Despite their low interaction cross-section, their large flux means that neutrinos can constitute a significant background in large-scale $0\nu\beta\beta$ detectors, like TPCs and organic scintillators. Considering the relevant energy range, it can be seen that $^8$B solar neutrinos constitute by far the most significant neutrino background contribution (see Figure 3.25)**{fn3.2j}**. Beyond $\sim$10 MeV the neutrino flux drops considerably, so that the critical neutrino energy range is $\sim$1 - 10 MeV. Figure 3.26 shows the cross-sections for different interactions with low-energy neutrinos, showing that the most important are neutrino capture on xenon ($^A\text{Xe} + \nu \rightarrow {}^A\text{Cs}^* + e^-$) and electron-neutrino elastic scattering ($e^- + \nu \rightarrow e^- + \nu$) **{fn3.2k}**.

**{fn3.2j}** — Note that Figure 3.25 shows that the terrestrial antineutrino flux is comparable to the flux of $^8$B solar neutrinos. Footnote **{fn3.2k}** shows that charged-current antineutrino interactions (namely IBD from Figure 3.26) produce a signal that will almost always be rejected. Experimental data also suggests that neutral-current electron antineutrino interactions have a significantly smaller cross-section than do neutral-current electron neutrino interactions [136]. For these reasons, the contribution due to terrestrial antineutrinos was not considered.

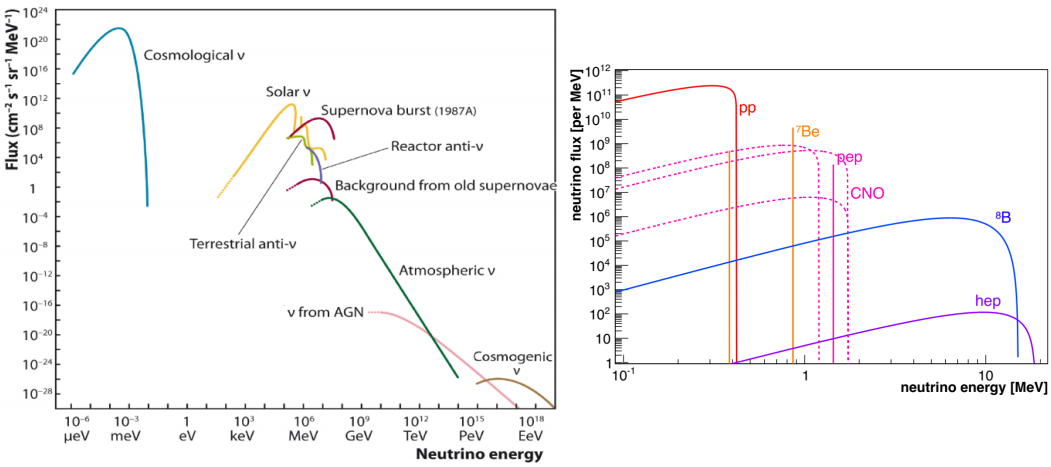


FIGURE 3.25: *Left panel:* Measured and expected fluxes from different contributions. *Right panel:* Contributions to solar neutrino flux. Note that only $^8$B $\beta^+$-decay neutrinos and hep neutrinos (from the interaction $^3\text{He} + p^+ \rightarrow {}^4\text{He} + e^+ + \nu$) achieve $\gtrsim Q_{\beta\beta}$ energy, and the $^8$B flux is $\sim 10^2$ times higher than the hep flux.
Image sources: [137] and [138]

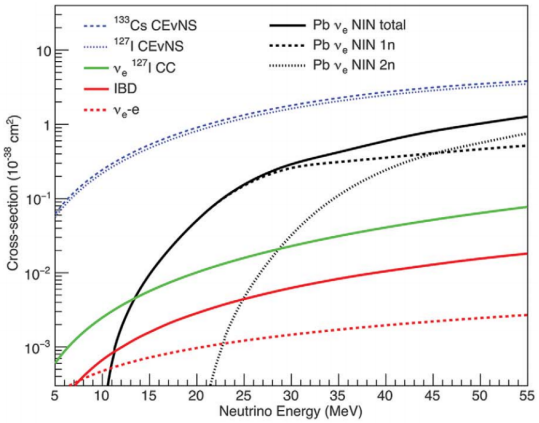


FIGURE 3.26: Cross-sections of low energy neutrino interactions. "CEvNS" corresponds to coherent elastic neutrino-nucleus scattering ($^A_Z\text{X} + \nu \rightarrow {}^A_Z\text{X} + \nu$). Its cross-section scales with $(A - Z)^2$, so the cross-section for Xe will be between those for $^{127}$I and for $^{133}$Cs. "NIN" corresponds to neutrino-induced neutrons. NIN has two variants: a neutral-current variant, $^A_Z\text{X} + \nu \rightarrow {}^A_Z\text{X}^* + \nu$, followed by either $^A_Z\text{X}^* \rightarrow {}^{A-1}_{Z}\text{X} + n$ or $^A_Z\text{X}^* \rightarrow {}^{A-2}_{Z}\text{X} + 2n$; and a charged-current variant, $^A_Z\text{X} + \nu \rightarrow {}^A_{Z+1}\text{X}^* + e^- + \nu$, followed by either $^A_{Z+1}\text{X}^* \rightarrow {}^{A-1}_{Z+1}\text{X} + n$ or $^A_{Z+1}\text{X}^* \rightarrow {}^{A-2}_{Z+1}\text{X} + 2n$. NIN is due to an interaction that affects the nuclear structure of the atom, so the $N$-dependence of the cross-section can be expected to be smaller than in CEvNS, and so the Xe NIN cross-section can be expected to be comparable to the Pb NIN cross-section. "$\nu$ $^{127}$I CC" refers to neutrino capture on $^{127}$I ($^{127}\text{I} + \nu \rightarrow {}^{127}\text{Xe}^* + e^-$). How this cross-section compares to the cross-section of neutrino capture on $^{136}$Xe will be discussed further. "IBD" refers to inverse beta decay ($^A_Z\text{X} + \overline{\nu} \rightarrow {}^A_{ZS-1}\text{X}^* + e^+$). "$\nu$-e" refers to neutrino-electron elastic scattering ($\nu + e^- \rightarrow \nu + e^-$).
Image source: [139]

Figure 3.27 shows a simplified energy level diagram for neutrino capture, showing the processes with the largest background contribution in Xe: electron emission and prompt de-excitation photon emission; and possible subsequent $\beta$ decay to the Cs atom. The interaction cross-section can vary in an irregular manner by 2-3 orders of magnitude between nuclides of similar $A$, but little information is available on the cross-section for natural xenon. For pure $^{136}$Xe, the neutrino capture interaction rate is estimated to be $\sim 3.4 \times 10^{-5}$counts/(keV kg yr) [140] The cross-section in neutrino-electron elastic scattering more consistent between isotopes: the estimated interaction rate for natural xenon is $1.70^{+0.19}_{-0.21} \times 10^{-7}$counts/(keV kg yr), and for 90% $^{136}$Xe-enriched xenon it is $1.65^{+0.19}_{-0.20} \times 10^{-7}$ [141]. Figure 3.28 shows the estimated energy spectra for $^{136}$Cs de-excitation and $\beta$ decay following neutrino capture, and for neutrino-electron elastic scattering (from all solar neutrinos).

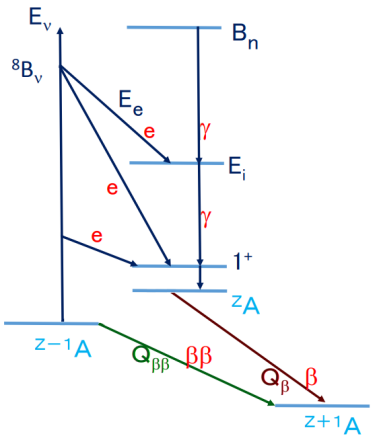


FIGURE 3.27: Energy level diagram for neutrino capture, showing the charged-current electron release and the subsequent nuclear de-excitation, as well as the $\beta$ decay of the resulting atom.
Image source: [142]

**{fn3.2k}** — CEvNS (Figure 3.26) deposits only a few keV in the sensitive volume, and its interaction resembles that of heavy particles [143] so the interaction can easily be rejected. As the incident neutrino's energy increases, the neutrino's de Broglie wavelength becomes small enough for it to affect a target atom's nuclear structure through an interaction, exciting the nucleus in the process and causing particle emission upon nuclear de-excitation ($n$, $p$, $\gamma$ or $\alpha$ particles). NINs are a product of this type of interaction. As shown in Figure 3.26, NINs only become important for energies $>$ 10 MeV, and since neutrino-induced nuclear excitation is primarily dependent on the neutrino's de Broglie wavelength, it can be expected that the release of other de-excitation particles will only become important for energies $>$ 10 MeV as well. Figure 3.25 shows that the neutrino flux cuts off for energy $\gtrsim$ 10 MeV, so neutrino-induced nuclear de-excitation events (including NINs) will not constitute a very significant background source, and therefore they can be discarded as well. The positron released in IBD will annihilate with the sensitive medium, releasing two annihilation $\gamma$ particles. By the Klein-Nishina formula, each annihilation $\gamma$ particle will be $\gtrsim$5 times more likely to produce more than one scatter site than to produce a single one, and their mean interaction length in LXe will be $\gtrsim$ 3 cm, making it extremely unlikely for an IBD event to be mistaken for $0\nu\beta\beta$, meaning that this interaction can also be discarded, leaving only neutrino capture on $^{136}$Xe and electron-neutrino elastic scattering as significant background contributions.

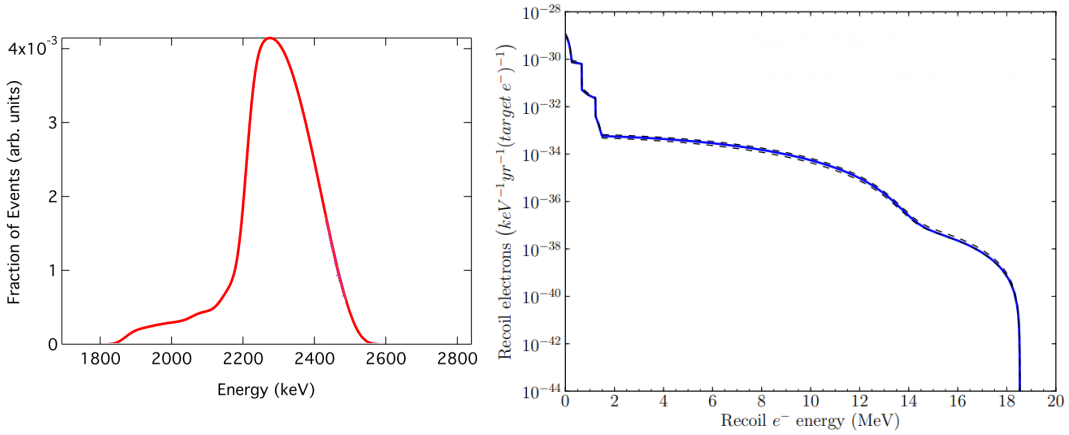


FIGURE 3.28: Energy spectra for charged-current neutrino capture on $^{136}$Xe and neutrino-electron elastic scattering. *Left panel:* Approximate energy spectrum for $^{136}$Cs de-excitation and $\beta$ decay following neutrino capture. *Right panel:* Energy spectrum for elastically-scattered electrons due to solar neutrinos (all contributions, not just $^{8}$B).
Image source: [140] (left) and [141] (right)

### Events producing background count-generating particles

In general, the types of events to consider are the ones that can cause a single-site deposition of an energy $\gtrsim Q_{\beta\beta}$ in the sensitive medium, either directly or indirectly, by way of daughter interactions. In the case of LXe and two-phase TPCs, the *direct deposition* must occur specifically through incident light particles, namely electrons and $\gamma$ particles (neutrinos will not be considered here due to their low cross-section). There are several types of events to account for, but they can be organized into two large groups, based on whether they themselves emit light particles; or whether they produce particles that can cause direct deposition events (termed here as *indirect deposition* events). This subsection lists the different types of event that can contribute to the background, and the subsection on background sources will present concrete examples.

#### *Direct deposition events — single electron emission*

Given the very low electron interaction length, single electron-emitting direct deposition events are typically only considered background sources if they occur inside the sensitive medium [43] [39] [123]. The processes typically considered in background models are listed in the box below:

**Naked $\beta$ decays ($\beta$ decays with no de-excitation $\gamma$)**
$^{A}_{Z}\mathrm{X}^{\mathrm{g.s.}} \rightarrow {}^{A}_{Z+1}\mathrm{X}^{\mathrm{g.s.}} + e^- + \overline{\nu}$
*(where $^{\mathrm{g.s.}}$ indicates that the nucleus is in the ground state)*

**Compton scattering**
$\gamma + e^- \rightarrow \gamma + e^-$

**Neutrino-electron elastic scattering**
$\nu + e^- \rightarrow \nu + e^-$

### *Direct deposition events — single $\gamma$ emission*

The large interaction length of $\gamma$ particles of energies $\gtrsim Q_{\beta\beta}$ (see Figure 3.24) allows single $\gamma$-emitting nuclides to be a background source whether they are inside or outside the sensitive medium [43] [39] [123]. The emission of single $\gamma$ particles is typically caused by nuclear de-excitation following particle capture (${}^{A}_{Z}\mathrm{X}^{*} \rightarrow {}^{A}_{Z}\mathrm{X}^{\mathrm{g.s.}} + \gamma$). The capture processes typically considered in background models are muon capture, $\alpha$ capture, neutron capture. The capture processes will be discussed further in the indirect deposition event paragraphs.

Another single $\gamma$-emitting process that can constitute a significant background contribution is radiative $\alpha$ capture ($\alpha$ capture with prompt $\gamma$ emission: $\alpha + {}^{A}_{Z}\mathrm{X} \rightarrow {}^{A+4}_{Z+2}\mathrm{X} + \gamma$, commonly referred to as the $(\alpha, \gamma)$ reaction). The $\gamma$ particles yielded by this reaction in typical rock can have a high energy (with a peak flux in the 4-10 MeV range, see Figure 3.29), and consequently a high penetrative power.

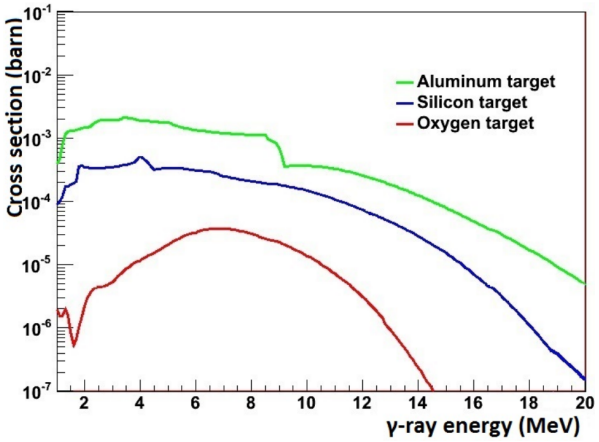


FIGURE 3.29: Calculated cross-sections of the $(\alpha, \gamma)$ reaction for varying $\gamma$ energies and with 4 MeV incident $\alpha$ particles. The three targets are elements present in large quantities in rock.
Image source: [144]

### *Direct deposition events — multiple $e^{-}/\gamma$ emission*

Some interactions result in the emission of more than a single light particle. There are three cases with different implications: multiple electron emission; multiple $\gamma$ emission; and mixed emission (some number of both electrons and $\gamma$ particles). For the same reason as in the case of single electron emission, multiple electron emission events are only a background contribution if they occur inside the sensitive medium. Multiple $\gamma$ emission and mixed emission events can contribute to the background provided that they either still produce only a single deposition site, or that multiple deposition sites are all sufficiently close together to be impossible to discriminate by the instrumentation. As such, a multiple $\gamma$ emission event can contribute to the background if it occurs inside the sensitive medium and yields a large number of low energy $\gamma$ particles, or if it occurs outside the sensitive medium and yields at least one $\gamma$ particle of energy $\gtrsim Q_{\beta\beta}$. Mixed emission events from outside the sensitive medium with $\gamma$ particles of energy $\gtrsim Q_{\beta\beta}$ are an important background contribution as well, whereas the same type of events occurring inside the sensitive medium can usually be rejected with high efficiency [39].

The multiple emission processes typically considered in background models are listed below:

**$2\nu\beta\beta$ events**

$$^{A}_{Z}\mathrm{X}^{\mathrm{g.s.}} \rightarrow {}^{A}_{Z+2}\mathrm{X}^{*} + 2e^{-} + 2\overline{\nu}$$

**Non-naked $\beta$ decays**

$$^{A}_{Z}\mathrm{X}^{\mathrm{g.s.}} \rightarrow {}^{A}_{Z+1}\mathrm{X}^{*} + e^{-} + \overline{\nu}$$

*followed by:*

$$^{A}_{Z+1}\mathrm{X}^{*} \rightarrow {}^{A}_{Z+1}\mathrm{X}^{\mathrm{g.s.}} + x \times \gamma \quad , \quad x = 0, 1, 2, \ldots$$

**Neutrino capture**

$$\nu + {}^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z+1}\mathrm{X}^{*} + e^{-}$$

*followed by:*

$$^{A}_{Z+1}\mathrm{X}^{*} \rightarrow {}^{A}_{Z+1}\mathrm{X}^{\mathrm{g.s.}} + x \times \gamma \quad , \quad x = 0, 1, 2, \ldots$$

***Indirect deposition events — activation by heavy particles***

Although heavy particle energy depositions are rejected by the sensitive medium with high efficiency (see Figure 3.23), heavy particles can still constitute a significant background contribution indirectly, by activating materials and thus causing subsequent direct deposition events. The particles of concern in this category are alphas and neutrons. Alphas were already emphasized in reference to the $(\alpha, \gamma)$ direct deposition event, however they are also a significant indirect background contributor through the $(\alpha, n)$ reaction ($\alpha$ capture with prompt neutron release: $\alpha + {}^{A}_{Z}\mathrm{X} \rightarrow {}^{A+3}_{Z+2}\mathrm{X}^{*} + n$). Neutrons have a high penetrating power, traveling an average distance in the $\mathcal{O}(1\text{ m})$ range through concrete before nuclear capture **{fn3.2l}**, and result in de-excitation $\gamma$ emission and in the production of radioactive nuclides.

***Indirect deposition events — activation by muons***

Cosmogenic muons typically originate from the decay of charged mesons produced in interactions of cosmic rays with the air. The average muon energy and flux at sea level is $\sim$4 GeV and $\approx 70\,\mathrm{m^{-2}s^{-1}sr^{-1}}$ respectively [149]. Although incident muons can be efficiently rejected by the sensitive medium (Section 1.6 of [109]), they can significantly contribute to background by way of muon capture ($\mu^{-} + {}^{A}_{Z}X \rightarrow {}^{A}_{Z-x}X^{*} + \nu_{\mu} + x \times n$, $x = 0, 1, 2, \ldots$). This process can activate stable nuclides, typically yields neutrons in the $10 - 20$ MeV range, and some rock de-excitation $\gamma$ particles have energies $\gtrsim Q_{\beta\beta}$ [150].

**Background sources for LXe and two-phase TPCs**

It can be seen that, in an above-ground experiment, background counts caused by activation of materials via cosmogenic muon capture would mask any actual $0\nu\beta\beta$ counts obtained in an LXe TPC. To counteract this, TPCs are typically constructed deep underground. At the depths used, the muon flux is low enough (calculated as $5.357\,\mathrm{m^{-2}day^{-1}}$ for LZ, at a water equivalent depth of $\sim 4.3\times10^{3}$ m.w.e. [109]; and projected to be $(0.326 \pm 0.035)\,\mathrm{m^{-2}day^{-1}}$ for nEXO, at a

**{fn3.2l}** — There are two important sources of neutrons to consider: cosmogenic neutrons due to cosmogenic muon capture (discussed in the next paragraph) and $(\alpha, n)$ reaction neutrons. It is predicted that neutrons from both sources will typically have energies in the $\sim$10 MeV range [145] [146]. Interacting with matter, fast neutrons typically lose energy to collisions with nucleons until they thermalize, at which point they are quickly captured by a nucleus. At this energy range, they are expected to have a mean free path of 5-10 cm [147] in concrete and to undergo an $\sim$15 collisions before absorption [148], hence the $\mathcal{O}(1\text{ m})$ range value.

water equivalent depth of $\sim 6 \times 10^3$ m.w.e. [123]) that muon products no longer constitute the principal background contribution. Instead, most background counts occur due to radioactive decays in the surrounding rock and in components of the detectors themselves. The majority of these decays can be attributed either directly or indirectly to the natural decay chains of $^{238}$U and $^{232}$Th: trace amounts of these isotopes are present in all materials [3]. Other contributions include $2\nu\beta\beta$ decay, long-lived radionuclides from outside the uranium and thorium decay chains, namely $^{60}$Co, and solar neutrinos. It should be noted that the majority of background counts are due to single site, single $\gamma$ Compton interaction events (from now on referred to as *1e events*, due to their production of a single electron with energy $\sim Q_{\beta\beta}$ in the sensitive volume). Table 3.4 shows the principal contributions in order of typical fraction of the total background counts (the rates of incidence may compare differently). The background rate for these contributions in LZ is listed in Section 4.2, and projected background rates for nEXO are provided in [123]. There is less information available on the projected $0\nu\beta\beta$ background rates for XENONnT.

| Background source | Event type | Process | Energy | Location | Parent Event |
|---|---|---|---|---|---|
| **$^{214}$Bi ($^{238}$U chain)** | Single $\gamma$ | De-excitation | 99.5 %$Q_{\beta\beta}$ | Ext. → D.C.<br>Ext. → Rock | $\beta$ decay of $^{214}$Pb |
| | Mixed $e^-/\gamma$ | $\beta$ decay | 99.5 %$Q_{\beta\beta}$ | Int. → Cath. | LXe contamination by $^{222}$Rn |
| **$^{208}$Tl($^{232}$Th chain)** | Single $\gamma$ | De-excitation | 106.3 %$Q_{\beta\beta}$ | Ext. → D.C.<br>Ext. → Rock | $\alpha$ decay of $^{212}$Bi |
| **$^{137}$Xe** | Single $e^-$ | Naked $\beta$ decay | 169.7 %$Q_{\beta\beta}$ | Int. → LXe | Muon capture by $^{136}$Xe<br>Neutron capture by $^{136}$Xe |
| **$^{8}$B solar neutrinos** | Mixed $e^-/\gamma$<br>Single $e^-$ | Capture<br>Elastic scattering | 103.6 %$Q_{\beta\beta}$<br>566.2 %$Q_{\beta\beta}$ | Int. → LXe | N/A |
| **$2\nu\beta\beta$** | Multiple $e^-$ | Double $\beta$ decay | ~100 %$Q_{\beta\beta}$ | Int. → LXe | N/A |
| **$^{60}$Co** | Mixed $e^-/\gamma$ | $\beta$ decay | 101.9 %$Q_{\beta\beta}$ | Ext. → D.C. | Neutron capture by $^{59}$Co |
| **Si and O** | Single $\gamma$ | $(\alpha, \gamma)$ reaction | 357.7 %$Q_{\beta\beta}$ | Ext. → Rock | $\alpha$ decay of rock |

TABLE 3.4: LXe and 2-phase TPC background sources. The "Energy" column indicates the maximum energy that can be deposited in the medium by the given background source (namely, for the case of $^{60}$Co, the $\beta$ decay electron energy was ignored, as the event occurs outside the sensitive medium). In the "Location" column, "Int." and "Ext." mean "Internal" and "External" respectively, "D.C." stands for "Detector Components", referring to components that are not in direct contact with the LXe, and "Cath." stands for "Cathode": some decay products of $^{222}$Rn are positively charged, and so they accumulate at the cathode. "Rock" refers to the underground rock surrounding the detector. The sources highlighted in green provide a minor contribution due to efficient background reduction. For a detailed description of the sources, see [39], [43] and [123].

### Background reduction methods

There are four principal approaches to background reduction, which are typically employed in conjunction: detector material *purification*; detector *shielding*; signal *vetoing*; and signal *rejection*. Detector material purification methods reduce the concentration of radioactive nuclides present in the detector components. Shielding methods consist of placing a barrier separating the detector from the exterior, preventing cosmogenic or rock background sources from inciding on the sensitive medium. In the case of vetoing and rejection, the background event successfully incides on the detector medium: vetoing consists of discarding events based on the position and timing of deposition events; and rejection consists of discarding events based on the pulse shapes captured by the PMT arrays. Table 3.5 lists the background reduction methods typically employed in LXe and 2-phase TPCs. For the implementation in LZ specifically, see Chapter 4.

| Method | Type | Backgrounds | Description / Comment |
|---|---|---|---|
| **Radiopure material usage** | Purification | From detector components | *Fabrication of detector components materials as radiopure as possible.* |
| **Store materials underground** | Purification | Cosmogenic capture daughters | *Store detector materials underground before construction, allow capture radioactivity to end.* |
| **Radon trapping** | Purification | $^{222}$Rn daughters inside LXe | *Xenon is purified on-line with a recirculation system. Activated charcoal in this system can be used to remove radon from the xenon.* |
| **LXe self-shielding** | Shielding | External $\gamma$'s | *Outer LXe layers shield inner LXe (Fig. 3.24).* |
| **Passive shielding layers** | Shielding | External sources | *Examples: steel plates; water; refrigerant, Pb. Inner layers more radiopure than outer ones.* |
| **Outer detector** | Veto & Shielding | Muons (*veto*); Neutrons (*shield*); Rock $\gamma$'s (*v. & s.*). | *Thick water + liquid scintillator layers instrumented with PMTs. Muons yield Cherenkov radiation. Neutrons are moderated.* |
| **Fiducialization *(and skin veto)*** | Veto | External and internal sources | *Define an inner volume of LXe as the fiducial volume (FV). Simultaneous energy depositions in and outside the FV are discarded.* |
| **Coincident event discrimination** | Veto | Internal radiation chains | *Events in quick succession are discarded (e.g. $^{214}$Bi $\gamma$ followed by $^{214}$Po $\alpha$).* |
| **Multisite event discrimination** | Veto | Single- and multi-$\gamma$ events | *If two or more deposition sites are sufficiently far away from each other, signal is discarded.* |
| **Energy Region of Interest (ROI)** | Rejection | All sources, but mainly $2\nu\beta\beta$ | *Signal energy window. If sum pulse energy is above or below window, signal is discarded.* |
| **S1/S2 ratio** | Rejection | Nuclear recoils | *Nuclear recoils have a higher S1/S2 ratio.* |
| **Daughter ion tagging** *(nEXO)* | Rejection | All sources except $2\nu\beta\beta$ | *$^{136}$Xe $\beta\beta$ decay yields stable $^{136}$Ba$^{+}$ daughter ion, identifiable via atomic spectroscopy.* |
| **Topology Discrimination** | Rejection | Single $\gamma$ events | *This work concerns this background rejection method.* |

TABLE 3.5: Table of background reduction methods in LXe and 2-phase TPCs for $0\nu\beta\beta$ search. Information compiled from [3], [43], [109], [123] and [39]. The xenon recirculation system referred in the "Radon trapping" entry also serves to remove electronegative impurities from the LXe, increasing the free electron lifetime (see pgs. 24.

### 3.2.4 Current state of the discipline

The currently ongoing generation of $0\nu\beta\beta$ experiments aims to probe the inverted-ordering Majorana neutrino mass, namely the $\langle m_{\beta\beta}\rangle \sim 0.01$ — 0.05 eV range, using low-background, tonne-scale experiments. Table 3.6 lists the $\beta\beta$-active masses, live-times and sensitivities of recently completed (until 2017), on-going and planned $0\nu\beta\beta$ experiments, listed in order of the lowest Majorana mass that they can probe. As shown in Tables 3.1, 3.2 and 3.3, different isotopes imply the use of distinct detection technologies, and Table 3.6 indicates that although no single technology or isotope is decisively superior, sufficiently sensitive detectors are more successfully being achieved using $^{136}$Xe, $^{130}$Te or $^{76}$Ge (Figure 3.30), operating a technology with very good scalability: TPCs, semiconductors and organic scintillators.

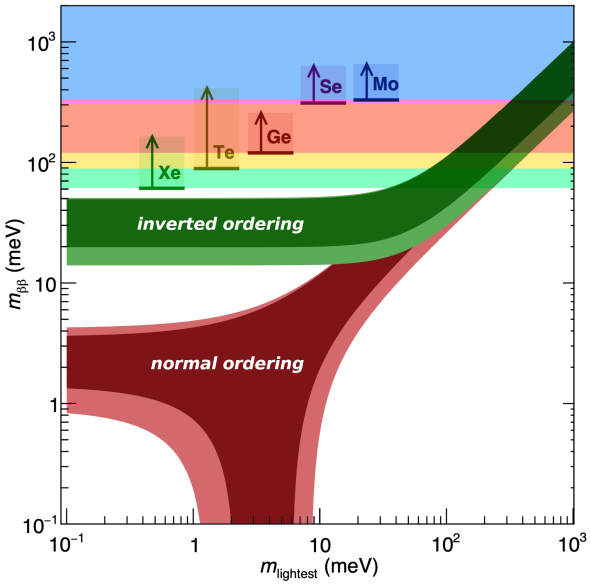


FIGURE 3.30: Currently attained sensitivity ranges for different isotopes (not counting projected sensitivities). As can be seen, $^{136}$Xe, $^{130}$Te or $^{76}$Ge have successfully attained superior sensitivities.
Image source: [151]

| Experiment | Technology | Isotope | $M_{\beta\beta}$ (kg) | $t$ (yr) | $T_{1/2}^{0\nu\beta\beta}$ (yr) | $\langle m_{\beta\beta}\rangle$ (eV) | $T_{1/2}^{0\nu\beta\beta}/\sqrt{M_{\beta\beta}}$ | Ref. | Result |
|---|---|---|---|---|---|---|---|---|---|
| nEXO | LXe TPC | $^{136}$Xe | 4500 | 10.0 | $> 9.2 \times 10^{27}$ | 0.0057 — 0.0177 | $> 1.4 \times 10^{26}$ | [125] | No |
| *LEGEND-200* | Semiconductor | $^{76}$Ge | 200 | 5.0 | $> 10^{27}$ | *0.035* — *0.112** | $> 7.1 \times 10^{25}$ | [70] | No |
| LZ | 2-phase TPC | $^{136}$Xe | 623 | 2.74 | $> 1.06 \times 10^{26}$ | 0.053 — 0.164 | $> 4.3 \times 10^{24}$ | [39] | No |
| KamLAND-Zen 400 | Organic Scint. | $^{136}$Xe | 291.2 | 2.04 | $> 1.07 \times 10^{26}$ | 0.061 — 0.165 | $> 6.3 \times 10^{24}$ | [83] | Yes |
| *PandaX-III* | HPXe TPC | $^{136}$Xe | 180 | 3.0 | $> 1 \times 10^{26}$ | 0.065 — 0.165 | $> 7.5 \times 10^{24}$ | [152] | No |
| SNO+ | Organic Scint. | $^{130}$Te | 800 | 5.0 | $> 9 \times 10^{25}$ | $< 0.067$ | $> 3.2 \times 10^{24}$ | [153] | No |
| CUORE | Bolometer | $^{130}$Te | 201.9 | 0.503 | $> 1.7 \times 10^{25}$ | 0.075 — 0.350 | $> 1.2 \times 10^{24}$ | [154] | Yes |
| NEXT-100 | HPXe TPC | $^{136}$Xe | 91 | 3.30 | $> 6 \times 10^{25}$ | 0.080 — 0.160 | $> 6.3 \times 10^{24}$ | [130] | No |
| GERDA-II | Semiconductor | $^{76}$Ge | 40 | 3.75 | $> 1.5 \times 10^{26}$ | 0.090 — 0.290 | $> 2.4 \times 10^{25}$ | [155] | No |
| ZICOS | Organic Scint. | $^{96}$Zr | 45 | 2.00 | $> 4 \times 10^{25}$ | *0.109* — *0.296** | $> 6.0 \times 10^{24}$ | [156] | No |
| *LUMINEU* | Scint. Bolometer | $^{100}$Mo | 0.676 | 5 | $> 5.3 \times 10^{24}$ | 0.167 — 0.476 | $> 6.3 \times 10^{24}$ | [157] | No |
| MJD | Semiconductor | $^{76}$Ge | 27.252 | 0.95 | $> 2.7 \times 10^{25}$ | 0.200 — 0.430 | $> 5.2 \times 10^{24}$ | [158] | Yes |
| SuperNEMO | Tracko-Calo | $^{82}$Se | 7 | 2.5 | $> 5.9 \times 10^{24}$ | *0.240* — *0.491** | $> 2.2 \times 10^{24}$ | [159] | Yes |
| *CUPID-Mo* | Scint. Bolometer | $^{100}$Mo | 2.264 | 0.96 | $> 1.5 \times 10^{24}$ | 0.3 — 0.5 | $> 1.5 \times 10^{24}$ | [151] | Yes |
| *CUPID-0* | Scint. Bolometer | $^{82}$Se | 4.65 | 1.14 | $> 3.5 \times 10^{24}$ | 0.311 — 0.638 | $> 1.6 \times 10^{24}$ | [160] | Yes |
| NEMO-3 | Tracko-Calo | $^{100}$Mo | 6.914 | 3.86 | $> 1.1 \times 10^{24}$ | 0.33 — 0.62 | $> 4.2 \times 10^{23}$ | [161] | Yes |
| NEMO-3 | Tracko-Calo | $^{130}$Te | 0.454 | 3.34 | $> 1.0 \times 10^{23}$ | *0.98* — *4.56** | $> 1.5 \times 10^{23}$ | [161] | Yes |
| *AURORA* | Inorg. Scint. | $^{116}$Cd | 0.305 | 5.0 | $> 2.2 \times 10^{23}$ | 1.0 — 1.7 | $> 4.0 \times 10^{23}$ | [87] | Yes |
| *AmoRE-Pilot* | Bolometer | $^{100}$Mo | 0.893 | 0.34 | $> 9.5 \times 10^{27}$ | 1.2 — 2.1 | $> 1.0 \times 10^{23}$ | [162] | Yes |
| NEMO-3 | Tracko-Calo | $^{150}$Nd | 0.037 | 2.54 | $> 1.8 \times 10^{22}$ | 1.6 — 5.3 | $> 9.4 \times 10^{22}$ | [161] | Yes |
| *COBRA* | Semiconductor | $^{116}$Cd | 0.01 | 1.70 | $> 1.1 \times 10^{21}$ | *14.1* — *24** | $> 1.1 \times 10^{22}$ | [76] | Yes |
| ELEGANT VI | Inorg. Scint. | $^{48}$Ca | 0.008 | 1.71 | $> 5.8 \times 10^{22}$ | 3.5 — 22 | $> 6.5 \times 10^{23}$ | [163] | Yes |
| NEMO-3 | Tracko-Calo | $^{116}$Cd | 0.405 | 0.21 | $> 1.6 \times 10^{22}$ | *3.7* — *6.3** | $> 2.5 \times 10^{22}$ | [161] | Yes |
| NEMO-3 | Tracko-Calo | $^{96}$Zr | 0.009 | 3.49 | $> 9.2 \times 10^{21}$ | 7.2 — 19.5 | $> 9.7 \times 10^{22}$ | [161] | Yes |
| NEMO-3 | Tracko-Calo | $^{48}$Ca | 0.007 | 2.58 | $> 1.3 \times 10^{22}$ | *7.4* — *46.5** | $> 1.6 \times 10^{23}$ | [161] | Yes |
| CANDLES-III | Inorg. Scint. | $^{48}$Ca | 0.57 | 0.17 | $> 0.8 \times 10^{22}$ | *9.4* — *59.2** | $> 1.1 \times 10^{22}$ | [164] | No |

TABLE 3.6: Table of 90% CL sensitivities of recently completed (until 2017), on-going and planned $0\nu\beta\beta$ experiments, arranged in ascending order of the achieved $\langle m_{\beta\beta}\rangle$ lower limit. The 10 lowest limits are highlighted in green. The "$T_{1/2}^{0\nu\beta\beta}/\sqrt{M_{\beta\beta}}$" column is the time sensitivity (Eq. 3.3) scaled against the active mass. The "Result" column indicates if the listed values are from a measurement or from a projected result. The italicized $\langle m_{\beta\beta}\rangle$ values with a blue asterisk were not provided in the reference, and were instead calculated based on the $k_{\beta\beta}^{m}$ and $k_{\beta\beta}^{T}$ values inferred from another experiment using the same isotope. NEMO-3 results were included due to the small amount of sensitivity data for certain isotopes.

# Chapter 4

# $0\nu\beta\beta$ Detection in LZ

This chapter is a continuation on the discussion of TPCs presented in Chapter 3, now focusing specifically on the LUX-ZEPLIN TPC. Section 4.1 summarizes the physics and specification relevant for $0\nu\beta\beta$ detection, and Section 4.2 discusses the background environment and hypothesis testing approach.

## 4.1 Detector physics and technical specifications

The LUX-ZEPLIN (LZ) experiment is a 2-phase, liquid-gas xenon time projection chamber (TPC) in the final stages of commissioning (Figure 4.3) at Davis cavern (at a depth of $\sim$1.5 km underground) of the Sanford Underground Research Facility in Lead, South Dakota, USA. Despite its primary focus being weakly interacting massive particle *(WIMP)* detection, it also has competitive capabilities for $0\nu\beta\beta$ search compared to current generation $0\nu\beta\beta$ experiments (see Table 3.6). It has a $\sim$7 tonne active detection mass of natural Xe (almost 623 kg of $^{136}$Xe), nearly 1 tonne of which constitutes a fiducial volume (see pg. 31) with background index of $\sim 5.4 \times 10^{-4}$ counts / (keV kg year) [39] (see Section 4.2 for more information). By combining both the S1 and S2 signals, LZ achieves an energy resolution at $Q_{\beta\beta}$ of $\sim 2.36\%$ (FWHM) [39]. Counting across the entire active volume over a 2.74 year live-time, it is expected to be capable of probing $\langle m_{\beta\beta} \rangle$ to within 53-164 meV [39].

Figure 4.1 shows a schematic of the LZ TPC. The gate grid, TPC field cage and cathode grid define the top, side and bottom of the cylindrical active region (or *drift region*), respectively. It has an equal height $d_{\text{max}}$ and diameter of 1456 mm and is filled with 7 tonnes of natural liquid xenon (LXe), kept at a temperature of 175.8 K. The gate grid, cathode grid and field cage produce and shape a vertically-oriented *drift field* of strength $E_{\text{drift}} = 310$ V/cm. 13 mm above the gate grid is the anode, placed inside the gaseous Xe at a distance above the LXe surface of 8 mm. Electroluminescence occurs in this thin gas volume, and it is referred to as the *S2 region*, or *gas gap*. The anode and gate produce an *extraction field* of strength $E_{\text{S2}} = 10$ kV/cm. Below the cathode is a 137.5 mm thick volume containing $\sim$1 tonne of LXe, where the electric field has the opposite direction and a field strength of $\sim$ 3-6 kV/cm (Section 3.2.1 from [109]), referred to as the *reverse-field region*. Above and below the three regions are readout planes: photomultiplier (PMT) arrays using Hamamatsu R11410-22 3-inch diameter low-radioactivity PMTs [39], with 253 units in the top array and 241 units in the bottom array. Outside the field cage and PMT array enclosure is an additional $\sim$ 2 tonnes of 2-phase xenon, forming a *Xe skin* of width 4-8 cm. This outer region is instrumented with PMTs as well (93 Hamamatsu R8520 1-inch PMTs), and serves to veto against multisite events. In total, the TPC has $\sim$10 tonnes of Xe. The field cage and PMT arrays are covered with polytetrafluoroethylene (PTFE) panels that serve to maximize the light collection efficiency by reflecting the scintillation light: immersed in LXe, they achieve a reflectivity of $\geq$ 97.3% [165]; while in gaseous Xe the reflectivity falls to 85% [165].

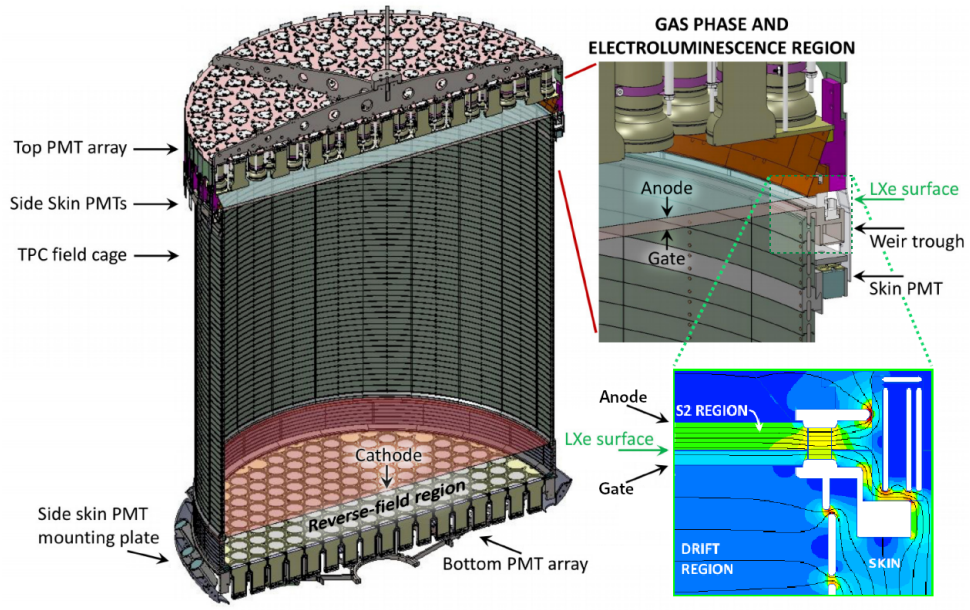


FIGURE 4.1: Schematic of the LZ TPC, with nested details of the electroluminescence region (gas gap). The subdetail shows an electrostatic modeling of the electric field strength near the S2 region, with low field strength in blue, and ranging through to red for high field strength. Technical details in main text.
Image source: [165] and [109]

Figure 4.2 shows a schematic of the TPC placed within the detector systems. The TPC is mounted inside a two-vessel vacuum-insulated cryostat, itself surrounded by an outer detector (OD), containing 17.3 tonnes of gadolinium-loaded liquid scintillator and instrumented with 120 Hamamatsu R5912 8-inch PMTs. The OD enclosure is placed inside a water tank containing ∼300 tonnes of ultra-pure water. For $0\nu\beta\beta$, the OD together with the water tank serve as an additional active veto against $\gamma$-ray scatters [39]. LXe is continuously circulated to a purification system outside the water tank: LXe flows out of the detector at the Weir trough (top right panel of Figure 4.1), and into the LXe tower (Section 6.4 of [109]). Two of the tasks of the LXe tower are to control the concentration of electronegative impurities and radon removal.

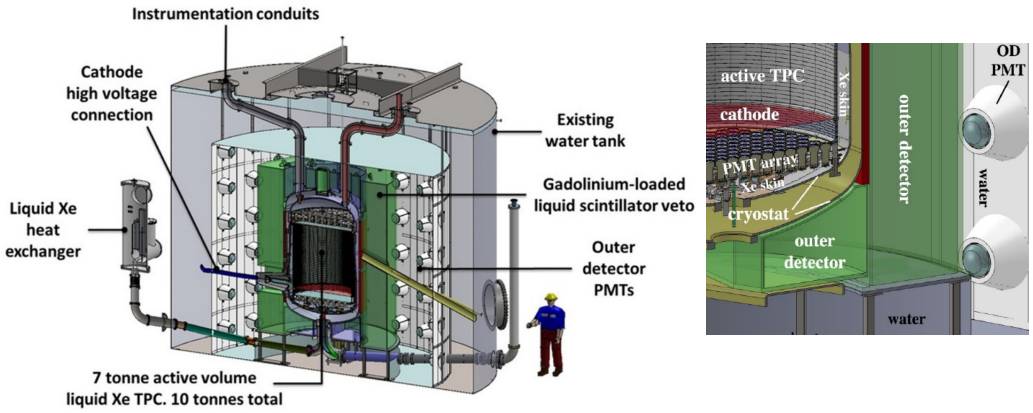


FIGURE 4.2: *Left panel:* Schematic of the LZ detector subsystems. The liquid Xe heat exchanger is part of the LXe tower (Section 6.4 of [109]). *Right panel:* Detail showing the inner and outer cryostat vessels.
Image source: [39] and [165]

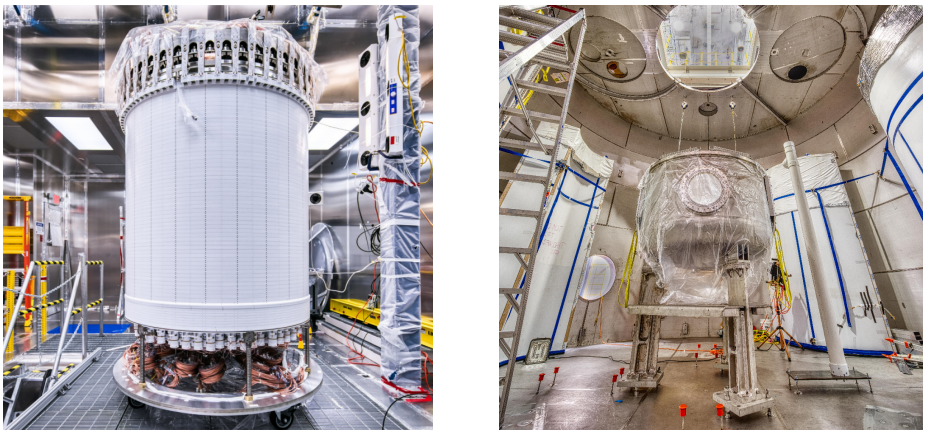

FIGURE 4.3: *Left panel:* The LZ TPC before being mounted into the cryostat (taken Apr. 2019). *Right panel:* Photo taken from inside the water tank, showing the bottom half of the outer cryostat vessel (taken May. 2019). Photo credit: Matthew Kapust, Sanford Underground Research Facility

The principle of operation of a TPC is described in Section 3.2.2 (pgs. 24 - 37). Summarily, the two $0\nu\beta\beta$ primary electrons deposit energy into the medium, in the form of scintillation (the S1 signal), ionizations and unretrievable phonons (pgs. 24 - 26). The ionization electrons create a topology (left panel in Figure 4.4) with average blob-to-blob distance $\sim$5-6 mm, which drifts upwards due to $E_{\text{drift}}$ at a constant average *drift velocity* $v_d$ (pgs. 28 - 30). For LZ, this velocity is $v_d \sim 2.0$ mm/$\mu$s (Table 3.3.1 of [109]). As they reach the LXe surface, 95% of the electrons are successfully extracted into the gas [39], where they are subjected to the much stronger extraction field $E_{\text{S2}}$, resulting in electroluminescence, also referred to as secondary scintillation (S2, pgs. 32 - 33). The S2 photons hit the top PMT array, resulting in a pulse output from each PMT. The collection of top-array time series PMT outputs for some event is referred to as a *waveform* (example pictured in the right panel of Figure 4.4). The PMT time series response for a single electroluminescence photon is pictured in Figure 4.5. Position reconstruction is performed along $z$ using the timing data, and along $xy$ based on statistical methods according to the distribution of PMT hits along the array [166]. The obtained spatial resolutions are $\sigma_z = 0.2$ mm and $\sigma_{xy} = 0.2$ - 0.5 mm, respectively [39], however it should be noted that the $xy$ resolution was estimated for the case of reconstruction by statistical methods of a single event position.

The energy deposited by an event, $E_{\text{dep}}$, is estimated in LZ using the total amplitudes of the S1 and S2 signal according to the following expression:

$$E_{\text{dep}} = W\left(S1/g_1 + S2/g_2\right), \tag{4.1}$$

where, $S1$ and $S2$ are in units of "photons detected" (phd), $W = 13.7$ eV (c.f. Eq. 3.7), and where $g_1$ and $g_2$ represent the average number of phd in each type of scintillation, with values $g_1 = 0.12$ phd/photon and $g_2 = 79.2$ phd/e$^-$ [39] [165]. This approach makes use of the anticorrelation of S1 and S2 signals (see pgs. 27 - 28) to improve energy resolution. The resolution is defined in this case as the estimated standard deviation $\sigma$ of the distribution of output $E_{\text{dep}}$ for a fixed event energy $E$ [39]. It is predicted that a resolution $(\sigma/E)$ at $E_{\text{dep}} \sim Q_{\beta\beta}$ of 0.88% is achievable, although sensitivity projections were calculated supposing the resolution is inferior, of 1%: an energy window, or *region of interest* (ROI), of $(\sigma/E) = \pm 1\%$, meaning that if 2433.3 keV $< E_{\text{dep}} <$ 2482.4 keV, then the event may be counted as a possible $0\nu\beta\beta$, and otherwise rejected [39].

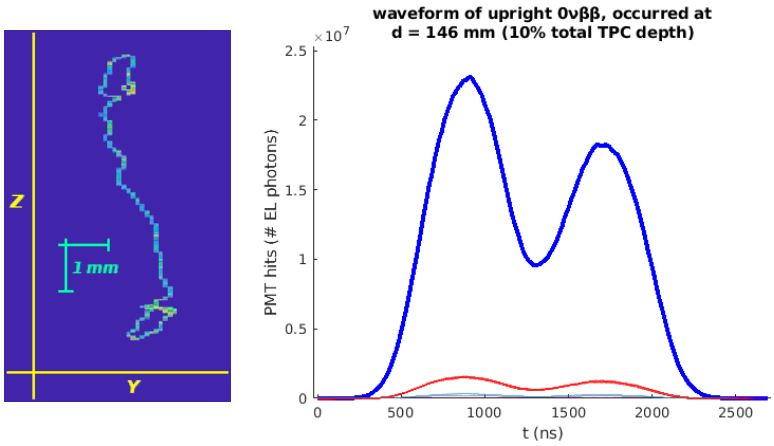


FIGURE 4.4: *Left panel:* Spatial electron density of a `Geant4` energy-deposition-only simulation of the topology of a $0\nu\beta\beta$ event in LXe (see Section 6.3, b2b scenario for simulation details). The yellow regions have a larger electron density. *Right panel*: Simulated PMT array output (*waveform*) (Section 6.3) due to a $0\nu\beta\beta$ event with the same topology as the event in the left panel, occurring 145 mm below the LXe surface (10% of the height of the active region) and directly above a PMT. The blue curve is the output from the PMT directly above the $0\nu\beta\beta$ event, the red curves (6 of them) are the outputs of the PMTs adjacent to the PMT directly above the $0\nu\beta\beta$ event, and the rest of the curves are from PMTs further away. The only distortion considered is due to diffusion.

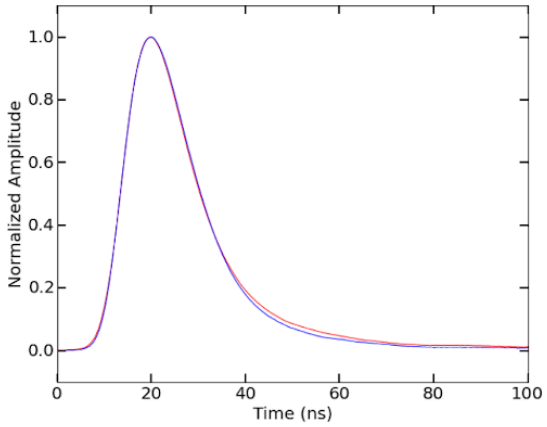


FIGURE 4.5: Time series output for a single electroluminescence photon on the top array PMTs in LZ.
Source: *internal communications*

There are three effects that will significantly distort a produced waveform. The most crucial effect is *diffusion*, described on the next page. The remaining two effects are *gate grid distortion* and *PMT saturation*. Gate grid distortion refers to the funnelling of the drift field lines toward the spaces between the spokes of the gate grid, resulting in a "quantization" of the $xy$ position of the ionization electrons along one of the planes, limiting the position reconstruction precision. It is pictured in the left panel of Figure 4.6. PMT saturation occurs because of the extremely large number of S2 photons produced in events of this energy (estimated in this work to be $\gtrsim 8 \times 10^7$ photons per event, see Section 6.3.2). Even accounting for the photon detection efficiency being $\sim$ 10%, it is expected that between 7 and 22 PMTs of the top array would

saturate [39], at which point the energy can only be reconstructed using the S2 hits in the non-saturated PMTs and in the bottom array, drastically reducing the value of $g_2$ in Equation 4.1 and thus reducing the energy resolution.

Diffusion (right panel of Figure 4.6), "smears out" the topology as the ionization electrons drift up toward the LXe surface and is inherent to the use of Xe as a sensitive medium (see pgs. 28 - 30). Equation 3.10 models the spatial smearing that occurs due to diffusion. The longitudinal and transverse diffusion coefficients, $D_L$ and $D_T$, are given in the LZ Technical Design Report (TDR) [109]. There, they were calculated by simulating the FWHM spread of cathode events (Table 3.3.1 [109]), listing the longitudinal and transverse values $\text{FWHM}_L^{\text{cath.}} = 2.0\ \mu\text{s} \approx 4.4$ mm and $\text{FWHM}_T^{\text{cath.}} = 1.4$ mm. To obtain $D_L$ and $D_T$ from these values, the following expressions can be used [167] [168]:

$$D_L = \frac{1}{2\ t^{\text{cath.}}} \times \left(\frac{\text{FWHM}_L^{\text{cath.}}}{2.355}\right)^2 \quad , \quad D_T = \frac{1}{t^{\text{cath.}}} \times \left(\frac{\text{FWHM}_T^{\text{cath.}}}{2.355}\right)^2 \tag{4.2}$$

where $t^{\text{cath.}} = d_{\text{max}}/v_{\text{d}} \approx 0.728$ ms is the drift time of cathode events. This results in the values $D_L = 1.98\,\text{mm}^2/\text{ms}$ and $D_T = 0.48\,\text{mm}^2/\text{ms}$. These were the coefficient values based on which the diffusion simulation was performed (Section 6.3.2). It was not noticed until after the simulations were completed that $D_L > D_T$, in conflict with measurements (Figure 3.15) and with theoretical predictions ($D_L \approx 0.1 D_T$ [167]). The value of $D_L$ is comparable to measured data, but $D_T$ is significantly smaller than expected. Fortunately, this does not affect the validity of the obtained results (Chapter 7), as due to the effect of gate grid distortion on the waveform, it was chosen to perform the analysis on the sum of all the PMT channels, leaving discrimination with $xy$ data for future work. This means that only the valid $D_L$ value affects the analyzed pulses.

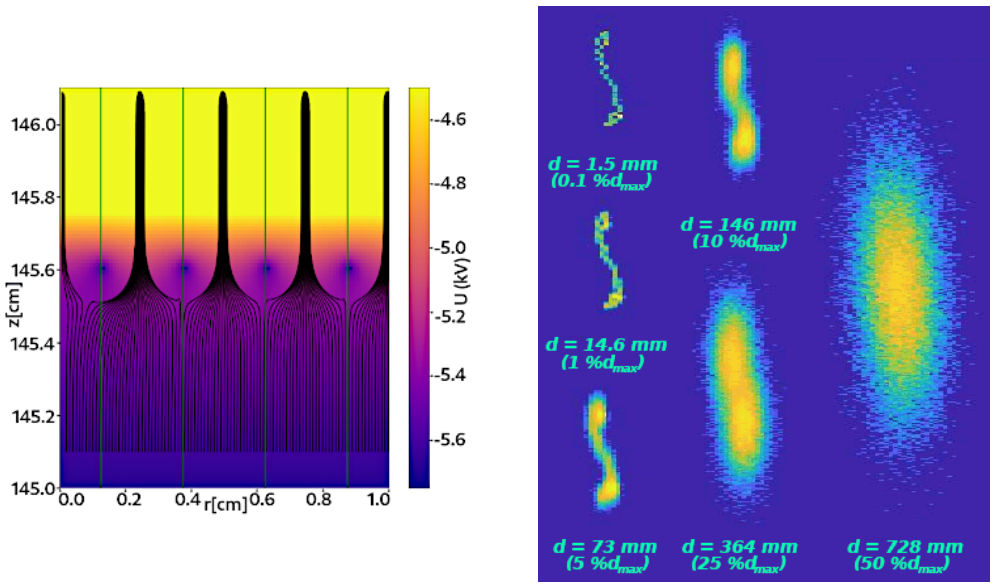


FIGURE 4.6: *Left panel:* Gate grid funneling the electron drift lines (in black) to specific bands. *Right panel:* Spatial electron density of the topology from Figure 4.4 captured at the LXe surface, assuming the $0\nu\beta\beta$ event occurred at depth $d$ and experienced diffusion during the drift as specified in the LZ Technical Design Report [109].
Right panel image source: [169]

The parameters of LZ relevant for this work are listed in Table 4.1.

| Parameter | Value | Unit | Description / Comment | Reference |
|---|---|---|---|---|
| $d_{\mathrm{max}}$ | 1456 | mm | *height of active region* | Table 1.2.1 of [109] |
| **diameter** | 1456 | mm | *diameter of active region* | Table 1.2.1 of [109] |
| **PMT diameter** | 64 | mm | *PMT cathode diameter* | Section 3.4 of [109] |
| **PMT $\leftrightarrow$ PMT** | 93 | mm | *top array PMT center-to-center spacing* | Section 3.5.3 of [109] |
| **gate grid depth** | 5.0 | mm | *depth of gate below LXe surface* | Section 3.6.1 of [109] |
| **gate grid pitch** | 5.0 | mm | *distance between spokes of the gate grid* | *internal communications* |
| **gas gap thickness** | 8.0 | mm | *height of anode above LXe surface* | Section 3.6.1 of [109] |
| $M_{0\nu\beta\beta}$ | 623 | kg | *mass of $^{136}$Xe in the active region* | [39] |
| $Q_{\beta\beta}$ | 2.459 | MeV | *total energy of $^{136}$Xe $0\nu\beta\beta$ primaries* | [43] |
| ***ROI*** | $\pm$1% | $Q_{\beta\beta}$ | *defined region of interest around $Q_{\beta\beta}$* | [39] |
| $\Delta E_{\mathrm{ROI}}$ | 58.75 | keV | *size of ROI energy window* | [39] |
| $\Delta POS_{\mathrm{xy}}$ | 0.2 - 0.5 | cm | *xy position resolution in energy ROI* | [39] |
| $\Delta POS_{\mathrm{z}}$ | 0.1 | $\mu$s | *PMT time resolution* | *internal communications* |
| $E_{\mathrm{drift}}$ | 310 | V/cm | *drift field* | [39] |
| $E_{\mathrm{S2}}$ | 10 | kV/cm | *extraction field* | Section 3.2.3 of [109] |
| $v_{\mathrm{d}}$ | $\sim$2.0 | mm/$\mu$s | *electron drift velocity* | Table 3.3.1 of [109] |
| $D_{\mathrm{L}}$ | 1.98 | mm$^2$/ms | *longitudinal diffusion coefficient* | Table 3.3.1 of [109] |
| $D_{\mathrm{T}}$ | 0.48 | mm$^2$/ms | *transverse diffusion coefficient (TDR)* | Table 3.3.1 of [109] |
| **photon yield** | 820 - 910 | eph/e$^-$ | *S2 photons per extracted electron* | Table 3.6.1 of [109] |
| $\varepsilon_{e^-}^{\mathrm{extract}}$ | 95% | – | *electron extraction efficiency* | [39] |
| **S2 PDE** | 10% | – | *S2 photon detection efficiency* | [165] |
| $REFL_{\mathrm{LXe}}$ | $\geq$ 97.3% | – | *PTFE reflectivity in LXe* | [165] |
| $REFL_{\mathrm{GXe}}$ | 85% | – | *PTFE reflectivity in GXe* | [165] |

TABLE 4.1: Table of LZ specifications relevant to the scope of this thesis.

## 4.2 LZ background environment and hypothesis testing

The background sources for 2-phase TPCs have been examined in detail in Section 3.2.3: Table 3.4 summarizes the information, with background contributors listed in order of importance. The low activity from the detector materials and LXe, together with the active vetos and efficient self-shielding of the LXe, allow to define two fiducial regions: a very clean 967 kg center region at $26 < z < 96$ cm with 78 cm diameter, termed the *fiducial volume* (FV) [39]; included inside a 5613 kg region with a higher average background at $2 < z < 132.6$ cm with 137.6 cm diameter [39], termed in this work the *extended fiducial volume* (XFV). The two regions are pictured in Figure 4.7: the left panel shows the radial and height distributions of background events after cuts due to the energy ROI, vetoing, S1/S2 ratio (see Table 3.5) and single scatter (SS, maximum accepted vertical separation between distinct deposition sites from the same event is set at 3 mm) [39]; and the right panel shows the integrated total background *counts* (events remaining after cuts) for a 1000 live day run [165], giving a value of 35.57 counts in the FV, and 6116 counts in the XFV.

A projection of the background rate and contributions inside the FV was performed in [39]: Figure 4.8 shows the background contributions after all cuts. Near the ROI, it can be seen that the cavern walls and detector components contribute most of the backgrounds that survived rejection, accounting together for 98.1% of the total $2.73 \times 10^{-4}$ counts/keV/kg/year background rate in the FV [39]. These backgrounds originate from the natural $^{232}$Th and $^{238}$U decay chains. Of particular importance in these chains are two $\gamma$-decays: the 2.615 MeV line of $^{208}$Tl, from the former chain; and the 2.448 MeV line of $^{214}$Bi, from the latter (see pgs. 45 - 46). The gammas from these decays that survive the cuts will most commonly interact via photoelectric effect, losing all their energy at once and releasing a single electron with energy

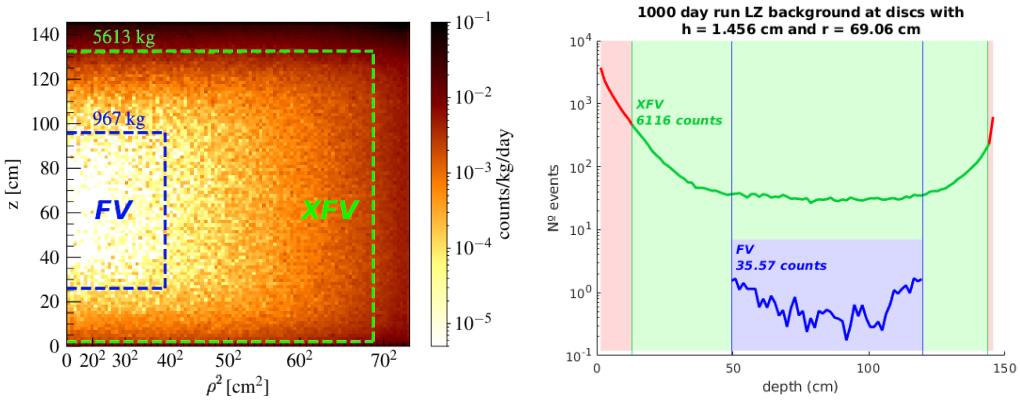


FIGURE 4.7: *Left panel:* Background counts/kg/day following cuts, as a function of the the vertical ($z$) and radial ($\rho$) directions. *Right panel:* Data from the left panel, integrated along the radial direction, calculated at discs of height 1%$d_{\max}$ and the same diameter as the XFV.
Left panel image source: [39]

within the ROI. The photoelectric effect mean interaction length for energy $\sim Q_{\beta\beta}$ is $\sim 1$ m (Figure 3.24). A significantly more rare way for the $^{208}$Tl gammas to survive the cuts is for them to transfer most of their energy via a single Compton interaction in the XFV (at most they can transfer 2.377 MeV this way) and then depositing the rest of their energy via photoelectric effect, at a $z$ no further than 3 mm away from the original deposition site height, so as to survive the SS cut. Of these two options, the former will consistute a significantly larger background contribution, and for the rest of this work they will be referred to as *single electron*, or $1e$ *events*.

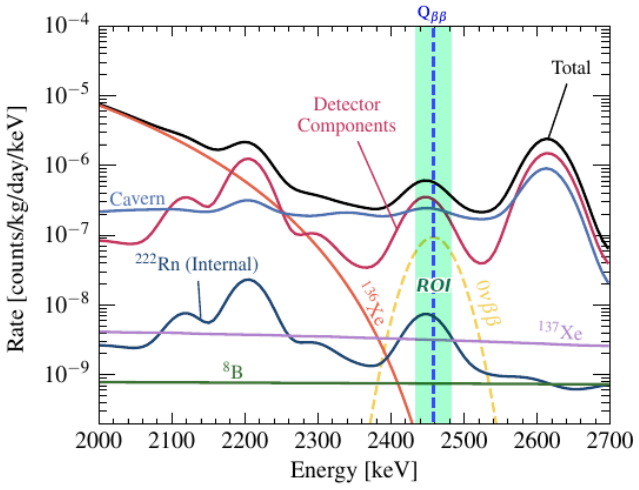


FIGURE 4.8: Background sources and rates inside the FV, after all of the available selection cuts. The yellow dashed line represents the expected $0\nu\beta\beta$ energy spectrum, with a rate estimated by considering $T_{1/2} = 1.06 \times 10^{26}$ years and does not contribute to the total background. More information on each of the individual background contributions is available in Section 3.2.3. Of note are the two most prominent background sources near $Q_{\beta\beta}$ and the comparatively low contribution from $2\nu\beta\beta$ (orange line marked $^{136}$Xe).
Source: [39]

The topology of $1e$ events differs from that of $0\nu\beta\beta$ events: instead of the two primary electrons in $0\nu\beta\beta$, there is just one, and hence the track features only one blob (Figure 4.9). The goal of this workis to determine whether it is possible to improve the background by way of a topology-based cut.

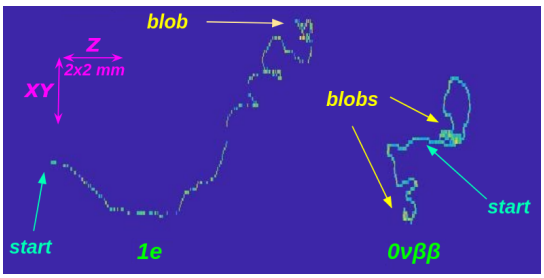


FIGURE 4.9: 2D projection of a `Geant4` LXe energy-deposition-only simulation of a $1e$ event with a primary electron of energy $Q_{\beta\beta}$ and a $0\nu\beta\beta$ event, both with no diffusion. The yellow regions indicate a larger number of ionization electrons.

Hypothesis testing is performed as described in detail in [165]. The LZ detector is projected to have a 1000 live day science run, and during that time, a number $c_{\text{sci}}$ of $0\nu\beta\beta$ candidate counts will be registered at positions $(x, y, z)$ along the detector, forming the collection of counts $\boldsymbol{\mathcal{C}} = \{(x_i, y_i, z_i)\}_{i=1}^{c_{\text{sci}}}$. It is possible for all of these counts to be due to background events that survived the applied cuts: this constitutes the *null hypothesis*, $H_0$. It has an associated probability $Pr(\boldsymbol{\mathcal{C}}|H_0)$, which can be calculated with the developed background model [39]. If the counts along the detector exceed the background model prediction, then $Pr(\boldsymbol{\mathcal{C}}|H_0)$ decreases, and if $\boldsymbol{\mathcal{C}}$ exceeds 99.87% of count collections predicted by the model, then an *alternative hypothesis*, $H_A$, is taken to be true: that $\boldsymbol{\mathcal{C}}$ is due to background counts and $0\nu\beta\beta$ counts together.

The *likelihood function* of $\boldsymbol{\mathcal{C}}$, $\mathcal{L}(H_0)$, is constructed from two components:

$$\mathcal{L}(H_0) = f(\boldsymbol{\mathcal{C}}|H_0) \times g(\text{nuis}|H_0) \tag{4.3}$$

The first component, $f(\boldsymbol{\mathcal{C}}|H_0)$, is called the *total probability model* [165], and represents the probability of $\boldsymbol{\mathcal{C}}$. It is given as:

$$f(\boldsymbol{\mathcal{C}}|H_0) = \text{Pois}(c_{\text{sci}}|H_0) \times \prod_{i=1}^{c_{\text{sci}}} \mathfrak{f}(\text{count } i|H_0), \tag{4.4}$$

where $\text{Pois}(c_{\text{sci}}|H_0)$ is a Poisson distribution with average $\mu_c$ given by the background model, and $\mathfrak{f}(\text{count } i|H_0)$ is the distribution for each count, and is referred to as the *event probability model* [165]. The event probability model is a weighted sum of the distributions due to *parameters of interest* $\boldsymbol{\alpha}$ and due to $N_c$ *nuisance parameters* $\boldsymbol{\nu}$. The second component of Eq. 4.3, $g(\text{nuis}|H_0)$, is a set of constraints on the nuisance parameters, consisting of a product of $N_c$ Gaussian distributions.

To eliminate the nuisance parameters, the profile likelihood ratio (PLR) is used. The PLR, $\lambda(\boldsymbol{\alpha})$, is defined as:

$$\lambda(\boldsymbol{\alpha}) = \frac{\mathcal{L}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\nu}|\boldsymbol{\alpha}})}{\mathcal{L}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\nu}})}, \tag{4.5}$$

where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\nu}}$ are the maximum likelihood estimators (MLE) of $\boldsymbol{\alpha}$ and $\boldsymbol{\nu}$ respectively, and $\widehat{\boldsymbol{\nu}|\boldsymbol{\alpha}}$ is the MLE of $\boldsymbol{\nu}$ given $\boldsymbol{\alpha}$, and $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\nu}) \equiv \mathcal{L}(H_0)$. This function effectively "normalizes out" the nuisance parameters. Additionally, for large data samples, the value of $-2\log\lambda(\boldsymbol{\alpha})$ approximates a chi-square distribution for the same number of degrees of freedom as the number of parameters of interest [170]. From here, a p-value can be assigned to $\boldsymbol{\mathcal{C}}$.

# Chapter 5

# Application of Machine Learning Methods to $0\nu\beta\beta$ Search in LZ

The complexity and small size of the primary electron tracks, together with diffusion and the gate grid funneling of the detector (see Figure 4.6), mean that not only is it difficult to analytically characterize the waveforms of each of the two *classes* of event ($0\nu\beta\beta$ or $1e$), but the morphological differences between the two classes of waveforms will not be obvious (Figure 5.1). With these difficulties in consideration, a statistical solution to the problem becomes attractive. Namely, a machine learning (ML) algorithm can be capable of recognizing nontrivial patterns in the morphology of the two classes, improving the LZ background discrimination with an additional background rejection step. For example, a boosted decision tree classifier was used to discriminate between $0\nu\beta\beta$ events and $\gamma$ backgrounds in the EXO-200 experiment, yielding a $\sim$15% increase in sensitivity [171].

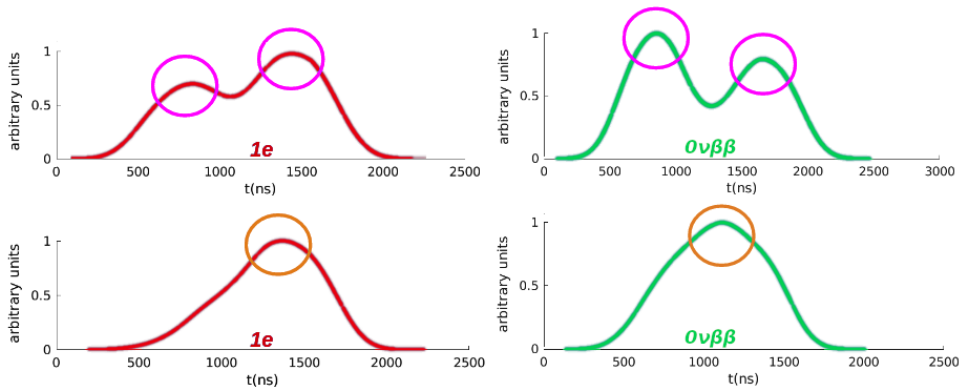


FIGURE 5.1: Two examples of GEANT4+ANTS2-simulated $1e$ (left) and $0\nu\beta\beta$ (right) LZ waveforms, with initially vertical primary electrons at $d = 145.6$ mm and no saturation or gate electric field distortion (see Chapter 6). As is exemplified here for the case of the number of peaks (circles) in the waveform, there are no prominent features that are mutually exclusive for $0\nu\beta\beta$ or for $1e$ waveforms.

The approach taken in this work was to parameterize the waveforms, converting a waveform $i$ from a signal representation $s_i[t]$ **{fn5a}** to an equivalent $\mathcal{W}$-dimensional *parameter vector* representation (see Section 6.4.2) $\boldsymbol{W}_i$. Ideally, the two classes would form two disjoint clusters of *datapoints* in the $\mathcal{W}$ *parameter space.* Then, supervised binary classification algorithms were trained on the resulting processed *dataset* to construct a *decision boundary*: all test events would be labeled $1e$ on one side of the boundary, and labeled $0\nu\beta\beta$ on the other.

**{fn5a}** — In this representation the signal amplitude is sampled at discrete times $t$.

This chapter presents the theoretical background on the implementation of a binary classifier to these two classes of events. Section 5.1 describes the algorithms selected for testing in this work, and Section 5.2 describes the classifier performance assessment tools that were used.

## 5.1 Supervised binary classification algorithms

A supervised binary classifier can be mathematically represented in the following manner:

Suppose there is a *training set* consisting of an $n$ pair sequence $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$, where:

- $\boldsymbol{X}_i \in \boldsymbol{\chi}$ is the *feature vector* for the training object $i$, $\boldsymbol{\chi}$ being the space of all possible feature vectors: the *feature space*.
- $Y_i \in \{\mathsf{N}, \mathsf{P}\}$ is the *label* corresponding to $\boldsymbol{X}_i$, $\mathsf{N} \equiv -1$ and $\mathsf{P} \equiv +1$ standing for "negative" and "positive" respectively.

Then a *binary classifier* is a function $b$ that takes as input a *test set* consisting of an $m$ pair sequence $(\boldsymbol{X}_{n+1}, Y^{\text{actual}}_{n+1}), \ldots, (\boldsymbol{X}_{n+m}, Y^{\text{actual}}_{n+m})$ **{fn5.1b}** and performs on it the transformation $b : \boldsymbol{\chi} \to \{\mathsf{N}, \mathsf{P}\}$ using the training set as reference, producing the new pairs $(\boldsymbol{X}_{n+j}, Y^{\text{guess}}_{n+j})$.

Four supervised binary classification algorithms were selected for this work based on their performance in a comparison made in [172] for the `scikit-learn` package [173], shown in Figure 5.2. Table 5.1 has a basic description of each compared algorithm. Since the feature vector *embedding* on the *feature manifold* (see Figure 5.3 for explanation) is not expected to be strictly linear (see Section 6.4.3), the performance of the classifiers for nonlinear and concentric data (the *"a)"* and *"b)"* sets in Figure 5.2) was given priority, with special importance attributed to the accurate mapping of the nonlinear portion in input data *"a)"*. The selected classifiers were the ones that best distinguished the more difficult characteristics of the assessment data:

- K-nearest neighbors (*kNN*, Section 5.1.1);
- Support vector machine with radial basis function kernel (*RBF SVM*, Section 5.1.2);
- Gaussian process classifier (*GP*, Section 5.1.3);
- Random forest classifier (*RF*, Section 5.1.4).

Because the mathematical equivalencies in the RBF SVM algorithm are what provides the classifier with its advantages, and because the GP classifier was the best-performing one in this work (Section 7.4), it was chosen to describe their algorithms in more detail than for the kNN and RF classifiers.

**{fn5.1b}** — The actual class of each object in this set, $Y_i^{\text{actual}}$, $i \in n+1, \ldots, n+m$, may or may not be known in advance. When testing the classifier's performance, the labels are known, and the set is referred to as a *test set*. In deployment, the labels are not known.

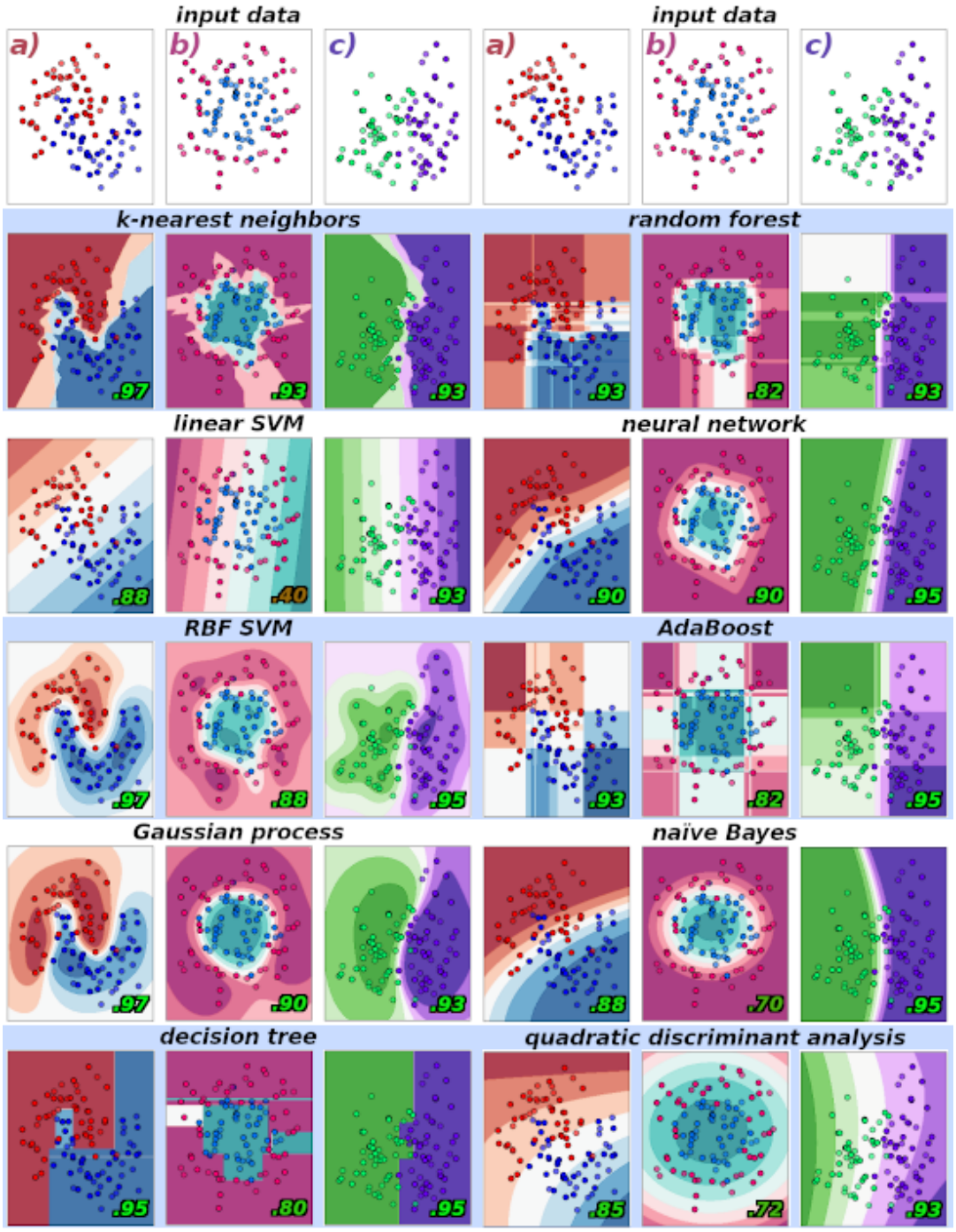


FIGURE 5.2: Performance of different classifiers and their corresponding accuracy scores (the fractions of correctly labeled test datapoints). In the images, the training set are displayed as opaque circles, while the test sets are shown semi-transparent. The shapes of the input data serve to test the classifier performance in the case of: a) a nonlinear boundary; b) concentric data; c) a quasi-linear boundary.
Source: [172].

| Algorithm | Summary description | Advantages | Disadvantages | References |
|---|---|---|---|---|
| **k-Nearest Neighbors** | *Assign to the test point the same label as the majority of the $k$ nearest neighbors* | • Very simple algorithm<br>• Effective on dense data | • Sensitive to irrelevant attributes<br>• Computationally expensive | [174] |
| **Linear SVM** | *Separate the two classes with a flat hyperplane, set furthest away from nearest points of either class* | • Effective in high-dimensional spaces<br>• Memory-efficient<br>• Effective on sparse data | • Unstable to mild parameter adjustments<br>• Class likelihoods not given by algorithm | Section 1.4 of [173], [175] |
| **RBF SVM** | *Same as linear SVM, but hyperplane can be curved* | • More versatile than linear SVM | • High risk of overfitting | Same as above |
| **Gaussian Process** | *Fit likelihoods of datapoints belonging to class* P *across* $\chi$ *via Bayesian regression, use a likelihood threshold to construct decision boundary* | • Prediction is probabilistic<br>• Automatic adjustment of fit parameters | • Computation time scales with the cube of the number of datapoints<br>• Low efficiency in high-dimensional spaces | [176] [177] |
| **Decision Tree** | *Define rules dividing feature space into regions balanced strongly in favor of one class* | • Computationally inexpensive<br>• Results are explicit | • Unstable to variations<br>• Biased results for unbalanced data | Section 1.10 of [173] |
| **Random Forests** | *Decision boundary made using average of several similarly trained trees* | • Perform implicit feature selection<br>• Suitable for large datasets | • Difficult to interpret results | Section 1.11.2 of [173] |
| **Neural Network** | *Fit likelihood of a datapoint belonging to either class for any* $\chi$ *via a weighted, biased sum of activation functions* | • Capable of learning nonlinear models | • Results are not explicit *(black box)*<br>• Requires difficult tuning<br>• Can fall into local minima | Section 1.17 of [173] |
| **Adaptive Boosting *(AdaBoost)*** | *Weighted sum of weak classifiers, with each later classifier adapted to the errors of the previous one* | • Resistant to overfitting | • Sensitive to outliers | Section 1.11.3 of [173] |
| **Naive Bayes** | *Likelihood of test point belonging to class based on product of Bayes theorem outputs from each feature* | • High accuracy if all features independent<br>• Effective with high dimensionality | • Bad performance if features not independent | Section 1.9 of [173] |
| **Quadratic Discrim. Analysis** | *Fit multivariate Gaussians to both classes, decision boundary constructed by comparing likelihoods* | • Simple algorithm<br>• Computationally inexpensive | • Ineffective for non-Gaussian data | Section 1.2 of [173] |

TABLE 5.1: Basic description of the algorithms in the comparison in Figure 5.2.

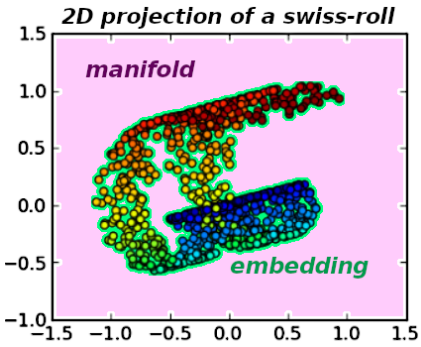


FIGURE 5.3: Depiction of a flat 2D object rolled up into a "swiss-roll" (highlighted in green), so that three spacial coordinates are necessary to describe the position of each point. However, not every point in 3D space corresponds to a point on the swiss-roll. The 3D space is referred as the manifold and the space occupied by the swiss-roll on the manifold is called the embedding.
Image source: `Everipedia.org`

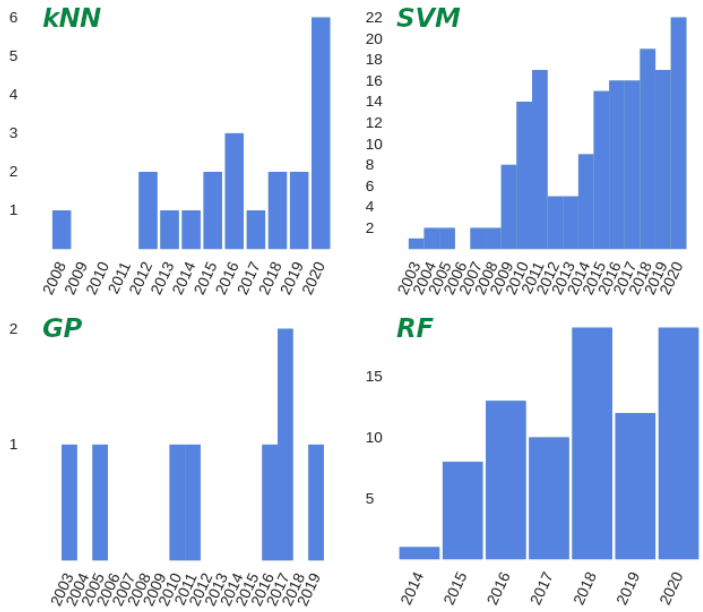


FIGURE 5.4: Number of papers published per year before 2021, related to each of the selected classification algorithms **{fn5.1c}**, present in the SAO/NASA Astrophysics Data System.

*Top left:* k-Nearest neighbors. *Top right:* Support vector machines.
*Bottom left:* Gaussian process. *Bottom right:* Random forests.

Image source: `adsabs.harvard.edu` (*retrieved 12 Apr. 2021*)

**{fn5.1c}** — The search terms used for each algorithm were the following:

**kNN** → `(keyword:"k-nearest neighbors" OR keyword:"kNN") AND (keyword:"classification" OR keyword:"classify")`

**SVM** → `(keyword:"support vector machine") AND (keyword:"classification" OR keyword:"classify")`

**GP** → `(keyword:"gaussian processes" OR keyword:"gaussian process") AND (keyword:"classification" OR keyword:"classify")`

**RF** → `(keyword:"random forests" OR keyword:"random forest" OR keyword:"decision tree ensemble" OR keyword:"decision tree ensembles") AND (keyword:"classification" OR keyword:"classify")`

### 5.1.1 The k-nearest neighbors classifier

The kNN algorithm is based on the assumption that the likelihood of each label varies smoothly along the feature space, and that the labels of nearby points are a good measure of that likelihood. For a $j$-th test datapoint, $\boldsymbol{X}_{n+j}$, and given an input parameter number $k$, it performs the following operations:

1. Find the $k$ training points that are nearest to $\boldsymbol{X}_{n+j}$ in feature space;
2. $Y_{n+j}^{\text{guess}}$ is assigned the same value as most of the $k$ nearest neighbors.

The algorithm is very simple and uses no explicit training or model [178]. The decision boundary in this case is not explicitly defined, but instead arises from the possible label vote results. The distance metric used for finding the $k$ training points in step 1 and the vote weighting applied in step 2 for selecting the majority label depend on the chosen implementation. Figure 5.5 shows an example of kNN classification using a Euclidean distance metric and equal weighting of nearest neighbors, with $k = 3$, applied on three test datapoints.

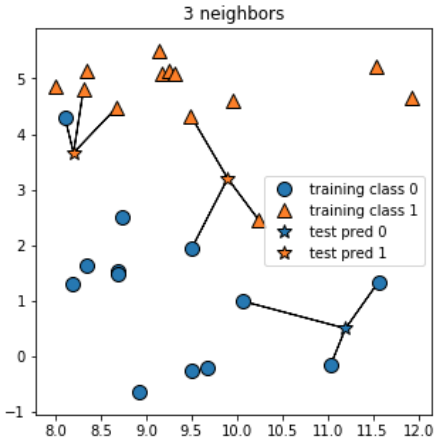


FIGURE 5.5: Example binary classification of three test datapoints (stars) using a kNN classifier with $k = 3$ and a 2-dimensional training dataset. In the leftmost test datapoint, note that even though the closest training datapoint is from the "blue" class (0), the test datapoint is still classified as belonging to the "red" class (1) because of the next two closest training datapoints.
Source: [179]

Classification using kNN is typically applied to problems where similar objects need to be identified, namely recommender systems and concept search. In astronomy and experimental physics, the use of kNN classifiers is comparatively uncommon (Figure 5.4), but they have been applied namely to the automatic classification of celestial objects [180], to neutron / $\gamma$-ray discrimination in plastic scintillators [181], solar wind categorization [182], spectral and luminosity classification of stellar spectra [183] and classification of different cosmic ray primaries [184].

The simplicity of the kNN classifier makes it an attractive baseline against which to compare the performance of other classification algorithms. It was seen to be very good in the test cases observed in Figure 5.2. Although it is sensitive to noise at low values of $k$ (Figure 5.6), it becomes robust against noise with large $k$ and denser data. Although the computational speed of the classifier decreases with the size of the dataset as $\mathcal{O}(n)$ due to the number of distance comparisons for increasingly large datasets, the current goals do not require that the classifier be particularly fast, so it is not a significant issue for this work. For these reasons, it was chosen to include the kNN algorithm in this work. It should be noted, however, that the kNN algorithm also has disadvantages relevant to this work. Namely, it loses effectiveness with high-dimensional, sparse data (curse of dimensionality, see Section 6.1.3), and is easily deceived by irrelevant attributes of the embedding [174].

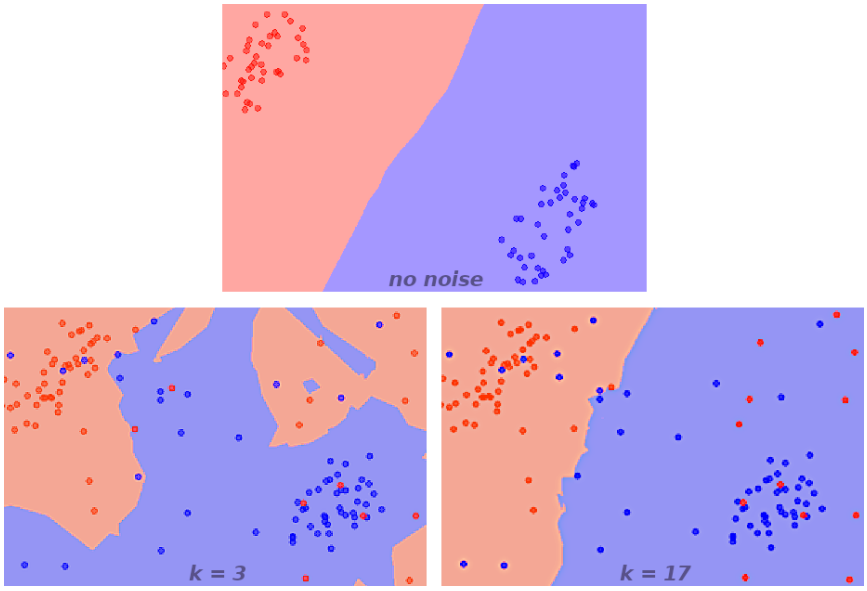


FIGURE 5.6: Simulation of two populations separated by kNN without the presence of noise (top panel) and when noise is present, for a small (bottom left panel) and a large (bottom right panel) value of $k$. Note that the presence of noise results in a very chaotic border for the case of small $k$.
Source: [185]

### 5.1.2 The support vector machine classifier with radial basis function kernel

In binary classification problems, support vector machines (SVMs) construct a "flat" decision boundary separating the two classes, the boundary is positioned in such a way as to maximize the *function margin*: the distance between the boundary and the training datapoints that it is closest to **{fn5.1d}** (the *support vectors*, see Figure 5.7). The decision boundary is positioned by the following:

1. Define a vector $\boldsymbol{w} \in \boldsymbol{\Phi}$ and intercept $\rho$ describing a hyperplane;
   - *The space $\boldsymbol{\Phi}$ can have the same number of dimensions (dimensionality) as the feature space $\boldsymbol{\chi}$, but it is common to define $\boldsymbol{\Phi}$ as being higher-dimensional than $\boldsymbol{\chi}$.*
   - *The decision boundary is a hyperplane in the space $\boldsymbol{\Phi}$. In effect, it may be flat in feature space, but it also may appear curved.*
2. Each training datapoint $\boldsymbol{X}_i$ has a distance to the hyperplane $(\boldsymbol{w}, \rho)$: the *functional margin* $\hat{\gamma}_{\boldsymbol{w},\rho}(\boldsymbol{X}_i, Y_i)$. The function margin of the hyperplane, $\widehat{\gamma_{\boldsymbol{w},\rho}}$, is defined as $\widehat{\gamma_{\boldsymbol{w},\rho}} \equiv \min_i \{\hat{\gamma}_{\boldsymbol{w},\rho}(\boldsymbol{X}_i, Y_i)\}$.
   - *The functional margin is defined as the value $\hat{\gamma}_{\boldsymbol{w},\rho}(\boldsymbol{X}_i, Y_i) \equiv (\boldsymbol{w} \cdot \boldsymbol{\phi}(\boldsymbol{X}_i) + \rho) \times Y_i$, where $\boldsymbol{\phi}(\boldsymbol{X}_i)$ is some function that performs the mapping $\boldsymbol{\phi} : \boldsymbol{\chi} \to \boldsymbol{\Phi}$.*
3. The decision boundary $(\hat{\boldsymbol{w}}, \hat{\rho})$ maximizes $\widehat{\gamma_{\boldsymbol{w},\rho}}$ relative to $||\boldsymbol{w}||$.

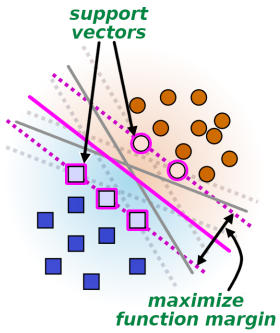


FIGURE 5.7: Illustration of the concepts of function margin, support vectors and of binary classification using an SVM. One class is depicted as blue squares, and the other one is depicted as red circles. The resulting decision boundary is highlighted in magenta, and the gray lines are test hyperplanes that produce smaller function margins than the obtained decision boundary. The support vectors are highlighted in magenta, and correspond to the datapoints equidistantly closest to the decision boundary.
Image source: [186]

**{fn5.1d}** — Excluding (possibly) outliers. The treatment of outliers is discussed in pg. 68.

SVMs are fast both in training with nontrivial embedding shapes and in testing with high-dimensional feature spaces. This is due to reformulations that are performed on the maximization in step 3. These are explained in detail in [187], but summarily they are achieved in four stages:

**1.** The maximization is converted into a convex optimization problem, referred to as the *primal problem*:

$$\begin{array}{lll} & \textit{maximization} & \\ \underset{\boldsymbol{w},\,\rho}{\text{maximize}} & \dfrac{\widehat{\gamma_{\boldsymbol{w},\rho}}}{||\boldsymbol{w}||} & \\ \text{subject to} & \hat{\gamma}_{\boldsymbol{w},\rho}(\boldsymbol{X}_i, Y_i) \geq \widehat{\gamma_{\boldsymbol{w},\rho}} & \\ & i = 1, \ldots, n. & \end{array} \quad \Rightarrow \quad \begin{array}{lll} & \textit{primal problem} & \\ \underset{\boldsymbol{w},\,\rho}{\text{minimize}} & \dfrac{1}{2}||\boldsymbol{w}||^2 & \\ \text{subject to} & \hat{\gamma}_{\boldsymbol{w},\rho}(\boldsymbol{X}_i, Y_i) \geq 1 & \\ & i = 1, \ldots, n. & \end{array} \tag{5.1}$$

*(Implying that $\widehat{\gamma_{\boldsymbol{w},\rho}} = 1$, but the parameters can be rescaled with no losses after optimization)*

**2.** The primal problem is expressed as finding a local minimum in a Lagrangian:

An equivalent Lagrangian $\mathcal{L}(\boldsymbol{w}, \rho, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a vector of Lagrange multipliers $\alpha_i \geq 0, i = 1, \ldots, n$, is defined:

$$\mathcal{L}(\boldsymbol{w}, \rho, \boldsymbol{\alpha}) = \tfrac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{n} \alpha_i \times \{Y_i \times (\boldsymbol{w} \cdot \boldsymbol{\phi}(\boldsymbol{X}_i) + \rho) - 1\}. \tag{5.2}$$

The primal problem can then be retrieved as:

$$\min_{\boldsymbol{w},\rho} \left\{ \max_{\boldsymbol{\alpha}\,:\,\alpha_i \geq 0} \{\mathcal{L}(\boldsymbol{w}, \rho, \boldsymbol{\alpha})\} \right\}. \tag{5.3}$$

**3.** The Lagrangian primal problem is converted into an equivalent *dual problem*:

$$\begin{array}{cc} \textit{primal problem} & \textit{dual problem} \\ \min\limits_{\boldsymbol{w},\rho} \left\{ \max\limits_{\boldsymbol{\alpha}\,:\,\alpha_i \geq 0} \{\mathcal{L}(\boldsymbol{w}, \rho, \boldsymbol{\alpha})\} \right\} & \Rightarrow \quad \max\limits_{\boldsymbol{\alpha}\,:\,\alpha_i \geq 0} \left\{ \min\limits_{\boldsymbol{w},\rho} \{\mathcal{L}(\boldsymbol{w}, \rho, \boldsymbol{\alpha})\} \right\} \end{array} \tag{5.4}$$

Solving the primal or the dual problem is equivalent: they yield the same set of values $(\hat{\boldsymbol{w}}, \hat{\rho}, \hat{\boldsymbol{\alpha}})$ as solutions.

**4.** The dual problem is expressed as an optimization problem in terms of the *kernels*:

$$\begin{array}{ll} \underset{\boldsymbol{\alpha}}{\text{maximize}} & \sum\limits_{i=1}^{n} \alpha_i - \dfrac{1}{2} \sum\limits_{\imath,\iota=1}^{n} Y_\imath Y_\iota \times \alpha_\imath \alpha_\iota K_{\imath\iota} \\ \text{subject to} & \alpha_i \geq 0 \\ & \quad i = 1, \ldots, n, \\ \text{and} & \sum_{i=0}^{n} Y_i \times \alpha_i = 0, \end{array} \tag{5.5}$$

where $K_{\imath\iota} \equiv K(\boldsymbol{X}_\imath, \boldsymbol{X}_\iota) \equiv \boldsymbol{\phi}(\boldsymbol{X}_\imath) \cdot \boldsymbol{\phi}(\boldsymbol{X}_\iota)$ is the *kernel*, a function with the mapping $K : \boldsymbol{\Phi}, \boldsymbol{\Phi} \to \mathbb{R}$.

---

*Solving the dual problem is faster than solving the primal problem for two reasons. First is the "kernel trick": the use of the inner products of $\boldsymbol{\phi}(\boldsymbol{X}_i)$, instead of $\boldsymbol{\phi}(\boldsymbol{X}_i)$ directly, allows calculations on a space $\boldsymbol{\Phi}$ of any dimensionality to incur roughly the same computational cost as calculations on the space $\boldsymbol{\chi}$. Second is that, because the solution $(\hat{\boldsymbol{w}}, \hat{\rho}, \hat{\boldsymbol{\alpha}})$ to the optimization problem satisfies the Karush-Kuhn-Tucker conditions [188], the decision boundary is defined solely in terms of the support vectors.*

If the dimensionality of the space $\mathbf{\Phi}$ is the same as that of the feature space $\boldsymbol{\chi}$, then the decision boundary constructed by the SVM is necessarily hyperplanar. However, if the dimensionality of $\mathbf{\Phi}$ is larger than that of $\boldsymbol{\chi}$, then it is possible for the SVM to construct a curved decision boundary. That is desirable, as it is not expected for the feature embeddings obtained in this work to be separable by a single hyperplane in feature space (see Section 6.4.3). The dimensionality of $\mathbf{\Phi}$ in the dual optimization problem is set by the kernel $K_{\imath\iota}$ (see Eq. 5.5). One example of a higher-dimensionality kernel is the radial basis function (RBF):

$$K_{\imath\iota} = e^{-\gamma||\boldsymbol{X}_\imath - \boldsymbol{X}_\iota||^2}, \tag{5.6}$$

where $\gamma > 0$ is a *shape parameter* used for scaling. The RBF returns a value nearer to 1 the closer together in feature space the points $\imath$ and $\iota$ are. In practice, RBF is a common choice of kernel for when there is no prior knowledge of the shape of the embedding of the data, and as such it is an appropriate initial choice for the given problem. If the shape parameter is defined as $\gamma = 1/(2\sigma^2)$, then $K_{\imath\iota}$ is a *Gaussian kernel*. The kernel used in this work is similar to the Gaussian kernel, except the $1/2$ factor is replaced for a factor that scales according to the dimensionality of $\boldsymbol{\chi}$. Figure 5.8 shows an example of a decision boundary generated by a Gaussian kernel.

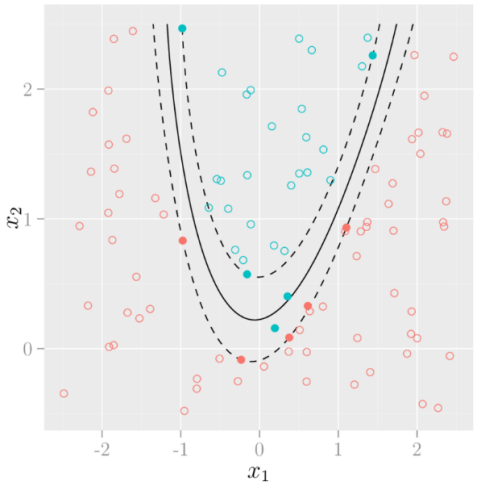


FIGURE 5.8: Curved decision boundary (continuous curve) generated by a Gaussian kernel SVM. The support vectors are along the dashed curves and filled in. Note that some points are closer to the decision boundary than the support vectors themselves.
Source: [189]

Note that the dual optimization problem in Eq. 5.5 is not tolerant to outliers. As seen by the datapoints nearer to the decision boundary than the support vectors themselves in Figure 5.8, this leaves the classifier susceptible to overfitting. To produce the final form of the SVM algorithm, a tolerance to outliers, $C$, is introduced into the dual optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{maximize}} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{\imath,\iota=1}^{n} Y_\imath Y_\iota \times \alpha_\imath \alpha_\iota K_{\imath\iota} \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C \\ & i = 1, \ldots, n, \\ \text{and} \quad & \textstyle\sum_{i=0}^{n} Y_i \times \alpha_i = 0. \end{aligned} \tag{5.7}$$

In practice, classification using SVMs is commonly applied to pattern recognition problems, namely face and speech detection / recognition, or text categorization [175]. In astronomy and experimental physics, SVM classification has had increased adoption since 2013, to where it is currently one of the most popular methods (Figure 5.4). It has been used namely for neutron/$\gamma$-ray discrimination in organic liquid scintillator [190], vetoing in gravitational wave analysis [191], and SVM variants have been applied to classification of stellar spectra [192] [193].

Favorable characteristics of SVM classifiers include their good performance with small or sparse datasets and in high-dimensional spaces, as well as when generalizing to new samples [175]. Unlike in certain algorithms, namely neural networks, there is no possibility of the training converging on local minima, because the SVM optimization problem is convex, so the training is guaranteed to arrive at the optimal solution given the dataset and for the particular chosen kernel and parameter settings. However the construction of the algorithm also makes it difficult to train SVMs well. Crucially, SVMs are sensitive to small parameter adjustments, and the likelihood of the datapoints belonging to either class have to be extracted indirectly via cross-validation, as they are not provided directly by the classifier [173].

### 5.1.3 The Gaussian process classifier

The Gaussian process classifier is a Bayesian algorithm that fits to the training data a distribution of functions representing the possible decision boundaries, one of which is then selected with a user-defined *threshold parameter* $\tau$. Like in the SVM algorithm (Section 5.1.2), where hyperplanes in a space $\boldsymbol{\Phi}$ of a higher dimensionality than the feature space $\boldsymbol{\chi}$ were used to create curved decision boundaries with the help of kernels, the Gaussian process classifier applies Bayesian linear regression to the training data in $\boldsymbol{\Phi}$, which then produces a curved decision boundary distribution in $\boldsymbol{\chi}$ with the help of a mathematical tool called a *Gaussian process*.

A Gaussian process (GP) is defined as a set of random variables where any finite subset follows a multivariate Gaussian distribution **{fn5.1f}**. Similar to how a multivariate Gaussian distribution is specified for any vector $\mathbf{x} \in \mathbb{R}^{\mathfrak{q}}$ by a mean vector $\boldsymbol{m} \in \mathbb{R}^{\mathfrak{q}}$ and a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{\mathfrak{q}\times\mathfrak{q}}$:

$$\underset{\text{Gaussian}}{\text{multivariate}} \quad \rightarrow \quad \mathbf{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}) \quad \rightarrow \quad p(\mathbf{x}) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{\sqrt{2^{\mathfrak{q}}\pi^{\mathfrak{q}}}} \exp\left[-(\mathbf{x}-\boldsymbol{m})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{m})/2\right],$$

a GP is specified for any random variable $\boldsymbol{x} \in \mathfrak{X}$ (where $\mathfrak{X}$ is an index set **{fn5.1g}**) by a *mean function* $m(\boldsymbol{x})$ and a covariance function $K(\boldsymbol{x}, \boldsymbol{x}')$, typically the Gaussian kernel $K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left[-||\boldsymbol{x} - \boldsymbol{x}'||^2/2\sigma^2\right]$ (where $\sigma$ is a *length parameter*) [194]:

$$\underset{\text{process}}{\text{Gaussian}} \quad \rightarrow \quad \mathcal{GP}(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}')). \tag{5.8}$$

A GP can be used to assign a set of probability distributions to a family of functions $\mathfrak{f} : \boldsymbol{x} \rightarrow \mathbb{R}$ with function outputs $\mathfrak{f}_{\boldsymbol{x}} \equiv \mathfrak{f}(\boldsymbol{x})$, where any vector $\mathbf{f} = [\mathfrak{f}_{\boldsymbol{x}_1} \ \ldots \ \mathfrak{f}_{\boldsymbol{x}_{\mathfrak{V}}}]^T$ (with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\mathfrak{V}}$ being any $\mathfrak{V}$ points from the index set $\mathfrak{X}$) has a multivariate Gaussian distribution [195]:

$$\begin{bmatrix} \mathfrak{f}_{\boldsymbol{x}_1} \\ \vdots \\ \mathfrak{f}_{\boldsymbol{x}_{\mathfrak{V}}} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} m(\boldsymbol{x}_1) \\ \vdots \\ m(\boldsymbol{x}_{\mathfrak{V}}) \end{bmatrix}, \begin{bmatrix} K_{11} & \ldots & K_{1\mathfrak{V}} \\ \vdots & \ddots & \vdots \\ K_{\mathfrak{V}1} & \ldots & K_{\mathfrak{V}\mathfrak{V}} \end{bmatrix} \right) \quad , \quad K_{\imath\imath'} \equiv K(\boldsymbol{x}_\imath, \boldsymbol{x}_{\imath'}) \tag{5.9}$$

To denote this, the following notation is used:

$$\mathfrak{f}(\cdot) \sim \mathcal{GP}(m(\cdot), K(\cdot, \cdot)). \tag{5.10}$$

The ability of GPs to assign probability distributions over function families makes them a versatile tool for fitting functions to data (an example fit is shown in Fig. 5.9). Unlike the kNN and SVM algorithms, the trained GP classifier does not perform the $\boldsymbol{\chi} \to \{\mathsf{N}, \mathsf{P}\}$ mapping directly. Instead, assigning to the test datapoint $\boldsymbol{X}_{n+j}$ a label $Y^{\text{guess}}_{n+j}$ is divided into three stages:

$\boldsymbol{\chi} \to \mathbb{R}$: A trained *latent function* $\hat{\varphi}$, obtained with the $n$ training datapoints, is used to map $\boldsymbol{X}_{n+j}$ to a Gaussian-distributed random variable $f^*_{n+j}$;

$\mathbb{R} \to [0, 1]$: A sigmoidal *squashing function* $\pi$ converts $f^*_{n+j}$ into a likelihood $L_{n+j}$ that the datapoint $\boldsymbol{X}_{n+j}$ would correspond to an object of the class $\mathsf{P}$, according to $\hat{\varphi}$;

$[0, 1] \to \{\mathsf{N}, \mathsf{P}\}$: A *threshold* $\tau$ is used for labeling: if $L_{n+j} > \tau$, then $Y^{\text{guess}}_{n+j} = \mathsf{P}$, else $Y^{\text{guess}}_{n+j} = \mathsf{N}$.

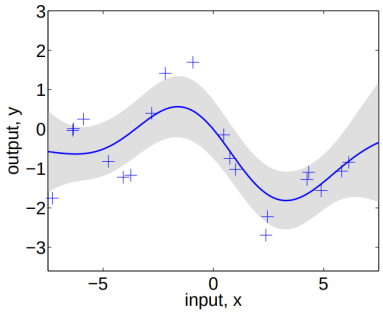


FIGURE 5.9: Function fit to datapoints (blue crosses) using GPs. The 95% confidence band of the GP is highlighted in gray. Note that for each value of input, x, the GP defines a Gaussian distribution with average $m(\mathsf{x})$, and so any set of fit values will follow a multivariate Gaussian distribution. The blue curve corrseponds to the mean function $m(\,\cdot\,)$.
Source: [177]

The trained latent function $\hat{\varphi}$ is the function in the GP that is most likely to have been obtained by the training dataset $\mathcal{D}_T \equiv (\boldsymbol{X}_i, Y_i), i = 1, \ldots, n$. The most likely function is determined by a Bayesian regression method:

Given a datapoint $\boldsymbol{X}_i$ as input, the latent function $\varphi$ outputs a Gaussian-distributed number $f_i \equiv \varphi(\boldsymbol{X}_i)$. The GP is initially assumed to have zero mean, with a Gaussian kernel, so the numbers outputted by the entire training dataset define a vector $\boldsymbol{f} \equiv [f_1 \ \ldots \ f_n]^T \sim \mathcal{N}(\boldsymbol{0}, \mathbf{K})$, where $\mathbf{K}$ is an $n \times n$ *covariance matrix*, with $K_{\imath\iota} = \exp[-||\boldsymbol{X}_\imath - \boldsymbol{X}_\iota||^2/2\sigma^2]$ (with unknown $\sigma$) for all values $\imath, \iota = 1, \ldots, n$. The vector $\boldsymbol{f}$ can then be used to represent the latent function itself: namely, the estimated test latent function output $\hat{f}^*_{n+j}$ is implied by the conditional distribution $f^*_{n+j}|\hat{\boldsymbol{f}}$, where $\hat{\boldsymbol{f}} \equiv [\hat{\varphi}(\boldsymbol{X}_1) \ \ldots \ \hat{\varphi}(\boldsymbol{X}_n)]$ is the latent function output vector most likely to have been obtained by $\mathcal{D}_T$. Bayesian linear regression is a method that can

**{fn5.1f}** — For example, the outcomes $\mathfrak{y}$ of $\mathfrak{n}$ measurements, $\boldsymbol{\mathfrak{y}} = [\mathfrak{y}_1, \ldots, \mathfrak{y}_\mathfrak{n}]$, at parameter values $\mathfrak{x}_\mathfrak{i}$ , $\mathfrak{i} = 1, \ldots, \mathfrak{n}$ with added statistical noise $\varepsilon_\mathfrak{i} \sim \mathcal{N}(0, \sigma^2)$, so $\mathfrak{y}_\mathfrak{i} = \mathfrak{f}(\mathfrak{x}_\mathfrak{i}) + \varepsilon_\mathfrak{i}$, will constitute a GP.

**{fn5.1g}** — An index set is one that indexes another set. For example it can be used to represent the indices of the training datapoints.

calculate the estimators for the length parameter, $\hat{\sigma}$, and for the latent function vector, $\hat{\boldsymbol{f}}$, without requiring prior information. It is based on Bayes' theorem:

Consider two arbitrary sets of outcomes $\mathfrak{a}$ and $\mathfrak{b}$, with probabilities of occurring $Pr(\mathfrak{a})$ and $Pr(\mathfrak{b})$ (termed *marginal probabilities*). The conditional probabilities $Pr(\mathfrak{a}|\mathfrak{b})$ and $Pr(\mathfrak{b}|\mathfrak{a})$ represent the likelihood of $\mathfrak{a}$ *(of* $\mathfrak{b}$*)* occurring if $\mathfrak{b}$ *(if* $\mathfrak{a}$*)* is known to have occurred. Then *Bayes' theorem* states:

$$Pr(\mathfrak{a}|\mathfrak{b}) = \frac{Pr(\mathfrak{b}|\mathfrak{a}) \times Pr(\mathfrak{a})}{Pr(\mathfrak{b})} \quad or \quad posterior = \frac{\text{likelihood} \times prior}{\text{marginal likelihood}}. \tag{5.11}$$

Applying Bayes' theorem to the training data $\mathcal{D}_T$ and latent function $\boldsymbol{f}$ yields:

$$Pr(\boldsymbol{f}|\mathcal{D}_T) = \frac{Pr(\mathcal{D}_T|\boldsymbol{f}) \times Pr(\boldsymbol{f})}{Pr(\mathcal{D}_T)}, \tag{5.12}$$

where:

- $Pr(\boldsymbol{f})$ is the *prior* model of $\boldsymbol{f}$, not accounting for the training dataset $\mathcal{D}_T$. Intuitively, $Pr(\boldsymbol{f}) = (2\pi)^{-n/2}\,|\mathbf{K}|^{-1/2}\exp[-\frac{1}{2}\boldsymbol{f}^T\mathbf{K}^{-1}\boldsymbol{f}]$;
- $Pr(\mathcal{D}_T)$ is a normalizing constant: $Pr(\mathcal{D}_T) = \int \mathrm{d}\boldsymbol{f}\,[Pr(\mathcal{D}_T|\boldsymbol{f}) \times Pr(\boldsymbol{f})]$;
- $Pr(\mathcal{D}_T|\boldsymbol{f})$ is the *likelihood*: the probability of the dataset $\mathcal{D}_T$ to have been produced by the function $\boldsymbol{f}$. It can work as a measure of goodness-of-fit between the data and the latent function;
- $Pr(\boldsymbol{f}|\mathcal{D}_T)$ Is the *posterior* model of $\boldsymbol{f}$, after accounting for $\mathcal{D}_T$.

The estimator $\hat{\boldsymbol{f}}$ is whichever vector $\boldsymbol{f}$ that maximizes the posterior probability, and $\hat{\sigma}$ is whichever $\sigma$ value that maximizes the marginal likelihood $Pr(\mathcal{D}_T)$, so:

$$\hat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{Pr(\boldsymbol{f}|\mathcal{D}_T)\} \quad , \quad \hat{\sigma} = \arg\max_{\sigma} \{Pr(\mathcal{D}_T)\} \tag{5.13}$$

The two estimators are interdependent, so they must be determined in tandem. The calculation of $\hat{\boldsymbol{f}}$ is discussed first. The box below details how the left-hand equation above reduces to:

$$\hat{\boldsymbol{f}} = \mathbf{K}\boldsymbol{\nabla}_{\boldsymbol{f}} \log\big(Pr(\mathcal{D}_T|\hat{\boldsymbol{f}})\big)\,. \tag{5.14}$$

The maximum of $Pr(\boldsymbol{f}|D_T)$ is the value of $\boldsymbol{f}$ for which $\boldsymbol{\nabla}_{\boldsymbol{f}} Pr(\boldsymbol{f}|D_T) = 0$. However, in these problems the logarithm of the probability is used instead of the probability itself, because a distribution and its logarithm share the same argmaxes. So:

$$\log(Pr(\boldsymbol{f}|\mathcal{D}_T)) = \log(Pr(\mathcal{D}_T|\boldsymbol{f})) + \log(Pr(\boldsymbol{f})) - \log(Pr(\mathcal{D}_T))\,. \tag{5.15}$$

Since $\log(Pr(\mathcal{D}_T))$ is a constant:

$$\boldsymbol{\nabla}_{\boldsymbol{f}} \log(Pr(\boldsymbol{f}|\mathcal{D}_T)) = \boldsymbol{\nabla}_{\boldsymbol{f}}\left[\log(Pr(\mathcal{D}_T|\boldsymbol{f})) + \log(Pr(\boldsymbol{f}))\right], \tag{5.16}$$

and so finding the argmax in Eq. 5.13 consists of solving:

$$\boldsymbol{\nabla}_{\boldsymbol{f}} \log(Pr(\mathcal{D}_T|\boldsymbol{f})) = -\boldsymbol{\nabla}_{\boldsymbol{f}} \log(Pr(\boldsymbol{f}))\,. \tag{5.17}$$

The prior distribution of $\boldsymbol{f}$ is multivariate Gaussian, $\boldsymbol{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$, so:

$$\log(Pr(\boldsymbol{f})) = -\left[\tfrac{1}{2}\boldsymbol{f}^T\mathbf{K}^{-1}\boldsymbol{f} + \tfrac{1}{2}\log(|\mathbf{K}|) + \tfrac{1}{2}\log(2\pi)\right], \tag{5.18}$$

meaning that:

$$\boldsymbol{\nabla}_{\boldsymbol{f}} \log(Pr(\boldsymbol{f})) = -\mathbf{K}^{-1}\boldsymbol{f}. \tag{5.19}$$

Substituting Eq. 5.19 into Eq. 5.17 results in:

$$\boldsymbol{\nabla}_{\boldsymbol{f}} \log(Pr(\mathcal{D}_T|\boldsymbol{f})) = \mathbf{K}^{-1}\boldsymbol{f}, \tag{5.20}$$

which is then refactored into Eq. 5.14.

The value of the latent function estimator $\hat{\boldsymbol{f}}$ is found by solving Eq. 5.14. It can be solved iteratively by providing an initial guess: for example, $\boldsymbol{f} = \boldsymbol{0}$ works well for this purpose [194]. To solve the equation iteratively, the value of $Pr(\mathcal{D}_T|\boldsymbol{f})$ is calculated on each iteration as:

$$Pr(\mathcal{D}_T|\boldsymbol{f}) = \prod_{i=0}^{n} \pi(Y_i \times f_i), \tag{5.21}$$

as explained in the box below.

Given that each datapoint is independent, $Pr(\mathcal{D}_T|\boldsymbol{f})$ is obtained from the individual datapoints as:

$$Pr(\mathcal{D}_T|\boldsymbol{f}) = \prod_{i=0}^{n} Pr((\boldsymbol{X}_i, Y_i)|f_i), \tag{5.22}$$

where $Pr((\boldsymbol{X}_i, Y_i)|f_i)$ is the probability that the datapoint/label pair $(\boldsymbol{X}_i, Y_i)$ could have been produced by the latent function $f_i \equiv \varphi(\boldsymbol{X}_i)$. Since $\boldsymbol{f}$ relates to the labels assigned to the datapoints, and not the datapoints themselves:

$$Pr((\boldsymbol{X}_i, Y_i)|f_i) \equiv Pr(Y_i|\boldsymbol{X}_i, \varphi(\boldsymbol{X}_i)), \tag{5.23}$$

which is the probability that the latent function would have assigned the correct label to the $i$-th training datapoint. If $Y_i = \mathsf{P} = 1$, then $Pr(Y_i|\boldsymbol{X}_i, \varphi(\boldsymbol{X}_i))$ is the probability that $\boldsymbol{X}_i$ would correspond to an object of the class $\mathsf{P}$ according to the current iteration $\varphi$, which is the same calculation that the squashing function $\pi$ performs on the test data during the $\mathbb{R} \to [0, 1]$ stage. So:

$$Pr(\mathsf{P} \equiv +1|\boldsymbol{X}_i, \varphi(\boldsymbol{X}_i)) = \pi(f_i) \quad , \quad Pr(\mathsf{N} \equiv -1|\boldsymbol{X}_i, \varphi(\boldsymbol{X}_i)) = 1 - \pi(f_i). \tag{5.24}$$

As $\pi$ is a sigmoid function, $\pi(-\xi) = 1 - \pi(\xi)$, so Eq. 5.21 is obtained.

To obtain the latent function output due to a test datapoint $f^*_{n+j}$, the distribution of the posterior model $Pr(\boldsymbol{f}|\mathcal{D}_T) \propto Pr(\mathcal{D}_T|\boldsymbol{f})Pr(\boldsymbol{f})$ must be known. For this purpose, typically a *Laplace approximation* is performed: the distribution is assumed to be Gaussian, with mean and variance given by:

$$\boldsymbol{f}|\mathcal{D}_T \sim \mathcal{N}\Big( \hat{f} \ , \ \Big[ -\nabla^2_{\boldsymbol{f}} \Big\{ \log\left(Pr(\boldsymbol{f}|\mathcal{D}_T)\right) \Big|_{\boldsymbol{f}=\hat{\boldsymbol{f}}} \Big]^{-1} \Big), \tag{5.25}$$

where the variance can be retrieved by (Section 3.4.1 of [177]):

$$-\nabla^2_{\boldsymbol{f}} \Big\{ \log\left(Pr(\boldsymbol{f}|\mathcal{D}_T)\right) \Big|_{\boldsymbol{f}=\hat{\boldsymbol{f}}} = \mathbf{K}^{-1} - \nabla^2_{\boldsymbol{f}} \Big\{ \log\left(Pr(\mathcal{D}_T|\boldsymbol{f})\right) \Big|_{\boldsymbol{f}=\hat{\boldsymbol{f}}} \equiv \boldsymbol{\mathcal{K}}^{-1}. \tag{5.26}$$

It should be noted that, although this is the typical approach **{fn5.1h}**, it is not ideal. It is not guaranteed that the true posterior distribution would be actually Gaussian: it can be significantly broader or narrower than indicated by the variance given by Laplace approximation, or even be asymmetrical [177]. There is the risk of this approach yielding poorly-performing latent function estimators, in which case other methods must be considered, namely expectation propagation (detailed in Section 3.6 of [177]).

The covariance matrix $\mathbf{K}$ in Eq. 5.14 is dependant on the length parameter $\sigma$, which is not set prior to regression. Its optimal value, $\hat{\sigma}$, should be used. The optimal value is obtained by solving the right-hand equation in Eq. 5.13:

$$\hat{\sigma} = \arg\max_{\sigma} \{Pr(\mathcal{D}_T)\} = \arg\max_{\sigma} \left\{ \int \mathrm{d}\boldsymbol{f} \ Pr(\mathcal{D}_T|\boldsymbol{f})Pr(\boldsymbol{f}) \right\}. \tag{5.27}$$

**{fn5.1h}** — Namely, it is the approach used in the Gaussian process classifier implementation used in the `scikit-learn` package [196], the package employed for binary classification in this work (Section 6.6.2).

To solve this equation, two manipulations are performed. First, similar to what was done in Eq. 5.25, a Laplace approximation of $Pr(\mathcal{D}|\boldsymbol{f})Pr(\boldsymbol{f})$ is used. Second, similar to what was done in Eq. 5.15, the quantity searched for is the argmax of $\log(Pr(\mathcal{D}_T))$, instead of $Pr(\mathcal{D}_T)$ directly. The resulting equation is (Section 3.4.4 of [177]):

$$\hat{\sigma} = \arg\max_{\sigma} \left\{ \log\left(Pr(\mathcal{D}_T|\hat{\boldsymbol{f}})\right) - \frac{1}{2}\left[\hat{\boldsymbol{f}}^T \mathbf{K}^{-1}\hat{\boldsymbol{f}} + \log\left(|\mathbf{K}|\cdot|\boldsymbol{\mathcal{K}}|\right)\right]\right\}, \tag{5.28}$$

where the same $\boldsymbol{\mathcal{K}}$ was used as in the definition in Eq. 5.26. To solve this equation, a multivariate optimization algorithm is used [194].

Note that finding $\hat{\boldsymbol{f}}$ requires that $\hat{\sigma}$ be known, and $\hat{\sigma}$ requires that $\hat{\boldsymbol{f}}$ be known. The Gaussian process classifier implementation used in this work (Section 6.6.2) is based on iterative algorithms from [177], where Eqns. 5.14 and 5.28 are both solved in each iteration, this procedure repeating until convergence of results.

Once the Bayesian regression method produces the trained latent function $\hat{\varphi}$, represented by the posterior function output vector $\boldsymbol{f}|\mathcal{D}_T \sim \mathcal{N}(\hat{\boldsymbol{f}}, \boldsymbol{\mathcal{K}})$ (Figure 5.10, $\boldsymbol{\mathcal{K}}$ defined in Eq. 5.26), the trained latent function can be used to estimate the latent function output $\hat{f}^*_{n+j}$ due to some test datapoint $\boldsymbol{X}_{n+j}$. This is not achieved explicitly by an operation of the form $\hat{\varphi}(\boldsymbol{X}_{n+j})$, however, as the function's representation in terms of a Gaussian-distributed vector does not allow for an operation of this kind. Instead, the distribution of $f^*_{n+j}|\hat{\boldsymbol{f}}$ is calculated, where $\hat{\boldsymbol{f}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{K}})$, which can then be used together with the distribution of the squashing function to yield the likelihood $L_{n+j} \equiv Pr(\mathsf{P}|f^*_{n+j}, \hat{\boldsymbol{f}})$ that $\boldsymbol{X}_{n+j}$ would be an object of the class $\mathsf{P}$ given $\hat{\varphi}$.

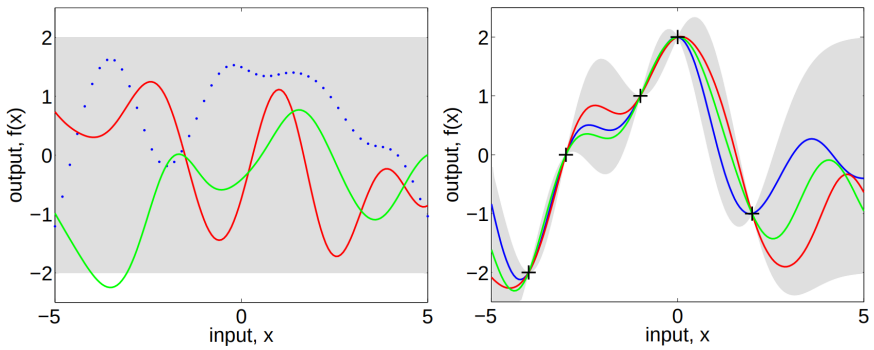


FIGURE 5.10: Illustrations of the prior and posterior GP with a one-dimensional feature space $\boldsymbol{\chi}$. *Left panel:* illustration of the GP corresponding to the prior latent function $\varphi$, with $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{K})$. The 95% confidence band of the GP is highlighted in gray. Three random example functions drawn from the GP are also shown. *Right panel:* illustration of the GP corresponding to the posterior latent function $\hat{\varphi}$, with $\boldsymbol{f}|\mathcal{D}_T \sim \mathcal{N}(\hat{\boldsymbol{f}}, \boldsymbol{\mathcal{K}})$, obtained from a noiseless dataset (indicated as black "+" points). The 95% confidence band of the GP is highlighted in gray. Three random example functions drawn from the GP are also shown.
Source: [177]

The distribution $f^*_{n+j}|\hat{\boldsymbol{f}}$ is typically approximated as a Gaussian with mean and variance given by:

$$f^*_{n+j}|\hat{\boldsymbol{f}} \sim \mathcal{N}\left( \boldsymbol{K}_*^T \mathbf{K}^{-1} \hat{\boldsymbol{f}} \ , \ K^*_{n+j} - \boldsymbol{K}_*^T \boldsymbol{\mathcal{K}}^{-1} \boldsymbol{K}_* \right), \tag{5.29}$$

where $\boldsymbol{K}_* \equiv [K(\boldsymbol{X}_{n+j}, \boldsymbol{X}_1) \ \ldots \ K(\boldsymbol{X}_{n+j}, \boldsymbol{X}_n)]^T$ and $K^*_{n+j} \equiv K(\boldsymbol{X}_{n+j}, \boldsymbol{X}_{n+j})$. The explanation of this expression is given in the box below:

It is assumed that both the training and the test latent function outputs belong to a GP. Namely, the prior distribution for both is multivariate Gaussian [194]:

$$\begin{bmatrix} \boldsymbol{f} \\ f^*_{n+j} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \boldsymbol{K}_* \\ \boldsymbol{K}_*^T & K^*_{n+j} \end{bmatrix} \right), \tag{5.30}$$

which corresponds to the definition of a multivariate Gaussian from a GP in Eq. 5.9. Section 9.3 of [197] indicates the following identity:

$$\begin{gathered} \text{Given } \begin{bmatrix} \mathfrak{u} \\ \mathfrak{d} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \boldsymbol{\mu}_\mathfrak{u} \\ \boldsymbol{\mu}_\mathfrak{d} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix} \right), \text{ then:} \\ \mathfrak{d}|\mathfrak{u} \sim \mathcal{N}\left( \boldsymbol{\mu}_\mathfrak{d} + \mathbf{C}^T\mathbf{A}^{-1}(\mathfrak{u} - \boldsymbol{\mu}_\mathfrak{u}) \ , \ \mathbf{B} - \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C} \right). \end{gathered} \tag{5.31}$$

Applying this identity to Eq. 5.30 gives the distribution of $f^*_{n+j}|\boldsymbol{f}$:

$$f^*_{n+j}|\boldsymbol{f} \sim \mathcal{N}\left( \boldsymbol{K}_*^T \mathbf{K}^{-1} \boldsymbol{f} \ , \ K^*_{n+j} - \boldsymbol{K}_*^T \mathbf{K}^{-1} \boldsymbol{K}_* \right). \tag{5.32}$$

The typical approach for obtaining $f^*_{n+j}|\hat{\boldsymbol{f}}$ is then to substitute $\boldsymbol{f}$ for $\hat{\boldsymbol{f}}$ in the mean, and substitute $\mathbf{K}$ for $\boldsymbol{\mathcal{K}}$ in the variance [177] [194]. This produces Eq. 5.29, but it should be noted that the mean in Eq. 5.29 no longer conforms to the identity given above. Nonetheless, the values of $\mathbf{K}$ and $\boldsymbol{\mathcal{K}}$ are similar enough for this inconsistency to not be very significant [194].

In the second stage of the GP classification algorithm, the $\mathbb{R} \to [0, 1]$ mapping, the likelihood $L_{n+j}$ is obtained via a weighted average of the distribution of $f^*_{n+j}|\hat{\boldsymbol{f}}$ with the outputs of the squashing function given $f^*_{n+j}$, $\pi(f^*_{n+j})$. The squashing function can be any sigmoidal function [177], a common choice being the Gaussian cumulative distribution function (*cdf*) [194]. So $L_{n+j}$ is given as:

$$L_{n+j} = \int \mathrm{d}f^*_{n+j} \ \pi(f^*_{n+j}) \times Pr(f^*_{n+j}|\hat{\boldsymbol{f}}). \tag{5.33}$$

In practice, GP classification is not a very prevalent method. In astronomy and experimental physics, GP regression has become increasingly adopted since 2015, however the use of GP classification has remained rare (Figure 5.11). The fact that its parameters are adjusted algorithmically during the $\boldsymbol{\chi} \to \mathbb{R}$ training stage makes GP classification an attractive "out-of-the-box" solution, as it can be expected to perform comparatively well even without prior calibration [176]. The main difficulty in applying GPs to machine learning purposes is that the training stage becomes very slow for large datasets: the computation time increases with the cube of the number of datapoints [198].

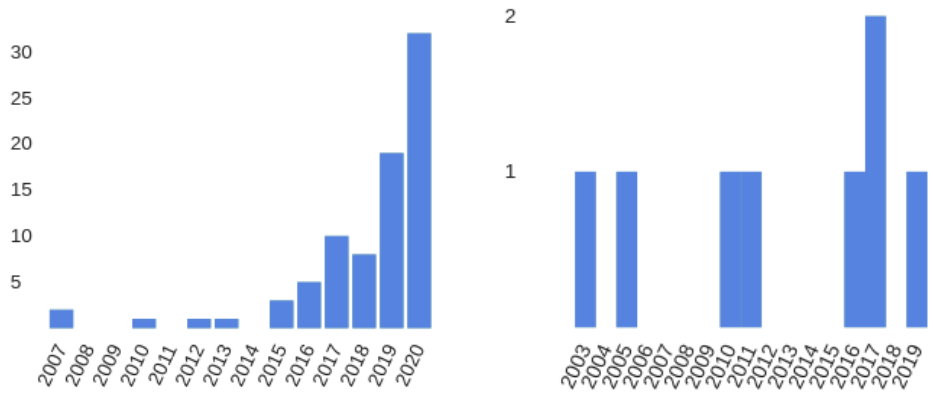


FIGURE 5.11: Number of papers published per year before 2021, related GP regression (*left panel*) and GP classification (*right panel*), present in the SAO/NASA Astrophysics Data System.
Source: `adsabs.harvard.edu` (*retrieved 12 Apr. 2021*)

### 5.1.4 The random forests classifier

The random forests (RF) classifier is an ensemble method of decision trees. *Ensemble methods* assign the same task to several models and combine their results together in order to obtain a better classification performance (Chapter 50 of [199]). Trained *decision trees* (DTs) assign a label $Y_{n+j}^{\text{guess}}$ to a test datapoint $\boldsymbol{X}_{n+j}$ according to a sequence of branching decisions based on feature values. DTs are constructed using training data by algorithmically identifying regions in feature space that are strongly balanced in favor of one of the classes. An example of a decision tree is shown in Figure 5.12 **{fn5.1i}**. DTs on their own suffer from *instability*: small variations in the training data can lead to significantly different classifications of the test data [200]. The RF classifier is an improvement on DTs, with improved stability.

Given an original training dataset $\mathcal{D}_T^0 \equiv (\boldsymbol{X}_i, Y_i), i = 1, \ldots, n$, the RF classifier typically constructs the decision boundary in the following manner:

1. Construct a set of $r$ *bootstrap samples* $\boldsymbol{\mathcal{D}}_T^{\text{boot}} \equiv \left\{\mathcal{D}_T^1, \ldots, \mathcal{D}_T^r\right\}$;
    - *Each individual bootstrap sample* $\mathcal{D}_T^k, k = 1, \ldots, r$ *is constructed from* $\mathcal{D}_T^0$ *by sampling* $n$ $(\boldsymbol{X}_i, Y_i)$ *pairs with replacement.*
2. Construct $r$ decision trees, $\mathrm{T}^k, k = 1, \ldots, r$, by training each one on a different bootstrap sample (this approach is termed *bagging*);
3. The class $Y_{n+j}^{\text{guess}}$ assigned to a test datapoint $\boldsymbol{X}_{n+j}$ is given by an average of the classes assigned by the $r$ DTs.

**{fn5.1i}** — The example presented is for discrete feature values, but DTs are also applicable to continuous features, by assigning decisions to ranges instead of discrete values. For example, the "High" and "Normal" values of humidity can be defined according to some percentage: "High" for humidity $> 85\%$ and "Normal" for humidity $\leq 85\%$.

| Day | Outlook | Temperature | Humidity | Wind | PlayTennis |
|---|---|---|---|---|---|
| $D_1$ | Sunny | Hot | High | Weak | No |
| $D_2$ | Sunny | Hot | High | Strong | No |
| $D_3$ | Overcast | Hot | High | Weak | Yes |
| $D_4$ | Rain | Mild | High | Weak | Yes |
| $D_5$ | Rain | Cool | Normal | Weak | Yes |
| $D_6$ | Rain | Cool | Normal | Weak | No |
| $D_7$ | Overcast | Cool | Normal | Strong | Yes |
| $D_8$ | Sunny | Mild | High | Strong | No |
| $D_9$ | Sunny | Cool | Normal | Weak | Yes |
| $D_{10}$ | Rain | Mild | Normal | Weak | Yes |
| $D_{11}$ | Sunny | Mild | Normal | Weak | Yes |
| $D_{12}$ | Overcast | Mild | High | Strong | Yes |
| $D_{13}$ | Overcast | Hot | Normal | Weak | Yes |
| $D_{14}$ | Rain | Mild | High | Strong | No |

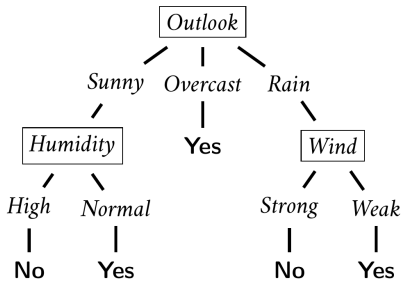


FIGURE 5.12: *Left panel:* Example dataset where the feature space represents the weather and the labels indicate whether someone played tennis on that day or not (the "PlayTennis" column). *Right panel:* A DT constructed on the above dataset. Note that the subset for the "Overcast" "Outlook" was not subdivided further, as all the "Overcast" entries had the same "PlayTennis" value. Note also that the "Temperature" does not significantly affect the outcome, and so was not included in the DT.
Example source: [201]

Training a DT is typically achieved with a recursive algorithm that splits the training dataset into subsets based on the feature values, until each dataset is balanced strongly in favor of one class. Typically the ID3 algorithm is used for this purpose [202]. This algorithm selects the best decision attribute based on some criterion comparing the randomness in the labels of the original set against the randomness in the labels of the new subset. To avoid overfitting, some splits are later removed: a procedure called *pruning*. For more information, see [203]. Figure 5.13 shows a comparison between the decision boundary constructed by a single DT and the decision boundary constructed by a vote of 25 DTs in the case of a problem difficult for DTs: decision boundaries along the diagonal of two features.

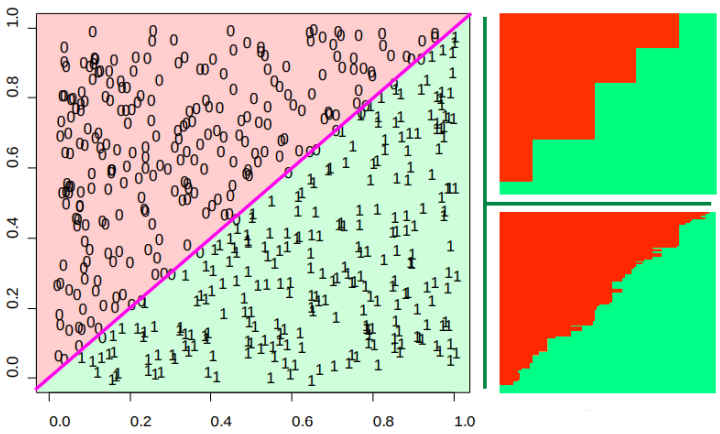


FIGURE 5.13: *Left panel:* 2D classification training dataset where the correct decision boundary (the magenta diagonal line) is not strictly defined on either of the feature values individually. This type of problem is difficult for DTs due to the pruning procedure. *Right upper panel:* Decision boundary constructed by a single DT. *Right lower panel:* Decision boundary contructed by a RF with 25 bootstrap samples.
Source: [204]

In practice, the RF algorithm is considered among the most versatile for classification of tabular data (*i.e.* where each datapoint is a feature vector, and not raw data, like images, audio or text), due to its robustness to minor changes in the training set and ease of handling missing information. RF classification is commonly applied in banking and medicine, namely for fraud detection and disease identification, respectively. In astronomy and experimental physics, RF classification has had increased adoption since 2014, and is currently a comparatively popular method (Figure 5.4). It has been used namely for classifying stellar spectra [205], galaxy and supernova classification [206] and quasar discrimination [207], background discrimination on the TACTIC TeV $\gamma$-ray telescope [208], and particle identification in the ALICE experiment [209]. The main drawback of RF classification is the difficulty in interpreting the obtained results.

## 5.2 Performance assessment of a binary classifier

This section serves as a brief introduction to the tools and approach for performance assessment of a binary classifier used in this work. For a more in-depth explanation of the implementation, see Sections 6.6.3 and 6.6.4.

When a classifier is given a binary test set pair ($Y_k \in \{\mathsf{N}, \mathsf{P}\}$ , $k \in 1, \ldots, n+m$), for an object $j$ from the test set with true label $Y_{n+j}^{\text{actual}}$, the classifier guesses a label $Y_{n+j}^{\text{guess}}$. There are four possible outcomes for the classification of the object:

- $Y_{n+j}^{\text{guess}} = \mathsf{N} = Y_{n+j}^{\text{actual}}$, a *true negative,* TN;
- $Y_{n+j}^{\text{guess}} = \mathsf{P} = Y_{n+j}^{\text{actual}}$, a *true positive,* TP;
- $Y_{n+j}^{\text{guess}} = \mathsf{N} \neq Y_{n+j}^{\text{actual}}$, a *false negative,* FN;
- $Y_{n+j}^{\text{guess}} = \mathsf{P} \neq Y_{n+j}^{\text{actual}}$, a *false positive,* FP.

For the entire test set, the total number of instances of each outcome are referred to as $TN$, $TP$, $FN$ and $FP$, respectively. From these numbers of instances, four *performance parameters* can be calculated:

- *True negative rate*, tn, defined as $\mathsf{tn} \equiv \frac{TN}{TN+FP}$;
- *True positive rate*, tp, defined as $\mathsf{tp} \equiv \frac{TP}{TP+FN}$;
- *False negative rate*, fn, defined as $\mathsf{fn} \equiv \frac{FN}{TP+FN}$;
- *False positive rate*, fp, defined as $\mathsf{fp} \equiv \frac{FP}{TN+FP}$.

These performance parameters can be arranged in a *confusion matrix* as per Table 5.2. Note that $\mathsf{tn} = 1 - \mathsf{fp}$ and $\mathsf{fn} = 1 - \mathsf{tp}$, so only two performance parameters will be used in this work. Sets of fp and tp values from different classifier samples will be organized in various manners to produce a more general picture of the classifier performance than is possible with just a single confusion matrix.

| | *guess label* | | |
|---|---|---|---|
| *actual label* | tn | fp | *1e* |
| | fn | tp | $0\nu\beta\beta$ |
| | *1e* | $0\nu\beta\beta$ | |

TABLE 5.2: The performance parameters organized into a confusion matrix. When the guess label coincides with the actual label, the output is a "true" outcome, else it is a false outcome.

Summarily, the performance assessment tests must clearly display two aspects: how well a classifier is expected to work, namely in deployment; and the evolution of the performance with sample size. Section 5.2.1 describes these aspects in detail. Unfortunately there is no single test that expresses them both in an easily readable manner, and different tests will be better suited for different applications. So to draw a complete picture of the classifier performance, this work primarily uses three methods:

- **Convergence tests;**
  - *This test serves to compare the behavior of the performance parameters of different classifier configurations for increasing sample size;*
- **Receiver operating characteristic (ROC) curves;**
  - *Serve to assess a classifier's performance in absolute terms;*
- **Cost curves.**
  - *Effective for comparing the performance of different classifiers during deployment.*

The fundamentals of each test are described in Sections 5.2.2-5.2.4. For more practical information, see Section 6.6.

### 5.2.1 Aspects to consider during assessment

A classifier that performs well for one usage is not guaranteed to perform well for others. For example, a classifier trained with a dataset balanced in favor of one class may then underperform when given a test dataset with the opposite balancing. Or even a trained classifier that performs well on the test data may then underperform with deployment data. These are two examples of the performance being affected by the balancing of the dataset. Considering that this work focuses on seeing a previously undetected rare event, the balancing is not known exactly, making it especially important to predict the classifier's performance for a balancing other than that of the train / test data. The balancing will be quantified with $Pr(\mathsf{P}) \in [0, 1]$, representing the proportion of the deployment data that is P [210].

But even if the balancing remains the same between different applications, a classifier can still be appropriate in one context but unacceptable in another. The two different types of misclassification, FP and FN, can incur different penalties depending on the context. For example in a cancer study an overzealous diagnosis (FP) is inconvenient but a failure to detect (FN) is fatal, while in this work a false discovery (FP) would be worse than failing to detect $0\nu\beta\beta$ (FN). The study of imbalanced binary classification includes methods for *cost-sensitive learning* (CSL) [211], which introduce the concept of different *misclassification costs* for FP and FN: $Co(\mathsf{FP})$ and $Co(\mathsf{FN})$. The misclassification cost is a relative unitless measure of the penalty incurred

by the two types of misclassification, a higher cost given to the worse type. In CSL, the goal is to construct a binary classifier that minimizes the total cost, $Co$, defined as:

$$Co \equiv Co(\mathsf{FN}) \times FN + Co(\mathsf{FP}) \times FP. \tag{5.34}$$

The ideal values of $Co(\mathsf{FP})$ and $Co(\mathsf{FN})$ for this work depend on the statistical model (Section 4.2) and their calculation is left for future work, but it is assumed that $Co(\mathsf{FP}) > Co(\mathsf{FN})$. With cost imbalance, a higher priority will be put on minimizing $FP$ than on maximizing $TP$.

Finally, given two identically distributed training / test pairs, their outputted performance parameters will likely differ due to statistical fluctuations: at some sample size, they will be on a distribution with some mean and width. As the sample size increases, the distribution will likely narrow as the mean converges on some asymptotically best performance (Figure 5.14). If this does not happen (Figures 5.15 and 5.16), it can be because the dataset is too sparse to construct an accurate decision boundary, or because some datapoints are very far away from all the others in feature space, or it can be an indication of overfitting. If a classifier algorithm consistently behaves in a more predictable manner than another one, it can be a reason to prefer it over the unpredictable one. Given an ensemble of training / test sample pairs of a given sample size, the statistics of its performance parameters serve as an estimate of the actual performance distribution for that sample size. Depending on the method used for splitting the data into test and training sets, the estimate will have a larger variance or bias relative to the actual distribution (the bias-variance tradeoff), so the approach must be chosen according to the intended goal (see Section 6.6).

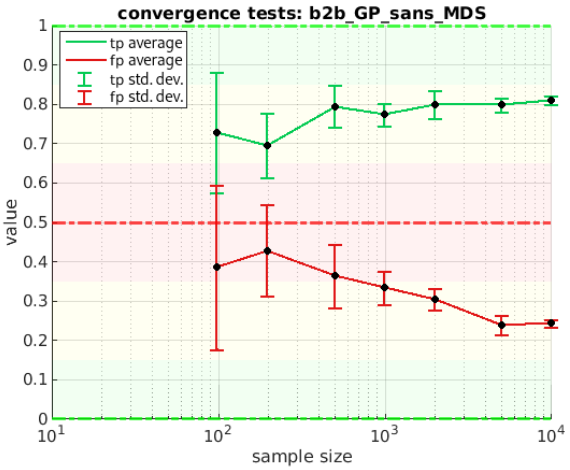


FIGURE 5.14: An example of a classifier whose performance parameters display good behavior (for the meaning of `b2b_GP_sans_MDS` , see Sections 6.4.3 and 6.6.1). The "error bars" here are in fact standard deviation estimates: the real error bars are $\sqrt{10}$ times narrower. Both performance parameters converge for increasing sample size, with fp lowering as tp does the opposite, the performance distribution consistently narrowing as this happens.

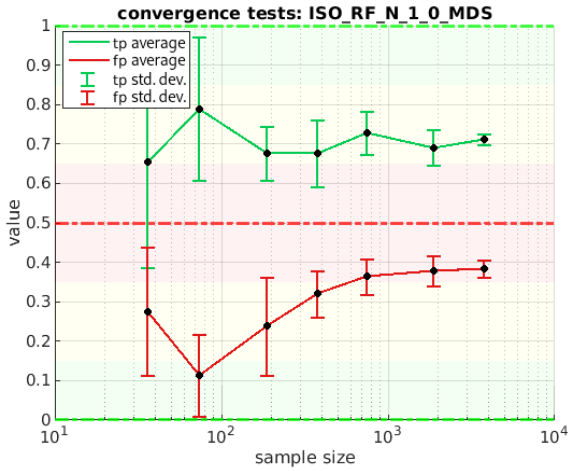


FIGURE 5.15: An example of a classifier whose performance parameters display bad behavior (for the meaning of `ISO_RF_N_1_0_MDS` , see Sections 6.4.3 and 6.6.3). The "error bars" here are in fact standard deviation estimates: the real error bars are $\sqrt{10}$ times narrower. The variance of both parameters decreases with sample size, but fp is increasing, clearly indicating that the classifier is failing to identify the shape of the class embeddings.

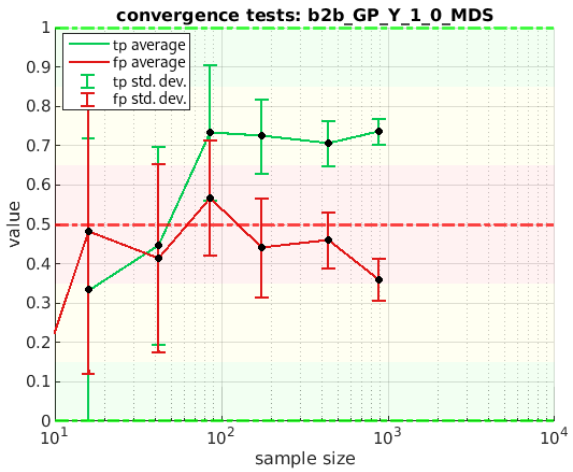


FIGURE 5.16: An example of a classifier whose performance parameters display erratic behavior (for the meaning of `ISO_RF_N_1_0_MDS` , see Sections 6.4.3 and 6.6.3). The "error bars" here are in fact standard deviation estimates: the real error bars are $\sqrt{10}$ times narrower. The parameters do not converge and instead just vary chaotically.

### 5.2.2 Convergence test

In a convergence test, the performance parameters are gauged over varying sample sizes for a given training / test pair **{fn5.2j}**: for each sample size, an ensemble of some number of samples is obtained, and as the sample size increases, the parameter variance estimate is expected to diminish as the parameter averages converge on two asymptotes. This test assesses both the classifier's performance and its reliability: the performance of some classifier can be very consistent (see Figure 5.14) or vary erratically (see Figure 5.16) depending on the training and test sets, and it is preferable that it be consistent. Considering the number of different configurations that had to be compared in this work **{fn5.2k}**, when some classifier misbehaved ultimately little attention was paid to understanding *why* that happened, opting instead for merely selecting the best-behaving ones out of the large available pool.

### 5.2.3 Receiver Operating Characteristic curves

The ROC curve is a popular performance test that plots the tp values obtained by a classifier at varying likelihood thresholds for positive classification, $\tau \in [0, 1]$, against the corresponding fp values (see Figure 5.17). A classifier's decision boundary can be shifted by changing $\tau$: a low value yields a high rate of positive classifications, so both high fp and tp. As $\tau$ increases, fp and tp both decrease. But one of the two performance parameters will decrease faster than the other one, and the lower the ratio fp/tp is as $\tau$ increases, the better the performance of the classifier. As such, the ROC curve expresses explicitly the range of perfomance that is expected to be achievable so the options are to either adjust $\tau$ for some more appropriate pair of performance parameters given the dataset, or to just opt for another classifier entirely.

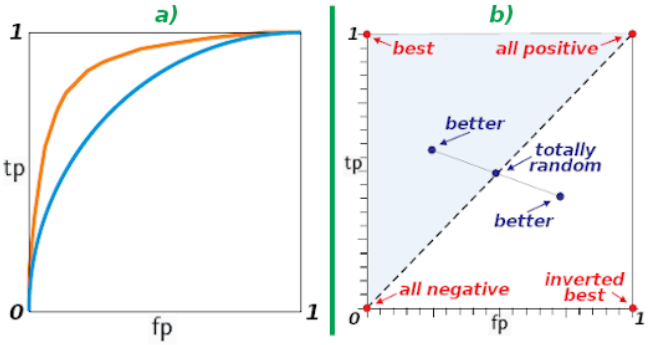


FIGURE 5.17: a) Two ROC curves for different classifiers. The convex hull of the orange classifier contains the entirety of the blue classifier, so the orange classifier is better than the blue one [212] b) Description of different regions in ROC space. The worst is for tp = fp, where the classification is random. On either side of this diagonal, the quality of the classifier improves.
Source: [212]

Using the convex hull of the ROC curve allows to quantify the performance in absolute terms with a single measure: the area under the curve *(AUC)* [213]. The ratio fp/tp for some classifiers will naturally be smaller than for others, resulting in convex hulls that are consistently further away from the tp = fp diagonal line (as is the case when comparing the orange and blue

**{fn5.2j}** — In this work, a sampling method was chosen that would prevent nuisance parameters from affecting the performance as the sample size changed. More information in Section 6.6.3.

**{fn5.2k}** — 24 configurations, half of which imply 5 convergence tests instead of just one, all for 2 different situations. In total 144 different convergence tests were performed. See Section 6.6.1 for more details.

classifiers in Figure 5.17 panel a) ). So the larger the value of the AUC, the better the classifier will be in absolute terms, provided that the curves do not intersect. A lower AUC for some classifier compared to another one is generally indication that in no case does it ever perform better **{fn5.2l}**.

However if the curves do intersect, then this simple analysis with the AUC is no longer accurate: for some $\tau$ range one of the classifiers is preferable, else it's the other one. In this case the choice of classifier has to be based on the balancing and costs in deployment. For this purpose the ROC curve is somewhat capable — it is possible, but unintuitive, to make a cost-sensitive assessment by manipulating the data in the ROC curve according to techniques shown in [210] — but a better tool exists: cost curves.

### 5.2.4 Cost curves

The cost curve plots the likely misclassification cost incurred by a classifier against the balancing of the data, so it can be used to make a cost-sensitive comparison of the performance of different classifiers in deployment [210]. It is easy to construct using the ROC datapoints and is good for selecting the best classifier out of an assortment, but not so good for representing the classifier performance in absolute terms.

The vertical axis is the normalized expected total misclassification cost $E[Co] \in [0, 1]$, given as:

$$E[Co] = \frac{\varepsilon[Co]}{\max_{\mathsf{fn},\mathsf{fp}} \{\varepsilon[Co]\}}, \tag{5.35}$$

where the unnormalized expected total misclassification cost $\varepsilon[Co]$ is built using $Co$ from Section 5.2.1 as a basis:

$$\varepsilon[Co] = Co(\mathsf{FN}) \times \mathsf{fn} \times Pr(\mathsf{P}) + Co(\mathsf{FP}) \times \mathsf{fp} \times (1 - Pr(\mathsf{P})) \tag{5.36}$$

The test was designed so that a single curve would show the performance for any $Co(\mathsf{FP})/Co(\mathsf{FN})$ ratio, so the horizontal axis is defined as:

$$PC(\mathsf{P}) = Pr(\mathsf{P}) \times \frac{Co(\mathsf{FN})}{\max_{\mathsf{fn},\mathsf{fp}} \{\varepsilon[Co]\}}. \tag{5.37}$$

$PC(\mathsf{P}) \in [0, 1]$ is termed the *probability cost,* and $\max_{\mathsf{fn},\mathsf{fp}} \{\varepsilon[Co]\}$ is calculated by plugging $\mathsf{fp} = \mathsf{tp} = 1$ into $\varepsilon[Co]$. When $PC(\mathsf{P}) = 0$ we get $E[Co] = \mathsf{fp}$, and when $PC(\mathsf{P}) = 1$ we get $E[Co] = \mathsf{fn} = 1 - \mathsf{tp}$ so, considering that $E[Co]$ is linear along $PC(\mathsf{P})$ **{fn5.2m}**, each point on the ROC curve will correspond to a straight line on the probability cost space . Assuming that the best possible $\tau$ is selected at each $PC(\mathsf{P})$, the cost curve of the classifier is the lower envelope of all the straight lines (blue curve in Figure 5.18 panel b) ).

Given that the misclassification cost is an *"ad-hoc"* unitless measure, the cost curve is not very useful for performance assessment in absolute terms, but it is much better than the ROC curve for comparisons. In Figure 5.18 panel b), the two red diagonal lines represent the expected cost when not using any classifier at all: the $E[Co] = PC(\mathsf{P})$ line corresponds to labelling all test datapoints as $\mathsf{N}$; and the other line is for the opposite. If the classifier gives a lower $E[Co]$ for some operating point (*i.e.* some $PC(\mathsf{P})$ value), then at that operating point the classifier is better than nothing by some undetermined amount. This can be extended to comparing two

**{fn5.2l}** — Section 6.6.4 has a more in-depth explanation of how to interpret the value of the AUC.

**{fn5.2m}** — $1 - PC(\mathsf{P}) = Co(\mathsf{FP})(1 - Pr(\mathsf{P}))/\max_{\mathsf{fn},\mathsf{fp}} \{\varepsilon[Co]\}$, so $E[Co]$ can be expressed in terms of $PC(\mathsf{P})$ as $E[Co] = PC(\mathsf{P}) \times \mathsf{fn} + (1 - PC(\mathsf{P})) \times \mathsf{fp}$.

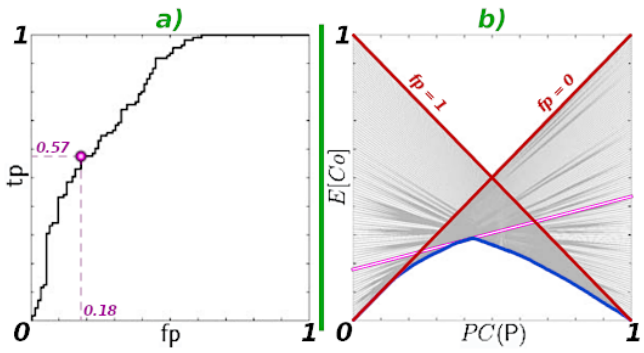


FIGURE 5.18: An example ROC curve ( a) ) being transferred into cost curve space ( b) ). Each point in the ROC curve corresponds to a straight line in the cost curve space, as exemplified by the purple point in a) being transferred to b). The cost curve itself is highlighted in blue, constructed out of the cost minima at each value of $PC(\mathsf{P})$. Note that the bottom red triangle in b) represents the cost curve of a classifier that accepts or rejects all events, meaning that the classifier in this case is no better than no classifier at all when the datasets are strongly imbalanced in favor of the N class.
Source: [210]

classifiers: if some classifier has a lower $E[Co]$ than another one for some set of operating points, then it is better than the other one for that range, also by an undetermined amount. Since $PC(\mathsf{P})$ relates explicitly to misclassification costs and to the balancing of datasets, the cost curve can be used to choose which classifier is better for some usage. With some rough guess of $Co(\mathsf{FP})$, $Co(\mathsf{FP})/Co(\mathsf{FN})$ and $PC(\mathsf{P})$, a region of interest on the cost curve can be chosen: for example, for analyzing the data in this work (Section 7.2), it is taken to be $PC(\mathsf{P}) < 0.5$.

# Chapter 6

# Description of the Simulation and Classification Procedure

The work summarized in this chapter consists of testing how well varying configurations of the kNN, SVM, GP and RF binary classifiers (see Section 5.1) perform at discriminating the waveforms produced in LZ by neutrinoless double beta decay, $0\nu\beta\beta$, events (see Sections 3.1 and 4.1) from the waveforms produced by the undesired single electron ($1e$) LZ background events with similar energy (see Section 4.2). Since there is no reference experimental $0\nu\beta\beta$ data for LZ, all the training data had to be simulated first. Developing an accurate but sufficiently fast simulator of $0\nu\beta\beta$ and $1e$ data and subsequently implementing it was by far the most time-consuming part of this work. Different processing approaches were then attempted on the resulting datasets, the best approach selected through comparative methods and then the best possible performance was assessed quantitatively.

The chapter is structured in the following way:

1. Section 6.1 serves to explain the challenges that led to the adoption of the decided-upon simulation and classification procedure;
2. Section 6.2 gives an overview of the structure of the procedure, explaining the design philosophy and describing the purpose of each step;
3. Sections 6.3 - 6.6 detail the implementation of each step of the procedure in sequence.

## 6.1 Preliminary considerations

Designing the simulation and classification procedure requires at least a rough prior sense of the capability of the LZ detector to discriminate event topology in the produced waveforms: the base simulation approach (Section 6.1.1) must be compatible with the quality of information that the waveform can provide. The most critical concerns in this regard are the following:

***Topology discrimination***

- *Topology is blurred due to longitudinal and transverse diffusion (Figure 4.6). In preliminary testing, it was determined that longitudinal diffusion blurred out all recognizable detail in the topology of the electron depositions beyond a relatively shallow depth of ∼ 400-500 mm. (Figures 6.1 and 6.2 show the progression of the blurring at different depths);*
- *The gate grid funnels the drift electrons into horizontal bands 5 mm apart (Figure 4.6). This drastically reduces the efficiency of transverse topology discrimination.*

***Dataset optimization***

- *The optimal representation of the dataset for the classifier must be used (Section 6.1.2);*
- *The dataset must be kept from becoming sparse due to high dimensionality (Section 6.1.3).*

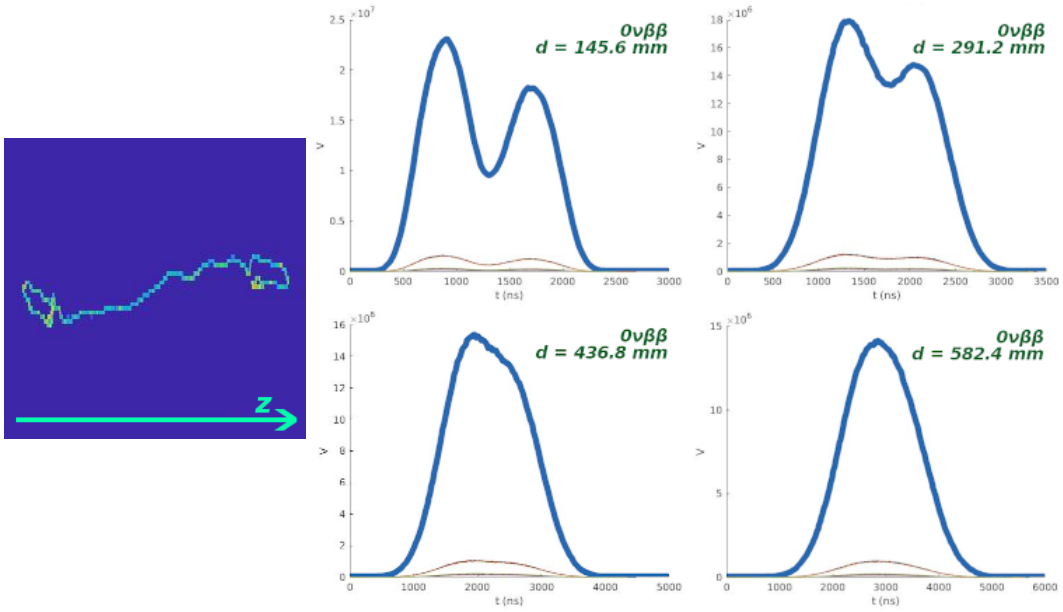


FIGURE 6.1: Example waveforms outputted by the PMT arrays for the pictured $0\nu\beta\beta$ secondary electron deposition (left) when emitted at increasing depths $d$ below the gas gap. The thicker line is the pulse outputted by the PMT directly above the event, and the thin lines are due to other PMTs in the array. The $z$ axis in the left image is along the horizontal direction.

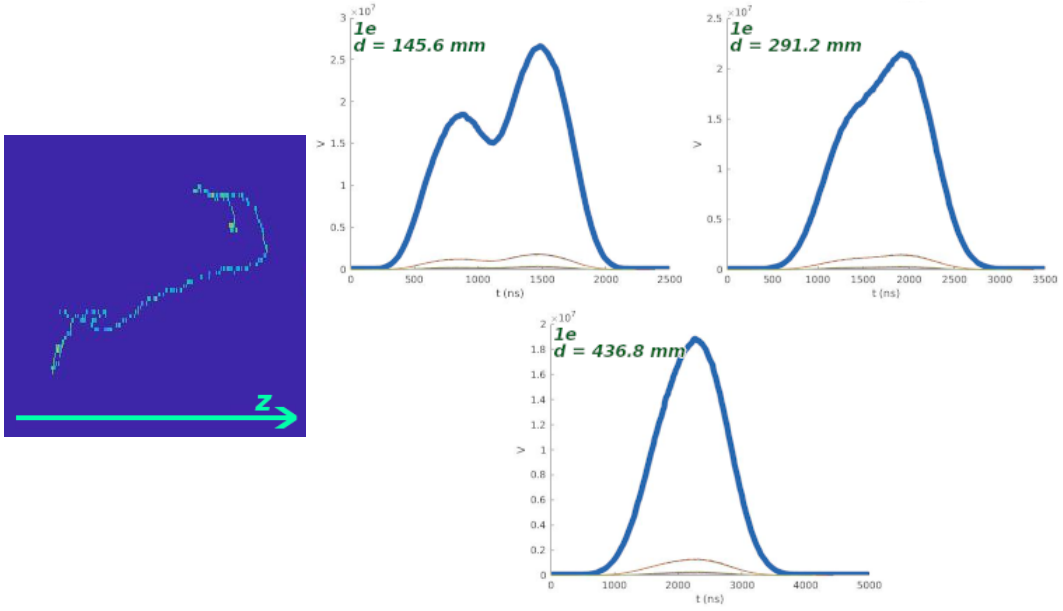


FIGURE 6.2: Example waveforms outputted by the PMT array for the $1e$ secondary electron deposition (left) when emitted at increasing depths $d$ below the gas gap. The thicker line is the pulse outputted by the PMT directly above the event, and the thin lines are due to other PMTs in the array. The $z$ axis in the left image is along the horizontal direction.

### 6.1.1 Base simulation and classification approach

The base approach was to begin with simpler, more optimistic scenarios and progress to more complex and realistic ones. Given this approach, the current form of the classification procedure does not compare the pulse outputs from different PMT channels, opting instead for summing all the PMT outputs together resulting in a 1-dimensional waveform rather than a 2-dimensional one.

The simulation was performed with more favorable datasets first, and adverse contributions would then be added iteratively until the simulation either completely recreated the real-life scenario, or the classifier performed too poorly to continue. This approach would then give insight into the most crucial factors hindering performance. The full scenario plan is listed in Table 6.1, but it was determined that simulations beyond the ISO scenario were not necessary for the current form of the classification procedure, as the performance assessment concluded that in deployment the addition of the classifier as a background rejection cut would neither improve nor worsen hypothesis testing counting statistics (see Section 7.3), so only the b2b and ISO scenarios were simulated.

| ***Scenario number*** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| ***Scenario name*** | **b2b** | **ISO** | **DEC0** | **GRID** | **MAT** | **SAT** |
| ***ASPECTS OF SIMULATION*** | *Settings* | | | | | |
| ***1. Realistic primary energy distribution*** | No | No | Yes | Yes | Yes | Yes |
| ***2. Isotropic primaries*** | No | Yes | Yes | Yes | Yes | Yes |
| ***3. Realistic energy deposition into LXe*** | No | Yes | Yes | Yes | Yes | Yes |
| ***4. Electron drift*** | Yes | Yes | Yes | Yes | Yes | Yes |
| ***5. Diffusion*** | Yes | Yes | Yes | Yes | Yes | Yes |
| ***6. Grid field distortion*** | No | No | No | Yes | Yes | Yes |
| ***7. Realistic gas gap*** | No | No | No | No | Yes | Yes |
| ***8. Electrophoton trajectories*** | Yes | Yes | Yes | Yes | Yes | Yes |
| ***9. Realistic detector size*** | No | No | No | No | Yes | Yes |
| ***10. Realistic detector materials*** | No | No | No | No | Yes | Yes |
| ***11. PMT array saturation*** | No | No | No | No | No | Yes |
| ***SUMMARY*** | | | | | | |
| **b2b** | best case scenario, initial primary direction always vertical | | | | | |
| **ISO** | isotropic initial primaries, equal energy $0\nu\beta\beta$ primaries | | | | | |
| **DEC0** | $0\nu\beta\beta$ primaries have realistic energy | | | | | |
| **GRID** | gate grid distorts electric field | | | | | |
| **MAT** | PTFE walls, detector dimensions correspond to LZ ones | | | | | |
| **SAT** | PMTs saturate for high luminescence | | | | | |

Table 6.1: Full scenario plan. In this work, the scenarios are employed in sequence, beginning from the smallest scenario numbers.

Furthermore, time constraints, together with the very slow simulation time (Section 6.3.3), meant that at most the events would have to be simulated only in a small number of $xy, z$ positions. It was decided that the best use of the available resources was to generate all the events at radius $\rho^2 = 0$ and depth $d = 0.1 \times d_{\text{max}} = 145.6$ mm below the gas gap (see Section 4.1). At this depth the diffusion already blurs near-horizontal events enough for discrimination to be difficult, but not enough to render it impossible for near-vertical events. This choice has benefits and drawbacks, listed on the next page. It was decided that the benefits of this choice outweigh the drawbacks, as those can be compensated for to a certain degree, while the drawbacks of the alternative would mean that little information could be confidently gleaned from this work.

*Benefits*

**Allows for a detailed analysis of the classifier behavior at that depth**

Improved statistics give a better sense of the region in feature space corresponding to each class (see Section 5.1), and so it becomes easier to understand which subspaces are more critical for the classifier performance.

**A larger dataset results in a better training of the classifier**

The average classifier performance improves as the dataset becomes larger. This means that using small datasets for several depths would result in an underestimation of the classifier's performance. A large dataset at one position makes it clearer how to interpret the performance results.

*Drawbacks*

**Cannot model the dependence of the performance on the event position**

Due to the diffusion and gate grid funnelling, it can be assumed that the waveform will not vary significantly with the $xy$ position of the event relative to the PMT array, it is still detrimental that this effect is not quantified. Regarding the variation of the classifier performance with depth, intuitively an event closer to the gas gap will be easier to discriminate than one further away, but again it cannot be quantified.

**Does not allow to calculate the depth beyond which the classifier is detrimental to the background rejection cuts**

Beyond a certain depth the classifier performance will degrade to where it becomes an additional counting error source, however the lack of a model of the performance variation with position means that the depth at which this happens is unknown.

### 6.1.2 Mapping from parameter space to feature space

As described in Chapter 5, each waveform is converted into a point in *parameter space*, and a deployed classifier will instead take as input a point in feature space. These spaces are different. The aim of parameterization is to be as descriptive as possible about the shape of the waveform using a numerical representation of its characteristics, but many of the extracted parameters will either not be decisive for the classification, or will not be in the ideal representation for the classifier **{fn6.1e}**. The distinction between parameter space and feature space serves to remove the non-decisive characteristics of the waveform and to optimize the representation of the decisive ones.

### 6.1.3 Curse of dimensionality

Another concern relates to the number of dimensions (the *dimensionality*) of the feature space. To construct an accurate decision boundary, the classifier must be given a space where the class datasets produce dense clusters of datapoints. Without this, it becomes difficult to discern which datapoints belong to a trend and which ones are outliers. A consequence of increasing the dimensionality of a dataset is that it becomes sparser — the so-called *curse of dimensionality* (Figure 6.3) — which would imply that great care must be taken to choose only uncorrelated features, so as to not compromise the discriminability of the class distributions. This would make feature choice a very laborious process. However there is an alternative — employed in this work — which is to liberally add dimensions to feature space when analysing the waveforms, and then apply a dimensionality reduction step to the dataset after the fact, thus recovering

**{fn6.1e}** — A simple example of this can be seen in Figure 5.13. Decision trees notoriously underperform for diagonal decision boundaries (Figure 5.13 Panel b) ), however by subtracting the two axes one can obtain a decision boundary with which the algorithm works much better.

the density of datapoints. This option seemed ideal, as it allowed for greater versatility when adding new features without the risk of compromising the classifier performance.

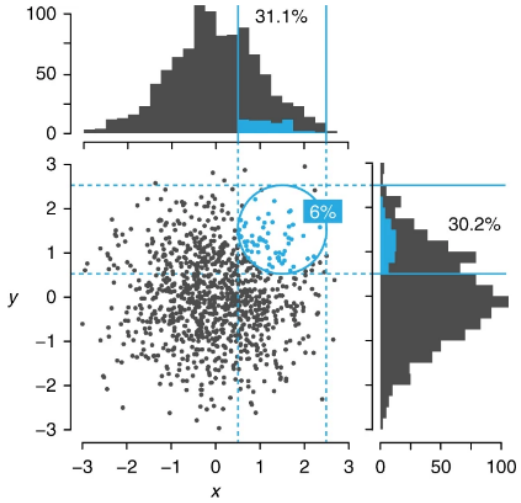


FIGURE 6.3: 1000 datapoints on a bivariate Gaussian distribution, along with the observed distributions of the datapoints when projected along $x$ and $y$. Selecting a region with only 6% of the datapoints in the dataset, its projections along a single variable both include over 30% of the datapoints.
Image source: [214]

## 6.2 Structure of simulation and classification procedure

Taking in consideration the comments in Section 6.1, Figure 6.4 shows a schematic of the chosen simulation and classification procedure, consisting of several steps divided into four stages. Since many of the components were expected to be subject to frequent changes, the procedure was designed to be modular and easily expandable, with placeholders set for functionality to be included in future work. Every depicted step directly uses the data obtained on the previous one. A summary of each stage follows. The next Sections describe each stage in detail.

The first stage (**simulation**, Section 6.3) takes as input the settings corresponding to a chosen scenario (Table 6.1), and produces a set of $0\nu\beta\beta$ and $1e$ waveforms generated by a simulation of the TPC. For each $0\nu\beta\beta$ or $1e$ event, the trajectory and energy deposition of the primary electrons in liquid xenon (LXe) is simulated, and the energy deposition is converted into the release of secondary electrons (Section 6.3.1). The `x,y,z` positions of the secondary electrons are then taken as input when simulating their drift up the TPC and diffusion, and then electroluminescence and S2 photon transport are simulated. As the S2 photons incide on the PMT array, the waveform is simulated, each PMT having its own channel (Section 6.3.2). Because the simulation is very resource-intensive (Section 6.3.3), it was run both locally and on remote machines (Section 6.3.4). The output of this stage is a set of $0\nu\beta\beta$ and $1e$ waveforms conforming to a certain scenario.

The second stage (**data extraction**, Section 6.4) takes as input the set of $0\nu\beta\beta$ and $1e$ waveforms, and converts them into points in feature space. The PMT channels are joined together (as discussed in Section 6.1) and the resulting 1-dimensional signal is subjected to some basic

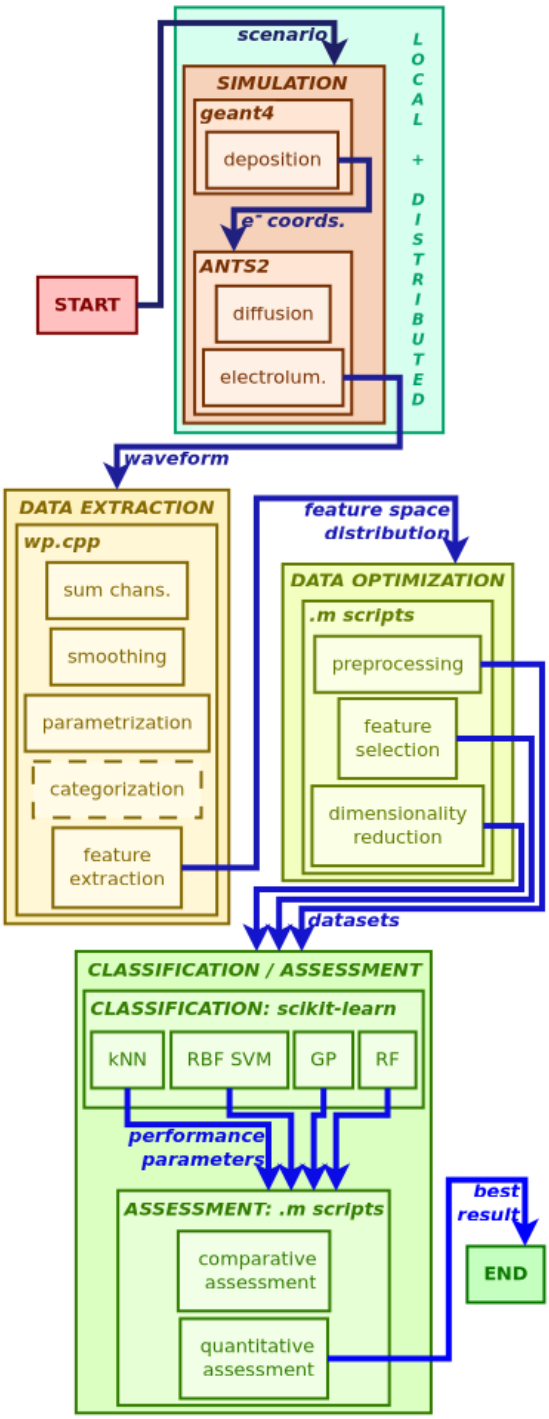


FIGURE 6.4: Schematic of the implemented procedure. The four stages correspond to the four colored rectangles. The first stage required the use of distributed computing, as indicated by the turquoise rectangle. The small off-white rectangles represent the steps in each stage, and the stage subtitles indicate the main software used. The arrows indicate the transfer of data between parts or stages, and the arrow subtitles indicate the form the data takes at the given point in the procedure.

preprocessing to aid in parameterization (Section 6.4.1). A parameterization step is then performed (Section 6.4.2), and before converting the signals from parameter space into feature

space (*feature extraction*, Section 6.4.4), there was an optional step where the dataset was broken up into categories subsets based on the signal morphology (*categorization*, Section 6.4.3)**.** The output of this stage is two sets of $0\nu\beta\beta$ and $1e$ positions in feature space, one with categorized data, and one with uncategorized data.

The third stage (**data optimization**, Section 6.5) takes as input the sets of $0\nu\beta\beta$ and $1e$ positions in feature space and finds the most optimal representation for the manifold embeddings (see Figure 5.3) corresponding to the datasets. It consists of a dimensionality reduction step (Section 6.5.3) preceded by two preparation steps: a preprocessing step where the dataset is adjusted to best work with the dimensionality reduction method (Section 6.5.1); and a feature selection step, in which irrelevant features are removed (Section 6.5.2). For each input position set, this stage outputs three datasets, representing the embedding following preprocessing, feature selection and dimensionality reduction.

The final stage (**classification / performance assessment**, Section 6.6) takes as input the datasets produced by all the preceding stages and outputs the best obtained performance measure. It gives each dataset as input to the four selected classification algorithms (procedure described in detail in Section 5.1, practical implementation in Section 6.6.2). The result is that the waveform sets obtained in the first stage can be processed by the defined procedure in 24 different ways, referred to in this work as *configurations* (Section 6.6.1). The best-performing configuration is then selected using a comparative performance assessment (Section 6.6.3), and then the result of the quantitative performance assessment (Section 6.6.4) of the best-performing configuration is given as output of the final stage.

## 6.3 Simulation

The simulation stage of the procedure for the two tested scenarios consists of the steps listed in the box below:

**Step 1.** `Geant4` simulation of energy deposition in LXe for $0\nu\beta\beta$ and $1e$ events with $Q = 2.458$ MeV. Different for each scenario (Section 6.3.1):

- **b2b:** *Only vertical events, with $Q/2$ energy $0\nu\beta\beta$ primary electrons (pgs. 92 - 94);*
- **ISO:** *Isotropic events, same energy on both $0\nu\beta\beta$ primary electrons (pg. 95).*

**Step 2.** Conversion of energy deposition into drift electrons. Different for each scenario (Section 6.3.1):

- **b2b:** *Coarse parameterization using experimental data (pgs. 94 - 95);*
- **ISO:** *Drift electron production model as in the LZ experiment (pgs. 95 - 97).*

**Step 3.** Setting event position: depth ($\texttt{d} = 145.6$ mm) and radius ($\texttt{r} = 0$ mm);

**Step 4.** Electron drift diffusion up to LXe surface (Section 6.3.2, pgs. 101 - 103);

**Step 5.** Electroluminescence and PMT array signal (Section 6.3.2, pgs. 103 - 104).

Section 6.3.3 describes the computational performance and baked-in limitations of the simulations. Section 6.3.4 describes the implementation of the simulation procedure for distributed computing and shows the final tally of simulations.

### 6.3.1 $0\nu\beta\beta$ and $1e$ deposition in LXe

This subsection corresponds to **Steps 1** and **2** from the previous box.

This part of the procedure was the only one that was approached very differently for the two studied scenarios. The procedures are described below in the sequence in which they were implemented. For each event, the final output of these two implementations was a text file containing a list of positions, and the number of drift electrons corresponding to that position. Two example excerpts (start and end) of a deposition file are presented in [215]. This output represents a track of point origins of large clusters of drift electrons (sometimes reaching into the thousands), not accounting for the spatial distribution of released electrons around each recoil site. However, diffusion very quickly washes out the initial positions, meaning that the expected effect on the produced waveforms is negligible. Additionally, these two implementations do not account for statistical fluctuation in the electron yields per energy loss: it is slated for future work.

#### Step 1: b2b scenario implementation

***Simulation physics list: selection and validation***

The $0\nu\beta\beta$ and $1e$ event primary electron tracks and energy depositions were simulated with `Geant4` version `10.4`, using the `FTFP_BERT_PEN` physics list (the default high-energy hadronic interaction physics list for this version [216]), together with the `Penelope` $e^{\pm}$ and photon electromagnetic models **{fn6.3a}**. `Penelope` is implemented according to the `PENELOPE-2008` code system [217] [218]. The `Geant4` implementation of `Penelope` includes models for ionization, photoelectric effect and bremsstrahlung, Compton scattering, Rayleigh scattering, inelastic scattering, pair production, and positron annihilation.

The `PENELOPE-2008` code system yields a description of electron transport for energies from 100 eV to 6 GeV that is in strong agreement with experimental data [219]. The elastic and multiple scattering model used in `Geant4` for electrons of energy $> 10$ eV is based on [220], an improvement on the `Penelope` code. Although `Geant4` multiple scattering simulations are shown to not agree with experimental data in backscattering experiments [221] [222], for cases of electron transport inside a single medium they are seen to output fairly accurate spatial distributions and depth dose profiles when given a sufficiently small maximum `Geant4` step size [223] [220], and hence are expected to be sufficiently faithful to experimental data for the purposes of this work.

Table 6.2 displays the relative frequency of the different processes accounted by the `Geant4` simulations, showing that for about three quarters of the event, `Penelope` is controlling the electron transport. Figure 6.5 displays the energy loss spectra for the different processes for 500 $1e$ and 500 $0\nu\beta\beta$ events together **{fn6.3b}**, showing that ionization processes transfer about 2 orders of magnitude more energy into electron recoil (ER) than does multiple scattering, meaning that the expected effect on the produced waveforms is negligible. The chosen maximum `Geant4` step size was 0.01 mm **{fn6.3c}**

**{fn6.3a}** — These $0\nu\beta\beta$ simulations were performed prior to this work, by Cedric Pereira.

**{fn6.3b}** — This was done because the spectra are seen to not be significantly different. The spectra for 500 $1e$ events and 500 $0\nu\beta\beta$ events individually are displayed in [224]

**{fn6.3c}** — Prior to this work, `Geant4` tests were performed by Cedric Pereira using `Option 3` (default), `Livermore` and `Penelope` EM models for maximum step sizes 0.05 mm, 0.03 mm, and 0.01 mm. $10^5$ b2b $0\nu\beta\beta$ events were simulated for all different configurations of model and step size, determining their energy deposition profile along $z$. This showed that any of the maximum `Geant4` step sizes outputted consistent results when using either the `Penelope` the `Livermore` EM model, but not `Option 3`.

| Process | Average Proportion | | Standard Deviation | | Model |
|---|---|---|---|---|---|
| ***Ionization*** | 74.30% | 74.10% | 0.89% | 0.79% | Penelope |
| ***Multiple Scattering*** | 24.67% | 24.75% | 0.66% | 0.48% | Geant4 |
| ***Photoelectric Effect*** | 0.68% | 0.75% | 0.37% | 0.40% | Penelope |
| ***Bremsstrahlung*** | 0.30% | 0.32% | 0.25% | 0.28% | Penelope |
| ***Compton Scattering*** | 0.02% | 0.05% | 0.05% | 0.11% | Penelope |
| ***Rayleigh Scattering*** | 0.03% | 0.03% | 0.06% | 0.06% | Penelope |
| | $0\nu\beta\beta$ | $1e$ | $0\nu\beta\beta$ | $1e$ | |

TABLE 6.2: Average relative frequencies **{fn6.3d}** of the different EM processes, as well as their corresponding standard deviations, calculated for a sample of 1300 $1e$ depositions and 1000 $0\nu\beta\beta$ depositions. Of note are the small values of the standard deviation, indicating that these relative frequencies are quite consistent.

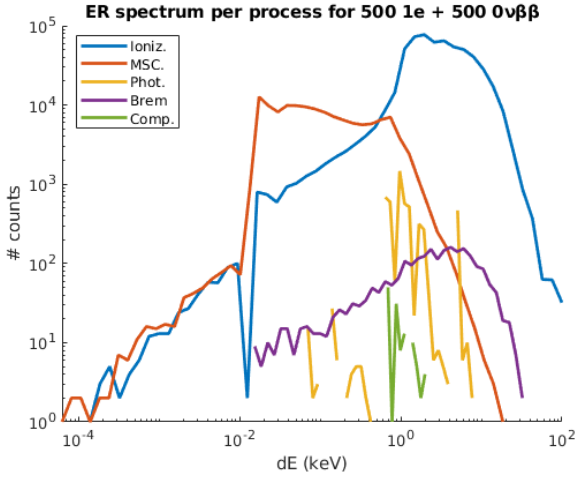


FIGURE 6.5: Spectrum of the energy loss from the primary electron that goes into ER. Of note is that not all of the energy lost by the primary necessarily goes into recoil. For example an ionization process will often lead to a child primary electron track. In these cases, part of the energy goes into ER, and part of the energy becomes the kinetic energy of the child electron. Rayleigh scattering is not displayed in this Figure because it doesn't transfer recoil energy.

### *Event simulation: sensitive volume definition, primary electron emission*

The simulations were performed by emitting the primary electrons from the center of a cube of width 100 cm, made of the material `G4_lXe`, defined as liquid xenon with density 2.953 g/cm$^3$ [225], and natural isotope composition [226] defined according to the NIST atomic weight and isotope composition database [227]. Since the applied drift field $E_{\text{drift}}$ of 31.0 V/mm (see Table 4.1) results in a mere 0.12 keV $\approx 5 \times 10^{-5} Q_{\beta\beta}$ energy increase across a 4 mm vertical track **{fn6.3e}**, it was seen as an insignificant contribution to the topology, and hence the electric field was not considered in the simulation.

**{fn6.3d}** — Relative frequencies by number of processes, not by energy loss. If it were by energy loss, the multiple scattering relative frequency would have been vanishingly small.

**{fn6.3e}** — The expected length of a $0\nu\beta\beta$ track.

For each $0\nu\beta\beta$ event, two primary electrons with energy 1.224 MeV were emitted simultaneously, one directly upward *(along the $z$ direction)* and one directly downward. For $1e$ events, one primary electron with energy 2.448 MeV was emitted directly downward. Later during data extraction approximately half of the $1e$ dataset, selected at random, was mirrored vertically (Section 6.4.1).

Each event returned the `x,y,z` position of the primary (and bremsstrahlung child primaries) at the end of each `Geant4` step, and the incident energy `E` and energy loss to ER, `dE`, during that `Geant4` step.

### Step 2: b2b scenario implementation

The energy loss to ER was converted into integer numbers of drift electrons per position using the ionization yield data from [228] as reference. Figure 6.6 shows the result of the parameterization used to recreate the results from the article. The parameterization is shown in [229]. The left panel of Figure 6.7 shows example topologies for $1e$ and $0\nu\beta\beta$ due to this implementation, and the right panel of Figure 6.7 shows statistics for the total number of S2 electrons produced per event of each class.

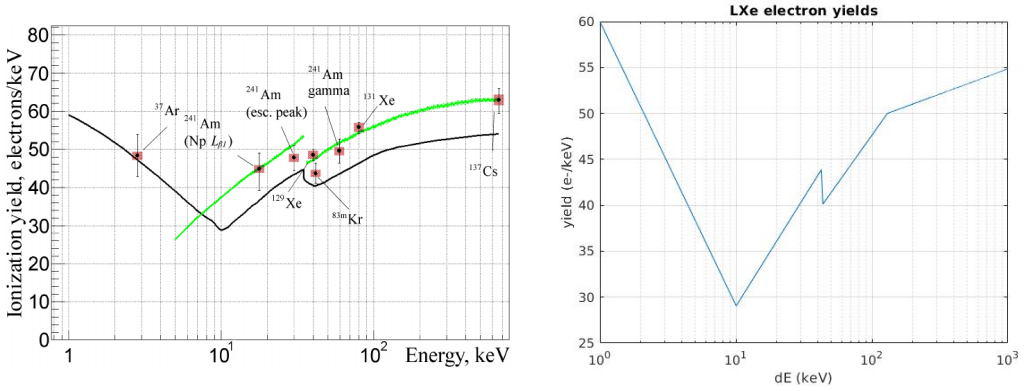


FIGURE 6.6: Comparison of experimental data for the ionization yield of LXe for electron recoils (left panel) with the electron yield model used in the b2b scenario, following a parameterization of [228] (right panel).
Left panel source: [228]

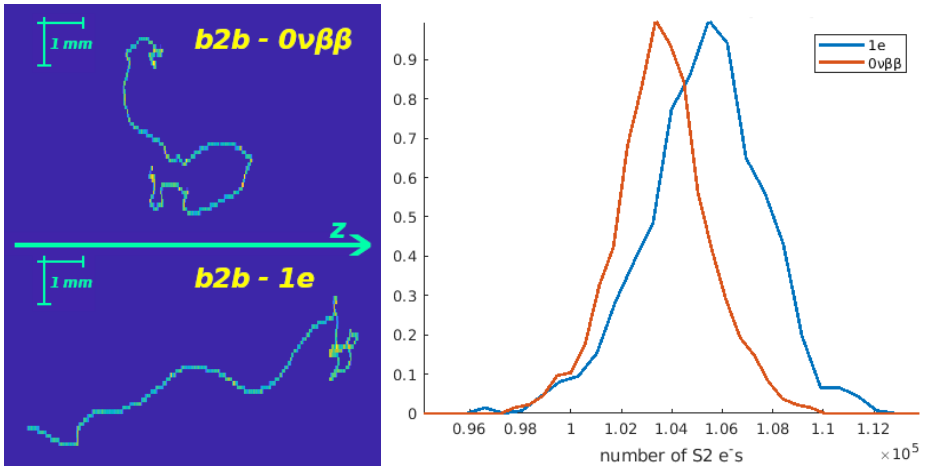


FIGURE 6.7: *Left panel:* Example $0\nu\beta\beta$ and $1e$ topologies generated using the b2b scenario implementation. *Right panel:* Statistics for the total number of drift electrons produced per event of each class for the b2b scenario.

**Goals of the ISO scenario implementation**

- Emit primary electrons isotropically;
- Implement drift electron production model used in the LZ experiment;
- Migrate the deposition step of the procedure into the same software as the rest of the simulation for improved portability.

**Step 1: ISO scenario implementation**

The simulation of the physics after deposition, namely drift diffusion, electroluminescence and waveform production, was performed on the `ANTS2` simulation and data processing package [4] (see Section 6.3.2). While in the b2b scenario the energy deposition information from `Geant4` was stored in files and loaded into `ANTS2`, for the ISO scenario `ANTS2` was interfaced directly with `Geant4` using the `G4Ants` package [230]. This way the depositions were generated using `Geant4` member functions, but they acted upon `ANTS2`-defined materials and geometry, and the deposition information was immediately available in `ANTS2` in each event.

***Chosen physics list***

The `Geant4` version was `10.5`, and the used physics list was the one available to `ANTS2` functionally equivalent to the one used for the b2b implementation: `QGSP_BERT_PEN`, which also uses the `Penelope` model for electromagnetic interactions The `FTFP`, `QGSP` and `BERT` components of the physics lists relate to different hadronic physics models [231], pertinent to energies in the GeV range, and hence the choice of `QGSP_BERT` over `FTFP_BERT` does not affect the results of this work in a significant manner, as $Q_{\beta\beta}$ is too low to be affected.

***Event simulation: sensitive volume definition, primary electron emission***

The target was a LXe cylinder of height 450 mm and a diameter of 500 mm, with the LXe material defined as in Section 6.3.2. The $1e$ and $0\nu\beta\beta$ primaries were emitted isotropically (with the two $0\nu\beta\beta$ primaries emitted in opposite directions) according to the procedures described in [232] in batches of 100 events, and afterwards each deposition was separated into its own file **{fn6.3h}**. [234] shows two example excerpts of a deposition file as they were obtained for this scenario.

**Step 2: ISO scenario implementation**

The LZ collaboration models drift electron production using a purpose-built set of semi-empirical models based on detector calibration and science data, collectively referred to as the Noble Element Simulation Technique (`NEST`) [41] [5]. Given an incident electron of energy $\mathcal{E}$ [235], this software package calculates the corresponding charge yield, $\mathtt{Qy}(\mathcal{E})$, by parameterizing it as a sum of sigmoidal functions [39] using as parameters the drift field and the density of the material, with additional free parameters calculated by fitting to experimental data [41].

**{fn6.3h}** — The use of batches of 100 was done to compensate for an issue that occurred on some machines with the chosen build of `ANTS2`, where even though a `Geant4` script could be run any number of consecutive times from within `ANTS2`, only up to 14 of the runs would be saved to disk. Along with the conversion from primary energy loss to number of drift electrons produced, [233] shows the procedure used for breaking up the output files with 100 deposition batches into single deposition files.

The NEST v2 member function NESTcalc::GetYields() (from file NEST.cpp) was used as reference for implementing the charge yield function (see left panel of Figure 6.8), using 310 V/cm for the drift field and 2.869 g/cm$^3$ for the density **{fn6.3i}**. The number of electrons released, $N_e$, is obtained for a given incident energy as $N_e(\varepsilon) = \varepsilon \times Qy(\varepsilon)$. In order to account for the trajectory of the primary electron, a dependency on the energy lost during the interaction, $d\varepsilon$, is introduced:

$$N_{loss}(\varepsilon, d\varepsilon) = d\varepsilon \times \frac{dN_e(\varepsilon)}{d\varepsilon}. \tag{6.1}$$

The shape of $dN_e(\varepsilon)/d\varepsilon$ is shown in the right panel of Figure 6.8.

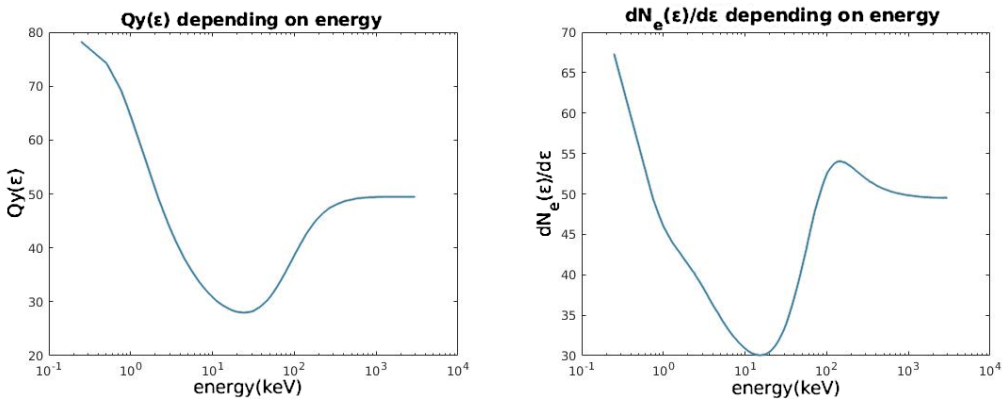


FIGURE 6.8: *Left panel:* Obtained drift electron yield function. [233] shows the used code. *Right panel:* Obtained $dN_e(\varepsilon)/d\varepsilon$. [233] shows the used code.

Each $1e$ and $0\nu\beta\beta$ event returned track information, namely the position of the primary electron (and bremsstrahlung child primary electrons) at the end of each Geant4 step, the kinetic energy of the (child) primary electron $\varepsilon$ and the energy loss to ER during that Geant4 step $d\varepsilon$. The energy information was converted to a number of produced diffusion electrons using the $N_{loss}$ function, as shown in [233]. Figure 6.9 shows example topologies for$1e$ and $0\nu\beta\beta$ events due to this implementation, and the statistics for the total number of diffusion electrons per event of each class.

**{fn6.3i}** — This value was calculated using the NEST v2 member function NESTcalc::SetDensity() (from file NEST.cpp) as reference. It calculates the LXe density using the temperature of the sensitive volume as a parameter, and the value of 2.869 g/cm$^3$ results from setting the temperature to 175.8 K, as per the LZ Technical Design Report (TDR, [109], page 57). [236] shows the procedure used for the density calculation. The obtained value of the density is more accurate than the one used in the b2b scenario, as that one did not account for the temperature, and the decreased density is expected to result in an increase in the average size of the deposition topologies by about 3% (i.e. an increase of about 0.1 mm in the average distance between the two blobs). Additionally, according to the NEST v2 model, the lower density is expected to increase the charge yield by about 3% as well. This means that the average number of drift electrons in the ISO model is expected to be larger than in the b2b one. Unfortunately the density of the LXe in ANTS2 mistakenly remained set to 3 g/cm$^3$ for this scenario, so the tracks are expected to suffer a 3% average decrease in size relative to those of the b2b scenario (i.e. a reduction of about 1 mm in the average distance between two blobs). The effect of this on the classifier performance is expected to be minor.

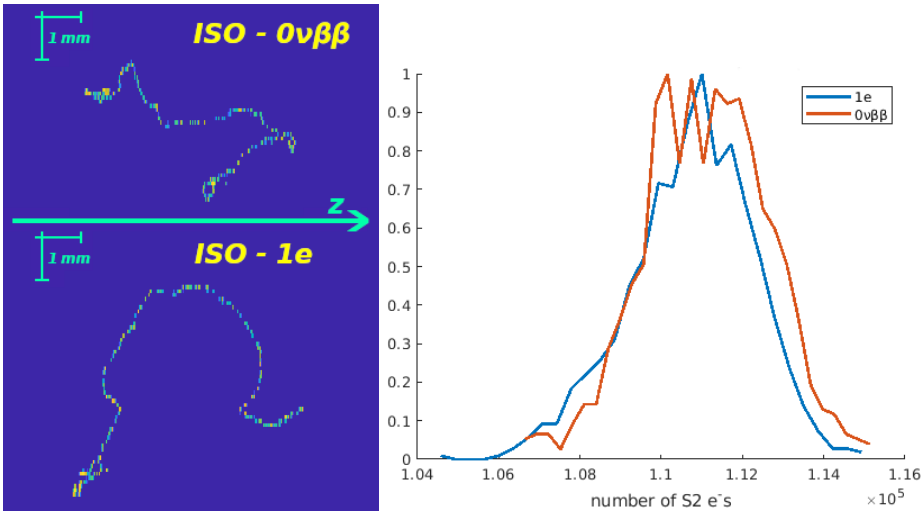


FIGURE 6.9: *Left panel:* Example $0\nu\beta\beta$ and $1e$ topologies generated using the ISO scenario implementation. *Right panel:* Statistics for the total number of drift electrons produced per event of each class for the ISO scenario.

### 6.3.2 ANTS2 script (diffusion, electroluminescence and waveform)

This subsection corresponds to **Steps 4** and **5** from the box on page 91.

The right panels on Figures 6.7 and 6.9 show that the number of drift electrons per event is consistently above $10^5$. With an electroluminescence (EL) photon yield of 820 photons per electron (Table 4.1), this means that each waveform will be the product of $\gtrsim 8 \times 10^7$ photons. Simulating the photon transport in `Geant4` is unfeasible: the simulation time can be estimated to be $\gtrsim$ 10 hours/event. An alternative for simulating the photon transport had to be chosen, and the one opted for was to use the `ANTS2` package [4]. This software was built to produce fast simulations of Anger-type camera **{fn6.3k}** readouts, and in particular it performs very fast photon tracing [237], making it ideal for this problem. `ANTS2` version `4.22` was used to produce the waveform sets, except for $\sim 10^3$ simulated $0\nu\beta\beta$ waveforms, which were generated using an older version of the program.

**{fn6.3k}** — Anger-type cameras are position-sensitive scintillation detectors that use 2-dimensional arrays of photodetectors (e.g. PMTs, or silicon photomultipliers) to produce a 3-dimensional event reconstruction by way of statistical methods. For example TPCs like the LZ detector are a type of Anger camera.

For portability, diffusion, electroluminescence and the generation of the resulting waveform in the PMT array were all simulated in a single `ANTS2` script. The script is listed in [238], and its algorithm is summarized in the box below:

***Configuration***

- Load *detector model* (*pgs. 98 - 101*);
- Set *simulation parameters* (*pgs. 98 - 101*).

***Simulation***

**`for` each track file**:

1. Clean up data containers, import track file and initialize output file;
2. Simulate *diffusion* on drift $e^-$ positions from track file (*pgs. 101 - 103*);
3. Simulate *electroluminescence* (*pg. 103*);
4. Collect EL photon PMT hit times into corresponding data containers;
5. Set the earliest hit time as the start of the signal;
6. *Convolve* the data containers with the *single EL photon pulse shape of the PMTs* (*pg. 103*);
7. Transfer data containers to JSON object, *output to file* (*pg. 103*).

### Detector model / simulation parameters

Figure 6.10 shows the implemented detector model. The detector's size was greatly reduced to speed up the electroluminescence simulations, with the PMT array modeled after the central region of the LZ top array (*Figure 6.11*). The material surrounding the detector is "gray matter": a material designed to absorb the electroluminescence photons (*i.e.* it has a reflectivity of zero). The specifications of all the used materials are given in Table 6.3, and Table 6.4 compares the specifications of the detector model to those defined in the LZ TDR [109].

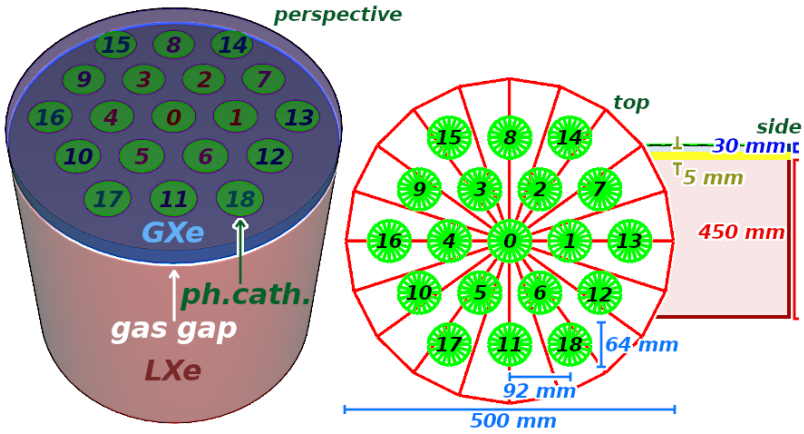


FIGURE 6.10: Persepective, top and side views of the detector model used, with materials and dimensions used. The gas gap is highlighted in yellow in the side view. The photocathodes are highlighted in green, with the numbers being their IDs. Once a photon hits the interior of the photocathode, it is considered as having been detected.

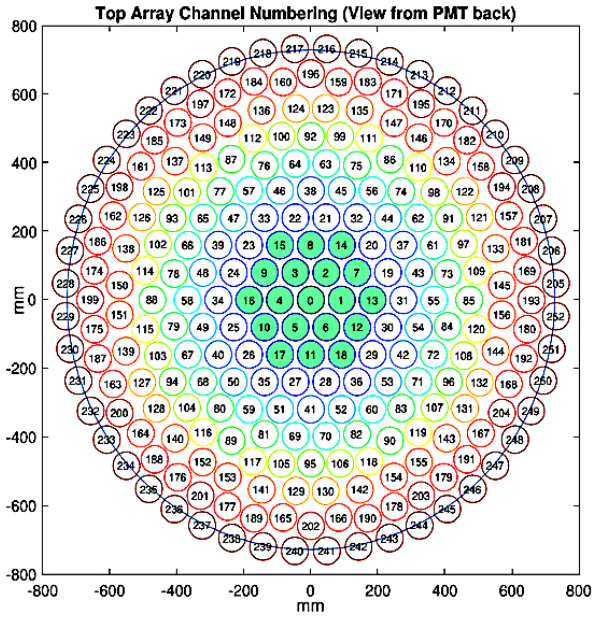


FIGURE 6.11: Top PMT array of the of the LZ TPC. The part of the PMT array that was recreated in the simulation is highlighted in green.
Source: [109]

| Material | Density (g/cm$^3$) | | Refractive Index | | Attenuation Length (mm) | | Rayleigh Length (cm) | |
|---|---|---|---|---|---|---|---|---|
| ***Gray matter*** | 2 | **N/A** | 1.6 | **N/A** | 0.333 | **N/A** | $\infty$ | **N/A** |
| ***GXe*** | 0.005 | 0.017 | 1 | $\sim$1 | $\infty$ | $\infty$ | $\infty$ | $5\times10^3$ |
| ***LXe*** | 3 | 2.869 | 1.7 | 1.67 | $3\times10^4$ | $3\times10^4$ | 30 | 30-50 |
| ***Photocathode*** | **N/A** | **N/A** | 1.6 | 1.57 * | 0 | $\sim$15 * | $\infty$ | *no info* |
| | **Model** | **Actual** | **Model** | **Actual** | **Model** | **Actual** | **Model** | **Actual** |

* *values indicated for PMT window material*

TABLE 6.3: Material definitions for the detector model. The actual values are taken from [109], except for the density of GXe, obtained from [239] for 1.8 bar(a) and 175.8 K, and for the attenuation length of the PMT window material (borosilicate glass), obtained from [240].

| Specification | Model | | Actual | | Ref. in [109] | |
|---|---|---|---|---|---|---|
| ***drift region height*** | 450 | mm | 1456 | mm | *Table* | *3.1.1* |
| ***diameter*** | 500 | mm | 1456 | mm | *Table* | *3.1.1* |
| ***Electroluminescence region (gas gap)*** | 5 | mm | 8 | mm | *Table* | *3.6.1* |
| ***PMT cathode diameter*** | 64 | mm | 64 | mm | *Section* | *3.4* |
| ***PMT center-to-center spacing*** | 92 | mm | 93 | mm | *Section* | *3.5.3* |
| ***PMT quantum efficiency*** | 100 | % | $\lesssim$ 28 | % | *Table* | *3.5.1* |
| ***long. diffusion (LZ cathode events)*** | 0.85 | $\mu$s | 0.849 | $\mu$s | *Table* | *3.3.1* |
| ***tran. diffusion (LZ cathode events)*** | 0.59 | mm | 0.594 | mm | *Table* | *3.3.1* |
| ***Electroluminescence photon yield*** | 860 | eph/$e^-$ | 820-910 | eph/$e^-$ | *Table* | *3.6.1* |
| ***PMT time resolution*** | 5 | ns | 10 | ns | *intern. comm.* | |

TABLE 6.4: Comparison of the specifications of the detector model used in the simulation to the specifications of the LZ detector, with indication of the sections of the technical design report used. The chosen electroluminescence yield is a midpoint between the two cited values in [109]. More information in Section 4.1.

The chosen values of most specifications are very similar to their true counterparts, however there are some differences whose effects must be considered. They are listed below, from the most major to the most minor:

**Smaller gas gap**

The electroluminescence photons from a single electron are emitted at a range of heights above the LXe surface. This means that a thicker gas gap will smear the PMT pulses more intensely than a thinner gas gap. Given that a two-phase TPC necessarily includes a gas gap, it was not done away with entirely, but it was shortened in order to reduce the longitudinal smearing. This should improve the performance of the classifier.

**Absorbing detector walls**

The walls of the LZ detector are coated with PTFE, which serves to reflect the electroluminescence photons, increasing the sensitivity of the detector to the low energy recoils of WIMP interactions with LXe. In the case of $0\nu\beta\beta$ , the recoils are much more energetic, and so the reflecting PTFE walls would only add a floor to the waveforms. Replacing the PTFE coating with an absorbing wall removes a source of distortion in the waveform, improving the performance of the classifier.

**No high-field region in the LXe**

The gate grid (see Section 4.1) is submerged in LXe at a depth of 5 mm, resulting in a thin LXe region above it where the electric field is much stronger than the drift field below. This region will result in some additional longitudinal smearing. It was not included in the current scenarios.

**Much smaller detector**

The main consequence of this is that it shortens the photon transport, speeding up the simulation. This choice further justifies the use of absorbing detector walls, as the PTFE coating is expected to cause more transverse smearing of the waveform in the case of a small detector. The effect on the time series signal is expected to be imperceptible.

**Different electroluminescence yield**

The chosen electroluminescence yield is a midpoint between the two values referred to in [109]. Considering that all three are very high numbers of photons per electron, and considering the extremely large number of photons per event, the effect of this choice should be minor.

**Very low density of GXe**

The reduced density of the GXe in the model detector is expected to result in a smaller number of interactions of the electroluminescence photons with the space between the PMT array and the gas gap, presumably reducing distortions in the waveform. However this effect is expected to be so small that it is imperceptible, considering the very large photon interaction lengths in the gas.

**Model density of LXe higher than its actual value**

The diffusion simulation is not sensitive to the density of the LXe, and neither are the electroluminescence photons, as they are only sensitive to the attenuation length, which is correct.

As listed above, the differences in the detector model relative to the actual detector are either expected to not affect the waveform in a significant manner, or at most remove distortions. In conclusion, it is expected that in the worst case they slightly improve the classifier's performance, in agreement with the chosen base simulation approach (Section 6.1.1): if the classifier fails to discriminate the two classes for the simulated waveforms, then it is guaranteed to fail for real data.

### Diffusion simulation

The used version of `ANTS2` did not include tools for simulating diffusion, so they had to be added into the source code. [241] shows the added functions, and below is a description of the used procedure: first the stages are summarily listed; then certain details are explained. The stages are as follows:

1. Receive as input parameters the depth `d` and the deposition text file `f_d`.
2. Calculate the longitudinal and transverse diffusion pdf widths `S_L` and `S_T` at the selected depth **{fn6.3l}**;
3. Read `f_d` and convert its data into a matrix **`P_e`** of the positions of all the drift electrons;
4. Generate a matrix of displacements **`D_e`** with the distribution of the diffusion pdf having `S_L` and `S_T` as parameters;
5. The output is the result of the operation **`P_e`** + **`D_e`**.

The widths of the diffusion pdf in LZ are seen to increase linearly with the depth of the event (Figure 3.6.4 in [109]), and so the values of the diffusion at different depths, $\sigma_L, \sigma_T$ **{fn6.3m}** **{fn6.3n}**, were calculated using as basis the diffusion values at cathode depth, referenced in Table 6.4:

$$\sigma_L = (0.85\ \mu\text{s}) \times v_d \times \frac{\mathtt{d}}{\mathtt{h}} \quad , \quad \sigma_T = (0.59\ \text{mm}) \times \frac{\mathtt{d}}{\mathtt{h}},$$

where $v_d = 2.0$ mm/$\mu$s is the drift speed and $\mathtt{h} = 1456$ `mm` is the depth of the cathode. Replacing $tD_L$ and $tD_T$ for $\sigma_L$ and $\sigma_T$, Equation 3.10 can be rewritten and broken up into two separate pdfs, one for longitudinal diffusion and one for transverse diffusion, which are now standard Gaussian pdfs with zero average and standard deviations $\mathtt{S\_L} = \sqrt{2} \times \sigma_L$ and $\mathtt{S\_T} = \sqrt{2} \times \sigma_T$ respectively. The effect of diffusion on the drift electron positions was simulated by displacing their `z` position by a random distance $\mathtt{d\_z} \sim \mathcal{N}(v_d, \mathtt{S\_L}^2)$ and their `x,y` positions by random distances $\mathtt{d\_x}, \mathtt{d\_y} \sim \mathcal{N}(0, \mathtt{S\_T}^2)$. Figure 6.12 compares the diffusion pdf with the resulting distribution due to $10^4$ displacements, showing good agreement between both.

**{fn6.3l}** — Even though the scenarios studied in this work do not use the transverse diffusion, the simulation already includes some adaptations to ease its inclusion in future work.

**{fn6.3m}** — The diffusion values relate to the diffusion coefficients as $\sigma_{L,T} = \sqrt{tD_{L,T}}$.

**{fn6.3n}** — Note that in the current version of the implementation the diffusion value is the same for all drift electrons in a deposition, *i.e.* the variation in $z$ value between the different electrons was not considered. This approximation was done because the depositions are quite small, and hence the diffusion value does not vary significantly. The corresponding adaptation for improved accuracy can be done in future work.

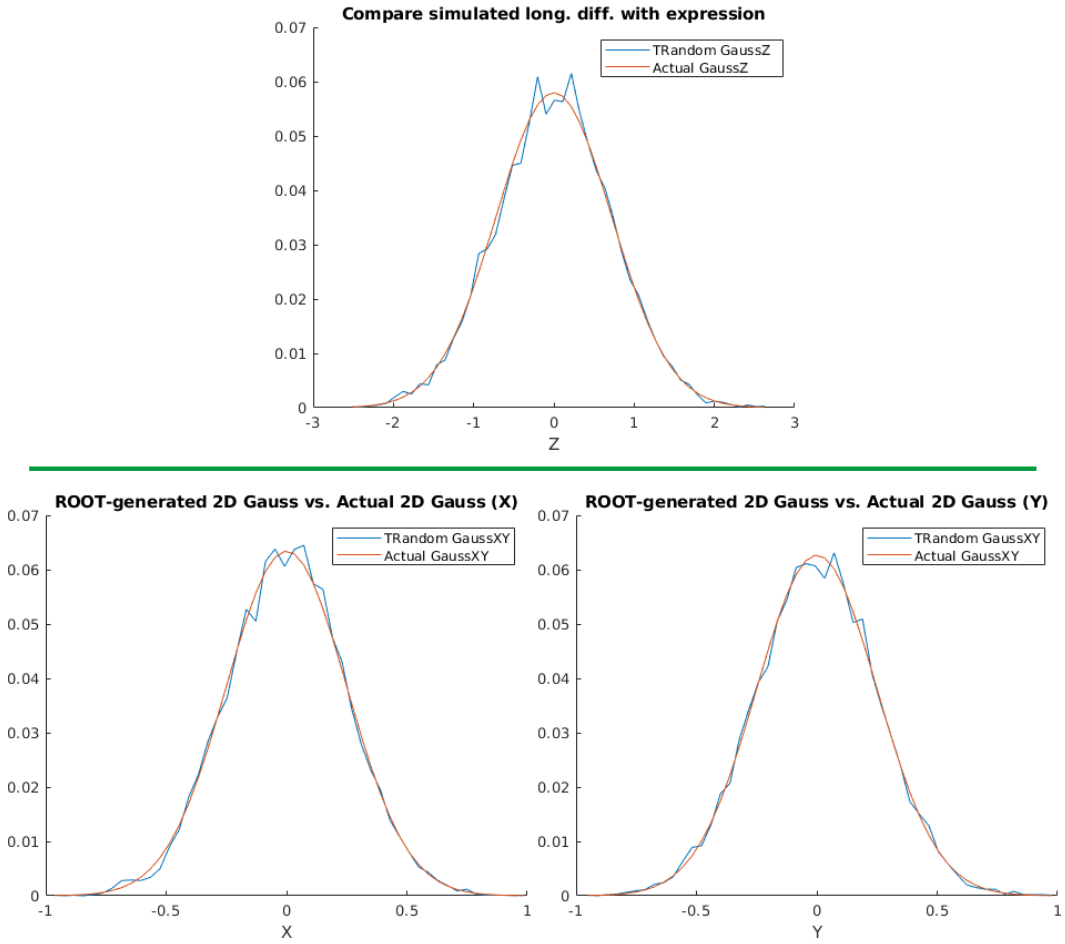


FIGURE 6.12: Comparison of diffusion-originating electron displacement statistics for $10^4$ electrons with the diffusion pdf along the $z$ (top panel), $x$ and $y$ (bottom panel) directions, showing that they are consistent.

The random number generator class in `CERN ROOT` [242], `TRandom`, was used to compute the displacements, as its member `Gaus()` produces Gaussian-distributed random numbers very fast [243] [244] **{fn6.3o}**. In order to perform the matrix addition **`P_e`** + **`D_e`** fast, the Eigen template library [246] was used, however it has not yet been fully integrated with ANTS2, currently requiring a slow final conversion from an Eigen matrix back into a type of container recognized by the ANTS2 script parser. Nevertheless the diffusion simulation is the fastest part of the ANTS2 script, due to the comparatively small number of particles.

Of note also is that the diffusion simulation saves computation time by not tracing the electrons as they drift along the detector, and instead just calculating their expected final configuration upon reaching the gas gap. As long as the electric field along the LXe of the detector is assumed to be uniform, this approximation is accurate, however in future work it will require modifications.

**{fn6.3o}** — The underlying uniform pdf random number generator is the rudimentary BSD linear congruential generator, of period $2^{31} \approx 2.1 \times 10^9$ [244] [245]. It boasts excellent speed, and its period is still large enough for the given usage, as the number of computations per deposition should at most reach the order of $10^6$, and each track file sets a new seed.

The output from this part of the ANTS2 script algorithm to the next one is a list of all the electron positions following displacement due to diffusion. The deposition is assumed to reach the gas gap starting at the electron with the lowest $z$ value after diffusion, and then proceeding to the higher ones **{fn6.3p}**.

### Electroluminescence simulation

ANTS2 provides tools for simulating the emission of S2 electroluminescence photons from the gas gap of a TPC. Each electron produces a column with the same vertical thickness as the gas gap and consisting of 860 isotropically-emitted photons, all originating at the same horizontal position and at random heights along the gas gap. ANTS2 calculates the emission time of a given photon $\gamma$, $t_\gamma$, as:

$$t_\gamma = \frac{z_{e^-}}{v_d} + \frac{z_\gamma}{v_g},$$

where, $z_{e^-}$ is the height of $\gamma$'s parent electron after diffusion; $z_\gamma$ is the height within the gas gap at which the photon was emitted (with the surface of the LXe being at $z_\gamma = 0$); and $v_g$ is the velocity of the electron in the gas gap. $v_g$ was set to 10 mm/$\mu$s, an approximation of $(8/5) \times \overline{v_{gLZ}}$, where $\overline{v_{gLZ}} \approx 6.7$ mm/$\mu$s is the average velocity of the electrons in the LZ gas gap (Section 3.2.3 of [109]) and $8/5$ is the ratio of the LZ gas gap thickness vs. the detector model gas gap thickness.

When tracing the photon trajectories through the detector, ANTS2 accounts for material absorption and Rayleigh scattering, as well optical processes at material interfaces — diffuse reflection, specular reflection and refraction — calculated according to the Fresnel equations and Snell's law [4]. If a photon is captured by the cathode of one of the PMTs, its detection time is stored in a list of "PMT hits" for that given PMT. Once the electroluminescence simulation finishes, the emission times in these lists are corrected, setting the earliest emission time to zero and adjusting all others accordingly.

The output from this part of the ANTS2 script algorithm to the next one is a set of lists, one per PMT, each indicating the emission times of the photons that hit the corresponding photocathode, such that the earliest photon emission time is zero.

### Convolution with the single photoelectron pulse shape

This part of the ANTS2 script builds the waveform resulting from the $0\nu\beta\beta$ or $1e$ event out of the list set obtained in the part above, taking into account the single photoelectron pulse shape of the PMT array as well as the selected PMT time resolution. The pulse shape was imported from a text file with signal values based on the data in Figure 4.5. The convolution was performed after quantizing both signals to match the PMT time resolution, with 5 ns bins. For the pulse shape, the matching was done through linear interpolation; and for the list set it was done by binning, forming a histogram corresponding to the "signal" at that resolution, without the pulse shape. The convolution was then performed on these equal-resolution signals. The procedures for performing these steps were added into the ANTS2 source code, and are listed in [238].

### Output file

The final output of the ANTS2 script is a text file where the PMT channels are listed in sequence, with the amplitude at multiples of the read time shown in arbitrary units. No units are necessary

**{fn6.3p}** — This means that the waveforms are produced up-side down, but given that the deposition pdfs are vertically symmetric, and considering what was discussed in **{fn6.3n}**, this does not affect the accuracy of the resulting dataset, while at the same time simplifying the implementation.

as they do not influence the performance of the classifier as long as the single photoelectron pulse amplitude is consistent. The zero time is the first instance of a nonzero amplitude, for any of the PMTs in the array: namely if a PMT #0 has the earliest photon hit and PMT #4 only has a photon hit a microsecond later, then first few entries *(text file lines)* of the PMT #4 channel will be zero.

### 6.3.3 Computational performance and limitations of the ANTS2 script

The output files of the ANTS2 script are fairly small, only $\sim$ 300 kB in size, however the simulation itself is very resource-intensive. Following optimizations to the source code **{fn6.3q}**, it still requires $\sim$ 3 GB RAM memory to run, and takes $\sim$ 5-7 minutes to produce one waveform. The main bottleneck is tracing the $\gtrsim 8 \times 10^7$ electroluminescence photons per event, as the rest of the procedure only takes a few seconds. The script runs on a single CPU thread.

Some further optimizations are possible, but would require significant modifications to the ANTS2 script interface and tracer module. The huge memory footprint is due to the large number of lists containing stepwise modifications to the PMT photon hit data. They are each cleared once the data becomes redundant, but the memory is not freed, because the scripting engine used in this work **{fn6.3r}** lacks garbage collection. At the same time, reusing the same containers is impractical, as their structures vary throughout the script. In order to reduce the memory footprint, it would therefore be necessary to either introduce garbage-collected containers into the current script interface, or to build a different script interface, which is outside the scope of this work.

On the other hand, the long computation time could be mitigated using parallel computing. GPU acceleration seems like an attractive solution to the problem of tracing photons, as their trajectories are independent from each other. However this would require adapting the tracing algorithm to allow for GPU computation. Additionally, the way the photon trajectory is stored would have to be significantly reworked. CPU multithreading was attempted, but no time benefit was noticed, while making the script more unstable, and so the choice was made to keep the script single-threaded.

An alternative to tracing the photons is to use statistical reconstruction data (PMT light response functions) [247] to produce statistically realistic waveforms, however it is not clear whether this approach will maintain the same classifier performance relative to tracing. If the performance were to worsen due to this approach, then it would not correspond to the chosen base simulation approach. A change from photon tracing to statistical methods is attractive, but would require sample tracing data to be simulated first, as well as validation.

**{fn6.3q}** — Most notably, the tracer algorithm in ANTS2 stores the interactions during photon transport into a "historian" container, from where the emission times and PMT IDs were extracted. This led to a significantly larger memory footprint, exceeding 10 GB in some events. The source code was therefore modified to only write the last interaction into the historian, resulting in a more tolerable memory footprint.

**{fn6.3r}** — ANTS2 provides two script interfaces, a JavaScript one and a Python one, however the Python interface was still under construction during the making of this work.

### 6.3.4 Distributed computing and final tally of the simulations

Considering the large memory footprint and long computation time, it became clear while constructing the b2b scenario waveform dataset that it is impractical to run the simulation from a single computer. At the same time it was known that installing the ISO scenario version of `ANTS2` (the one bundled with the `G4Ants` interface) on several computers would prove challenging. `ANTS2` is a `Qt` application [248], and `Qt` packages tend to enter in conflict with those of both `Geant4` and CERN `ROOT` during installation. Furthermore it was uncertain whether the simulation would remain stable over multiple platforms, resulting in a very long predicted configuration time.

This motivated the use of a portable, preconfigured version of `ANTS2`, adapting it for distributed computing. For this purpose, the `ants2-docker` image [249] was modified to install the `ANTS2` branch with the source code additions used in the simulation [250], instead of the master branch. The `Dockerfile` is listed in [251]. The ISO scenario simulations were run locally, as well as on the Lisbon and Minho nodes of the Infraestrutura Nacional de Computação Distribuída [6].

The final numbers of simulated waveforms were the following:

*b2b scenario*

$0\nu\beta\beta$ 5075 events

$1e$ 4994 events

*ISO scenario*

$0\nu\beta\beta$ 6451 events

$1e$ 4721 events

## 6.4 Data extraction

This stage of the classification procedure consists of converting the waveforms produced in the simulation stage into points in parameter space, and then mapping those points into feature space. A `C++` application was written to perform this task, listed in [252]. Its algorithm is summarized in the box below, and the details of interest for different parts of the algorithm are described in the Subsections that follow:

*Initialization*

- Read input filenames;
- Make output files;

*Data extraction*

**`for` each input file**:

**Step 1.** Read and store into container;

**Step 2.** Apply *dataset preprocessing* (*pgs. 106*);

**Step 3.** Run *parameterization* (*pgs. 107 - 108*);

**Step 4.** *Optional* → Break up dataset by *categories* based on waveform pulse shape (*pgs. 109 - 114*);

**Step 5.** *Extract features* from parameters (*pgs. 114 - 116*).

### 6.4.1 Dataset preprocessing

Each waveform underwent three preprocessing operations:

1. Summation together of all the PMT channels into a single signal;
2. *For b2b scenario only* $\rightarrow$ 50% chance of time reversal of the summated signal;
3. Smoothing of the summated signal.

The choice of summating the PMT channels is explained in Section 6.1. The time reversal is done because all the b2b scenario $1e$ event primary electrons are emitted directly downward. Time reversal prevents the classifier from using the direction of emission as a decision criterion. Both $0\nu\beta\beta$ and $1e$ events were time reversed so that the reversal of the PMT pulse shape would not be used as a decision criterion by the classifier. The smoothing is performed in order to aid in taking practically valuable derivatives of the amplitude (see Figure 6.13). The smoothing procedure is shown in [252].

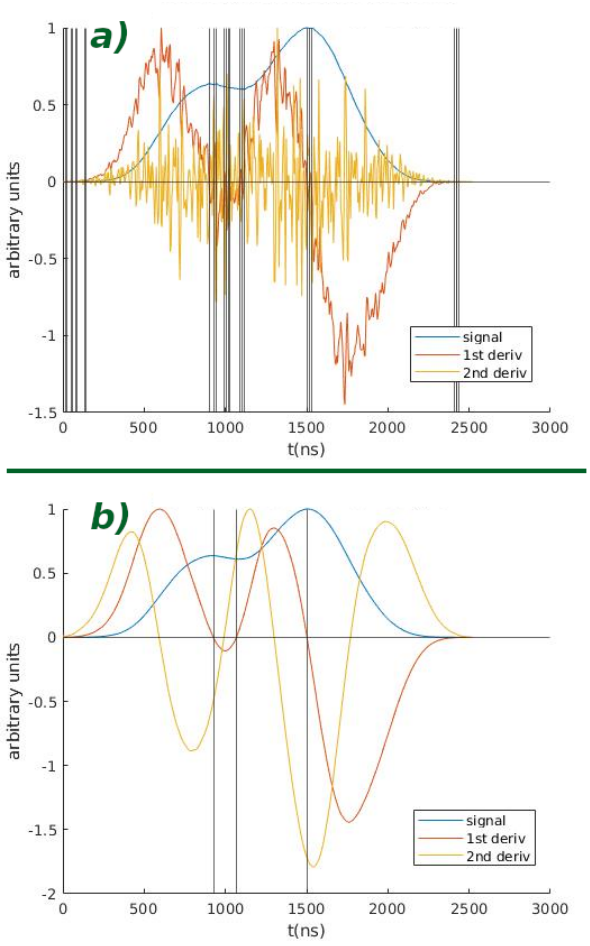


FIGURE 6.13: 1st and 2nd derivative of signal: a) without smoothing; b) with smoothing applied. The vertical lines are the zeros of the 1st derivative. Note that even though the signal (in blue) appears reasonably smooth even in the top graph, its derivatives become distorted beyond usability. Note also that the smoothing serves as a low-pass filter, removing the high-frequency component without affecting the "overall" shape of the signal.

### 6.4.2 Parameterization

The parameterization step in this work builds upon the waveform parameterizer of the LZ Analysis Package (`LZap`) [41], preserving the same operating structure but adapting the extracted parameter set to the given problem. The code is listed in [252]. The extracted parameters are summarized below, and detailed further on for more nuanced parameters:

- Waveform unique ID (for ease of identification);
- Clear presence of bremsstrahlung;
- Area, and area fraction times **{fn6.4a}**;
- Times and amplitudes of signal peaks and plateaus;
- *Optional* → Number of peaks and plateaus;
- Root mean square signal width and amplitude;

#### Clear presence of bremsstrahlung

It was seen that due to the higher energy of the $1e$ events, they underwent bremsstrahlung more frequently than $0\nu\beta\beta$ events, making its presence in the signal a potential discriminator. Many of the times, bremsstrahlung would transfer a very significant fraction of the primary electron's energy into a single photon, which would then continue the energy deposition at a separate location, often resulting in trails of zeros in the middle of the signal, (Figure 6.14), which were then used as an identifier of the process. For these events, the presence of bremsstrahlung was recorded, and the rest of the parameterization was performed on the side of the signal (relative to the zero trail) with the largest area. However this method does not account for all bremsstrahlung events. A significant proportion of waveforms show a bremsstrahlung pulse that is conjoined with the rest of the signal, and in these the presence of bremsstrahlung was not recorded. In future work it is worth addressing this problem.

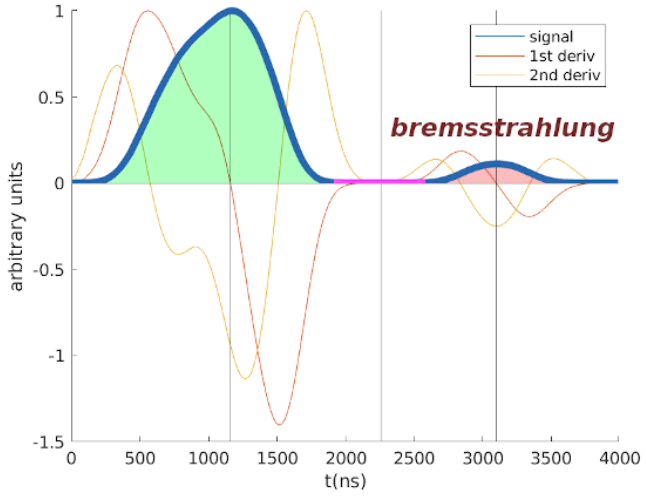


FIGURE 6.14: Example of a $0\nu\beta\beta$ event with a trail of zeros (highlighted in magenta). The part of the signal with the largest area was assumed to not result from bremsstrahlung,

**{fn6.4a}** — The time value at which the cumulative signal area reached certain percentages of the total area.

### Area, and area fraction times

One significant difference in the topologies of $0\nu\beta\beta$ and $1e$ events is the slightly larger length of the $1e$ ones (Figure 6.15). To characterize the length of the signal, along with some measures of its shape (like amplitude averages for various time windows, or skewness) the times at which the signal reaches different percentages of the total area (the area fraction times, or AFTs) were extracted from the signal's cumulative area function, $A(t) = \int_0^t \mathrm{d}t\ s(t)$, where $s(t)$ is the signal amplitude at time $t$. For more details, see [252].

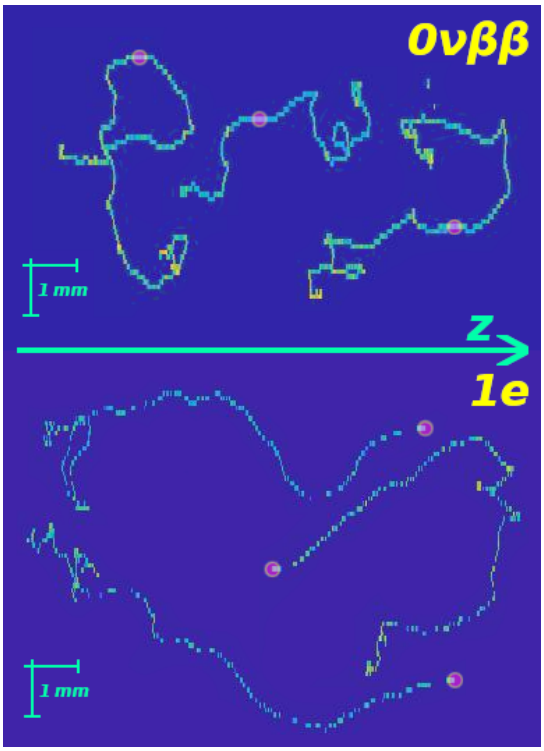


FIGURE 6.15: Comparison of topologies of $0\nu\beta\beta$ and $1e$, showing that $1e$ topologies are often longer than $0\nu\beta\beta$ ones. The faded magenta circles show the location from where the primary electrons were emitted.

### Times and heights of signal peaks and plateaus

Looking at a number of $0\nu\beta\beta$ and $1e$ waveforms, it is noticeable that $0\nu\beta\beta$ ones show two peaks (local maxima of the signal) more often than $1e$ ones, while $1e$'s tend to produce a waveform with a single peak next to a comparatively flat, mid-amplitude region, in this work termed a *plateau.* For an example, see Figure 6.16. This motivated the use of the number of peaks and plateaus, as well as the height and time of each, as parameters describing the signal. Peaks were defined as points in the signal $s(t)$ where the values of derivatives were $s'(t) = 0$ and $s''(t) < 0$, while plateaus were defined as points where $s''(t) = 0$ and $s'''(t) > 0$ **{fn6.4b}**. For more details, see [252].

**{fn6.4b}** — Since the focus was on the sign of $s'(t)$, $s''(t)$ and $s'''(t)$, the derivatives were taken in a very simple manner, merely calculating the slopes at $t$ for $s(t)$, $s'(t)$ and $s''(t)$, respectively, by taking the difference of two adjacent points divided by the time step. This explains the need for smoothing, as each successive derivative would highlight the coarsenesses in the original signal. Each successive derivative of the signal also underwent its own smoothing step.

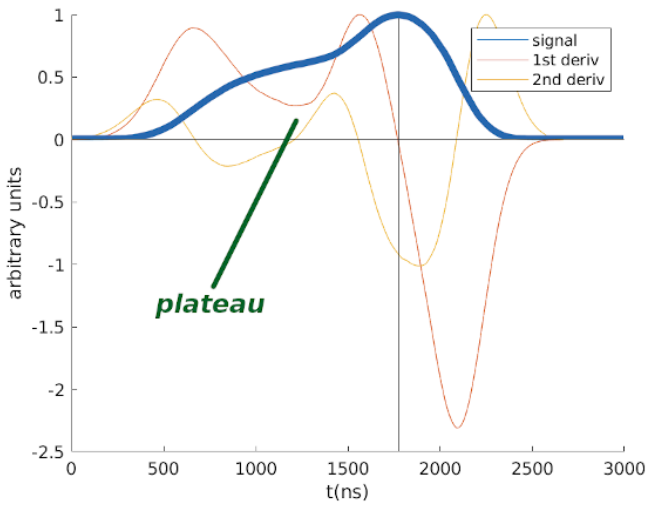


FIGURE 6.16: Example of a plateau in a $1e$ signal.

### 6.4.3 Categorization

#### Categorization — Motivation

The chaotic path taken by the primary electrons through the LXe results in a large array of different possible morphologies. For example, even though $1e$ events have a single *blob* (definition in pg. 15) in the topology while $0\nu\beta\beta$ events have two, it is not uncommon for the $1e$ waveform to show two peaks, or for a $0\nu\beta\beta$ waveform to have a single peak. Although rarer, there were even waveforms with a larger number of peaks (see Figure 6.17 for an example).

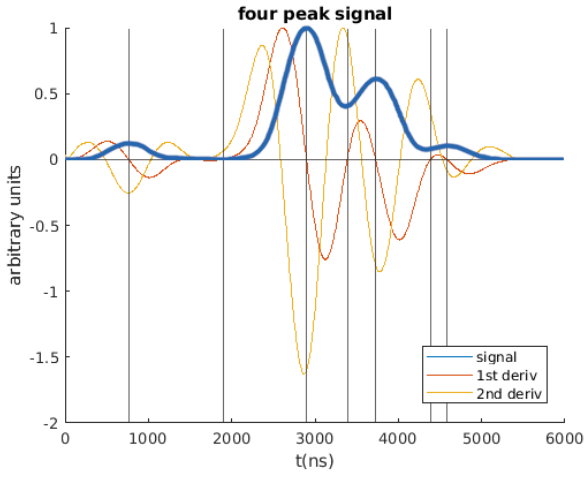


FIGURE 6.17: Example of a $0\nu\beta\beta$ signal showing four peaks. The two peaks in the middle are due to the primary electrons, while the peaks at the edges are due to bremsstrahlung.

Considering the number of factors affecting signal shape, it is unlikely for there to exist some linear axis change that would return a Gaussian distribution of parameter values. As an example, Figure 6.18 shows a scatter plot of a 25% area fraction time against the time of the first plateau, displaying obvious discontinuities. This is problematic for the dimensionality reduction step (*see Section 6.5.3*): it implies the need to use a nonlinear dimensionality reduction method that

does not assume a Gaussian distribution on the latent variables, but those methods tend to become unreliable beyond $\sim 5$ dimensions (Section 7.7 of [253]). Since the dataset given to the dimensionality reduction step is expected to be described by $\gtrsim 20$ features (*Section 6.5.2*), this is not acceptable. The most reliable dimensionality reduction methods for high dimensionalities are those with the most restrictive assumptions about the data, namely principle component analysis (PCA) and classical metric multidimensional scaling (classical MDS) (Section 7.7 of [253]), both of which perform linear projections assuming Gaussian-distributed data. For deployment it would also be necessary to retrieve the transformation operator applied to the training data when mapping from feature space to latent variable space: an operation possible for both PCA and classical MDS, but difficult or even impossible for many nonlinear methods.

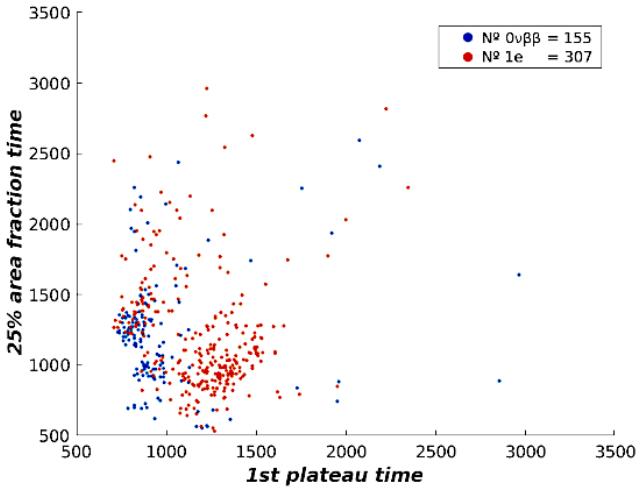


FIGURE 6.18: Scatter plot of a 25% area fraction time against the time of the first plateau. The scatter plot was obtained using a small dataset with 1029 $1e$ waveforms and 1029 $0\nu\beta\beta$ waveforms. Only the events with at least one plateau are shown here.

Although PCA and classical MDS are both good at preserving the shape even of non-Gaussian distributions on a linear manifold (see Figure 6.19 and Figure 6.21), they risk mapping distinct regions of the manifold to the same values of the projection coordinates for the case of a nonlinear manifold (Figure 6.20 and Figure 6.22). Considering the stringent dimensionality limits on nonlinear methods, it was decided that the most practical option was to divide the dataset into subdatasets based on some feature of the signals, termed *categories*, supposing that the parameter value distribution within one category would be less complex than the distribution for all categories combined, and so the embedding would be less likely to overlap with itself after a linear projection **{fn6.4c}**.

**{fn6.4c}** — Another option is to apply linear dimensionality reduction first, and then apply additional nonlinear dimensionality reduction on the resulting data, as suggested in [253], but this solution is deceptively complex and time-consuming to implement. First, it requires testing for the number of dimensions that can be cut by linear dimensionality reduction before manifold overlapping occurs. This can be achieved, for example, by comparing relative Mahalanobis distances *(see Section 6.5.2)* between different points on the manifold. Then the nonlinear dimensionality reduction method has to be tested for the same problem. Finally, there is the issue of retrieveing the transformation performed by the nonlinear method. This procedure can be more effective than the currently implemented solution, however, so it was left for future work.

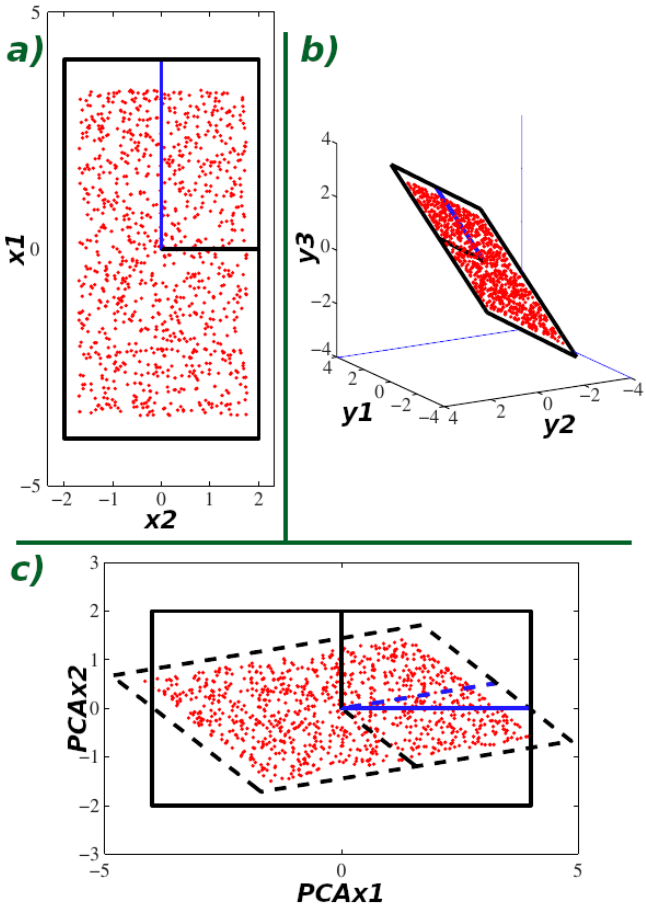


FIGURE 6.19: a) A non-Gaussian distribution of arbitrary datapoints. The embedding is described in the latent variables (`x1,x2`); b) The 3D input dataset onto which PCA is applied. The embedding is placed in a flat (linear) manifold, described by three coordinates: (`y1,y2,y3`). Even though in this coordinate system all three coordinates must be given to uniquely describe each datapoint on the manifold, there exists some transformation that uniquely maps all datapoints to only two coordinates. c) Projection of the input onto the first two principal components found by PCA, (`PCAx1,PCAx2`). Note that although the result is inverted, there is no overlapping, meaning that it should be preferable to feed a classifier the data in c), rather than the data in b), as it would be less prone to the curse of dimensionality.
Image source: [253]

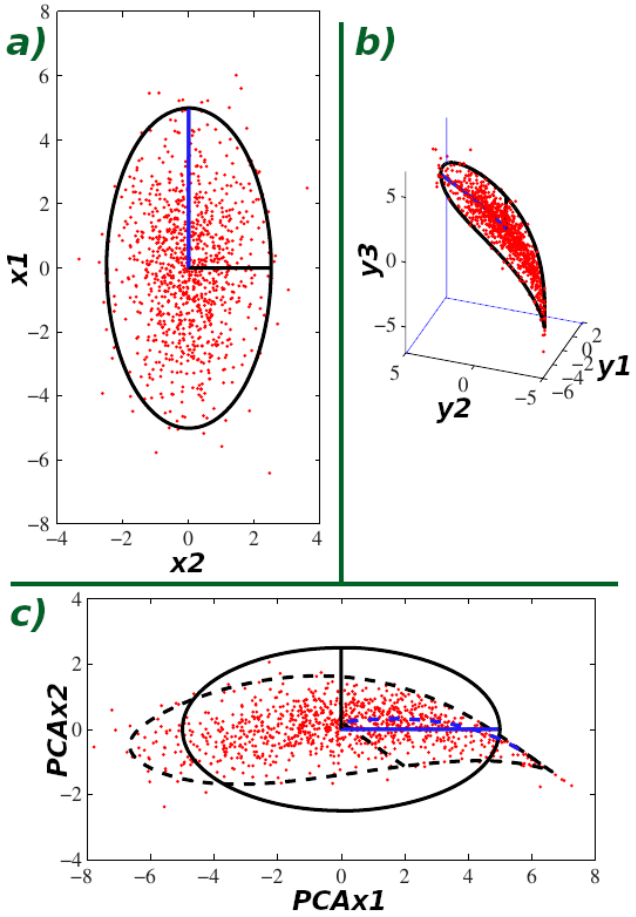


FIGURE 6.20: a) A bivariate Gaussian distribution of arbitrary datapoints. The embedding is described in the latent variables (`x1,x2`); b) The 3D input dataset onto which PCA is applied. The embedding is placed in a curved (nonlinear) manifold, described by three coordinates: (`y1,y2,y3`). Even though in this coordinate system all three coordinates must be given to uniquely describe each datapoint on the manifold, there exists some transformation that uniquely maps all datapoints to only two coordinates. c) Projection of the input onto the first two principal components found by PCA, (`PCAx1,PCAx2`). Note that the curved manifold caused a significant distortion to the shape of the projection. Even though there still does not appear to be significant overlapping of the embedding, it is alarming how even a small curvature of the manifold can greatly affect the PCA projection.
Image source: [253]

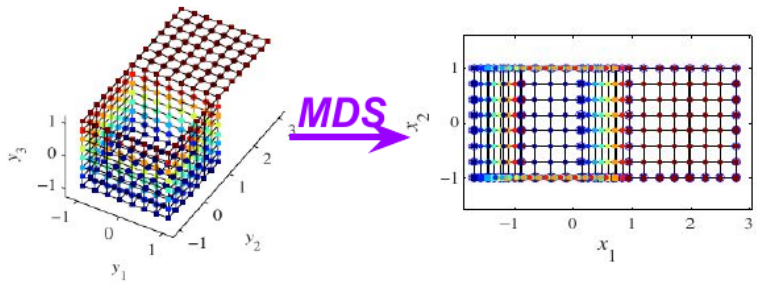


FIGURE 6.21: An "open box" 3D manifold of a 2D embedding, and the resulting 2D projection from applying classical MDS to the data. The manifold features six flat (linear) faces separated by creases, and there is no linear 2D projection where the embedding will not overlap with itself. Unsurpsrisingly, classical MDS does exactly that, but it is worth noting that the flat "lid" (the brown square) is undistorted.
Source: [253]

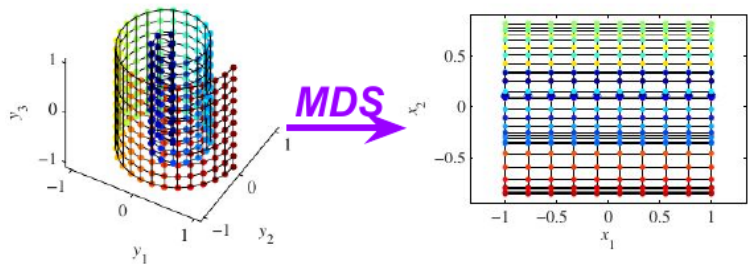


FIGURE 6.22: A "cinnamon bun" 3D manifold of a 2D embedding, and the resulting 2D projection from applying classical MDS to the data. The manifold is nonlinear, and a dimensionality reduction has to be able to "unfurl" the manifold in order to not have the embedding overlap with itself when projected. Classical MDS, being a linear method, is unable to do this. Furthermore, there is no flat shape to preserve. In this case, it would be clearly preferable to give a classifier the high-dimensional data, as the curse of dimensionality would not affect it as severely as would the overlapping.
Image source: [253]

### Categorization — Implementation

In practice, the dataset was subdivided by the number of peaks, the number of plateaus and the clear presence of bremsstrahlung in the signals (according to what was described in page 107). Accounting for the rarity of events with more than 2 peaks, or more than 1 plateau, there were categories for 1, 2 and $> 2$ peaks, and categories for 0, 1 and $> 1$ plateaus. The categorization has two issues, however. Firstly, the statistics is reduced, as the subdatasets corresponding to each category will not have as many datapoints as the full dataset. Secondly, although the categorization makes it easier for the classifier to draw a decision boundary within the category, it makes it impossible to use the number of peaks and plateaus as a decision criterion. Considering these issues, the categorization step was made optional. When not using categorization, the number of peaks and plateaus was taken as a parameter.

The used datasets are referred to in the following manner:

**sans** → no categorization *(the entire simulated dataset);*

**N10** → no clear presence of bremsstrahlung, one peak, no plateaus;

**N11** → no clear presence of bremsstrahlung, one peak, one plateau;

**N20** → no clear presence of bremsstrahlung, two peaks, no plateaus;

**Y10** → clear presence of bremsstrahlung, one peak, no plateaus;

**Z99** → datapoints from all other categories **{fn6.4d}**.

Table 6.5 shows the number of events that fell into each category, for the b2b and ISO scenarios respectively. Only four of the categories had statistically viable datasets given the small dataset, so the rest of the categories were not used for classification in this work. It is noteworthy that the ISO scenario affected the morphology of $0\nu\beta\beta$ events much more severely than that of $1e$ events. A possible explanation is that the shorter paths of the two primary electrons in $0\nu\beta\beta$ events made their b2b scenario events concentrate along a narrower range of positions, while the$1e$ events could spread out, meaning that the positions of the blobs for $0\nu\beta\beta$ events were more correlated to the initial orientation than the position of the blob for $1e$ events. The result of that would be that the morphologies of $0\nu\beta\beta$ events would deteriorate stronger than those of $1e$ events when switching from the b2b scenario to the ISO scenario, consistent with the data in the tables above: many $0\nu\beta\beta$ events became virtually amorphous in the ISO scenario, while roughly the same proportion of $1e$ events as before show some kind of structure.

### 6.4.4 Feature extraction

Following (optional) categorization, the points in parameter space were then mapped to points in feature space, as explained in Section 6.2. Table 6.6 lists the used features both for the case that categorization was used and for the case that it was not. When categorization was not used, the data for non-applicable features (e.g. the time of the second peak on a signal with only one peak) was appropriately zero-padded. The more nuanced features are explained below.

#### tSkew and aSkew

Skewness is a measure commonly used to test the asymmetry of a probability distribution. It is intuitive to suppose that the simulated $0\nu\beta\beta$ events will generally produce more symmetric signals than $1e$, so it was used as a feature for the classifier. There are several ways to measure skewness, and two simple ones are used in feature space: `tSkew` and `aSkew`. They are defined as follows:

$$\texttt{tSkew} = \frac{t_{50\%} - \texttt{pk\_Xns}}{\texttt{RMSw}} \quad , \quad \texttt{aSkew} = \frac{\texttt{M50\_95}}{\texttt{M5\_50}} - 1, \tag{6.2}$$

where $t_{50\%}$ is the $50\%$ area fraction time (AFT), `RMSw` is the root mean square width, `pk_Xns` is the time of the highest signal peak, `M50_95` is the [50% - 95%] AFT range signal mean, and `M5_50` is the [5% - 50%] AFT range signal mean. `tSkew` is based on Pearson's first skewness coefficient.

**{fn6.4d}** — Z99 datasets were produced by removing all N10, N11, N20 and Y10 datapoints from a sans dataset following data optimization.

*b2b scenario (%)*

| $0\nu\beta\beta$ | | | | |
|---|---|---|---|---|
| *No Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 44.77 | 25.95 | 1.52 | |
| 1 | 9.89 | 1.10 | 0.14 | |
| >1 | 0.39 | 0.08 | 0.06 | |
| # plateaus | 5075 events | | | |
| *W/ Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 8.67 | 4.16 | 0.43 | |
| 1 | 2.34 | 0.28 | 0.02 | |
| >1 | 0.16 | 0.02 | 0.00 | |
| # plateaus | | | | |

| $1e$ | | | | |
|---|---|---|---|---|
| *No Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 36.20 | 14.72 | 0.98 | |
| 1 | 17.84 | 2.50 | 0.16 | |
| >1 | 1.44 | 0.18 | 0.00 | |
| # plateaus | 4994 events | | | |
| *W/ Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 14.52 | 4.43 | 0.48 | |
| 1 | 5.23 | 0.80 | 0.08 | |
| >1 | 0.36 | 0.08 | 0.00 | |
| # plateaus | | | | |

*ISO scenario (%)*

| $0\nu\beta\beta$ | | | | |
|---|---|---|---|---|
| *No Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 61.17 | 12.63 | 0.91 | |
| 1 | 9.83 | 0.79 | 0.05 | |
| >1 | 0.28 | 0.05 | 0.00 | |
| # plateaus | 6451 events | | | |
| *W/ Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 10.31 | 1.88 | 0.14 | |
| 1 | 1.74 | 0.14 | 0.00 | |
| >1 | 0.08 | 0.02 | 0.00 | |
| # plateaus | | | | |

| $1e$ | | | | |
|---|---|---|---|---|
| *No Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 39.93 | 13.58 | 1.27 | |
| 1 | 19.19 | 1.89 | 0.17 | |
| >1 | 1.74 | 0.04 | 0.02 | |
| # plateaus | 4721 events | | | |
| *W/ Brem.* | 1 | 2 | >2 | # peaks |
| 0 | 12.43 | 3.81 | 0.38 | |
| 1 | 4.58 | 0.47 | 0.06 | |
| >1 | 0.40 | 0.04 | 0.00 | |
| # plateaus | | | | |

TABLE 6.5: Percentages of the dataset that fell into each category. The "***No Brem.***" tables correspond to signals where no bremsstrahlung was detected, while the "***W/ Brem.***" tables are for signals where bremsstrahlung was detected according to the method described in Section 6.4.2.

| name | type | sans | N10 | N11 | N20 | Y10 | Z99 | description |
|---|---|---|---|---|---|---|---|---|
| ID | string | yes | yes | yes | yes | yes | yes | unique identifier for each signal |
| bremTF | bool | yes | no | no | no | no | yes | clear presence of bremsstrahlung |
| nPks | int | yes | no | no | no | no | yes | number of peaks |
| nPlat | int | yes | no | no | no | no | yes | number of plateaus |
| A5_95 | ampNsec | yes | yes | yes | yes | yes | yes | 90% of the total signal area |
| M5_95 | amp_arb | yes | yes | yes | yes | yes | yes | [5%-95%] AFT range signal mean |
| M10_90 | amp_arb | yes | yes | yes | yes | yes | yes | [10%-90%] AFT range signal mean |
| M25_75 | amp_arb | yes | yes | yes | yes | yes | yes | [25%-75%] AFT range signal mean |
| M5_50 | amp_arb | yes | yes | yes | yes | yes | yes | [5%-50%] AFT range signal mean |
| M5_75 | amp_arb | yes | yes | yes | yes | yes | yes | [5%-75%] AFT range signal mean |
| M5_90 | amp_arb | yes | yes | yes | yes | yes | yes | [5%-90%] AFT range signal meanl |
| M10_95 | amp_arb | yes | yes | yes | yes | yes | yes | [10%-95%] AFT range signal mean |
| M25_95 | amp_arb | yes | yes | yes | yes | yes | yes | [25%-95%] AFT range signal mean |
| M50_95 | amp_arb | yes | yes | yes | yes | yes | yes | [50%-95%] AFT range signal mean |
| L5_95 | nanosec | yes | yes | yes | yes | yes | yes | [5%-95%] AFT range signal length |
| L10_90 | nanosec | yes | yes | yes | yes | yes | yes | [10%-90%] AFT range signal length |
| L25_75 | nanosec | yes | yes | yes | yes | yes | yes | [25%-75%] AFT range signal length |
| L5_50 | nanosec | yes | yes | yes | yes | yes | yes | [5%-50%] AFT range signal length |
| L5_75 | nanosec | yes | yes | yes | yes | yes | yes | [5%-75%] AFT range signal length |
| L5_90 | nanosec | yes | yes | yes | yes | yes | yes | [5%-90%] AFT range signal length |
| L10_95 | nanosec | yes | yes | yes | yes | yes | yes | [10%-95%] AFT range signal length |
| L25_95 | nanosec | yes | yes | yes | yes | yes | yes | [25%-95%] AFT range signal length |
| L50_95 | nanosec | yes | yes | yes | yes | yes | yes | [50%-95%] AFT range signal length |
| pk_Xh | amp_arb | yes | yes | yes | yes | yes | yes | maximum height of signal |
| pk_Hi0 | amp_arb | y/n | yes | yes | yes | yes | y/n | height of the 1st earliest peak |
| pk_Hi1 | amp_arb | y/n | no | no | yes | no | y/n | height of the 2nd earliest peak |
| pk_Hi2 | amp_arb | y/n | no | no | no | no | y/n | height of the 3rd earliest peak |
| pk_Xns | nanosec | yes | yes | yes | yes | yes | yes | time of the highest signal peak |
| pk_ns0 | nanosec | y/n | yes | yes | yes | yes | y/n | time of the 1st earliest peak |
| pk_ns1 | nanosec | y/n | no | no | yes | no | y/n | time of the 2nd earliest peak |
| pk_ns2 | nanosec | y/n | no | no | no | no | y/n | time of the 3rd earliest peak |
| pltXh | amp_arb | y/n | no | yes | no | no | y/n | height of the highest plateau |
| pltHi0 | amp_arb | y/n | no | yes | no | no | y/n | height of the 1st earliest plateau |
| pltHi1 | amp_arb | y/n | no | no | no | no | y/n | height of the 2nd earliest plateau |
| pltHi2 | amp_arb | y/n | no | no | no | no | y/n | height of the 3rd earliest plateau |
| pltXns | nanosec | y/n | no | yes | no | no | y/n | time of the highest plateau |
| pltNs0 | nanosec | y/n | no | yes | no | no | y/n | time of the 1st earliest plateau |
| pltNs1 | nanosec | y/n | no | no | no | no | y/n | time of the 2nd earliest plateau |
| pltNs2 | nanosec | y/n | no | no | no | no | y/n | time of the 3rd earliest plateau |
| RMSw | nanosec | yes | yes | yes | yes | yes | yes | root mean square width |
| RMSa | amp_arb | yes | yes | yes | yes | yes | yes | root mean square amplitude |
| tSkew | double | yes | yes | yes | yes | yes | yes | similar to Pearson 1st skewness |
| aSkew | double | yes | yes | yes | yes | yes | yes | see Equation 6.2 |

*DESCRIPTION OF CUSTOM TYPES*

| | |
|---|---|
| amp_arb | custom typedef of double, represents signal amplitude, in arbitrary units |
| nanosec | custom typedef of double, represents time, in nanoseconds |
| ampNsec | custom typedef of double, represents area |

TABLE 6.6: List of the features used in each of the datasets given to the classifier. In the features marked y/n, the irrelevant data (e.g. the time of the second peak in the case of a signal with a single peak) was replaced with zeros. All times (e.g. pk_ns0) are measured after the 5% area fraction time.

## 6.5 Data optimization and testing

This stage of the procedure serves to mitigate the curse of dimensionality by finding the projection of the feature space dataset that will maximize the classifier's performance. This is achieved using a dimensionality reduction method. These methods take as input a $n_{\text{vec}} \times n_{\text{feat}}$ *feature matrix* $\mathbf{F}$ where each datapoint in the dataset is represented by a feature vector $\boldsymbol{F}_i$ , $i = 1, \ldots, n_{\text{vec}}$, each with $n_{\text{feat}}$ features, and then the methods output a $n_{\text{vec}} \times n_{\text{red}}$ matrix $\mathbf{R}$, where the feature vectors are represented in only $n_{\text{red}} \leq n_{\text{feat}}$ dimensions. This analysis was made in `MATLAB`, a script was written that performs the data optimization stage: it is listed in [254].

### 6.5.1 Data Optimization — Preprocessing

Dimensionality reduction methods generally work best with centered datasets where the variance is the same for all coordinates (Section 2.4.1 of [253]). The preprocessing script is listed in [254]. It takes as input the $1e$ and the $0\nu\beta\beta$ feature matrices, $\mathbf{F}^{1e}$ and $\mathbf{F}^{0\nu\beta\beta}$, concatenates them to form a $(n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta}) \times n_{\text{feat}}$ matrix $\mathbf{F}^{\text{all}}$, centers and scales $\mathbf{F}^{\text{all}}$ and then breaks up the resulting preprocessed matrix into $\mathbf{F}^{1e}_{\text{prep}}$ and $\mathbf{F}^{0\nu\beta\beta}_{\text{prep}}$. The centering is done by subtracting the mean row vector from each row, and the scaling is done by divinding each column by its standard deviation. The preprocessing was done both for continuous features (for example `L5_95`) and for discrete ones (for example `bremTF`, where the `true / false` data was converted to `1 / 0`).

### 6.5.2 Feature Selection

It is recommended that irrelevant variables be removed prior to dimensionality reduction (Section 1.3.1 of [253]). The feature selection script is listed in [254]. The chosen algorithm was sequential forward floating selection (SFFS) due to its good performance at high dimensionalities relative to its ease of implementation (see Figure 6.23). SFFS begins by choosing two features out of the dataset, and then proceeds to add and remove features according to what maximizes the value of some criterion function [255]. Although an ideal criterion function would refer to the performance of the classifier given the dataset, it was chosen to leave performance-based criteria for future work due to concerns about ease of implementation and computation speed. Instead the Mahalanobis distance **{fn6.5a}** between the $0\nu\beta\beta$ and $1e$ datasets was used due to its sensitivity to correlated features.

**{fn6.5a}** — The Mahalanobis distance $J$ between two datasets is a measure of the distance of their centers that gives more precedence to displacements in uncorrelated variables. For two class feature matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ it is defined as $J = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{\text{com}}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ [256] , where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the mean vectors for $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively, and $\boldsymbol{\Sigma}_{\text{com}}$ is the common covariance matrix for $\mathbf{F}_1$ and $\mathbf{F}_2$. Specifically, in this work the common covariance matrix was defined as $\boldsymbol{\Sigma}_{\text{com}} \equiv (\boldsymbol{\Sigma}_{1e} + \boldsymbol{\Sigma}_{0\nu\beta\beta})$, where $\boldsymbol{\Sigma}_{1e}$ and $\boldsymbol{\Sigma}_{0\nu\beta\beta}$ are the covariance matrices for the $1e$ and $0\nu\beta\beta$ matrices, respectively.

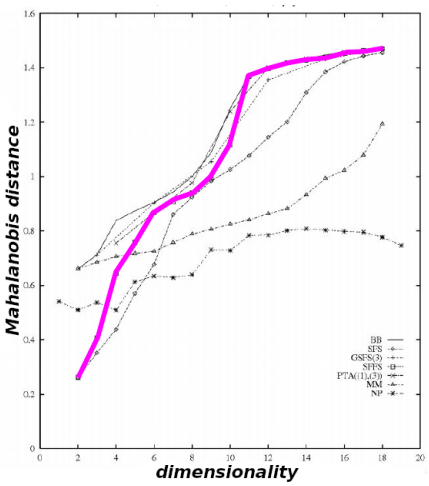


FIGURE 6.23: Performance comparison for different feature selection algorithms given a dataset with two multivariate Gaussian-distributed classes. SFFS is highlighted in magenta, and is shown to give the best performance out of all the algorithms for cases of high dimensionality. The performance criterion used in this analysis was the Mahalanobis distance **{fn6.5a}**.
Image source: [256]

SFFS was implemented according to the algorithm in the box below:

### *Inputs*

- $n_{\text{vec}}^{1e} \times n_{\text{feat}}$ preprocessed feature matrix $\mathbf{F}_{\text{prep}}^{1e}$;
- $n_{\text{vec}}^{0\nu\beta\beta} \times n_{\text{feat}}$ preprocessed feature matrix $\mathbf{F}_{\text{prep}}^{0\nu\beta\beta}$.

### *Outputs*

- $n_{\text{vec}}^{1e} \times n_{\text{SFFS}}$ , $n_{\text{SFFS}} \leq n_{\text{feat}}$ preprocessed, feature-selected matrix $\mathbf{F}_{\text{SFFS}}^{1e}$;
- $n_{\text{vec}}^{0\nu\beta\beta} \times n_{\text{SFFS}}$ , $n_{\text{SFFS}} \leq n_{\text{feat}}$ preprocessed, feature-selected matrix $\mathbf{F}_{\text{SFFS}}^{0\nu\beta\beta}$.

### *Initialization*

$\mathbf{F}_{\text{SFFS}}^{1e}$ and $\mathbf{F}_{\text{SFFS}}^{0\nu\beta\beta}$ begin the procedure having only two columns each, corresponding to the features least correlated between each other.

### *Selection procedure*

**while** (**Steps 1** and **2** together alter the columns in $\mathbf{F}_{\text{SFFS}}^{1e}$ and $\mathbf{F}_{\text{SFFS}}^{0\nu\beta\beta}$) **:**

**Step 1.** ***Inclusion***

Of the features not in $\mathbf{F}_{\text{SFFS}}^{1e}$ and $\mathbf{F}_{\text{SFFS}}^{0\nu\beta\beta}$, concatenate the one whose inclusion maximizes the Mahalanobis distance between the two classes;

**Step 2.** ***Conditional exclusion***

If there are features whose exclusion makes the Mahalanobis distance between the two classes even larger than the value obtained for Step 1, then remove the one for which the Mahalanobis distance is the largest.

*Note that even though $n_{SFFS} \leq n_{feat}$, $n_{SFFS}$ is still a very large number: datasets with more than 20 features after SFFS were common.*

### 6.5.3 Dimensionality Reduction

The dimensionality reduction script is listed in [254]. Section 6.4.3 explains the choice to use a linear dimensionality reduction method, but beyond this the chosen method must also be appropriate for the size of the dataset. PCA is the recommended option for large datasets with low dimensionality, while classical MDS is the preferred option for small datasets with large dimensionality (Section 4.2.2. of [253]). As such, the classical MDS method was chosen in this work.

Classical MDS supposes that a centered, scaled, high-dimensionality dataset can be losslessly projected on a lower-dimensional space by finding the orthogonal axis change that best preserves the scalar products of the datapoint vectors. In practice, the $(n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta}) \times n_{\text{SFFS}}$ preprocessed, feature-selected matrix $\mathbf{F}_{\text{SFFS}}^{\text{all}}$ was given as input, and two conditions were imposed:

$$\begin{cases} \mathbf{F}_{\text{SFFS}}^{\text{all}} = \mathbf{W}\mathbf{F}_{\text{MDS}}^{\text{all}} \\ (\mathbf{F}_{\text{SFFS}}^{\text{all}})^T \mathbf{F}_{\text{SFFS}}^{\text{all}} = (\mathbf{F}_{\text{MDS}}^{\text{all}})^T \mathbf{F}_{\text{MDS}}^{\text{all}} \end{cases}, \tag{6.3}$$

where $\mathbf{W}$ corresponds to the axis change itself. $\mathbf{F}_{\text{MDS}}^{\text{all}}$ is calculated by recognizing that the square matrix $(\mathbf{F}_{\text{SFFS}}^{\text{all}})^T \mathbf{F}_{\text{SFFS}}^{\text{all}}$ can be eigenvalue-decomposed:

$$(\mathbf{F}_{\text{SFFS}}^{\text{all}})^T \mathbf{F}_{\text{SFFS}}^{\text{all}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{6.4}$$

and so:

$$(\mathbf{F}_{\text{MDS}}^{\text{all}})^T \mathbf{F}_{\text{MDS}}^{\text{all}} = (\mathbf{\Lambda}^{1/2}\mathbf{U}^T)^T (\mathbf{\Lambda}^{1/2}\mathbf{U}^T). \tag{6.5}$$

Out of the resulting eigenvalues, the $n_{\text{red}}$ ones larger than or equal to $1\%$ of the maximum eigenvalue were used to form the output matrix $\mathbf{R}$:

$$\mathbf{R} = I_{n_{\text{red}} \times (n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta})} \mathbf{\Lambda}^{1/2} \mathbf{U}^T, \tag{6.6}$$

where $I_{n_{\text{red}} \times (n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta})}$ is a $n_{\text{red}} \times (n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta})$ matrix with ones along the diagonal and zeros everywhere else. An approximation of $\mathbf{W}$ can be retrieved as:

$$\mathbf{W} \approx \mathbf{F}_{\text{SFFS}}^{all} \mathbf{U} (\mathbf{\Lambda}^{-1/2}) I_{(n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta}) \times n_{\text{red}}}, \tag{6.7}$$

where $I_{(n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta}) \times n_{\text{red}}}$ is a $(n_{\text{vec}}^{1e} + n_{\text{vec}}^{0\nu\beta\beta}) \times n_{\text{red}}$ matrix with ones along the diagonal and zeros everywhere else.

## 6.6 Classifier implementation and performance assessment

Because of the decided-upon design of the classification procedure, beyond the choice of scenario, there were three other branching points that could make the datasets differ from each other. Each different set of decisions will be called a *configuration* in this work. The different dataset configurations are described in Section 6.6.1.

Following data optimization, the datasets for each configuration were transferred to the classifier, which:

1. Broke them up into training and test subsets;
2. Constructed a decision boundary using the training data;
3. Inferred for each test datapoint the likelihood that it had been produced by a $0\nu\beta\beta$ event;
4. Performed the classification according to if each likelihood exceeded a set threshold.

The implementation of the classification algorithms is described in Section 6.6.2.

This procedure was done with the intention of assessing how well the classifier performs with the data given for each scenario. It only makes sense to describe its performance according to the best configuration for each case. So prior to making a quantitative performance assessment, the performance of the different configurations had to be compared between each other to find the best one for b2b and for ISO scenarios. This brought along time concerns: it was determined that the GP classifier was very slow to calculate, and so the best use had to be made of a preferably small amount of test runs. The chosen solution for this was to subdivide the comparative performance assessment into two parts: first all the generated configurations were pitted against each other in a convergence test (denoted *P1A*); and then a cost curve comparison (denoted *P1B*) was done on the best-performing configurations from P1A. Again the configurations were compared and only the single best for each scenario was then passed over to the quantitative assessment (denoted *P2*). The comparative assessment procedure is described in Section 6.6.3, and the quantitative assessment procedure is described in Section 6.6.4.

### 6.6.1 Nomenclature for the dataset configurations

Below are listed the possible branching points:

**Branching point 1:** ***Categorization***

**cat** → Input dataset divided into categories;

**sans** → Input dataset left as is, without categorization **{fn6.6a}**.

**Branching point 2:** ***Optimization***

**prep** → Data optimization stops after preprocessing;

**SFFS** → Data optimization stops after feature selection;

**MDS** → Data optimization stops after dimensionality reduction.

**Branching point 3:** ***Algorithm***

**kNN** → Classification using k-nearest neighbors;

**RBF SVM** → Support vector machine classification with radial basis function kernel;

**GP** → Gaussian process classifier;

**RF** → Random forests classifier.

Additionally, as mentioned before, the datasets can be from the b2b or the ISO scenario. Table 6.7 shows the nomenclature used for all the possible configurations.

**{fn6.6a}** — Each sans configuration corresponds to a single dataset, while each cat configuration corresponds to 5 datasets: N10, N11, N20, Y10 and Z99.

***b2b scenario***

Branching point 1: ***Categorization*** → **cat**

| | | Branching point 2: ***Optimization*** | | |
|---|---|---|---|---|
| | | **prep** | **SFFS** | **MDS** |
| Branching point 3: ***Algorithm*** | **kNN** | `b2b_kNN_cat_prep` | `b2b_kNN_cat_SFFS` | `b2b_kNN_cat_MDS` |
| | **SVM** | `b2b_SVM_cat_prep` | `b2b_SVM_cat_SFFS` | `b2b_SVM_cat_MDS` |
| | **GP** | `b2b_GP_cat_prep` | `b2b_GP_cat_SFFS` | `b2b_GP_cat_MDS` |
| | **RF** | `b2b_RF_cat_prep` | `b2b_RF_cat_SFFS` | `b2b_RF_cat_MDS` |

Branching point 1: ***Categorization*** → **sans**

| | | Branching point 2: ***Optimization*** | | |
|---|---|---|---|---|
| | | **prep** | **SFFS** | **MDS** |
| Branching point 3: ***Algorithm*** | **kNN** | `b2b_kNN_sans_prep` | `b2b_kNN_sans_SFFS` | `b2b_kNN_sans_MDS` |
| | **SVM** | `b2b_SVM_sans_prep` | `b2b_SVM_sans_SFFS` | `b2b_SVM_sans_MDS` |
| | **GP** | `b2b_GP_sans_prep` | `b2b_GP_sans_SFFS` | `b2b_GP_sans_MDS` |
| | **RF** | `b2b_RF_sans_prep` | `b2b_RF_sans_SFFS` | `b2b_RF_sans_MDS` |

***ISO scenario***

Branching point 1: ***Categorization*** → **cat**

| | | Branching point 2: ***Optimization*** | | |
|---|---|---|---|---|
| | | **prep** | **SFFS** | **MDS** |
| Branching point 3: ***Algorithm*** | **kNN** | `ISO_kNN_cat_prep` | `ISO_kNN_cat_SFFS` | `ISO_kNN_cat_MDS` |
| | **SVM** | `ISO_SVM_cat_prep` | `ISO_SVM_cat_SFFS` | `ISO_SVM_cat_MDS` |
| | **GP** | `ISO_GP_cat_prep` | `ISO_GP_cat_SFFS` | `ISO_GP_cat_MDS` |
| | **RF** | `ISO_RF_cat_prep` | `ISO_RF_cat_SFFS` | `ISO_RF_cat_MDS` |

Branching point 1: ***Categorization*** → **sans**

| | | Branching point 2: ***Optimization*** | | |
|---|---|---|---|---|
| | | **prep** | **SFFS** | **MDS** |
| Branching point 3: ***Algorithm*** | **kNN** | `ISO_kNN_sans_prep` | `ISO_kNN_sans_SFFS` | `ISO_kNN_sans_MDS` |
| | **SVM** | `ISO_SVM_sans_prep` | `ISO_SVM_sans_SFFS` | `ISO_SVM_sans_MDS` |
| | **GP** | `ISO_GP_sans_prep` | `ISO_GP_sans_SFFS` | `ISO_GP_sans_MDS` |
| | **RF** | `ISO_RF_sans_prep` | `ISO_RF_sans_SFFS` | `ISO_RF_sans_MDS` |

TABLE 6.7: The nomenclature used for all the possible configurations.

### 6.6.2 Implementation of the classification algorithms

This subsection describes the aspects of the implementation that do not vary throughout the performance assessment. For details on the chosen threshold values, $\tau$, selection of test sets, and numbers of test iterations, see Sections 6.6.3 and 6.6.4.

The classification was done using the classes in `scikit-learn 0.21.3` [173] corresponding to each of the classification algorithms **{fn6.6b}**. The default parameter values were used for each class. In the output, the classifier mapped each datapoint in the test set onto a single dimension $L_{0\nu\beta\beta} \in [0, 1]$, representing the likelihood given by the classifier that the datapoint had been produced by a $0\nu\beta\beta$ event. Afterward, each test datapoint was given a label according to the threshold $\tau$: if $L_{0\nu\beta\beta} > \tau$, then the event was labeled as $0\nu\beta\beta$, and otherwise it was labeled as $1e$. The classifier labels were then compared to the actual labels to produce a final result: a pair of performance parameter values, fp and tp (see Section 5.1).

P1B tested a further algorithm, where the data was classified according to a vote done by the GP and RF algorithms. To implement this, the test data was first labeled by the two algorithms separately, and then it was labeled as $0\nu\beta\beta$ only if both algorithms had labeled it as such.

For more details on the implementation of the classification algorithms, see [261].

### 6.6.3 Comparative performance assessment

The objective of the comparative assessment is to clearly show which of the configurations yields the best performance. It is not so important to know exactly what the performance of a configuration is, as long as it is visibly better or worse than that of the others. During the design of this part, it was seen as preferable to favor low variance over low bias: if a performance measurement is made, and the performance assessment method biases the results equally across all configurations, then that is not a problem for a comparison because their performance relative to eachother is unaffected; but if the performance result of the configurations were to have strong statistical variation across each instance of the same configuration, then a comparative assessment becomes more difficult, as there is less confidence in the measurement outcome of the configurations' performance relative to eachother. Above all else, this preference manifested in opting for splitting samples into test and training data using Monte Carlo cross-validation in both tests, as it is a simple method with low variance [262].

#### P1A — Comparative performance assessment using convergence test

The theoretical fundamentals of this method are introduced in Section 5.2.2. Ideally the datasets $\mathbf{F}$ would have been large enough for a trend to emerge in the performance parameters even when the *sample size*, $n_{\text{in}}$, contained a small fraction of all the available datapoints. This would allow for ensembles with a statistically representative *number of samples* $\mathbf{n_{sam}}$ to still have all of them be disjoint from each other, minimizing the bias and variance of the convergence test results. However in this work the datasets were too small for a trend to emerge when using a small fraction of the data, so an alternative had to be chosen that would keep variance low, potentially at the expense of a higher bias.

**{fn6.6b}** — Specifically, the classes used were `KNeighborsClassifier()` [257] from the `sklearn.neighbors` module for kNN; `SVC()` [258] from the `sklearn.svm` module for RBF SVM; `GaussianProcessClassifier()` [196] from the `sklearn.gaussian_process` module for GP; and `RandomForestClassifier()` [259] from the `sklearn.ensemble` module for RF [260].

It was decided that, given a desired $n_{\text{in}}$ corresponding to a *fraction* $\phi$ of all the *available datapoints* $n_{\text{datapts}}$ **{fn6.6c}**, training and test data of all $\mathbf{n_{sam}}$ samples in the ensemble would always come from only the first $n_{\text{in}}/2$ datapoints in the dataset from each of the two classes (yielding $n_{\text{in}}$ datapoints with equal balancing). Although this approach increases the bias of the performance parameters, namely making the classifier more susceptible to outliers, it ensures that the variance will only change according to the successive differences in the distribution of the sample data along feature space, and not because of differences in the amount of overlap between samples **{fn6.6d}** For more information on the implementation of the test, see [263].

From each dataset, subdatasets were extracted with $n_{\text{in}}$'s given by $\phi = \{1\%, 2\%, 5\%, 10\%, 20\%, 50\%, 100\%\}$, with ensembles of size $\mathbf{n_{sam}} = 10$ each. Then the average and the standard deviation of fp and tp for each ensemble was calculated. When comparing the configurations, the fp and tp averages for the $\phi = 100\%$ ensemble were used as estimators $\widehat{\mathsf{fp}}$ and $\widehat{\mathsf{tp}}$ for the performance parameters of the dataset, and the trend of the fp and tp values for each dataset, $\widehat{\mathrm{tr}_{\mathsf{fp}}}$ and $\widehat{\mathrm{tr}_{\mathsf{tp}}}$, was estimated with an integer rating from 1 to 4, using the following standard:

**Rating 4:** ***Good trend***

*Performance clearly improves with larger subdatasets, standard deviation decreases as expected;*

**Rating 3:** ***No trend***

*Performance fluctuates without converging, standard deviation can fail to decrease;*

**Rating 2:** ***Questionable trend***

*Performance appears to worsen slightly with larger subdatasets, but the standard deviation is too large for an obvious trend to emerge;*

**Rating 1:** ***Bad trend***

*Performance clearly worsens with larger datasets, standard deviation either decreases as expected or increases.*

*See Figures 6.27, 6.28, 6.29 and 6.30 for examples of each trend.*

The values of the four estimators — $\widehat{\mathsf{fp}}$, $\widehat{\mathsf{tp}}$, $\widehat{\mathrm{tr}_{\mathsf{fp}}}$ and $\widehat{\mathrm{tr}_{\mathsf{tp}}}$ — can vary significantly between configurations, so it is more informative to compare configurations differing in one branching point at a time *(e.g. just compare* cat *and* sans **{fn6.6e}***, without interference from other branching points)* instead of comparing each configuration to all others simulatenously. The ranking of the configurations was done according to the *branching point test algorithm* on the next page.

**{fn6.6c}** — A variation in balancing from one dataset to another risked increasing the variance in the results. To ensure equal balancing, it was decided to define $n_{\text{datapts}} \equiv 2 \times \min(\{n_{\text{vec}}^{0\nu\beta\beta}, n_{\text{vec}}^{1e}\})$.

**{fn6.6d}** — This is in opposition to the case of sampling from the entire dataset every time. By doing so, the samples in an ensemble would initially be disjoint, but would become progressively more conjoint with the increase in sample size. This would add unpredictable behaviors to the test results.

**{fn6.6e}** — The estimators in the categorization branching point cannot be compared directly: while each sans configuration corresponds to a single dataset, each cat configuration has 5 datasets, summing up to the same number of datapoints as in the corresponding sans dataset. The estimators for each cat configuration were obtained with a weighted average of the estimator values of its corresponding datasets, with the weight being the dataset size.

1. Having selected a branching point to compare, group up the configurations into sample clusters so only the chosen branching point changes inside each cluster;

   *For example, to compare the performance of the three optimization options, the data is grouped into clusters of three configurations, where only the optimization changes, whereas the categorization and classification algorithm stay the same within the given cluster.*

2. Calculate the average value of $\widehat{\mathsf{fp}}$ , $\widehat{\mathsf{tp}}$, $\widehat{\mathsf{tr}_{\mathsf{fp}}}$ and $\widehat{\mathsf{tr}_{\mathsf{tp}}}$ for each cluster and calculate the displacement of each configuration's estimator values away from the cluster averages;

3. Sum the displacements in the estimator values of all the configurations with the same branching point choice in a given scenario;

   *Following the example given above, this would result in 6 final values, 3 for each scenario. The 3 final values in each scenario would correspond to the sum of deviations of all prep configurations, the sum of deviations of all SFFS configurations and the sum of deviations of all MDS configurations.*

4. For each scenario, decide on a ranking of the different branching point choices, according to the resulting displacement sums.

   *The final ranking according to the displacement sums was made usually giving precedence to good performance over a good performance trend* **{fn6.6f}**. *Regarding the performance parameters, precedence was generally given to a low* $\widehat{\mathsf{fp}}$ *value over a high* $\widehat{\mathsf{tp}}$*, due to the balancing in deployment.*

### P1B — Comparative performance assessment using cost curves

The configurations that performed best in P1A **{fn6.6g}** were then tested more thoroughly via their cost curves. The theoretical fundamentals of cost curves are introduced in Section 5.2.4. Again ensembles with $\mathsf{n_{sam}} = 10$ were used, but this time the sample size was fixed at $\phi = 100\%$. The ensemble averages for $\mathsf{fp}$ and $\mathsf{tp}$ were calculated for $\tau = \{0.1, 0.2, 0.3, \ldots, 0.9\}$, and these were then used to draw the straight lines in the cost space, whose lower envelope then forms the cost curve itself for the given configuration (see Figure 6.24 for a demonstration). For information on the script implementation, see [264].

To select a best classifier for each scenario, a sligtly different representation was used for the cost curve. The curves were normalized so that, at any value of $PC(\mathsf{P})$, the amplitude of the curve would be $E[Co]_{\text{norm}} = E[Co]/E[Co]_{\text{noClassifer}}$, where $E[Co]_{\text{noClassifier}}$ has the values of the red curve in Figure 6.24. This makes it easier to compare the misclassification costs of the different configurations for small $PC(\mathsf{P})$, the region of interest for this work **{fn6.6h}**. The classifier selected for quantitative performance assessment was the one with consistently the lowest misclassification cost in the region of interest. Although the goal of the comparative test

**{fn6.6f}** — This is because usually the configurations with the worst trends were the ones with smaller datasets, pointing to possibly a sparse dataset. Before deployment, a much larger training dataset would be necessary regardless.

**{fn6.6g}** — The number of configurations to send over to P1B was not decided upon in advance, instead opting for sending the best-performing configurations where the P1A results were already too similar to be decisive.

**{fn6.6h}** — This normalization will obviously increase the uncertainty near $PC(\mathsf{P}) = 0$ and $PC(\mathsf{P}) = 1$. In general the difficulty and time-expense in determining uncertainties for cost curves [210] was a contributing factor for choosing ROC curves over cost curves for the absolute test. However even with a high uncertainty the consistent underperformance of one classifier relative to another would still be a significant result, so for comparative tests the cost curve remains the most attractive solution.

may be to find the best configuration for each scenario, merely finding a very good configuration is also a satisfactory outcome: if two configurations show a near-identical performance in the cost curve, then it should be very similar in the receiver operator characteristic (ROC) curve as well. With this in consideration, although there are procedures suggested for constructing confidence bands in cost curves [210], they were not used here, instead relying on the assumption that if a cost curve made from 10 sample performance averages consistently appears below another one for a widerange of operating points, then that is a sufficiently significant result to be decisive.

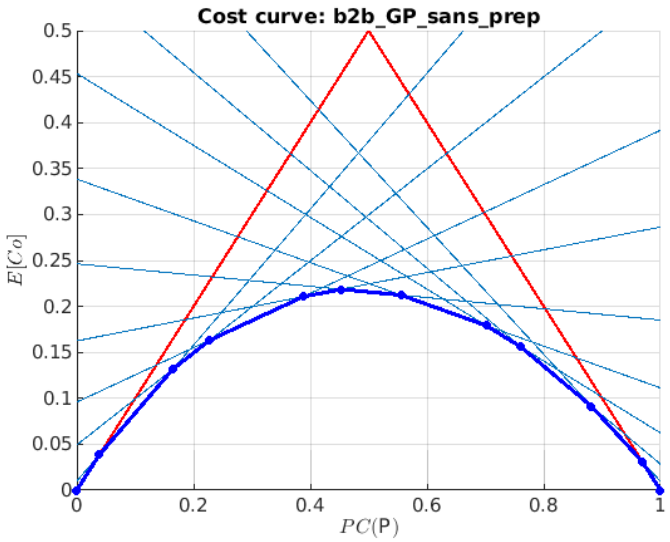


FIGURE 6.24: Cost curve (thick blue curve) obtained for the `b2b_GP_sans_prep` configuration. The two diagonal red lines are the "no classifier" case. The thin blue lines are constructed from the ensemble average `fp` and `tp` values for each studied $\tau$. The blue dots are the points where the lines intersect, forming the lower envelope. The top half of the cost space ($E[Co] > 0.5$) was cut out because there is never any relevant information there.

### 6.6.4 Quantitative performance assessment using ROC curves (P2)

With the best configurations for each scenario selected, the objective of the quantitative performance assessment is now to express the performance of these configurations in absolute terms, which will then be used as a surrogate for the performance of the classifier itself. The performance has to be expressed in quantities that have some physical significance, and the focus in this case is on ensuring that the estimators accurately portray the actual performance range. In this sense, unlike for the comparative assessment, this time it is the bias that needs to be minimized, even if at the expense of an increase in variance. These considerations dictated the design of the test. Summarily the chosen algorithm is listed on the next page.

**for each scenario**:

**Step 1.** Split the entire configuration dataset into test / training data in 10 different ways using 10-fold cross-validation;

**Step 2.** **for each split**:

***Step 2.a)*** Perform the classification for threshold values $\tau = \{0.01, 0.02, \ldots, 0.99\}$, obtaining the fp and tp values for each threshold;

***Step 2.b)*** Construct the ROC curve (see Section 5.2.3 for theoretical fundamentals);

***Step 2.c)*** For the obtained ROC curve, construct the confidence band with 99% confidence level (CL) using simultaneous joint confidence regions (SJR);

**Step 3.** Take the intersection of all the confidence bands to produce the 90% CL simultaneous confidence band for the classifier performance. Take the AUC 90% CL confidence interval to be the AUC of the upper and lower bounds of the ROC confidence band. Take the AUC estimate to be the AUC of the average ROC curve.

With this, there will be a confidence of 90% that the actual average ROC curve of the classifier will be inside the resulting confidence band. Below, aspects of the chosen algorithm are described in detail. For additional information, see [265].

### Splitting data using 10-fold cross-validation

10-fold cross-validation is a splitting method where successive disjoint 10% chunks of the full dataset are used as the test sets (see Figure 6.25) **{fn6.6i}**. This method was chosen for three reasons:

**Low bias and fairly low variance**

[262] *compares the performance of different splitting methods, concluding that of a number of assessed methods, leave-one-out and 10-fold cross-validation yield the smallest bias for a collection of tested classification algorithms.*

**All test sets are disjoint**

*This means that all test sets are independent samples of the same distribution: they are independent and identically distributed (iid). This is useful for constructing the final confidence band.*

**Ease of implementation**

*Another option could have been to use leave-one-out cross-validation. This algorithm is like 10-fold cross-validation, except that the test set is a single data point each time. Using it would be very time-consuming, considering the slowness of the classification algorithm.*

**{fn6.6i}** — To avoid having test sets with a single class, the $0\nu\beta\beta$ and $1e$ datasets were concatenated and then shuffled. For more information see [261].

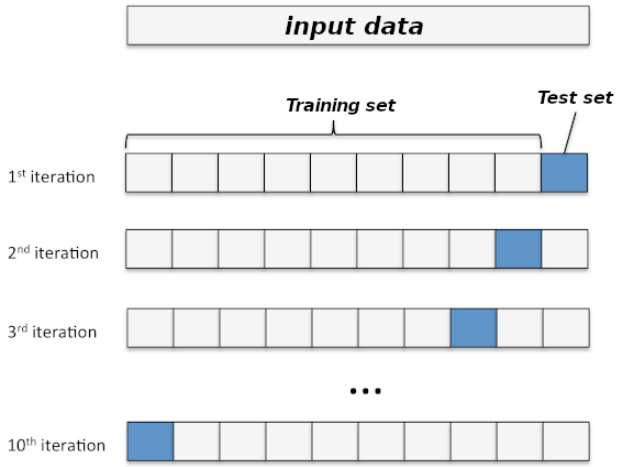


FIGURE 6.25: Schematic depiction of 10-fold cross-validation.
Image source: [266]

### Constructing ROC curve confidence bands using simultaneous joint confidence regions

Simultaneous joint confidence regions (SJR) is a method where confidence rectangles sized according to the Kolmogorov-Smirnov (KS) statistic are centered on each tested $\tau$ and then joined together to form a simultaneous confidence band (Figure 6.26) [267]. It reverses the standard usage of the KS statistic by supposing that if an ROC curve is constructed by sampling some performance distribution, then the actual curve could not be further away from the sample curve than some fixed amount **{fn6.6j}**. The dimensions of the confidence rectangles depend on the desired CL and on the number of datapoints from each class in the test dataset. For a 99% CL rectangle, the height $h_{99\%}$ and width $w_{99\%}$ are given as:

$$h_{99\%} = \frac{1.628}{\sqrt{n_{\text{test}}^{0\nu\beta\beta}}} \quad , \quad w_{99\%} = \frac{1.628}{\sqrt{n_{\text{test}}^{1e}}},$$

where $n_{\text{test}}^{0\nu\beta\beta}$ and $n_{\text{test}}^{1e}$ are the numbers of datapoints from each class in the current test set and 1.628 is obtained from the KS statistic [268].

**{fn6.6j}** — The reasoning here is based on interpreting the sample fp and tp values as complements to discrete cumulative distribution functions (cdf): when a test set is run through a classifier, it outputs two pdf samples — $p_s(L_{0\nu\beta\beta})$ for the $0\nu\beta\beta$ datapoints, and $n_s(L_{0\nu\beta\beta})$ for the $1e$ datapoints — and then when a $\tau$ value is chosen, the true positive classifications happen for the datapoints that fall into $p_s(L_{0\nu\beta\beta} > \tau)$, while the false positives happen for the ones that fall into $n_s(L_{0\nu\beta\beta} > \tau)$; because the corresponding discrete cdf's are defined as $P_s[\tau] \equiv \sum_{L_{0\nu\beta\beta}=0}^{\tau} p_s(L_{0\nu\beta\beta})$ and $N_s[\tau] \equiv \sum_{L_{0\nu\beta\beta}=0}^{\tau} n_s(L_{0\nu\beta\beta})$, for some $\tau$ the performance parameters can be expressed as $\text{fp} = 1 - N_s[\tau]$ and $\text{tp} = 1 - P_s[\tau]$. This being the case, fp and tp are complements to discrete cdfs for a sample of some distributions with continuous cdfs $P(\tau)$ and $N(\tau)$. The distributions themselves are unknown, but by Kolmogorov's theorem it *is* known with some % CL that $|P(\tau) - P_s[\tau]|$ and $|N(\tau) - N_s[\tau]|$ cannot exceed some value given by the Kolmogorov distribution, which does not depend on the distribution in question. For more information, see [268].

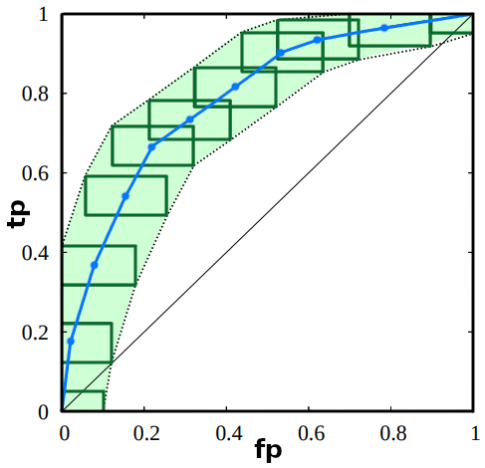


FIGURE 6.26: Example of the use of SJR. The sample ROC curve is highlighted in blue with the test points marked with large dots, and the simultaneous confidence band is highlighted in green. Confidence rectangles are placed centered on each test point.. Then to form the simultaneous confidence band the top left and bottom right vertices of all the confidence rectangles are connected together with straight lines.
Image source: [267].

SJR was chosen for its speed, ease of implementation and comparatively good performance. [267] lists a collection of methods for constructing ROC confidence bands, and then tests them empirically. SJR was one of only 2 methods whose purported CL was even remotely accurate, all others significantly overestimating the confidence. SJR was seen to actually slightly underestimate its own CL for fp and tp based on normal cumulative distribution functions and 99% CL confidence rectangles, while the other method (fixed-width bands) was more accurate in its confidence, but it required many more samples to produce a result, so it was opted against in favor of SJR.

### Using the intersection of all 99% CL sample confidence bands as the 90% CL final confidence band

Considering that all the sample curves are independent and identically distributed (iid), meta-analysis procedures recommend pooling the data to produce a final confidence band [269]. Then the same procedure can be done as for SJR, forming confidence rectangles centered at each $\tau$ performance parameter average with dimensions of roughly $1.22/\sqrt{10 \times n^{1e}_{\text{test}}}$. For the sake of simplicity, however, a different procedure was used: the intersection of all 10 99% CL confidence bands was used as a single 90% CL band, based on an argument about the joint probability of independent sets **{fn6.6k}**.

**{fn6.6k}** — The likelihood of the actual ROC curve falling within the confidence band made by one sample is $Pr(\text{ in } \mathtt{CB}_1 ) = 0.99$. The likelihood of it falling within the confidence band of that sample and also that of another one at the same time is $Pr(\text{ in } \mathtt{CB}_1 \cap \text{ in } \mathtt{CB}_2 ) = Pr(\text{ in } \mathtt{CB}_1 )Pr(\text{ in } \mathtt{CB}_2 ) = Pr(\text{ in } \mathtt{CB}_1 )^2 = 0.99^2$, since the two samples are iid. Extending this to all samples, the likelihood of the actual ROC curve falling within the intersection of all confidence bands is is $0.99^{10} \approx 0.9$.

**Interpretation of the AUC value**

This is an extension of the description of the ROC curve theoretical fundamentals in Section 5.2.3. Beyond the intuitive notion that larger AUC values are better, the interpretation of specific AUC values is more subtle: it is not a linear measure. Summarily the AUC is a sum comparative measure of how certain the classifier is that $0\nu\beta\beta$ test datapoints are in fact $0\nu\beta\beta$, vs. how certain it is that $1e$ datapoints are *not* $0\nu\beta\beta$. Below is an attempt at a statistical definition, along with a description of a few limitations of the measure.

Consider an infinitely large test dataset with the two classes $0\nu\beta\beta$ and $1e$. When fed to the classifier, the $0\nu\beta\beta$ datapoints will output one pdf of $L_{0\nu\beta\beta}$ values, $p(L_{0\nu\beta\beta})$, and the $1e$ will output another one, $n(L_{0\nu\beta\beta})$. Their corresponding cdfs are $P(\tau) \equiv \int_0^\tau dL_{0\nu\beta\beta}\ p(L_{0\nu\beta\beta})$ and $N(\tau) \equiv \int_0^\tau dL_{0\nu\beta\beta}\ n(L_{0\nu\beta\beta})$, and the performance parameters are defined as $\mathsf{tp} = \mathsf{tp}(\tau) = 1 - P(\tau)$ and $\mathsf{fp} = \mathsf{fp}(\tau) = 1 - N(\tau)$. After forming a continuous ROC curve, the AUC can be calculated as **{fn6.6l}**:

$$\text{AUC} = \int_0^1 d\tau\ n(\tau)\mathsf{tp}(\tau).$$

The two factors in the integral can be interpreted probabilistically. The pdf value $n(\tau)$ represents the probability that a $1e$ datapoint will output a $0\nu\beta\beta$ likelihood equal to $\tau$: $n(\tau) \equiv Pr(L_{0\nu\beta\beta} = \tau|1e)$. Meanwhile datapoints are labeled as $0\nu\beta\beta$ when $L_{0\nu\beta\beta}$ exceeds a thershold $\tau$, so the value of $\mathsf{tp}(\tau)$ is the probability that $0\nu\beta\beta$ datapoints will have a $0\nu\beta\beta$ likelihood larger than $\tau$: $\mathsf{tp}(\tau) \equiv Pr(L_{0\nu\beta\beta} > \tau|0\nu\beta\beta)$. So the AUC expression can be rewritten as:

$$\text{AUC} = \int_0^1 d\tau\ Pr(L_{0\nu\beta\beta} = \tau|1e)Pr(L_{0\nu\beta\beta} > \tau|0\nu\beta\beta),$$

so, because the distributions are independent:

$$\text{AUC} = E[\,Pr(\,L_{0\nu\beta\beta} = \tau|1e\ \cap\ L_{0\nu\beta\beta} > \tau|0\nu\beta\beta\,)\,] \quad , \quad \tau \in [0,1].$$

The AUC is the expected probability that the classifier will give a higher $L_{0\nu\beta\beta}$ to a $0\nu\beta\beta$ datapoint than a $1e$ datapoint [270]. Assuming that each of the two distributions either has one peak or a tight cluster of peaks, then if the two distributions have no $L_{0\nu\beta\beta}$ values in common, AUC $= 1$ if the classifier is correctly trained and AUC $= 0$ if it is inversely trained. The worst possible case is AUC $= 1/2$, which happens when the two distributions are identical, and so $0\nu\beta\beta$ has a higher $L_{0\nu\beta\beta}$ exactly half the time. The more similar the two distributions are, the more the AUC converges on $1/2$.

The expression above clearly shows some limitations of the AUC as a selection criterion. Since it is a nonlinear measure, it is hard to judge the relationship between the AUC value of different classifiers, other than knowing that typically a larger AUC is preferable. Additionally a wide range of distribution shapes can yield the same AUC value, so without knowing them in advance, the AUC will give little indication of how far apart the distributions are, and so it is a bad measure of how well the classifier is expected to perform in deployment, where the balancing can potentially drown out one of the distributions if it is too wide. Still, since the AUC is a popular measure of the performance of a classifier, it will be included.

**{fn6.6l}** — The ROC curve can be interpreted as the function $\mathsf{tp} = \mathsf{tp}(\mathsf{fp})$, so the AUC becomes AUC $\equiv \int_0^1 d\mathsf{fp}\ \mathsf{tp}(\mathsf{fp})$. Changing variables once with $d\mathsf{fp} = -d(N(\tau))$ gives AUC $= \int_0^1 d(N(\tau))\ \mathsf{tp}(N(\tau))$, and then the variables can be changed again with $d(N(\tau)) = n(\tau)d\tau$ [213].

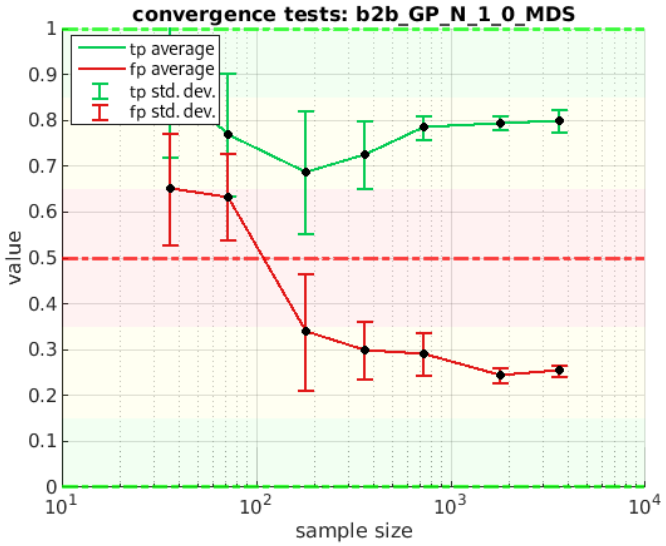


FIGURE 6.27: Example of a convergence test with a good trend. Note that the bars are not error bars, but instead show the $1\sigma$ width of the sample distribution.

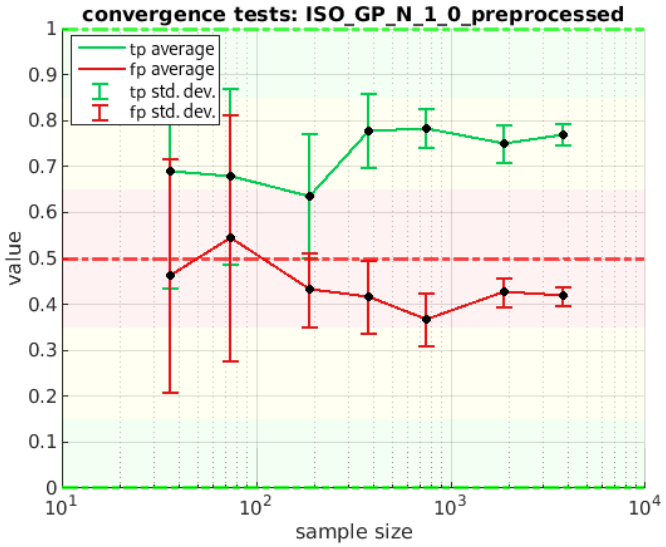


FIGURE 6.28: Example of a convergence test with no trend. Note that the bars are not error bars, but instead show the $1\sigma$ width of the sample distribution.

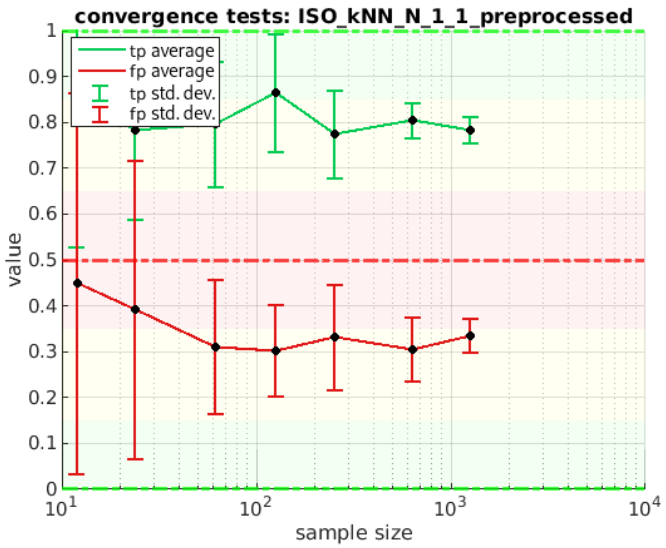


FIGURE 6.29: Example of a convergence test with a questionable trend. Note that the bars are not error bars, but instead show the $1\sigma$ width of the sample distribution.

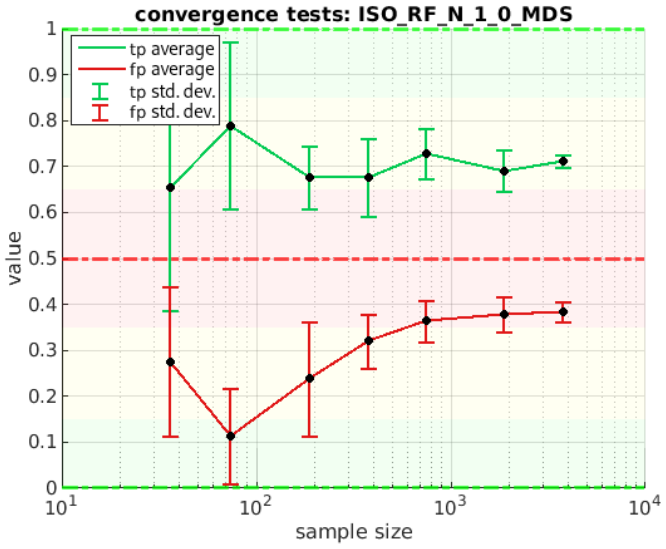


FIGURE 6.30: Example of a convergence test with a bad trend. Note that the bars are not error bars, but instead show the $1\sigma$ width of the sample distribution.

# Chapter 7

# Results and Discussion

This chapter presents and discusses the results obtained from the binary classification tests for each scenario (Section 6.1.1, pg. 87). Section 7.1 presents the results of the performance convergence test, indicating which configurations passed onto the cost curve test, and the results of that test are shown in Section 7.2. Finally the performance in absolute terms is presented for best configurations of each scenario in Section 7.3. An analysis of the results, discussion of methodology, the viability of binary classification in its current form for background discrimination in LZ, and other potential applications of the developed classifier are presented in Section 7.4.

## 7.1 Convergence comparative test (P1A)

The sampling procedure described in Section 6.6.3 (pgs. 122 - 124) was used on all the obtained datasets (see Table 6.7) to produce convergence graphs like those in Figures 6.27, 6.28, 6.29 and 6.30, which were then evaluated to obtain $\widehat{\mathsf{fp}}$, $\widehat{\mathsf{tp}}$, $\widehat{\mathsf{tr}_{\mathsf{fp}}}$ and $\widehat{\mathsf{tr}_{\mathsf{tp}}}$ (pg. 123) for each configuration (see Section 6.6.1). Then, by the decided-upon algorithm, the different branching points were compared, and the configurations that qualify for the cost curve test (P1B, pgs. 124 - 125) were selected. The obtained results for the b2b and ISO scenarios are shown in Section 7.1.1 and 7.1.2 respectively, and the selection of qualifying configurations is described in Section 7.1.3.

### 7.1.1 b2b scenario results

A 50 / 50 balancing of $0\nu\beta\beta$ and $1e$ datapoints was maintained for both the sans datasets (see box on pg. 114) as well as all the datasets corresponding to each cat configuration (Section 6.4.3). The numbers of datapoints are listed in Table 7.1. The four estimator values for each cat configuration were obtained by a weighted average of the resulting estimator values for each of their corresponding datasets. The results for all configurations are listed in Table 7.2.

| b2b | # datapoints | $W$ |
|---|---|---|
| **N10** | 3618 | 0.36 |
| **N11** | 1004 | 0.10 |
| **N20** | 1478 | 0.15 |
| **Y10** | 878 | 0.09 |
| **Z99** | 3010 | 0.30 |
| ***sans*** | *9988* | *1.00* |

TABLE 7.1: Numbers of datapoints for each category (see box on pg. 114) and for sans, together with the weights $W$ used when averaging the performance parameters to obtain the performance parameters of the cat configurations.

The sum displacements obtained in the branching point test are listed in Table 7.3, together with the decided-upon rankings of the branching point options. An example of a displacement calculation is shown in Table 7.4. When ranking, in all cases except one the decisions were made

based on the $\widehat{\text{fp}}$ sum displacements, as the $\widehat{\text{tp}}$ sum displacements usually corroborated them, and so the trend estimators $\widehat{\text{tr}_{\text{fp}}}$ and $\widehat{\text{tr}_{\text{tp}}}$ were not given much importance. The one exception to this was the decision to place GP in the 1st rank and RF in the 2nd, even though RF has a lower $\widehat{\text{fp}}$ sum displacement. This was chosen because RF obtained the worst $\widehat{\text{tp}}$ sum displacement out of the *Algorithm* branching point (see Section 6.6.1), whereas the sum displacements for GP are more well-balanced. Of note too is that for the *Optimization* and *Categorization* branching points, both the best branching options have the worst trends.

| **Configuration name** | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
|---|---|---|---|---|
| b2b_RF_sans_prep | 0.20 | 4 | 0.75 | 4 |
| b2b_GP_cat_MDS | 0.22 | 4 | 0.81 | 4 |
| b2b_RF_cat_prep | 0.22 | 3 | 0.77 | 3 |
| b2b_RF_cat_MDS | 0.22 | 4 | 0.75 | 4 |
| b2b_GP_cat_SFFS | 0.22 | 4 | 0.81 | 3 |
| b2b_GP_cat_prep | 0.23 | 4 | 0.81 | 4 |
| b2b_RF_cat_SFFS | 0.23 | 4 | 0.77 | 3 |
| b2b_SVM_cat_MDS | 0.24 | 4 | 0.81 | 3 |
| b2b_SVM_cat_prep | 0.25 | 4 | 0.79 | 2 |
| b2b_GP_sans_prep | 0.25 | 4 | 0.80 | 4 |
| b2b_GP_sans_SFFS | 0.25 | 4 | 0.80 | 4 |
| b2b_GP_sans_MDS | 0.25 | 4 | 0.80 | 4 |
| b2b_RF_sans_SFFS | 0.25 | 4 | 0.75 | 4 |
| b2b_RF_sans_MDS | 0.25 | 4 | 0.75 | 4 |
| b2b_SVM_cat_SFFS | 0.26 | 4 | 0.80 | 3 |
| b2b_kNN_cat_prep | 0.27 | 3 | 0,78 | 3 |
| b2b_kNN_cat_SFFS | 0.29 | 3 | 0.77 | 4 |
| b2b_kNN_cat_MDS | 0.30 | 3 | 0.78 | 3 |
| b2b_SVM_sans_SFFS | 0.30 | 4 | 0.85 | 4 |
| b2b_SVM_sans_prep | 0.30 | 4 | 0.85 | 3 |
| b2b_SVM_sans_MDS | 0.30 | 4 | 0.80 | 3 |
| b2b_kNN_sans_prep | 0.30 | 4 | 0.75 | 4 |
| b2b_kNN_sans_SFFS | 0.30 | 4 | 0.75 | 4 |
| b2b_kNN_sans_MDS | 0.30 | 4 | 0.75 | 3 |

TABLE 7.2: Resulting estimator values from graph evaluation and weighted averaging. The highlight colors are a simple visual indication of how "good" each of the values is, with blue being "better" than green.

| *b2b scenario* | | estimator sum displacement | | | |
|---|---|---|---|---|---|
| **rank** | ***Algorithm*** | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
| **1** | **GP** | -0.13 | 1.0 | 0.12 | 2.0 |
| **2** | **RF** | -0.18 | 0.0 | -0.17 | 1.0 |
| **3** | **SVM** | 0.11 | 0.8 | 0.19 | -3.0 |
| **4** | **kNN** | 0.21 | -2.0 | -0.14 | 0.0 |
| | | estimator sum displacement | | | |
| **rank** | ***Optimization*** | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
| **1** | **prep** | -0.05 | -0.7 | 0.02 | -1.0 |
| **2** | **MDS** | 0.02 | 0.3 | -0.04 | 1.0 |
| **3** | **SFFS** | 0.04 | 0.3 | 0.02 | 0.0 |
| | | estimator sum displacement | | | |
| **rank** | ***Categorization*** | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
| **1** | **cat** | -0.16 | -2.0 | 0.03 | -3.0 |
| **2** | **sans** | 0.16 | 2.0 | -0.03 | 3.0 |

TABLE 7.3: Rankings obtained from the branching point tests, together with the obtained sum displacements. Note that for $\widehat{\text{fp}}$ lower sum displacements are better, while for the rest of the estimators a higher sum displacement is preferable.

| Group | Group avg. | GP | RF | SVM | kNN |
|---|---|---|---|---|---|
| **sans_prep** | 0.263 | -0.013 | -0.063 | 0.038 | 0.038 |
| **cat_MDS** | 0.244 | -0.028 | -0.028 | 0.001 | 0.055 |
| **cat_prep** | 0.239 | -0.014 | -0.023 | 0.006 | 0.031 |
| **cat_SFFS** | 0.250 | -0.029 | -0.022 | 0.015 | 0.036 |
| **sans_SFFS** | 0.275 | -0.025 | -0.025 | 0.025 | 0.025 |
| **sans_MDS** | 0.275 | -0.025 | -0.025 | 0.025 | 0.025 |
| ***$\widehat{fp}$ sum displacement*** | | -0.13 | -0.18 | 0.11 | 0.21 |

TABLE 7.4: Example sum displacement calculation, for the $\widehat{\text{fp}}$ estimator, in the *Algorithm* branching point. The configurations are joined into groups where only the *Algorithm* differs between configurations and the group average is calculated for each. Then, the group average is subtracted from the $\widehat{\text{fp}}$ values of each configuration. The sum of the obtained deviations is then taken for each *Algorithm*.

Finally one more important detail is that MDS optimization fares slightly better for cat configurations than prep does (see Table 7.5), even though overall prep still outperforms MDS.

| Configuration name | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
|---|---|---|---|---|
| b2b_GP_cat_MDS | 0.22 | 4 | 0.81 | 4 |
| b2b_RF_cat_prep | 0.22 | 3 | 0.77 | 3 |
| b2b_RF_cat_MDS | 0.22 | 4 | 0.75 | 4 |
| b2b_GP_cat_prep | 0.23 | 4 | 0.81 | 4 |

TABLE 7.5: MDS optimization faring slightly better than prep for cat configurations. The kNN, SVM and SFFS configurations were removed for clarity. Since they performed worse, they should contribute less to the decision.

### 7.1.2 ISO scenario results

The approach to balancing and the performance of the cat configurations was the same here as the one described in Section 7.1.1. Table 7.6 lists the number of datapoints. The results for all configurations are listed in Table 7.7. Sum displacements and rankings listed in Table 7.8.

| b2b | # datapoints | $W$ |
|---|---|---|
| **N10** | 3774 | 0.40 |
| **N11** | 1266 | 0.13 |
| **N20** | 1282 | 0.14 |
| **Y10** | 1168 | 0.12 |
| **Z99** | 1952 | 0.21 |
| ***ISO*** | *9442* | *1.00* |

TABLE 7.6: Numbers of datapoints for each category and for sans, together with the weights $W$ used when averaging the performance parameters to obtain the performance parameters of the cat configurations.

| Configuration name | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
|---|---|---|---|---|
| ISO_RF_cat_MDS | 0.34 | 2 | 0.72 | 3 |
| ISO_RF_cat_SFFS | 0.34 | 2 | 0.69 | 3 |
| ISO_RF_cat_prep | 0.34 | 2 | 0.70 | 3 |
| ISO_GP_cat_prep | 0.35 | 2 | 0.77 | 3 |
| ISO_GP_sans_MDS | 0.35 | 2 | 0.80 | 4 |
| ISO_RF_sans_prep | 0.35 | 3 | 0.75 | 4 |
| ISO_RF_sans_SFFS | 0.35 | 3 | 0.75 | 4 |
| ISO_RF_sans_MDS | 0.35 | 1 | 0.75 | 4 |
| ISO_kNN_sans_SFFS | 0.35 | 4 | 0.70 | 3 |
| ISO_kNN_sans_MDS | 0.35 | 4 | 0.70 | 3 |
| ISO_GP_cat_SFFS | 0.35 | 2 | 0.78 | 3 |
| ISO_kNN_cat_SFFS | 0.36 | 3 | 0.70 | 2 |
| ISO_kNN_cat_MDS | 0.36 | 3 | 0.69 | 2 |
| ISO_GP_cat_MDS | 0.38 | 2 | 0.78 | 3 |
| ISO_kNN_cat_prep | 0.39 | 2 | 0.69 | 2 |
| ISO_GP_sans_prep | 0.40 | 1 | 0.80 | 4 |
| ISO_GP_sans_SFFS | 0.40 | 1 | 0.80 | 4 |
| ISO_kNN_sans_prep | 0.40 | 4 | 0.70 | 3 |
| ISO_SVM_cat_SFFS | 0.41 | 2 | 0.81 | 3 |
| ISO_SVM_cat_MDS | 0.41 | 2 | 0.79 | 3 |
| ISO_SVM_cat_prep | 0.43 | 2 | 0.80 | 3 |
| ISO_SVM_sans_prep | 0.45 | 3 | 0.85 | 4 |
| ISO_SVM_sans_SFFS | 0.45 | 3 | 0.85 | 4 |
| ISO_SVM_sans_MDS | 0.45 | 3 | 0.85 | 4 |

Table 7.7: Resulting estimator values from graph evaluation and weighted averaging. The highlight colors are a simple visual indication of how "good" each of the values is, with cyan being "better" than green.

*ISO scenario* — estimator sum displacement

| rank | *Algorithm* | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
|---|---|---|---|---|---|
| 1 | RF | -0.21 | -1.5 | -0.19 | 1.5 |
| 2 | GP | -0.04 | -4.5 | 0.18 | 1.5 |
| 3 | kNN | -0.06 | 5.5 | -0.37 | -4.5 |
| 4 | SVM | 0.31 | 0.5 | 0.39 | 1.5 |

estimator sum displacement

| rank | *Optimization* | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
|---|---|---|---|---|---|
| 1 | MDS | -0.05 | -0.3 | 0.00 | 0.0 |
| 2 | SFFS | -0.03 | 0.7 | 0.01 | 0.0 |
| 3 | prep | 0.08 | -0.3 | -0.01 | 0.0 |

estimator sum displacement

| rank | *Categorization* | $\widehat{\text{fp}}$ | $\widehat{\text{tr}_{\text{fp}}}$ | $\widehat{\text{tp}}$ | $\widehat{\text{tr}_{\text{tp}}}$ |
|---|---|---|---|---|---|
| 1 | sans | 0.01 | 1.0 | 0.07 | 2.0 |
| 2 | cat | -0.01 | -1.0 | -0.07 | -2.0 |

Table 7.8: Rankings obtained from the branching point tests, together with the obtained sum displacements.

When ranking, the same priorities were chosen as in Section 7.1.1. Again the $\widehat{\text{tp}}$ sum displacement for RF is one of the worst, even though it outperforms the other algorithms in regard to the $\widehat{\text{fp}}$ sum displacement. GP was ranked higher than kNN because GP outperforms it in regards to true positive rate by a very significant margin, whereas their difference in $\widehat{\text{fp}}$ is very minor. In the *Categorization* ranking, precedence was given to the fact that sans outperformed cat in every respect other than $\widehat{\text{fp}}$.

### 7.1.3 Transfer to cost curve comparative test (P1B)

Summarily, it can be seen that the biggest differences in performance appear when comparing the classification algorithms: RF and GP consistently outperform kNN and SVM. The choice between RF and GP is more nuanced: the GP performance is more consistent throughout the four estimators, while RF has very good $\widehat{\mathsf{fp}}$ and comparatively poor $\widehat{\mathsf{tp}}$. The optimization results are less disparate, but MDS seems to generally be the best optimization option, although for b2b scenario sans configurations the prep optimization is preferable. The trend performance of prep configurations is slightly worse than that of more optimized ones, unsurprisingly. More alarming is that MDS has the same trend performance for the ISO scenario as prep does: this will be elaborated upon in the discussion (Section 7.4). Finally for the *Categorization* branching point the results are not decisive: their performance is not consistent between the two scenarios.

With these points in consideration, the following configurations advanced to the cost curve test:

- `b2b_GP_cat_MDS`;
- `b2b_GP_sans_prep`;
- `b2b_RF_cat_MDS`;
- `b2b_RF_sans_prep`;
- `ISO_GP_cat_MDS`;
- `ISO_GP_sans_MDS`;
- `ISO_RF_cat_MDS`;
- `ISO_RF_sans_MDS`.

As mentioned in Section 6.6.2, an additional classification algorithm was passed over into P1B, where GP and RF vote on whether to classify the datapoints as $0\nu\beta\beta$. This algorithm will be referred to as GvR, and its addition yields 4 extra configurations:

- `b2b_GvR_cat_MDS`;
- `b2b_GvR_sans_prep`;
- `ISO_GvR_cat_MDS`;
- `ISO_GvR_sans_MDS`;

## 7.2 Cost curve comparative test (P1B)

Cost curves were generated for the 12 configurations *(8 selected + 4 extra)* transferred from P1A, according to the specifications referred to in Section 6.6.3. The obtained results for the b2b and ISO scenarios are shown on pages 138 and 139 respectively, and the selection of qualifying configurations is described on page 140.

### b2b scenario results

The resulting cost curves are displayed in Figure 7.1, together with the "no classifier" case. The cost curves with the alternative normalization described in Section 6.6.3 are displayed in Figure 7.2. The best classifier for each operating point $PC(\mathsf{P})$ is the one that matches the "lower envelope" curve.

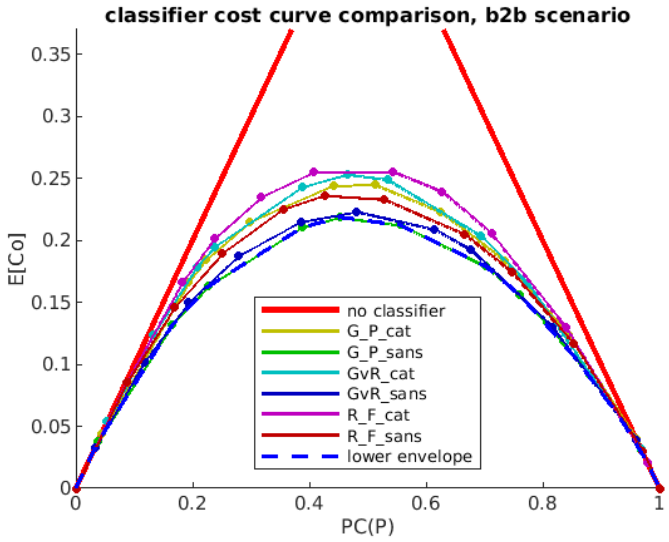


FIGURE 7.1: b2b scenario cost curves for the b2b configurations that qualified for P1B, normalized to $\max_{\mathsf{fn},\mathsf{fp}} \{\varepsilon[Co]\}$.

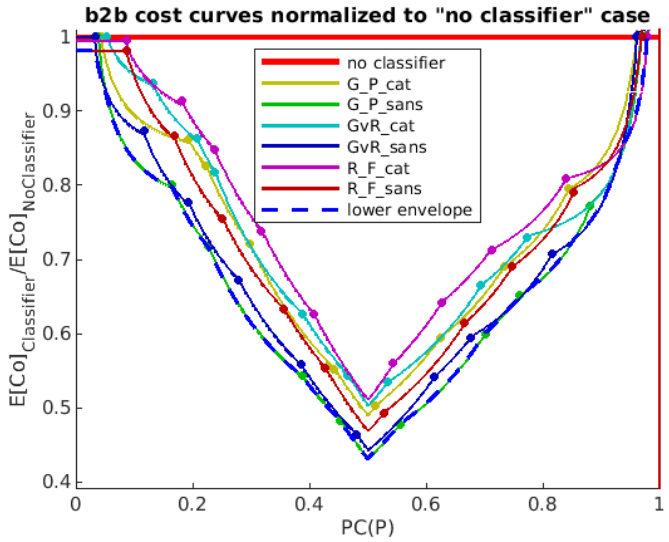


FIGURE 7.2: b2b scenario cost curves for the b2b configurations that qualified for P1B, normalized to the "no classifier" case.

### ISO scenario results

The resulting cost curves are displayed in Figure 7.3, together with the "no classifier" case. The cost curves with the alternative normalization described in Section 6.6.3 are displayed in Figure 7.4. The best classifier for each operating point is the one that matches the "lower envelope" curve.

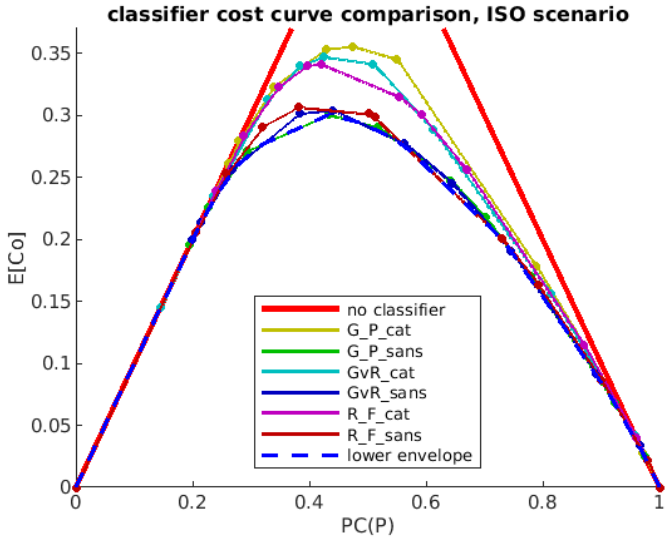


FIGURE 7.3: ISO scenario cost curves for the b2b configurations that qualified for P1B, normalized to $\max_{fn,fp}\{\varepsilon[Co]\}$.

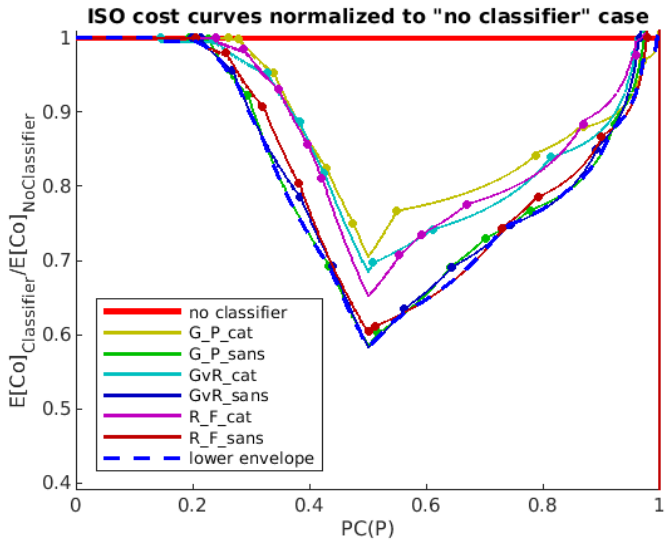


FIGURE 7.4: ISO scenario cost curves for the b2b configurations that qualified for P1B, normalized to the "no classifier" case.

**Transfer to ROC curve quantitative test (P2)**

As can be seen in the above figures, all the remaining configurations have roughly similar performance, but unlike in P1A, some are clear winners. The most striking aspect of the above figures is the very clear divide between the performance of the cat configurations and the sans ones, especially in the ISO case: every sans configuration outperforms every cat configuration for almost all $PC(\mathsf{P})$ values. It is also clear that the GvR algorithm contributes little to the comparison, as its performance mostly appears in between those of GP and RF. It is also interesting that RF performs better than GP for cat configurations, but the best configuration for both scenarios is the sans one with GP. Finally it can also be observed that for $PC(\mathsf{P}) \lesssim 0.2$ (low probability in deployment conditions of a $0\nu\beta\beta$ event compared to $1e$ events, see Section 5.2.4) all configurations display roughly the same performance as the "no classifier" case, implying that for the ISO scenario the choice of using a classifier to aid in counting statistics would not give any discernible increase in performance over not using any classifier at all.

With these points in consideration, the following configurations advanced to the ROC curve absolute performance test:

- `b2b_GP_sans_prep`;
- `ISO_GP_sans_MDS`;

## 7.3 ROC curve quantitative test (P2)

ROC curves were generated according to the procedure in Section 6.6.4 for non-categorized dimensionality-reduced data labeled with the Gaussian process classifier. The resulting ROC curve for the b2b scenario is shown in Figure 7.5, yielding $\text{AUC}_{\text{b2b}} = 0.87^{+0.10}_{-0.12}$ at 90% confidence, and the ROC curve for the ISO scenario is shown in Figure 7.6, yielding $\text{AUC}_{\text{ISO}} = 0.78^{+0.09}_{-0.11}$ at 90% confidence. Figure 7.7 shows the sample pdfs ($p_s(L_{0\nu\beta\beta})$ from **{fn6.6j}** on pg. 127) and smoothed average pdfs for the b2b scenario, and Figure 7.8 shows the sample pdfs and smoothed average pdfs for the ISO scenario. Note how in the ISO scenario a large part of the $1e$ pdf falls into the same likelihoods as the $0\nu\beta\beta$ pdf, around $0.6 < L_{0\nu\beta\beta} < 0.95$.

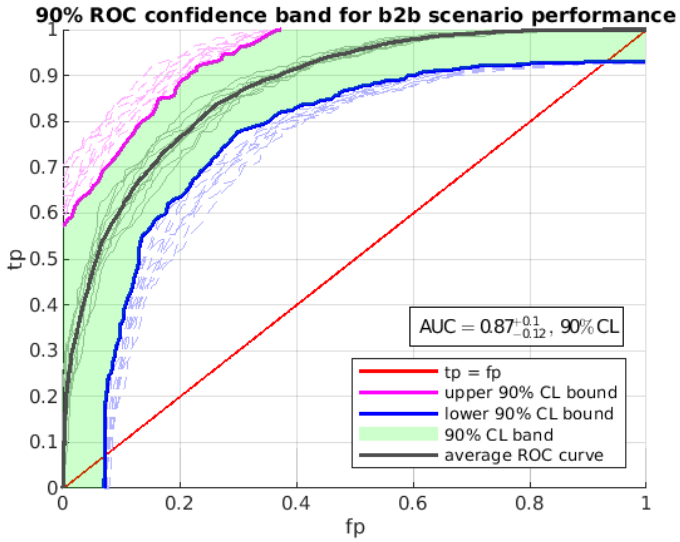


FIGURE 7.5: ROC curve for `b2b_GP_sans_prep`. The thin black curves are the sample ROC curves, and the thin dashed curves are the sample confidence bands. The bold lines are the estimates.

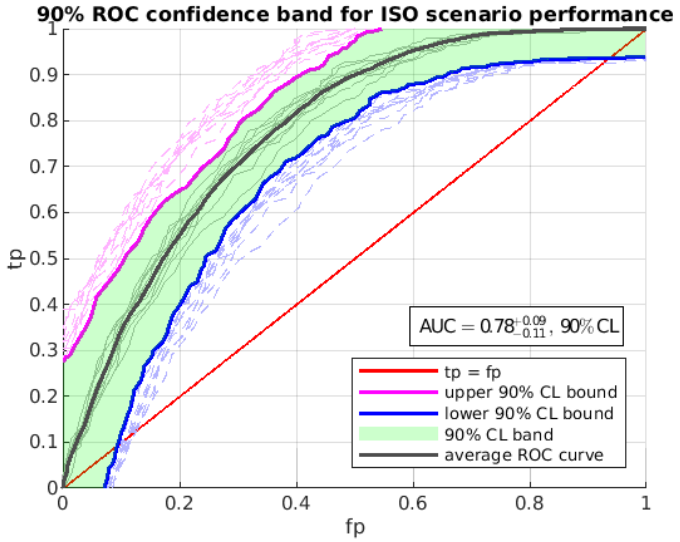


FIGURE 7.6: ROC curve for `ISO_GP_sans_MDS`. The thin black curves are the sample ROC curves, and the thin dashed curves are the sample confidence bands. The bold lines are the estimates.

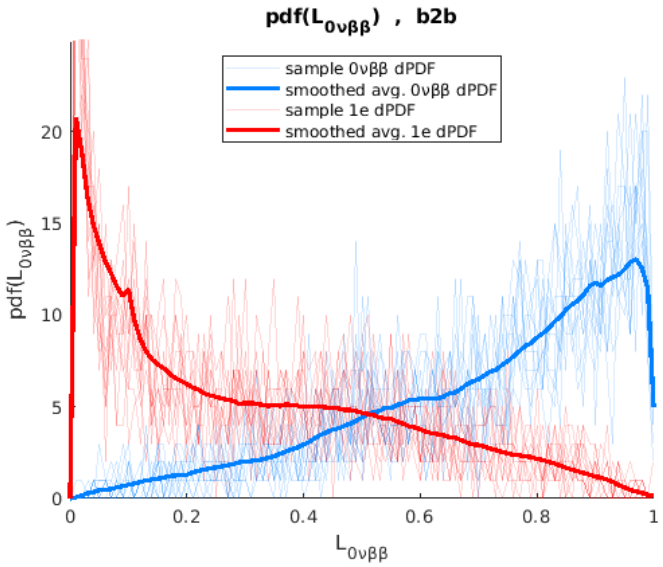


FIGURE 7.7: Sample pdfs and smoothed average pdfs for the b2b scenario.

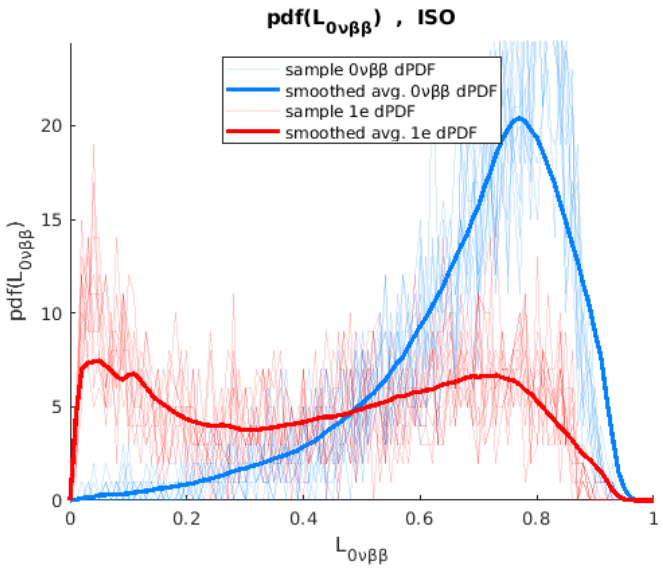


FIGURE 7.8: Sample pdfs and smoothed average pdfs for the ISO scenario.

## 7.4 Discussion

### 7.4.1 Analysis of the results

Several observations can be made from the obtained results. They are listed below, along with justifications and implications for each.

**There was a sharp drop in performance when changing from the b2b scenario to the more realistic ISO scenario.**

***Justifications***

- The average AUC fell by 0.09, from 0.87 to 0.78 (Figures 7.5 and 7.6), meaning that in the ISO scenario it is less probable that $0\nu\beta\beta$ will have a higher $L_{0\nu\beta\beta}$ than $1e$.
- In the ISO scenario the $L_{0\nu\beta\beta}$ pdf given by the classifier for the $1e$ class has two peaks, one of which overlaps with the peak for the $0\nu\beta\beta$ class (Figures 7.7 and 7.8).

***Comments***

A likely reason for this drop in performance is that the transition to isotropic initial emission of the primary electrons made the two blobs of the $0\nu\beta\beta$ events be nearly horizontally oriented more often than when the primary electrons were always emitted directly up or down.

While the $1e$ events have long tracks that allow the position of the blob to not depend very strongly on the primary electron emission direction, the primary electron tracks for $0\nu\beta\beta$ events are much shorter, meaning that the blob position will be less random than in the $1e$ case. While the initial emission direction was favorable, this led to a good distinction between the waveforms of the two classes of event, but when the emission direction became random, the characteristic details of the $0\nu\beta\beta$ waveforms will have disappeared, while mostly preserving the details of the $1e$ waveforms.

Evidence for this explanation can be seen in Table 6.5, where the category dataset sizes remain roughly the same for $1e$ events but are severely affected for $0\nu\beta\beta$. It may seem strange then, that the $0\nu\beta\beta$ classifier pdf only experiences a minor change in shape, whereas the $1e$ pdf becomes severely distorted, but this can simply indicate that the $0\nu\beta\beta$ datapoints migrated in feature space onto a well-defined territory that happens to be shared by many $1e$ datapoints. Then the classifier would be usually correct in identifying $0\nu\beta\beta$ events, but would also be forced to label $1e$ events as $0\nu\beta\beta$ ones.

**GP was the best-performing classifier algorithm out of the four tested.**

***Justifications***

- On the convergence test, GP and RF are seen to consistently outperform kNN and SVM (Table 7.3 and 7.8).
- On the cost curve test, GP is seen to consistently outperform RF (Figure 7.2 and 7.4).

***Comments***

It is interesting that even though RF tended to have better fp, GP is still the best option in deployment. A possible explanation for this is that GP has more ease with decision boundaries that bisect more than one axis, e.g. diagonal or curved decision boundaries, since RF is still constructed around decision trees.

**Categorized data underperformed relative to non-categorized data.**

***Justifications***

The expected cost of using categorized data was consistently higher than that of non-categorized data (Figure 7.2 and 7.4).

***Comments***

- In the convergence test this was not immediately apparent, although the poor trend estimator displacement sums already suggested that categorization may be a questionable approach.
- It should be noted that for the **N11**, **N20** and **Z99** categories the classifiers tended to ouperform the non-categorized data, while the other categories generally underperformed.
- It is not impossible that the categorized data underperformed merely due to poor statistics, considering that **N11**, **N20** and **Z99** tended to be fairly large classes. Before concluding that categorization is fundamentally unusable, tests with more datapoints or fewer categories should be made.

**Dimensionality reduction seemed to result in a minor improvement to the performance of the classifier when the dataset formed a simple geometry in feature space.**

***Justifications***

In the convergence test, the classifier performed slightly better following dimensionality reduction, except with non-categorized data in the b2b scenario (Tables 7.3, 7.5 and 7.8).

***Comments***

- The displacement sums of $\widehat{\mathsf{fp}}$ , $\widehat{\mathrm{tr}_{\mathsf{fp}}}$, $\widehat{\mathsf{tp}}$ and $\widehat{\mathrm{tr}_{\mathsf{tp}}}$ for the *Optimization* branching point were generally the closest to zero out of all the branching points, suggesting that the use of data optimization had only a very minor effect on performance.
- The fact that the ISO scenario showed a consistent preference for dimensionality reduction while in the b2b scenario that preference only appeared in categorized data may suggest that the chosen dimensionality reduction algorithm was removing some decisive details in the feature space geometry of the dataset. The reasoning is as follows. Given that the ISO scenario performed worse than b2b, it is intuitive to suppose that its feature space geometry would be simpler than that of the b2b scenario: as in, there would be fewer decisive details in the ISO dataset geometry than in the b2b dataset geometry. Obviously the geometry of categorized datasets is simpler, as that was the goal of categorization. By reducing the dimensionality of the dataset, the feature space geometry is projected onto just a subset of the original features, and so there is the risk that certain small details are lost during optimization.
- Nonlinear dimensionality reduction methods are capable of "exploding" geometry onto a small number of dimensions without having it overlap. If the problem was that decisive details were projected onto the same region as other data, then it could be useful to try using nonlinear dimensionality reduction methods in future work.

**For deployment with the ISO scenario, it makes no difference whether to use binary classification or not.**

***Justifications***

In the ISO scenario cost curve, the expected cost for low $PC(\mathsf{P})$ is the same when using a classifier as when labeling all test data as $1e$ (Figure 7.4).

***Comments***

- There is little purpose in advancing to the DEC0 (see Table 6.1) scenario before either making major modifications to the classifier design, or making minor modifications to the design of the detector.
- If the balancing between $1e$ and $0\nu\beta\beta$ in deployment were better, then the classifier would significantly improve the counting statistics of the detector.

### 7.4.2 Discussion about methodology

At many points in this work, choices were made that guided it down a specific path, and so it is important to discuss why certain decisions were made, considering the possible outcomes of the alternatives. Several stages in this work required the development of new tools or methods, so this Subsection discusses the taken methodological choices. It is presented as a listed discussion, separated into the categories of simulation, data extraction, classification algorithm, and performance assessment.

*SIMULATION*

**Does the simulation accurately represent the topologies and the LZ detector?**

It is accurate within the requirements of each studied scenario. There are some aspects that were either omitted or simplified, and some slight inaccuracies, but care was taken to ensure that they would not affect the classifier performance, the idea being that as the performance would be tested with increasingly accurate simulations, the classifier performance would worsen until the classifier was ineffective. For a more in-depth description of the inaccuracies, see Section 6.3.

**Would a larger dataset significantly improve the performance?**

- ***For uncategorized data:***

  No. The convergence test showed that the performance in uncategorized data was already varying very little for the larger sample sizes. Supposing that the performance would continue behaving the same way, then a sample size several orders of magnitude larger would be necessary before the performance changed significantly.

- ***For categorized data:***

  Possibly: the sample size for some of the categories was very small and they failed to converge.

**Would performance in deployment improve by using the estimated real-life balancing?**

Although it is true that classifiers generally perform best in deployment when the training data and deployment data have the same balancing [210], in this case, with $d = 145.6$ mm and assuming $T_{1/2}^{0\nu\beta\beta} \sim 10^{26}$ years, the deployment balancing would yield $\sim 10^3$ $1e$ events for each $0\nu\beta\beta$ event, so making a dataset with enough $0\nu\beta\beta$ events with the same time constraints and using the same methods would be unrealistic.

**Could a larger dataset be generated more quickly using PMT light response functions?**

The light response function is dependent on the detector geometry and materials, so before a dataset can be generated using light response functions, it must be generated via photon tracing first. Additionally, the obtained light response function must also be validated before being applied for testing of the classifiers (Section 6.3.3).

**Would the use of multichannel waveforms instead of channel sums improve performance?**

It was determined in a crude preliminary test that it was more common in the case of $1e$ events than in the case of $0\nu\beta\beta$ events for PMT channels an equal distance away from the event origin to display a small difference in pulse area (different at most by a factor of $\sim 0.85$). It should be noted that, like in the case of the event simulation described in Section 6.3, this preliminary test was performed using the diffusion coefficients calculated according to the diffusion data listed in the LZ TDR [109], and it was later determined that the listed cathode transverse diffusion value was much smaller than measurement and prediction data (see Section 4.1, pg. 55). With a more accurate transverse diffusion value, together with the funnelling effect of the gate grid, it is expected that the PMT channel distinction will wash out significantly, though possibly not completely. In future work, the option of using multichannel waveforms should be investigated further.

**Would it be viable to use the classifier closer to the surface of the LXe?**

Although the smaller diffusion would increase the performance of the classifier, above the fiducial volume the background increases exponentially, so deployment conditions would become even more unfavorable. Although the performed tests do not allow to answer this question confidently, most likely the counting statistics for $d < 145.6\,\mathrm{mm}$ would not be affected by the inclusion of the classifier either.

### *DATA EXTRACTION*

**Is the used feature set the best possible?**

The current parameterization is quite descriptive, the mapping onto feature space is lossless, and in case some features are redundant, feature selection is capable of removing them. So if the used feature set is not the best possible, then it is still very good. However it could be that some additional, nontrivial features would be useful, and once multiple channels are used instead of the channel sums, new parameters will have to be extracted. This consideration was recognized during the construction of the data extraction framework, and so it was designed to make the addition and removal of parameters and features quite easy.

### *CLASSIFICATION ALGORITHM*

**The algorithms were used with their default parameters. Would a different parameter set improve performance?**

It is very likely that the algorithm parameters are not ideal: notably, SVM and RF classifiers typically require extensive calibration procedures before achieving optimal performance. However it is also unlikely that a different choice of parameters would improve the performance dramatically enough to make the classifier affect the deployment counting statistics in the ISO scenario. Still, the parameters will have to be changed in future work, and the same performance assessment method from Section 6.6 can be used to select the best parameters. Adjusting parameters for the Gaussian Process classification may prove difficult due to it being a slow and resource-intensive algorithm, however it may be possible to circumvent this problem by using sparse GP approximations [271], which retain the favorable characteristics of GP, but with faster computation time. These should be investigated in future work.

**Is the selected algorithm the best possible?**

Referring to the classifier comparison in Figure 5.2, the selected classifiers were the ones that appeared to show the most promising behavior when constructing the decision boundary. However it is possible that some other algorithm could yield a better performance than the ones tested, and so in future work the usage of alternative algorithms should be investigated.

### *PERFORMANCE ASSESSMENT*

**Is the convergence test in its current design a legitimate way of comparing classifiers?**

Although it is not very reliable for catching subtle performance differences between different configurations, it is a fast method of highlighting the obvious differences, which is for what the test was used. It allowed to eliminate all kNN , SVM and SFFS configurations. The decision to also eliminate most of the prep datasets is reasonably well justified for ISO data, given that it consistently underperformed compared to both MDS and SFFS (Tables 7.7 and 7.8), but for cat configurations in b2b data it may seem like a more questionable choice, since the slightly better performance of MDS data compared to prep data in that case could have just been a statistical fluke.

**Are the ROC confidence bands correctly estimated?**

- ***99% CL sample confidence bands:***

  May be slightly wider than necessary [267], but should otherwise be accurate. The construction was originally designed for Gaussian likelihood pdfs [268], and the obtained likelihood pdfs are not Gaussian (Figures 7.7 and 7.8), but the construction is based on the Kolmogorov-Smirnov statistic, which does not depend on the distribution in question, so the shape of the likelihood pdfs should not significantly affect the effectiveness of the method.

- ***90% CL final confidence band:***

  Obviously too wide **{fn7.4a}**. The 90% CL is better interpreted as the minimal confidence level of the band. In future work, further research will be done on how to combine the sample confidence bands.

### 7.4.3 Implementation of the binary classifier in LZ

From the observations in Section 7.4.1 it can be concluded that if the detector construction remains as it is right now, then there is no benefit to implementing the classifier in its current form into the LZ data analysis pipeline: due to the high background, the counting statistics would remain roughly the same, with or without the classifier. In order to make the classifier improve the LZ counting statistics, either the performance has to become better, or the deployment data has to become more balanced.

Improving the balancing of the deployment data corresponds to increasing the mass of $0\nu\beta\beta$-active material, $M_{\beta\beta}$, or to decreasing the background event rate. Background reduction is not practicable: being LZ's primary goal to search for WIMPs, with background reduction as crucial there as it is for $0\nu\beta\beta$ discovery, it can be safely assumed that the background is already as low as it can realistically get. As for $M_{\beta\beta}$, although it could be increased significantly by using enriched Xe instead of natural Xe **{fn7.4b}**, acquiring the necessary Xe would be prohibitively expensive **{fn7.4c}**, as well as moving the schedule of the entire experiment back by at least a few years **{fn7.4d}**. With these considerations, the best solution seems to be to improve the performance of the classifier.

To improve the classifier performance, the overlap of the two classes in feature space must be smaller. As can be seen in Figure 5.1, the difference in shape between the $0\nu\beta\beta$ waveforms and the $1e$ waveforms is subtle, and often they look the same. Electron diffusion frequently made the resulting waveforms for both classes appear as just a single Gaussian peak with little distinguishing detail. Comparing the PMT channels in the array shows some potential, albeit limited, and is expected to worsen considerably in more realistic scenarios, especially when accounting for the gate grid. While these two crucial points — diffusion and $xy$ discrimination — are not addressed, any improvements to the classifier design will have a limited effect on its performance.

Regarding diffusion, it can be seen from the data in Figure 3.14 that the longitudinal diffusion coefficient $D_L$ decreases for increasing values of drift field $E_{\text{drift}}$. If the $E_{\text{drift}}$ value in LZ were increased above the design goal field of 650 V/cm, then it would be possible to obtain a $D_L$ below the 20 cm$^2$/s used in this work. This could be achieved by feeding the cathode grid (Figure 4.1) a HV even lower than the design goal $-100$ kV, however not only would that require a major redesign of the cathode HV delivery system, as the power supply is rated at $-120$ kV, but it would also cause spurious electroluminescence, which could blind the detector to WIMPs (*Section 3.3 of* [109]), making the LZ detector incapable of completing its primary science goal. This consideration renders this approach impracticable.

**{fn7.4a}** — This can be shown by supposing that two 99% CL independent sample confidence bands come out identical: according to the method used, the probability that the real ROC curve is inside some sample band $A$ is 0.99, but the probability that the real ROC curve is inside band $A$ and also inside band $A$ is $0.99 \times 0.99 = 0.9801$, clearly an absurd result.

**{fn7.4b}** — An enrichment of 90% in $^{136}$Xe, like for nEXO [123], would result in a roughly 10-fold increase in the $0\nu\beta\beta$ rate. Supposing $T_{1/2} = 1.06 \times 10^{26}$ years, this would place $PC(\mathsf{P})$ closer to $\sim 0.5$ in the fiducial volume (*see Section 4.2*).

**{fn7.4c}** — In 2019, the projected Xe market price was $\sim 3.70$ \$/g [121][122]. Supposing that an additional 63 t of Xe would need to be procured to obtain a 90% enrichment, then just the procurement alone would cost $\sim 240$ M\$.

**{fn7.4d}** — A procurement of this large a quantity of Xe would raise the same concerns of disrupting the market as were raised in the nEXO procurement, namely it forces the procurement to be done over a long period of time. Additionally, the world Xe enrichment capacity is in the order of only a few tonnes per year [123].

On the other hand, it makes attractive the use of LZ's reverse-field region *(RFR)* for $0\nu\beta\beta$ detection. It is a 137.5 mm thick gap between the cathode grid and the bottom PMT array (see Figure 4.4), filled with LXe, with an upward-directed drift field of 3-6 kV/cm **{fn7.4e}** (Figure 7.9, also Section 3.2.1 from [109]). The RFR has almost 60 kg of $^{136}$Xe and a background event rate slightly lower than that of the top of the detector. An option could be to install readout grids slightly above the bottom shield grid, thus forming a miniature liquid-phase TPC of similar design to EXO-200 [272] below the regular two-phase TPC. It should be practicable provided that the grids are sufficiently transparent to WIMP recoil photons. It should be noted that EXO-200 achieved a resolution at $Q_{\beta\beta}$ of $\sim$ 2.90% (FWHM) [43], and it would be expected that the energy resolution in the RFR would be similar or better. Considering that the value is already higher than the $\sim$ 2.36% (FWHM) resolution of LZ, the background rate can be expected to increase considerably, but the improvement in discrimination thanks to a much lower $D_L$ could be very significant, and so it may an interesting option to investigate.

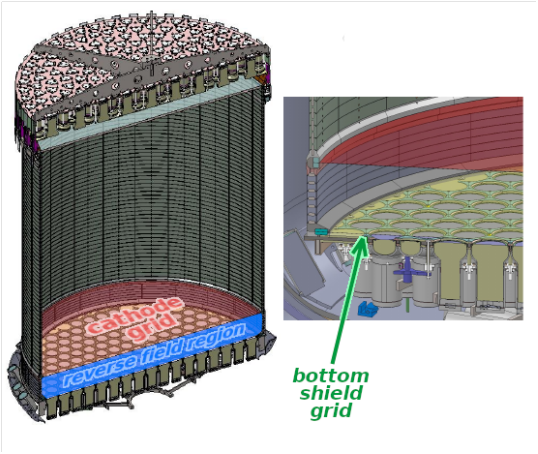


FIGURE 7.9: LZ reverse field region, together with its corresponding grids. Source of images: [109]

Regarding $xy$ discrimination, the design of the current PMT array and the presence of the gate grid hinder its effectiveness. Although the Hamamatsu low-radioactivity PMTs are essential to amplify the weak signals from WIMP recoils, the identification of the very strong $0\nu\beta\beta$ signals is instead worsened by the large PMT center-to-center spacing of 93 mm *(Section 6.3.2)*. By using an array of smaller and more compact photosensors, the $xy$ discrimination could be significantly improved. The MEG experiment is one that has had some success in improving position resolution by transitioning from using an array of 2-inch PMTs for detecting LXe photons to 6 mm $\times$ 6 mm Hamamatsu multi-pixel photon counters when moving to its 2nd generation design, although the obtained dark count rate of $10 - 100$ Hz could cause complications for WIMP searches [273]. But the problem remains of funnelling due to the gate grid. It could potentially be mitigated by using a different grid shape, but careful study of this matter is necessary.

It should be noted that the current classifier design is not ideal: there are many aspects in which it can still be improved. The design of its framework and underlying software was made to be modular, so that any step could be easily exchanged or improved in future work. Obvious avenues for improvement include the usage of different classification algorithms *(for example*

**{fn7.4e}** — This is competitive with the EXO-200 design maximum drift field of 3.7 kV/cm [272].

*boosted classifiers)* or more refined data optimization methods. One aspect that was not very expanded upon was filtering of the raw signal: only smoothing was applied. A promising approach to explore in future work is to try removing some of the diffusion blur via deconvolution, possibly even allowing for some limited topology reconstruction in the best case. As an example, by using track reconstruction based on the Richardson-Levy deconvolution algorithm, the NEXT collaboration was capable of improving their background rejection in the NEXT-White TPC roughly threefold compared to previous benchmarks [274].

### 7.4.4 Other applications of the developed classifier

Most straightforward is the application of the classifier to other TPC $0\nu\beta\beta$ experiments. The XENONnT experiment uses a very similar design to LZ [143], so the same classification framework can be translated directly, only requiring a new training dataset to account for the different PMT array and detector dimensions. The nEXO experiment does not have a gas gap and uses charge collecting tiles instead of a PMT array [123], but otherwise the classifier can also largely be reused for it as well. There is even an additional advantage here: since there are no electroluminescence photons to simulate in nEXO, simulating each event should be almost $\sim 10^3$ times faster, so making a large dataset would be much more straightforward. The sensitive elements of the charge collecting tiles are square strips with a 3 mm center-to-center distance, so $xy$ discrimination in nEXO should be very strong. NEXT-100 and PandaX-III both have a similar design, with a GXe sensitive volume coupled with a low-resolution detector array: SiPMs on NEXT-100 [132], and micromegas on PandaX-III [275]. Building a dataset for either would pose the same challenges experienced in LZ, as both produce a large number of particles to track. It should also be noted that nEXO, NEXT-100 and PandaX-III all already have projects in progress for background discrimination using machine learning.

The developed binary classification framework allows for the fast preparation of time-series training / test data and comparison of several different classification algorithms and data processing methods. It allows for the simultaneous testing of a broad range of approaches to a single classification problem, comparing advantages and assessing overall effectiveness. Additionally, it is capable of estimating classifier performance for balancings of the input data other than the ones used. As such, it can serve as a versatile tool for viability studies, provided that the input data is waveform-like.

# Chapter 8

# Conclusion

The original goal of this work was to test how well the LZ TPC can discriminate background events from $^{136}$Xe $0\nu\beta\beta$ events in the LXe sensitive volume using topology reconstruction: most background events are rejected due to some combination of incorrect energy, wrong type of recoil or coincident energy depositions (*e.g.* with the veto detectors), leaving almost only $\sim 2.459$ MeV single primary electron events due to photoelectric effect interactions by $^{232}$Th and $^{238}$U $\gamma$ particles as the remaining background, which has a different topological signature than that of $0\nu\beta\beta$. During this preliminary research it was found that the short $0\nu\beta\beta$ blob distance in LXe, coupled with the drift diffusion, grid distortion and the poor imaging resolution of the PMT array, would make LZ unable to reconstruct event topologies, requiring an alternative approach for rejecting $1e$ events. The arrived-at solution was to parameterize the accepted S2 waveforms and then use a binary classifier to attempt to reject events that are labeled as background, with an increased focus on the time series component of the waveforms (along $z$).

Constructing the classifier meant preparing a training dataset and selecting the most appropriate binary classification algorithm. To prepare the dataset, waveforms from a large number of $0\nu\beta\beta$ and $1e$ events were parameterized and then mapped onto feature space. To do this, the waveforms themselves had to be produced first via Monte Carlo simulation. To diagnose the $0\nu\beta\beta$ discrimination potential in LZ, best-case scenarios were simulated first, with iterative increases in statistical and detector accuracy added later, the starting point being events with vertical primary electrons. During preliminary testing it was also found that the signals from each PMT channel for the same event were usually roughly the same, so to simplify development the parameterization was performed on the channel sum, leaving multichannel parameterization for future work. In total two scenarios were tested before the classifier performance became too poor to continue.

The simulation first generated a deposition of drift electrons in LXe due to the $0\nu\beta\beta$ and $1e$ primaries using Geant4, and then electron diffusion and drift, as well as electroluminescence photon tracks and the resulting binned time series of photons detected by the PMTs were all produced in ANTS2. The large number of photon tracks, $\gtrsim 8 \times 10^7$ per event, meant that the simulation was slow to compute: $\sim 10$ min per event, using $\sim 3$ GB of RAM. This meant that a large enough dataset would take a very long time to produce on a single machine. To address the problem, a Docker image of the simulation suite was made and distributed among a number of computers, as well as the Infraestrutura Nacional de Computação Distribuída. In total each tested scenario had a dataset of $\sim 10^4$ waveforms, with roughly equal balancing of the two classes, always using a single depth: $d = 145.6$ mm.

The morphology of the obtained waveforms was unpredictable and the differences between the two classes are subtle. In order to reduce the overlap of the two classes in parameter space, a large collection of relevant morphological parameters was extracted from each waveform, expanding upon the parameterization performed by the LZ Analysis Package, leading to a large

number of extracted features. The amount of features risked reducing classifier performance due to the curse of dimensionality. To address this, dimensionality reduction was employed. Datasets without dimensionality reduction were also preserved, so as to compare the performance of the two options. It was supposed that dimensionality reduction would perform better on small subsets of the data, so the datasets were subdivided according to the morphology of the waveforms, again preserving the non-subdivided versions as well for performance comparison.

Binary classification algorithms were selected during preliminary research according to a classifier comparison done by the `scikit-learn` development team. The four chosen algorithms were: k-nearest neighbors; support vector machines with RBF kernel; Gaussian process; and random forests. They were used as defined in the performance comparison, with no changes to their parameters. A large set of configurations had to be tested, so a fast two-step comparison procedure was employed: first all configurations were compared with a performance convergence test, eliminating the underperforming ones, and then the remaining configurations were compared via their resulting cost curves, leaving just the best configuration for each scenario. Its performance was then assessed in absolute terms using an ROC curve.

The comparison procedure for both tested scenarios singled out the Gaussian process classifier as the best-performing one. It was also determined that the non-subdivided versions of the data performed consistently better than the subdivided ones, and dimensionality reduction was shown to produce a minor improvement in performance. The ROC curve of the vertical primary electron scenario yielded an AUC of $0.87^{+0.10}_{-0.12}$ at 90% confidence, with a minor improvement in background rejection predicted for the expected signal-to-background ratio at the tested depth in LZ. However, the second scenario tested, with equal-energy primaries emitted isotropically, yielded an AUC of just $0.78^{+0.09}_{-0.11}$ at 90% confidence, and clearly no predicted improvement in LZ background rejection. In effect, it can be concluded that the current implementation is not viable for application into the LZ background cuts.

The main goal of future work should be to make the two classes of event appear more distinct in feature space. For this purpose, the most promising approach is the addition of a waveform deconvolution step prior to parameterization. Next, improvements to the parameterization, feature extraction, data optimization and performance assessment framework should follow, namely with the introduction of multichannel ($xy$) parameterization and nonlinear dimensionality reduction, and with the automatization of certain parts of the performance convergence test. Finally, the classification algorithm parameters should be optimized for the task, and attempting the use of other classification algorithms may prove promising. Of special interest is the use of sparse Gaussian process approximations instead of the Gaussian process classifier, as it is expected to significantly reduce computation time without a performance penalty. If these improvements allow the classifier to be implemented into the LZ background cuts, then simulation representativity can be increased as planned: first by making the event statistics more accurate; and then by increasing the accuracy of the detector simulation.

# Bibliography


[1] Iason Baldes. "Early Universe Cosmology and the Matter-Antimatter Asymmetry". PhD thesis. Melbourne U., 2015.

[2] Palash B. Pal. "Dirac, Majorana, and Weyl fermions". In: *American Journal of Physics* 79.5 (2011), pp. 485–498. DOI: 10.1119/1.3549729. eprint: https://doi.org/10.1119/1.3549729. URL: https://doi.org/10.1119/1.3549729.

[3] J. J. Gomez-Cadenas et al. "The search for neutrinoless double beta decay". In: (2012). arXiv: 1109.5515v2 [hep-ex].

[4] Andrey Morozov et al. "ANTS2 package: Simulation and experimental data processing for Anger camera type detectors". In: *Journal of Instrumentation* 11 (Apr. 2016), P04022–P04022. DOI: 10.1088/1748-0221/11/04/P04022.

[5] M. Szydagis et al. *Noble Element Simulation Technique v2.0*. Version v2.0.0. July 2018. DOI: 10.5281/zenodo.1314669. URL: https://doi.org/10.5281/zenodo.1314669.

[6] *INCD homepage*. https://www.incd.pt/. Accessed: 8 June 2020. 2020.

[7] Simone Biondini. *Effective field theories for heavy Majorana neutrinos in a thermal bath*. 2016. arXiv: 1612.07933 [hep-ph].

[8] Utpal Sarkar. *Particle and Astroparticle Physics*. 2008.

[9] Gordon Gilmore. "Practical Gamma-Ray Spectrometry: Second Edition". In: *Practical Gamma-ray Spectrometry, 2nd Edition, by Gordon R. Gilmore. ISBN 978-0-470-86196-7. Published by Wiley-VCH Verlag, Wenheim, Germany, 2008.* (Apr. 2008). DOI: 10.1002/9780470861981.

[10] Gary Steigman. "Observational Tests of Antimatter Cosmologies". In: *Annual Review of Astronomy and Astrophysics* 14.1 (1976), pp. 339–372. DOI: 10.1146/annurev.aa.14.090176.002011. eprint: https://doi.org/10.1146/annurev.aa.14.090176.002011. URL: https://doi.org/10.1146/annurev.aa.14.090176.002011.

[11] J. Alcaraz et al. "Search for antihelium in cosmic rays". In: *Physics Letters B* 461.4 (1999), pp. 387–396. ISSN: 0370-2693. DOI: 10.1016/S0370-2693(99)00874-6. URL: https://www.sciencedirect.com/science/article/pii/S0370269399008746.

[12] Gary Steigman. "When Clusters Collide: Constraints On Antimatter On The Largest Scales". In: *Journal of Cosmology and Astroparticle Physics - JCAP* 10 (Aug. 2008). DOI: 10.1088/1475-7516/2008/10/001.

[13] Charlotte Edge et al. "An X-Ray Flux-Limited Sample of Clusters of Galaxies - Evidence for Evolution of the Luminosity Function". In: *Monthly Notices of the Royal Astronomical Society* 245 (June 1990), p. 559.

[14] O. Reimer et al. "EGRET Upper Limits on the High-Energy Gamma-Ray Emission of Galaxy Clusters". In: *The Astrophysical Journal* 588 (Feb. 2003). DOI: 10.1086/374046.

[15] S.L. Glashow A.G. Cohen A. De Rujula. "A Matter-Antimatter Universe?" In: (1997). arXiv: 9707087 [astro-ph].

[16] Itzhak Bars et al. *Extra Dimensions in Space and Time*. Jan. 2010. DOI: 10.1007/978-0-387-77638-5.

[17] Yvonne Y. Y. Wong. "Neutrino cosmology". Invisibles'19 School, LSC Canfranc, June 3 – 7, 2019. 2019. URL: https://indico.ific.uv.es/event/3550/sessions/1211/attachments/7122/8327/lecture1_YvonneW.pdf.

[18] J. M. Uson and D. T. Wilkinson. "Search for Small-Scale Anisotropy in the Cosmic Microwave Background". In: Phys. Rev. Lett. 49.19 (Nov. 1982), pp. 1463–1465. DOI: 10.1103/PhysRevLett.49.1463.

[19] Edward W. Kolb and Michael S. Turner. *The early universe*. Vol. 69. 1990.

[20] D. N. Spergel et al. "Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology". In: ApJS 170.2 (June 2007), pp. 377–408. DOI: 10.1086/513700. arXiv: astro-ph/0603449 [astro-ph].

[21] W. Weise. "Antinucleon - nucleon annihilation". In: *Nuclear Physics A* 558 (1993), pp. 219–233. ISSN: 0375-9474. DOI: 10.1016/0375-9474(93)90396-F. URL: https://www.sciencedirect.com/science/article/pii/037594749390396F.

[22] A. D. Sakharov. "Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe". In: *Soviet Journal of Experimental and Theoretical Physics Letters* 5 (Jan. 1967), p. 24.

[23] James M. Cline. "Baryogenesis". In: (2006). arXiv: 0609145v1 [hep-ph].

[24] Borut Bajc et al. "Threshold corrections to dimension-six proton decay operators in non-minimal SUSY SU(5) GUTs". In: *Nuclear Physics B* 910 (2016), pp. 1–22. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2016.06.017. URL: https://www.sciencedirect.com/science/article/pii/S0550321316301596.

[25] Mark Trodden. "Electroweak baryogenesis". In: *Rev. Mod. Phys.* 71 (5 1999), pp. 1463–1500. DOI: 10.1103/RevModPhys.71.1463. URL: https://link.aps.org/doi/10.1103/RevModPhys.71.1463.

[26] J. H. Christenson et al. "Evidence for the $2\pi$ Decay of the $K_2^0$ Meson". In: *Phys. Rev. Lett.* 13 (4 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[27] K. Jansen. "Status of the finite temperature electroweak phase transition on the lattice". In: *Nuclear Physics B - Proceedings Supplements* 47.1 (1996), pp. 196–211. ISSN: 0920-5632. DOI: 10.1016/0920-5632(96)00045-X. URL: https://www.sciencedirect.com/science/article/pii/092056329600045X.

[28] A.M. Sirunyan et al. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Physics Letters B* 805 (2020), p. 135425. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135425. URL: https://www.sciencedirect.com/science/article/pii/S037026932030229X.

[29] Francis Bernardeau, Christophe Grojean, and Jean Dalibard, eds. Vol. 86. Les Houches. Elsevier, 2007, pp. 1–558. DOI: 10.1016/S0924-8099(13)60038-4. URL: https://www.sciencedirect.com/science/article/pii/S0924809913600384.

[30] W. Bonivento et al. "Proposal to Search for Heavy Neutral Leptons at the SPS". In: (2013). arXiv: 1310.1762 [hep-ex].

[31] P Lipari. "Introduction to neutrino physics". In: (2003). DOI: 10.5170/CERN-2003-003.115. URL: http://cds.cern.ch/record/677618.

[32] Alessandro Strumia. "Baryogenesis via leptogenesis". In: *arXiv e-prints*, hep-ph/0608347 (Aug. 2006), hep–ph/0608347. arXiv: hep-ph/0608347 [hep-ph].

[33] O'Sullivan Kevin. "The Search for $0\nu\beta\beta$ in $^{136}$Xe with EXO-200". PhD thesis. Department of Physics, Stanford University, 2012, pp. I–104.

[34] A. S. Barabash. "Precise half-life values for two-neutrino double-$\beta$ decay". In: *Phys. Rev. C* 81 (3 2010), p. 035501. DOI: 10.1103/PhysRevC.81.035501. URL: https://link.aps.org/doi/10.1103/PhysRevC.81.035501.

[35] J. B. Albert et al. "Search for Majoron-emitting modes of double-beta decay of $^{136}$Xe with EXO-200". In: *Phys. Rev. D* 90 (9 2014), p. 092004. DOI: 10.1103/PhysRevD.90.092004. URL: https://link.aps.org/doi/10.1103/PhysRevD.90.092004.

[36] VLADIMIR I TRETYAK and YURI G ZDESENKO. "TABLES OF DOUBLE BETA DECAY DATA—AN UPDATE". In: *Atomic Data and Nuclear Data Tables* 80.1 (2002), pp. 83–116. ISSN: 0092-640X. DOI: 10.1006/adnd.2001.0873. URL: https://www.sciencedirect.com/science/article/pii/S0092640X01908733.

[37] V. I. Tretyak. "DECAY0 event generator for initial kinematics of particles in $\alpha$, $\beta$ and $2\beta$ decays". Talk given at Laboratori Nazionali del Gran Sasso, 17 March 2015. 2015. URL: https://agenda.infn.it/event/9358/attachments/57515/67844/Tretyak_2015.pdf.

[38] O. A. Ponkratenko, V. I. Tretyak, and Yu. G. Zdesenko. "Event generator DECAY4 for simulating double-beta processes and decays of radioactive nuclei". In: *Physics of Atomic Nuclei* 63.7 (2000), pp. 1282–1287. ISSN: 1562-692X. DOI: 10.1134/1.855784. URL: https://doi.org/10.1134/1.855784.

[39] D. S. Akerib et al. "Projected sensitivity of the LUX-ZEPLIN experiment to the $0\nu\beta\beta$ decay of $^{136}$Xe". In: *Phys. Rev. C* 102 (1 2020), p. 014602. DOI: 10.1103/PhysRevC.102.014602. URL: https://link.aps.org/doi/10.1103/PhysRevC.102.014602.

[40] J. Kotila and F. Iachello. "Phase-space factors for double-$\beta$ decay". In: *Phys. Rev. C* 85 (3 2012), p. 034316. DOI: 10.1103/PhysRevC.85.034316. URL: https://link.aps.org/doi/10.1103/PhysRevC.85.034316.

[41] D.S. Akerib et al. "Simulations of events for the LUX-ZEPLIN (LZ) dark matter experiment". In: *Astroparticle Physics* 125 (2021), p. 102480. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2020.102480. URL: https://www.sciencedirect.com/science/article/pii/S0927650520300529.

[42] NEXT Collaboration et al. "The NEXT-100 experiment for neutrinoless double beta decay searches (Conceptual Design Report)". In: (2011). arXiv: 1106.3630 [physics.ins-det].

[43] Michelle J. Dolinski, Alan W.P. Poon, and Werner Rodejohann. "Neutrinoless Double-Beta Decay: Status and Prospects". In: *Annual Review of Nuclear and Particle Science* 69.1 (2019), pp. 219–251. DOI: 10.1146/annurev-nucl-101918-023407. eprint: https://doi.org/10.1146/annurev-nucl-101918-023407. URL: https://doi.org/10.1146/annurev-nucl-101918-023407.

[44] J. Dawson et al. "Search for double-$\beta$ decays of tin isotopes with enhanced sensitivity". In: *Phys. Rev. C* 78 (3 2008), p. 035503. DOI: 10.1103/PhysRevC.78.035503. URL: https://link.aps.org/doi/10.1103/PhysRevC.78.035503.

[45] R. Arnold et al. "First Results of the Search for Neutrinoless Double-Beta Decay with the NEMO 3 Detector". In: *Phys. Rev. Lett.* 95 (18 2005), p. 182302. DOI: 10.1103/PhysRevLett.95.182302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.182302.

[46] P. F. de Salas et al. "2020 global reassessment of the neutrino oscillation picture". In: *Journal of High Energy Physics* 2021.2 (2021), p. 71. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2021)071. URL: https://doi.org/10.1007/JHEP02(2021)071.

[47] Andrej Babič and Fedor Šimkovic. "Majorana phases, CP violation, sterile neutrinos and neutrinoless double-beta decay". In: *AIP Conference Proceedings* 1572.1 (2013), pp. 7–10. DOI: 10.1063/1.4856537. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4856537. URL: https://aip.scitation.org/doi/abs/10.1063/1.4856537.

[48] X. Qian and P. Vogel. "Neutrino mass hierarchy". In: *Progress in Particle and Nuclear Physics* 83 (2015), pp. 1–30. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2015.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0146641015000307.

[49] Werner Maneschg. "Review of neutrinoless double beta decay experiments: Present status and near future". In: *Nuclear and Particle Physics Proceedings* 260 (2015). The 13th International Workshop on Tau Lepton Physics, pp. 188–193. ISSN: 2405-6014. DOI: 10.1016/j.nuclphysbps.2015.02.039. URL: https://www.sciencedirect.com/science/article/pii/S240560141500098X.

[50] S. Andringa et al. "Current Status and Future Prospects of the SNO+ Experiment". In: *Advances in High Energy Physics* 2016 (2016), p. 6194250. ISSN: 1687-7357. DOI: 10.1155/2016/6194250. URL: https://doi.org/10.1155/2016/6194250.

[51] Alexander Barabash. "Precise Half-Life Values for Two-Neutrino Double-$\beta$ Decay: 2020 Review". In: *Universe* 6.10 (2020). ISSN: 2218-1997. DOI: 10.3390/universe6100159. URL: https://www.mdpi.com/2218-1997/6/10/159.

[52] D. Fink et al. "Q-Value and Half-Lives for the Double-Beta-Decay Nuclide 110Pd". In: (2011). arXiv: 1112.5786 [nucl-ex].

[53] B Lehnert et al. "Double beta decays into excited states in 110Pd and 102Pd". In: *Journal of Physics G: Nuclear and Particle Physics* 43.11 (2016), p. 115201. ISSN: 1361-6471. DOI: 10.1088/0954-3899/43/11/115201. URL: http://dx.doi.org/10.1088/0954-3899/43/11/115201.

[54] D. Fink et al. "$Q$ Value and Half-Lives for the Double-$\beta$-Decay Nuclide $^{110}$Pd". In: *Phys. Rev. Lett.* 108 (6 2012), p. 062502. DOI: 10.1103/PhysRevLett.108.062502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.062502.

[55] O. Chkvorets et al. "A tin-loaded liquid scintillator approach for the 2 neutrino double-beta decay measurement of Sn-124". In: (2017). arXiv: 1707.08001 [physics.ins-det].

[56] Jouni Suhonen. "On the double-beta decays of 70Zn, 86Kr, 94Zr, 104Ru, 110Pd and 124Sn". In: *Nuclear Physics A* 864.1 (2011), pp. 63–90. ISSN: 0375-9474. DOI: 10.1016/j.nuclphysa.2011.06.021. URL: https://www.sciencedirect.com/science/article/pii/S0375947411004453.

[57] M.J. Hwang et al. "A search for $0\nu\beta\beta 0\nu\beta\beta$ decay of 124Sn124Sn with tin-loaded liquid scintillator". In: *Astroparticle Physics* 31 (July 2009), pp. 412–416. DOI: 10.1016/j.astropartphys.2009.05.001.

[58] Juan Gómez-Cadenas et al. "Sense and sensitivity of double beta decay experiments". In: *Journal of Cosmology and Astroparticle Physics* 2011 (June 2011), p. 007. DOI: 10.1088/1475-7516/2011/06/007.

[59] Denys Poda and Andrea Giuliani. "Low background techniques in bolometers for double-beta decay search". In: *International Journal of Modern Physics A* 32.30 (2017), p. 1743012. DOI: 10.1142/S0217751X17430126. eprint: https://doi.org/10.1142/S0217751X17430126. URL: https://doi.org/10.1142/S0217751X17430126.

[60] Vivek Singh. "Development of Cryogenic Bolometer for Neutrinoless Double Beta Decay in $^{124}$Sn". PhD thesis. Homi Bhabha National Institute, 2014, pp. I–110.

[61] HyangKyu Park. "The AMoRE: Search for neutrinoless double beta decay of $^{100}$Mo". In: *Workshop on Calculation of Double-Beta Matrix Elements (MEDEX'15)*. Vol. 1686. American Institute of Physics Conference Series. Oct. 2015, 020016, p. 020016. DOI: 10.1063/1.4934905.

[62] F. Alessandria et al. *Sensitivity and Discovery Potential of CUORE to Neutrinoless Double-Beta Decay*. 2013. arXiv: 1109.0494 [nucl-ex].

[63] D. R. Artusa et al. "Initial performance of the CUORE-0 experiment". In: *The European Physical Journal C* 74.8 (2014), p. 2956. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-2956-6. URL: https://doi.org/10.1140/epjc/s10052-014-2956-6.

[64] Dusan Budjas. "Germanium detector studies in the framework of the GERDA experiment". In: (Jan. 2009).

[65] Glenn F. Knoll. *Radiation Detection and Measurement, 3rd ed.* 3rd edition. New York: John Wiley and Sons, 2000. ISBN: 978-0-471-07338-3, 978-0-471-07338-3.

[66] J. Gruszko and J. A. Detwiler. "Surface Alpha Interactions in P-Type Point-Contact HPGe Detectors: Maximizing Sensitivity of $^{76}$Ge Neutrinoless Double-Beta Decay Searches". PhD thesis. UW Physics, 2017, pp. 1–192.

[67] Graham Giovanetti. "P-type point contact germanium detectors and their application in rare-event searches". PhD thesis. Mar. 2019.

[68] M. Agostini et al. "Background-free search for neutrinoless double-$\beta$ decay of 76Ge with GERDA". In: *Nature* 544.7648 (2017), pp. 47–52. ISSN: 1476-4687. DOI: 10.1038/nature21717. URL: https://doi.org/10.1038/nature21717.

[69] C. H. Yu et al. "The Majorana Demonstrator Status and Preliminary Results". In: *EPJ Web of Conferences* 178 (May 2018). DOI: 10.1051/epjconf/201817801006.

[70] Jordan Myslik. *LEGEND: The Large Enriched Germanium Experiment for Neutrinoless Double-Beta Decay*. 2018. arXiv: 1810.00849 [physics.ins-det].

[71] R. Martin. "The Majorana Demonstrator". Talk held at DBD '11, Osaka, Japan, November 2011. 2011. URL: http://www-eng.lbl.gov/~shuman/NEXT/Other_experiments/MJD.pdf.

[72] M. Agostini et al. "Final Results of GERDA on the Search for Neutrinoless Double-$\beta$ Decay". In: *Phys. Rev. Lett.* 125 (25 2020), p. 252502. DOI: 10.1103/PhysRevLett.125.252502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.252502.

[73] N. Abgrall et al. *The LEGEND double beta decay experiment*. Snowmass 2021, Letter of Intent. 2021.

[74] M. Buuck and J. A. Detwiler. "A Radiogenic Background Model for the MAJORANA DEMONSTRATOR". PhD thesis. UW Physics, 2019, pp. 1–396.

[75] W. Xu et al. "Testing the Ge Detectors for the MAJORANA DEMONSTRATOR". In: *Physics Procedia* 61 (2015). 13th International Conference on Topics in Astroparticle and Underground Physics, TAUP 2013, pp. 807–815. ISSN: 1875-3892. DOI: https://doi.org/10.1016/j.phpro.2014.12.104. URL: https://www.sciencedirect.com/science/article/pii/S1875389214007172.

[76] Joachim Ebert et al. "Results of a search for neutrinoless double- $\beta$ decay using the COBRA demonstrator". In: *Physical Review C* 94 (Sept. 2015). DOI: 10.1103/PhysRevC.94.024603.

[77] K Abe et al. "XMASS detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 716 (2013), pp. 78–85.

[78] Y. Kishimoto. "The XMASS Experiment". Slides for the international symposium on "Revealing the history of the universe with underground particle and nuclear research 2019", Tohoku University, March 8th 2019. 2019. URL: http://www-sk.icrr.u-tokyo.ac.jp/xmass/prelist/TheXmassExperiment-v2_Kishimoto2014.pdf.

[79] XMASS Collaboration. *XMASS homepage, "About" Section*. http://www-sk.icrr.u-tokyo.ac.jp/xmass/about/future/index-e.html. Accessed: Dec. 04 2020. 2020.

[80] N. Kudomi et al. "Double beta decays and ELEGANT V and VI at Oto Cosmo Observatory". In: *Nuclear Physics A* 629.1 (1998). Quark Lepton Nuclear Physics, pp. 527–530. ISSN: 0375-9474. DOI: 10.1016/S0375-9474(97)00732-X. URL: https://www.sciencedirect.com/science/article/pii/S037594749700732X.

[81] Yoshihito Gando. "Neutrinoless double beta decay search with KamLAND-Zen". In: *PoS* NOW2018 (2019), p. 068. DOI: 10.22323/1.337.0068.

[82] S. Ajimura et al. "Low background measurement in CANDLES-III for studying the neutrinoless double beta decay of $^{48}$Ca". In: (2021). arXiv: 2008.09288 [hep-ex].

[83] Junpei Shirai. "KamLAND-Zen experiment". In: *PoS* HQL2018 (2018), p. 050. DOI: 10.22323/1.332.0050.

[84] Edward Leming. "SNO+: Current Results and Future Prospects". In: *PoS* NOW2018 (2019), p. 027. DOI: 10.22323/1.337.0027.

[85] K Abe et al. "Search for two-neutrino double electron capture on 124Xe with the XMASS-I detector". In: *Physics Letters B* 759 (2016), pp. 64–68.

[86] S. Umehara. "Search for Neutrino-less Double Beta Decay with CANDLES". Presentation slides for talk given on 28 July at the the 19th Particles and Nuclei International Conference (PANIC11, 24-28 July 2011, MIT). 2011. URL: http://web.mit.edu/panic11/talks/thursday/PARALLEL-5E/5-1720/umehara/467-0-5E5_Umehara.pdf.

[87] A. S. Barabash et al. "Final results of the Aurora experiment to study $2\beta$ decay of $^{116}$Cd with enriched $^{116}CdWO_4$ crystal scintillators". In: *Phys. Rev. D* 98.9 (2018), p. 092007. DOI: 10.1103/PhysRevD.98.092007. arXiv: 1811.06398 [nucl-ex].

[88] SuperNEMO collaboration. *SuperNEMO Conceptual Design Report*. Tech. rep. Accessed: 04-27-2021. 2010, pp. 1–58. URL: http://www.hep.ucl.ac.uk/~saakyan/NEMO_Docs/SuperNEMO_Demonstrator_CDR.pdf.

[89] N. Ishihara et al. "Performance of the drift chamber beta-ray momentum analyzer for double beta decay experiments". In: *Progress of Theoretical and Experimental Physics* 2017.7 (July 2017). 073H01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptx091. eprint: https://academic.oup.com/ptep/article-pdf/2017/7/073H01/19375628/ptx091.pdf. URL: https://doi.org/10.1093/ptep/ptx091.

[90] C. Patrick and F. Xie. "Status of the SuperNEMO $0\nu\beta\beta$ experiment". In: (2017). arXiv: 1704.06670 [physics.ins-det].

[91] H. Kakuno. "Status of the DCBA experiment". International Workshop on "Double Beta Decay and Underground Science", DBD16, Osaka University, Japan, 2016. 2016. URL: http://www.rcnp.osaka-u.ac.jp/dbd16/Data/Prog/061_DCBA_HidekazuKakuno.pdf.

[92] R. Arnold et al. "Technical design and performance of the NEMO 3 detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 536.1 (2005), pp. 79–122. ISSN: 0168-9002. DOI: 10.1016/j.nima.2004.07.194. URL: https://www.sciencedirect.com/science/article/pii/S0168900204016821.

[93] A.S. Barabash et al. "Calorimeter development for the SuperNEMO double beta decay experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 868 (2017), pp. 98–108. ISSN: 0168-9002. DOI: 10.1016/j.nima.2017.06.044. URL: https://www.sciencedirect.com/science/article/pii/S0168900217306976.

[94] David R. Nygren. *Proposal to investigate the feasibility of a novel concept in particle detection*. LBL (Berkeley Lab), 1974.

[95] M. Nakamura et al. *PEP-4 - TIME PROJECTION CHAMBER*. LBNL Report #: LBID-128. Lawrence Berkeley National Laboratory, 1979. URL: https://escholarship.org/uc/item/8cx4h1cg.

[96] Denis Bernard et al. "HARPO: a TPC as a gamma-ray telescope and polarimeter". In: *Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray*. Ed. by Tadayuki Takahashi, Jan-Willem A. den Herder, and Mark Bautz. Vol. 9144. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2014, 91441M, p. 91441M. DOI: 10.1117/12.2055307. arXiv: 1406.4830 [astro-ph.IM].

[97] D.S. Akerib et al. "The LUX-ZEPLIN (LZ) experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 953 (2020), p. 163047. ISSN: 0168-9002. DOI: 10.1016/j.nima.2019.163047. URL: https://www.sciencedirect.com/science/article/pii/S0168900219314032.

[98] R.L. Platzman. "Total ionization in gases by high-energy particles: An appraisal of our understanding". In: *The International Journal of Applied Radiation and Isotopes* 10.2 (1961), pp. 116–127. ISSN: 0020-708X. DOI: 10.1016/0020-708X(61)90108-9. URL: https://www.sciencedirect.com/science/article/pii/0020708X61901089.

[99] Vitaly Chepel and Henrique Araújo. "Liquid noble gas detectors for low energy particle physics". In: *Journal of Instrumentation* 8 (July 2012). DOI: 10.1088/1748-0221/8/04/R04001.

[100] Steven P. Ahlen. "Theoretical and experimental aspects of the energy loss of relativistic heavily ionizing particles". In: *Rev. Mod. Phys.* 52 (1 1980), pp. 121–173. DOI: 10.1103/RevModPhys.52.121. URL: https://link.aps.org/doi/10.1103/RevModPhys.52.121.

[101] Juan José (dir.) Martín-Albo Simón Justo Gómez Cadenas. "The NEXT experiment for neutrinoless double beta decay searches". PhD thesis. Feb. 2015, pp. I–192.

[102] Felix Theeuwes and Richard J Bearman. "The p, V, T behavior of dense fluids V. The vapor pressure and saturated liquid density of xenon". In: *The Journal of Chemical Thermodynamics* 2.4 (1970), pp. 507–512. ISSN: 0021-9614. DOI: 10.1016/0021-9614(70)90100-X. URL: https://www.sciencedirect.com/science/article/pii/002196147090100X.

[103] R. C. Kemp, W. R. G. Kemp, and P. W. Smart. "The Triple Point of Xenon as a Possible Defining Point on an International Temperature Scale". In: *Metrologia* 21.2 (Jan. 1985), pp. 43–48. DOI: 10.1088/0026-1394/21/2/001.

[104] H. A. Koehler et al. "Vacuum-ultraviolet emission from high-pressure xenon and argon excited by high-current relativistic electron beams". In: *Phys. Rev. A* 9 (2 1974), pp. 768–781. DOI: 10.1103/PhysRevA.9.768. URL: https://link.aps.org/doi/10.1103/PhysRevA.9.768.

[105] William P. Jesse and John Sadaukis. "Absolute Energy to Produce an Ion Pair by Beta Particles from $S^{35}$". In: *Phys. Rev.* 107 (3 1957), pp. 766–771. DOI: 10.1103/PhysRev.107.766. URL: https://link.aps.org/doi/10.1103/PhysRev.107.766.

[106] Aleksey Bolotnikov and Brian Ramsey. "The spectroscopic properties of high-pressure xenon". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 396.3 (1997), pp. 360–370. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00784-5. URL: https://www.sciencedirect.com/science/article/pii/S0168900297007845.

[107] T. Takahashi et al. "Average energy expended per ion pair in liquid xenon". In: Phys. Rev. A 12.5 (Nov. 1975), pp. 1771–1775. DOI: 10.1103/PhysRevA.12.1771.

[108] A. Baldini et al. "Absorption of scintillation light in a 100l liquid xenon $\gamma$-ray detector and expected detector performance". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 545.3 (2005), pp. 753–764. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.02.029. URL: https://www.sciencedirect.com/science/article/pii/S0168900205008259.

[109] B. J. Mount et al. *LUX-ZEPLIN (LZ) Technical Design Report*. 2017. arXiv: 1703.09144 [physics.ins-det].

[110] T. H. V. T. Dias et al. "Full-energy absorption of x-ray energies near the Xe L- and K-photoionization thresholds in xenon gas detectors: Simulation and experimental results". In: *Journal of Applied Physics* 82.6 (1997), pp. 2742–2753. DOI: 10.1063/1.366105. eprint: https://doi.org/10.1063/1.366105. URL: https://doi.org/10.1063/1.366105.

[111] T. Doke et al. "Estimation of Fano factors in liquid argon, krypton, xenon and xenon-doped liquid argon". In: *Nuclear Instruments and Methods* 134.2 (1976), pp. 353–357. ISSN: 0029-554X. DOI: 10.1016/0029-554X(76)90292-5. URL: https://www.sciencedirect.com/science/article/pii/0029554X76902925.

[112] E. Conti et al. "Correlated fluctuations between luminescence and ionization in liquid xenon". In: *Phys. Rev. B* 68 (5 2003), p. 054201. DOI: 10.1103/PhysRevB.68.054201. URL: https://link.aps.org/doi/10.1103/PhysRevB.68.054201.

[113] David Nygren. "Can the "intrinsic" energy resolution in xenon be surpassed?" In: *Journal of Physics Conference Series*. Vol. 309. Journal of Physics Conference Series. Aug. 2011, 012006, p. 012006. DOI: 10.1088/1742-6596/309/1/012006.

[114] David Nygren. "High-pressure xenon gas electroluminescent TPC for 0-$\nu$ $\beta\beta$-decay search". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 603.3 (2009), pp. 337–348. ISSN: 0168-9002. DOI: 10.1016/j.nima.2009.01.222. URL: https://www.sciencedirect.com/science/article/pii/S0168900209002897.

[115] M. J. Jewell et al. "Search for neutrinoless double beta decay with EXO-200 and nEXO". PhD thesis. 2020.

[116] Yichen Li et al. "Measurement of longitudinal electron diffusion in liquid argon". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 816 (2016), pp. 160–170. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.01.094. URL: https://www.sciencedirect.com/science/article/pii/S0168900216001443.

[117] O. Njoya et al. "Measurements of electron transport in liquid and gas Xenon using a laser-driven photocathode". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 972 (2020), p. 163965. ISSN: 0168-9002. DOI: 10.1016/j.nima.2020.163965. URL: https://www.sciencedirect.com/science/article/pii/S0168900220304320.

[118] Sam S.-S. Huang and Gordon R. Freeman. "Electron mobilities in gaseous, critical, and liquid xenon: Density, electric field, and temperature effects: Quasilocalization". In: *The Journal of Chemical Physics* 68.4 (1978), pp. 1355–1362. DOI: 10.1063/1.435954. eprint: https://doi.org/10.1063/1.435954. URL: https://doi.org/10.1063/1.435954.

[119] E. Aprile et al. "Observation of two-neutrino double electron capture in 124Xe with XENON1T". In: *Nature* 568.7753 (2019), pp. 532–535. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1124-4. URL: https://doi.org/10.1038/s41586-019-1124-4.

[120] D S Akerib et al. "Search for two neutrino double electron capture of 124 Xe and 126 Xe in the full exposure of the LUX detector". In: *Journal of Physics G: Nuclear and Particle Physics* 47.10 (2020), p. 105105. ISSN: 1361-6471. DOI: 10.1088/1361-6471/ab9c2d. URL: http://dx.doi.org/10.1088/1361-6471/ab9c2d.

[121] Dick Betzendahl. "The rare gas market begins to stabilize". In: *gasworld (US Edition)* 53.3 (2019), pp. 50–53.

[122] D. Herman and Kenneth G. Unfried. "Xenon Acquisition Strategies for High-Power Electric Propulsion NASA Missions". In: 2015.

[123] nEXO Collaboration et al. *nEXO Pre-Conceptual Design Report*. 2018. arXiv: 1805.11142 [physics.ins-det].

[124] Andi Tan; Xiangdong Ji. "PandaX-II Dark Matter Detector and its First Results". PhD thesis. University of Maryland, 2019, pp. I–149.

[125] J. B. Albert et al. "Sensitivity and discovery potential of the proposed nEXO experiment to neutrinoless double-$\beta$ decay". In: *Phys. Rev. C* 97 (6 2018), p. 065503. DOI: 10.1103/PhysRevC.97.065503. URL: https://link.aps.org/doi/10.1103/PhysRevC.97.065503.

[126] A.M.F. Trindade et al. "Study of the loss of xenon scintillation in xenon-trimethylamine mixtures". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 905 (2018), pp. 22–28. ISSN: 0168-9002. DOI: 10.1016/j.nima.2018.07.020. URL: https://www.sciencedirect.com/science/article/pii/S0168900218308520.

[127] Shaobo Wang. "PandaX-III: Search for neutrinoless double beta decay of $^{136}$Xe at CJPL". Presentation for the 2nd TDLI Mini-Workshop on "New Physics at the Tera Scale", Shanghai, China, Aug. 5, 2019. 2019. URL: https://indico-tdli.sjtu.edu.cn/event/3/contributions/47/attachments/53/88/PandaX-III_TDLI2019_Shaobo.pdf.

[128] NEXT Collaboration et al. "Design and characterization of the SiPM tracking system of NEXT-DEMO, a demonstrator prototype of the NEXT-100 experiment". In: (2013). arXiv: 1206.6199 [physics.ins-det].

[129] Javier Galan et al. "Topological background discrimination in the PandaX-III neutrinoless double beta decay experiment". In: *Journal of Physics G: Nuclear and Particle Physics* 47 (Oct. 2019). DOI: 10.1088/1361-6471/ab4dbe.

[130] J. Martín-Albo et al. "Sensitivity of NEXT-100 to neutrinoless double beta decay". In: *Journal of High Energy Physics* 2016.5 (2016), p. 159. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2016)159. URL: https://doi.org/10.1007/JHEP05(2016)159.

[131] Xun Chen et al. "PandaX-III: Searching for neutrinoless double beta decay with high pressure 136Xe gas time projection chambers". In: *Science China Physics, Mechanics & Astronomy* 60.6 (2017), p. 061011. ISSN: 1869-1927. DOI: 10.1007/s11433-017-9028-0. URL: https://doi.org/10.1007/s11433-017-9028-0.

[132] Vicente Alvarez et al. "NEXT-100 technical design report (TDR). Executive summary". In: *Journal of Instrumentation* 7 (June 2012). DOI: 10.1088/1748-0221/7/06/T06001.

[133] S Andriamonje et al. "Development and performance of Microbulk Micromegas detectors". In: *Journal of Instrumentation* 5.02 (2010), P02001.

[134] Pompa Federica et al. "Neutrino physics with the XENONnT experiment". MA thesis. Dipartimento di Fisica e Astronomia, Università di Bologna, 2019, pp. I–108.

[135] S. Sukhoruchkin and Z. Soroko. "Excited Nuclear States for Ni-60 (Nickel)". In: Jan. 2012. ISBN: 978-3-642-22797-4. DOI: 10.1007/978-3-642-22798-1_519.

[136] K Abe et al. "Measurement of neutrino and antineutrino neutral-current quasielasticlike interactions on oxygen by detecting nuclear deexcitation $\gamma$ rays". In: *Physical Review D* 100.11 (2019), p. 112009.

[137] Christian Spiering. "Towards high-energy neutrino astronomy". In: *The European Physical Journal H* 37.3 (2012), pp. 515–565. ISSN: 2102-6467. DOI: 10.1140/epjh/e2012-30014-2. URL: https://doi.org/10.1140/epjh/e2012-30014-2.

[138] Michael Wurm. "Solar neutrino spectroscopy". In: *Physics Reports* 685 (2017). Solar Neutrino Spectroscopy, pp. 1–52. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2017.04.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157317300601.

[139] D. Akimov et al. "Observation of coherent elastic neutrino-nucleus scattering". In: *Science* 357.6356 (2017), pp. 1123–1126. ISSN: 0036-8075. DOI: 10.1126/science.aao0990. eprint: https://science.sciencemag.org/content/357/6356/1123.full.pdf. URL: https://science.sciencemag.org/content/357/6356/1123.

[140] Hiroyasu Ejiri and S. Elliott. "The charged current neutrino cross section for solar neutrinos, and background to $\beta\beta0\nu$ experiments". In: *Physical Review C* 89 (Sept. 2013). DOI: 10.1103/PhysRevC.89.055501.

[141] Nuno Barros and K. Zuber. "Solar neutrino-electron scattering as background limitation for double beta decay". In: *Journal of Physics G-nuclear and Particle Physics - J PHYS G-NUCL PARTICLE PHYS* 38 (Mar. 2011). DOI: 10.1088/0954-3899/38/10/105201.

[142] S. R. Elliott and H. Ejiri. "Solar neutrinos as background to neutrinoless double-beta decay experiments". In: *AIP Conference Proceedings* 1894.1 (2017), p. 020008. DOI: 10.1063/1.5007633. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.5007633. URL: https://aip.scitation.org/doi/abs/10.1063/1.5007633.

[143] E. Aprile et al. "Projected WIMP sensitivity of the XENONnT dark matter experiment". In: *Journal of Cosmology and Astroparticle Physics* 2020.11 (2020), pp. 031–031. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2020/11/031. URL: http://dx.doi.org/10.1088/1475-7516/2020/11/031.

[144] Ashok Tiwari et al. "Observation of annual modulation induced by $\gamma$ rays from $(\alpha, \gamma)$ reactions at the Soudan Underground Laboratory". In: *Phys. Rev. C* 96 (4 2017), p. 044609. DOI: 10.1103/PhysRevC.96.044609. URL: https://link.aps.org/doi/10.1103/PhysRevC.96.044609.

[145] A. S. Malgin. "On the energy spectrum of cosmogenic neutrons". In: *Journal of Experimental and Theoretical Physics* 125.5 (2017), pp. 728–740. ISSN: 1090-6509. DOI: 10.1134/S1063776117100089. URL: https://doi.org/10.1134/S1063776117100089.

[146] V.A. Kudryavtsev, P. Zakhary, and B. Easeman. "Neutron production in ($\alpha$,n) reactions". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 972 (2020), p. 164095. ISSN: 0168-9002. DOI: 10.1016/j.nima.2020.164095. URL: https://www.sciencedirect.com/science/article/pii/S0168900220305088.

[147] A El-Sayed Abdo. "Calculation of the cross-sections for fast neutrons and gamma-rays in concrete shields". In: *Annals of Nuclear Energy* 29.16 (2002), pp. 1977–1988.

[148] Sara Pozzi and Imre Pázsit. "The Statistics of the Number of Neutron Collisions Prior to Absorption". In: *Nuclear Science and Engineering* 153 (May 2006). DOI: 10.13182/NSE06-A2595.

[149] J. Beringer et al. "Review of Particle Physics". In: *Phys. Rev. D* 86 (1 2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001. URL: https://link.aps.org/doi/10.1103/PhysRevD.86.010001.

[150] D.F. Measday. "The nuclear physics of muon capture". In: *Physics Reports* 354.4 (2001), pp. 243–409. ISSN: 0370-1573. DOI: 10.1016/S0370-1573(01)00012-6. URL: https://www.sciencedirect.com/science/article/pii/S0370157301000126.

[151] D. Poda. "CUPID-Mo: new results on $^{100}$Mo $0\nu/2\nu2\beta$". presentation for GDR Neutrino remote meeting, 23-24 November 2020. 2020. URL: https://indico.in2p3.fr/event/22500/contributions/87750/attachments/60972/83027/GDR-Neutrino_20201123_Poda_v3.pdf.

[152] Xun Chen et al. "PandaX-III: Searching for Neutrinoless Double Beta Decay with High Pressure $^{136}$Xe Gas Time Projection Chambers". In: (2016). arXiv: 1610.08883 [physics.ins-det].

[153] V. Lozza. "The SNO+ Experiment for Neutrinoless Double-Beta Decay". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016). 37th International Conference on High Energy Physics (ICHEP), pp. 1836–1841. ISSN: 2405-6014. DOI: 10.1016/j.nuclphysbps.2015.09.296. URL: https://www.sciencedirect.com/science/article/pii/S2405601415007853.

[154] DQ Adams et al. "Improved Limit on Neutrinoless Double-Beta Decay in Te 130 with CUORE". In: *Physical review letters* 124.12 (2020), p. 122501.

[155] Anatoly Smolnikov and for Collaboration. "Status of the GERDA experiment aimed to search for neutrinoless double beta decay of 76Ge". In: (Jan. 2009).

[156] Yoshiyuki FUKUDA. *ZICOS – Neutrinoless double beta decay experiment using Zr-96 with an organic liquid scintillator-*. June 2018. DOI: 10.5281/zenodo.1300437. URL: https://doi.org/10.5281/zenodo.1300437.

[157] M. Tenconi. "LUMINEU: A Pilot Scintillating Bolometer Experiment for Neutrinoless Double Beta Decay Search". In: *Physics Procedia* 61 (2015). 13th International Conference on Topics in Astroparticle and Underground Physics, TAUP 2013, pp. 782–786. ISSN: 1875-3892. DOI: 10.1016/j.phpro.2014.12.099. URL: https://www.sciencedirect.com/science/article/pii/S1875389214007123.

[158] J. M. López-Castaño et al. "Recent results of the Majorana Demonstrator experiment". In: *WORKSHOP ON CALCULATION OF DOUBLE-BETA-DECAY MATRIX ELEMENTS (MEDEX'19)* (2019). DOI: 10.1063/1.5130979. URL: http://dx.doi.org/10.1063/1.5130979.

[159] Cheryl PATRICK, Lauren DAWSON, and Alessandro MINOTTI. *SuperNEMO $0\nu\beta\beta$ sensitivity studies*. July 2018. DOI: 10.5281/zenodo.1302919. URL: https://doi.org/10.5281/zenodo.1302919.

[160] O. Azzolini et al. "Final result of CUPID-0 phase-I in the search for the $^{82}$Se Neutrinoless Double-$\beta$ Decay". In: *Phys. Rev. Lett.* 123.3 (2019), p. 032501. DOI: 10.1103/PhysRevLett.123.032501. arXiv: 1906.05001 [nucl-ex].

[161] A. Barabash, V. Brudanin, and the Collaboration. "Investigation of double beta decay with the NEMO-3 detector". In: *Physics of Atomic Nuclei* 74 (Feb. 2010). DOI: 10.1134/S1063778811020062.

[162] V. Alenkov et al. "First results from the AMoRE-Pilot neutrinoless double beta decay experiment". In: *The European Physical Journal C* 79.9 (2019), p. 791. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7279-1. URL: https://doi.org/10.1140/epjc/s10052-019-7279-1.

[163] S. Umehara et al. "Neutrino-less double-$\beta$ decay of $^{48}$Ca studied by $CaF_2$(Eu) scintillators". In: *Phys. Rev. C* 78 (5 2008), p. 058501. DOI: 10.1103/PhysRevC.78.058501. URL: https://link.aps.org/doi/10.1103/PhysRevC.78.058501.

[164] T. Iida et al. "The CANDLES experiment for the study of Ca-48 double beta decay". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016). 37th International Conference on High Energy Physics (ICHEP), pp. 2633–2635. ISSN: 2405-6014. DOI: 10.1016/j.nuclphysbps.2015.10.013. URL: https://www.sciencedirect.com/science/article/pii/S2405601415009724.

[165] Ibles Olcina Samblas and Henrique Araújo. "A statistical framework for the characterisation of WIMP dark matter with the LUX-ZEPLIN experiment". PhD thesis. 2019, pp. 1–162.

[166] D. S. Akerib et al. "Position reconstruction in LUX". In: *Journal of Instrumentation* 13.02 (Feb. 2018). DOI: 10.1088/1748-0221/13/02/p02001.

[167] Peter Sorensen. "Anisotropic diffusion of electrons in liquid xenon with application to improving the sensitivity of direct dark matter searches". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 635.1 (2011), pp. 41–43. ISSN: 0168-9002. DOI: 10.1016/j.nima.2011.01.089. URL: https://www.sciencedirect.com/science/article/pii/S0168900211001756.

[168] E. Aprile and T. Doke. "Liquid xenon detectors for particle physics and astrophysics". In: *Rev. Mod. Phys.* 82 (3 2010), pp. 2053–2097. DOI: 10.1103/RevModPhys.82.2053. URL: https://link.aps.org/doi/10.1103/RevModPhys.82.2053.

[169] P. Rossiter and R. Watson. "Electron Drift Update". LZ 1.11 Offline computing meeting slides, Nov 12 2018. 2018.

[170] Wolfgang A. Rolke, Angel M. López, and Jan Conrad. "Limits and confidence intervals in the presence of nuisance parameters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 551.2 (2005), pp. 493–503. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.05.068. URL: https://www.sciencedirect.com/science/article/pii/S016890020501291X.

[171] J. B. Albert et al. "Search for Neutrinoless Double-Beta Decay with the Upgraded EXO-200 Detector". In: *Phys. Rev. Lett.* 120 (7 2018), p. 072701. DOI: 10.1103/PhysRevLett.120.072701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.072701.

[172] G. Varoquaux, A. Müller, and J. Grobler. *Classifier Comparison*. https://scikit-learn.org/0.22/auto_examples/classification/plot_classifier_comparison.html. Accessed: 12 Mar. 2020. 2020.

[173] Fabian Pedregosa et al. "Scikit-Learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12.null (2011), pp. 2825–2830. ISSN: 1532-4435.

[174] Nitin Bhatia and Vandana. "Survey of Nearest Neighbor Techniques". In: *International Journal of Computer Science and Information Security* 8 (July 2010).

[175] Sasan Karamizadeh et al. "Advantage and Drawback of Support Vector Machine Functionality". In: Sept. 2014. DOI: 10.1109/I4CT.2014.6914146.

[176] R. Frigola. "Bayesian Time Series Learning with Gaussian Processes". PhD thesis. Jan. 2015.

[177] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Nov. 2005. ISBN: 9780262256834. DOI: 10.7551/mitpress/3206.001.0001. URL: https://doi.org/10.7551/mitpress/3206.001.0001.

[178] L. Kozma. *k Nearest Neighbors Algorithm (kNN)*. Lecture notes for T-61.6020 Special Course in Computer and Information Science. 2008.

[179] Dmytro Nasyrov. *Machine Learning. k-Nearest Neighbors.* https://medium.com/pharos-production/learning-machine-learning-petals-32c68673d8fc. Accessed: 04-27-2021. 2017.

[180] LiLi Li, YanXia Zhang, and YongHeng Zhao. "k-Nearest Neighbors for automated classification of celestial objects". In: *Science in China Series G: Physics, Mechanics and Astronomy* 51.7 (2008), pp. 916–922. ISSN: 1862-2844. DOI: 10.1007/s11433-008-0088-4. URL: https://doi.org/10.1007/s11433-008-0088-4.

[181] Matthew Durbin et al. "K-Nearest Neighbors regression for the discrimination of gamma rays and neutrons in organic scintillators". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 987 (2021), p. 164826. ISSN: 0168-9002. DOI: 10.1016/j.nima.2020.164826. URL: https://www.sciencedirect.com/science/article/pii/S0168900220312237.

[182] Hui Li et al. "Machine Learning Approach for Solar Wind Categorization". In: *Earth and Space Science* 7.5 (2020). e2019EA000997 10.1029/2019EA000997, e2019EA000997. DOI: 10.1029/2019EA000997. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019EA000997. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019EA000997.

[183] Michael J. Brice and Răzvan Andonie. "Automated Morgan Keenan Classification of Observed Stellar Spectra Collected by the Sloan Digital Sky Survey Using a Single Classifier". In: AJ 158.5, 188 (Nov. 2019), p. 188. DOI: 10.3847/1538-3881/ab40d0.

[184] G. Rastegarzadeh and M. Nemati. "Primary mass discrimination of high energy cosmic rays using PNN and k-NN methods". In: *Advances in Space Research* 61.4 (2018), pp. 1181–1191. ISSN: 0273-1177. DOI: 10.1016/j.asr.2017.11.016. URL: https://www.sciencedirect.com/science/article/pii/S0273117717308116.

[185] Oliver Sutton. *Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction*. 2012.

[186] R. Berwick. "An Idiot's guide to Support vector". Lecture slides for the 2011 6.034 Artificial Intelligence course, MIT. 2011. URL: https://web.mit.edu/6.034/wwwbob/svm.pdf.

[187] A. Ng. *CS229 Lecture Notes, Part V: Support Vector Machines*. Lecture notes for CS229 Machine Learning course. 2018.

[188] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. ISBN: 978-0-521-83378-3.

[189] Clément Walter. "Using Poisson processes for rare event simulation". Theses. Université Paris Diderot / Sorbonne Paris Cité, Oct. 2016. URL: https://tel.archives-ouvertes.fr/tel-01575418.

[190] Xunzhen Yu et al. "Neutron–gamma discrimination based on the support vector machine method". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 777 (2015), pp. 80–84. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.12.087. URL: https://www.sciencedirect.com/science/article/pii/S0168900214015551.

[191] Zhen Wang et al. "SVM Multivariate Veto". Poster for The 14th Gravitational Wave Data Analysis Workshop (GWDAW-14), January 26th to 29th, 2010, Rome, Italy. 2010. URL: https://www.mit.edu/~caoj/pub/doc/jcao_j_svm.pdf.

[192] Changde Du, Ali Luo, and Haifeng Yang. "Adaptive stellar spectral subclass classification based on Bayesian SVMs". In: *New Astronomy* 51 (2017), pp. 51–58. ISSN: 1384-1076. DOI: 10.1016/j.newast.2016.08.015. URL: https://www.sciencedirect.com/science/article/pii/S1384107616300793.

[193] Liu Zhong-Bao. "Stellar Spectral Classification with Minimum Within-Class and Maximum Between-Class Scatter Support Vector Machine". In: *Journal of Astrophysics and Astronomy* 37.2 (2016), p. 9. ISSN: 0973-7758. DOI: 10.1007/s12036-016-9374-0. URL: https://doi.org/10.1007/s12036-016-9374-0.

[194] Mark Ebden. "Gaussian Processes: A Quick Introduction". In: (2015). arXiv: 1505.02965 [math.ST].

[195] C. Do and H. Lee. *Gaussian Processes*. Lecture notes for CS229: Machine Learning course. 2008.

[196] `scikit-learn` developers. *`GaussianProcessClassifier()` class documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessClassifier.html. Accessed: 20 Jul. 2020. 2020.

[197] R. von Mises. *Mathematical Theory of Probability and Statistics*. 2nd edition. Academic Press Inc., 1967. ISBN: 978-1-4832-3213-3.

[198] D. Foreman-Mackey. "A practical introduction to Gaussian processes for astronomy". Presentation slides for the 231th AAS Meeting (8-12 January 2018) special session on "Applied statistical methods in astronomy: Gaussian processes and machine learning". 2018. URL: http://hea-www.harvard.edu/AstroStat/aas231/DForeman-Mackey_20180110_aas231.pdf.

[199] Oded Maimon and Lior Rokach. *Data Mining and Knowledge Discovery Handbook, 2nd ed.* Jan. 2010. ISBN: 9780387098227.

[200] Mark Last, Oded Maimon, and Einat Minkov. "Improving Stability of Decision Trees." In: *IJPRAI* 16 (Mar. 2002), pp. 145–159. DOI: 10.1142/S0218001402001599.

[201] Tom M. Mitchell. *Machine Learning*. 1st ed. New York: McGraw Hill, 1997. ISBN: 978-0-07-042807-2.

[202] J. R. Quinlan. "Induction of decision trees". In: *Machine Learning* 1.1 (1986), pp. 81–106. ISSN: 1573-0565. DOI: 10.1007/BF00116251. URL: https://doi.org/10.1007/BF00116251.

[203] Floriana Esposito et al. "A Comparative Analysis of Methods for Pruning Decision Trees". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19 (June 1997), pp. 476–491. DOI: 10.1109/34.589207.

[204] A. Cutler. "Random Forests for Regression and Classification". Talk at Ovronnaz, Switzerland, September 15-17, 2010. 2010. URL: https://math.usu.edu/adele/RandomForests/Ovronnaz.pdf.

[205] Zhenping Yi and Jingchang Pan. "Application of random forest to stellar spectral classification". In: 7 (Oct. 2010). DOI: 10.1109/CISP.2010.5648041.

[206] A. Baldeschi et al. "Star formation and morphological properties of galaxies in the Pan-STARRS $3\pi$ survey- I. A machine learning approach to galaxy and supernova classification". In: (Apr. 2020).

[207] Dan Gao, Yanxia Zhang, and Yongheng Zhao. "Random forest algorithm for classification of multiwavelength data". In: *Research in Astronomy and Astrophysics* 9 (Mar. 2009), p. 220. DOI: 10.1088/1674-4527/9/2/011.

[208] Mradul Sharma et al. "Validation of a new background discrimination method for the TACTIC TeV $\gamma$-ray telescope with Markarian 421 data". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 770 (2015), pp. 42–47. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.10.012. URL: https://www.sciencedirect.com/science/article/pii/S0168900214011590.

[209] Tomasz Trzciński, Łukasz Graczykowski, and Michał Glinka. "Using Random Forest Classifier for Particle Identification in the ALICE Experiment". In: *Information Technology, Systems Research, and Computational Physics*. Ed. by Piotr Kulczycki et al. Cham: Springer International Publishing, 2020, pp. 3–17. ISBN: 978-3-030-18058-4.

[210] Chris Drummond and Robert C. Holte. "Cost curves: An improved method for visualizing classifier performance". In: *Machine Learning* 65.1 (2006), pp. 95–130. ISSN: 1573-0565. DOI: 10.1007/s10994-006-8199-5. URL: https://doi.org/10.1007/s10994-006-8199-5.

[211] Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme. "Cost-sensitive learning methods for imbalanced data". In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*. 2010, pp. 1–8. DOI: 10.1109/IJCNN.2010.5596486.

[212] Václav Hlaváč. "Classifier performance evaluation". Class slides for the Czech Technical University in Prague. 2018. URL: http://people.ciirc.cvut.cz/~hlavac/TeachPresEn/31PattRecog/13ClassifierPerformance.pdf.

[213] Donald Bamber. "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph". In: *Journal of Mathematical Psychology* 12.4 (1975), pp. 387–415. ISSN: 0022-2496. DOI: 10.1016/0022-2496(75)90001-2. URL: https://www.sciencedirect.com/science/article/pii/0022249675900012.

[214] Naomi Altman and Martin Krzywinski. "The curse(s) of dimensionality". In: *Nature Methods* 15.6 (2018), pp. 399–400. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0019-x. URL: https://doi.org/10.1038/s41592-018-0019-x.

[215] A. Solovov. *b2b Deposition File Example Excerpts*. https://github.com/water-h2o/01_depFile_b2bhttps://github.com/water-h2o/01_depFile_b2b. Accessed: 15 June, 2021. 2021.

[216] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.06.125. URL: https://www.sciencedirect.com/science/article/pii/S0168900216306957.

[217] Geant4 Collaboration. *Physics Reference Manual, Release 10.4*. https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsReferenceManual/BackupVersions/V10.4/html/index.html. Accessed: 12 May 2020. 2017.

[218] F. Salvat, José M. Fernández-Varea, and Josep Sempau. "PENELOPE-2008: A code system for Monte Carlo simulation of electron and photon transport". In: Workshop Proceedings, Barcelona, Spain, 30 June-3 July 2008. © OECD 2009 NEA No. 6416. NUCLEAR ENERGY AGENCY ORGANISATION FOR ECONOMIC CO-OPERATION and DEVELOPMENT, 2009. ISBN: 978-92-64-99066-1.

[219] J. Sempau et al. "Experimental benchmarks of the Monte Carlo code penelope". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 207.2 (2003), pp. 107–123. ISSN: 0168-583X. DOI: 10.1016/S0168-583X(03)00453-1. URL: https://www.sciencedirect.com/science/article/pii/S0168583X03004531.

[220] O. Kadri et al. "Incorporation of the Goudsmit–Saunderson electron transport theory in the Geant4 Monte Carlo code". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 267.23 (2009), pp. 3624–3632. ISSN: 0168-583X. DOI: 10.1016/j.nimb.2009.09.015. URL: https://www.sciencedirect.com/science/article/pii/S0168583X09009823.

[221] Sung Kim et al. "Validation Test of Geant4 Simulation of Electron Backscattering". In: *Nuclear Science, IEEE Transactions on* 62 (Feb. 2015). DOI: 10.1109/TNS.2015.2401055.

[222] Matej Batic et al. "Validation of Geant4 Simulation of Electron Energy Deposition". In: *IEEE Transactions on Nuclear Science* 60 (July 2013). DOI: 10.1109/TNS.2013.2272404.

[223] A.F. Bielajew and F. Salvat. "Improved electron transport mechanics in the PENELOPE Monte-Carlo model". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 173.3 (2001), pp. 332–343. ISSN: 0168-583X. DOI: 10.1016/S0168-583X(00)00363-3. URL: https://www.sciencedirect.com/science/article/pii/S0168583X00003633.

[224] A. Solovov. *Spectra for 500* $1e$ *Events and 500* $0\nu\beta\beta$ *Events*. https://github.com/water-h2o/02_1e_0vbb_spectrahttps://github.com/water-h2o/02_1e_0vbb_spectra. Accessed: 15 June, 2021. 2021.

[225] Geant4 Collaboration. *G4NistMaterialBuilder.cc,* `HepAndNuclearMaterials()` *and* `BuildMaterial()`. https://www.apc.univ-paris7.fr/~franco/g4doxy/html/G4NistMaterialBuilder_8cc-source.html. Accessed: 12 May 2020. 2012.

[226] Geant4 Collaboration. *G4NistElementBuilder.cc,* `BuildElement()` *source code*. https://www.apc.univ-paris7.fr/~franco/g4doxy/html/G4NistElementBuilder_8cc-source.html. Accessed: 12 May 2020. 2012.

[227] J. S. Coursey et al. *Atomic Weights and Isotopic Compositions with Relative Atomic Masses*. Accessed: 04-27-2021. 2015. URL: https://www.nist.gov/pml/atomic-weights-and-isotopic-compositions-relative-atomic-masses.

[228] Dmitry Akimov et al. "Experimental study of ionization yield of liquid xenon for electron recoils in the energy range 2.8 - 80 keV". In: *Journal of Instrumentation* 9 (Aug. 2014). DOI: 10.1088/1748-0221/9/11/P11014.

[229] A. Solovov. *b2b Scenario Conversion Procedure from Deposited Energy into Produced Drift Electrons*. https://github.com/water-h2o/14_dE2e_b2bhttps://github.com/water-h2o/14_dE2e_b2b. Accessed: 15 June, 2021. 2021.

[230] *et al.* Morozov A. *G4ants*. https://github.com/andrmor/G4ants. Github repository. 2019.

[231] Geant4 Collaboration. *Guide for Physics Lists, Release 10.5*. https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsListGuide/BackupVersions/V10.5-2.0/html/index.html. Accessed: 12 May 2020. 2019.

[232] A. Solovov. *ISO Scenario* $0\nu\beta\beta$ *Emission Script*. https://github.com/water-h2o/03_isoEmithttps://github.com/water-h2o/03_isoEmit. Accessed: 15 June, 2021. 2021.

[233] A. Solovov. *ISO Scenario Conversion Procedure from Deposited Energy into Produced Drift Electrons*. https://github.com/water-h2o/05_dE2e_ISOhttps://github.com/water-h2o/05_dE2e_ISO. Accessed: 15 June, 2021. 2021.

[234] A. Solovov. *ISO Deposition File Example Excerpts*. https://github.com/water-h2o/04_depFile_ISOhttps://github.com/water-h2o/04_depFile_ISO. Accessed: 15 June, 2021. 2021.

[235] Brian Lenardo et al. "A Global Analysis of Light and Charge Yields in Liquid Xenon". In: *IEEE Transactions on Nuclear Science* 62.6 (2015), 3387–3396. ISSN: 1558-1578. DOI: 10.1109/tns.2015.2481322. URL: http://dx.doi.org/10.1109/TNS.2015.2481322.

[236] A. Solovov. *NEST-based LXe Density Calculation*. https://github.com/water-h2o/06_NEST_denshttps://github.com/water-h2o/06_NEST_dens. Accessed: 15 June, 2021. 2021.

[237] F. Neves et al. "Measurement of the absolute reflectance of polytetrafluoroethylene (PTFE) immersed in liquid xenon". In: *Journal of Instrumentation* 12 (Dec. 2016). DOI: 10.1088/1748-0221/12/01/P01017.

[238] A. Solovov. *Developed ANTS2 Script*. https://github.com/water-h2o/15_ANTS2_scripthttps://github.com/water-h2o/15_ANTS2_script. Accessed: 15 June, 2021. 2021.

[239] Eric Lemmon. *THERMOPHYSICAL PROPERTIES OF FLUIDS*. en. CRC, CRC, CO, 2009.

[240] Corning Inc. *Properties of PYREX®, PYREXPLUS® and Low Actinic PYREX Code 7740 Glasses*. Accessed May 28 2020. 2020. URL: http://www.quartz.com/pxprop.pdf.

[241] A. Solovov. *Diffusion Simulation Source Code*. https://github.com/water-h2o/07_diff_simhttps://github.com/water-h2o/07_diff_sim. Accessed: 15 June, 2021. 2021.

[242] Rene Brun and Fons Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00048-X. URL: https://www.sciencedirect.com/science/article/pii/S016890029700048X.

[243] W. Hörmann and G. Derflinger. "The ACR method for generating normal random variables". In: *Operations-Research-Spektrum* 12.3 (1990), pp. 181–185. ISSN: 1436-6304. DOI: 10.1007/BF01719718. URL: https://doi.org/10.1007/BF01719718.

[244] CERN ROOT collaboration. *TRandomClass Reference*. https://root.cern.ch/doc/master/classTRandom.html. Accessed: 2 June, 2020. 2020.

[245] D. Soeder, C. Abad, and G. Acevedo. "Black-Box Assessment of Pseudorandom Algorithms". Black Hat USA 2013 security conference talks (Accessed Jun. 3 2020). 2013. URL: https://media.blackhat.com/us-13/US-13-Soeder-Black-Box-Assessment-of-Pseudorandom-Algorithms-WP.pdf.

[246] G. Guennebaud, B. Jacob, and *et al. Eigen v3*. http://eigen.tuxfamily.org. Accessed: May 28, 2020. 2010.

[247] Andrey Morozov et al. "Iterative reconstruction of the detector response for medical gamma cameras". In: (Jan. 2015).

[248] The Qt Company. *Qt Homepage*. https://www.qt.io/. Accessed: 8 June, 2020. 2020.

[249] V. Solovov. *ants2-docker Dockerfile commit*. https://github.com/vovasolo/ants2-dockerhttps://github.com/vovasolo/ants2-docker. Accessed: 8 June, 2020. 2018.

[250] A. Solovov. *Waveform simulation suite used for distributed computing*. https://github.com/water-h2o/gurgelSimhttps://github.com/water-h2o/gurgelSim. Accessed: 8 June, 2020. 2019.

[251] A. Solovov. *Used ANTS2 Dockerfile*. https://github.com/water-h2o/08_the_dockerfilehttps://github.com/water-h2o/08_the_dockerfile. Accessed: 15 June, 2021. 2021.

[252] A. Solovov. *Developed Waveform Parameterizer*. https://github.com/water-h2o/09_waveformParameterizerhttps://github.com/water-h2o/09_waveformParameterizer. Accessed: 15 June, 2021. 2021.

[253] Michel Verleysen John A. Lee. *Nonlinear Dimensionality Reduction*. 1st edition. Springer-Verlag New York, 2007. ISBN: 978-0-387-39350-6.

[254] A. Solovov. *Developed Optimization Procedure*. https://github.com/water-h2o/10_data_optimizationhttps://github.com/water-h2o/10_data_optimization. Accessed: 15 June, 2021. 2021.

[255] P. Pudil, J. Novovičová, and J. Kittler. "Floating search methods in feature selection". In: *Pattern Recognition Letters* 15.11 (1994), pp. 1119–1125. ISSN: 0167-8655. DOI: 10.1016/0167-8655(94)90127-9. URL: https://www.sciencedirect.com/science/article/pii/0167865594901279.

[256] A. Jain and D. Zongker. "Feature selection: evaluation, application, and small sample performance". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.2 (1997), pp. 153–158. DOI: 10.1109/34.574797.

[257] `scikit-learn` developers. *`KNeighborsClassifier()` class documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. Accessed: 20 Jul. 2020. 2020.

[258] `scikit-learn` developers. *`SVC()` class documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html. Accessed: 20 Jul. 2020. 2020.

[259] `scikit-learn` developers. *`RandomForestClassifier()` class documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. Accessed: 20 Jul. 2020. 2020.

[260] Lars Buitinck et al. "API design for machine learning software: experiences from the scikit-learn project". In: (2013). arXiv: 1309.0238 [cs.LG].

[261] A. Solovov. *Classification Algorithm Implementation*. https://github.com/water-h2o/11_jupyterScripthttps://github.com/water-h2o/11_jupyterScript. Accessed: 15 June, 2021. 2021.

[262] Annette M. Molinaro, Richard Simon, and Ruth M. Pfeiffer. "Prediction error estimation: a comparison of resampling methods". In: *Bioinformatics* 21.15 (May 2005), pp. 3301–3307. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bti499. eprint: https://academic.oup.com/bioinformatics/article-pdf/21/15/3301/6236941/bti499.pdf. URL: https://doi.org/10.1093/bioinformatics/bti499.

[263] A. Solovov. *Convergence Test Script*. https://github.com/water-h2o/16_convergenceTesthttps://github.com/water-h2o/16_convergenceTest. Accessed: 15 June, 2021. 2021.

[264] A. Solovov. *Developed Cost Curve Script*. https://github.com/water-h2o/12_costCurveScripthttps://github.com/water-h2o/12_costCurveScript. Accessed: 15 June, 2021. 2021.

[265] A. Solovov. *Developed ROC Script*. https://github.com/water-h2o/13_ROCscripthttps://github.com/water-h2o/13_ROCscript. Accessed: 15 June, 2021. 2021.

[266] Juan Buhagiar et al. "Automatic Segmentation of Indoor and Outdoor Scenes from Visual Lifelogging". PhD thesis. June 2017.

[267] Sofus Macskassy and Foster Provost. "Confidence Bands for ROC Curves: Methods and an Empirical Study". In: Jan. 2004, pp. 61–70.

[268] Gregory Campbell. "Advances in statistical methodology for the evaluation of diagnostic and laboratory tests". In: *Statistics in Medicine* 13.5-7 (1994), pp. 499–508. DOI: 10.1002/sim.4780130513. eprint: https://www.onlinelibrary.wiley.com/doi/pdf/10.1002/sim.4780130513. URL: https://www.onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780130513.

[269] Jonathan J Deeks, Douglas G Altman, and Michael J Bradburn. "Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-Analysis". In: *Systematic Reviews in Health Care*. John Wiley & Sons, Ltd, 2001. Chap. 15, pp. 285–312. ISBN: 9780470693926. DOI: 10.1002/9780470693926.ch15. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470693926.ch15. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470693926.ch15.

[270] Tom Fawcett. "An introduction to ROC analysis". In: *Pattern Recognition Letters* 27.8 (2006). ROC Analysis in Pattern Recognition, pp. 861–874. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2005.10.010. URL: https://www.sciencedirect.com/science/article/pii/S016786550500303X.

[271] Matthias Bauer, Mark van der Wilk, and Carl Edward Rasmussen. "Understanding Probabilistic Sparse Gaussian Process Approximations". In: (2017). arXiv: 1606.04820 [stat.ML].

[272] M. Auger et al. "The EXO-200 detector, part I: Detector design and construction". In: *Journal of Instrumentation - J INSTRUM* 7 (Feb. 2012). DOI: 10.1088/1748-0221/7/05/P05010.

[273] A. M. Baldini et al. "The design of the MEG II experiment". In: *The European Physical Journal C* 78.5 (2018), p. 380. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5845-6. URL: https://doi.org/10.1140/epjc/s10052-018-5845-6.

[274] A. Simón et al. "Boosting background suppression in the NEXT experiment through Richardson-Lucy deconvolution". In: (2021). arXiv: 2102.11931 [physics.ins-det].

[275] Shaobo Wang. "PandaX-III high pressure xenon TPC for Neutrinoless Double Beta Decay search". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 958 (2020). Proceedings of the Vienna Conference on Instrumentation 2019, p. 162439. ISSN: 0168-9002. DOI: 10.1016/j.nima.2019.162439. URL: https://www.sciencedirect.com/science/article/pii/S0168900219310034.